The discrete inverse scattering problem in one dimension*

K. M. Case

The Rockefeller University, New York, New York 10021
(Received 18 June 1973)

A discrete version of the inverse scattering problem in one dimension is considered. While the natural formulation is somewhat different from the three-dimensional problem with spherical symmetry, the equations of solution turn out to be almost identical. Indeed, in the continuous limit (Schrödinger equation) even the slight differences disappear. Two equivalent treatments corresponding to considering incidence from left or right are given. For actual computation a combination of the two seems most efficient.

I. INTRODUCTION

Previously[1] we have discussed discrete versions of the inverse scattering problem corresponding to a spherically symmetric potential. Here the purely one-dimensional discrete problem is discussed. There are two reasons: First there are illuminating differences between this case and the earlier treated one. Second, the problem is somewhat intermediate between that for the three-dimensional isotropic and anistropic potentials.

Explicitly we consider the following equation:

$$\tfrac{1}{2}\{\phi(\lambda,n+1)+\phi(\lambda,n-1)\}=\lambda g(n)\,\phi(\lambda,n), \qquad -\infty<n<\infty. \tag{I.1}$$

For simplicity (only) we assume that

$$g(n)\equiv 1, \quad |n|>N. \tag{I.2}$$

If we put

$$\lambda=1-E\Delta^2, \quad g(n)=e^{q(n)\Delta^2},$$

and pass to the limit

$$\Delta\to 0, \quad n\Delta\to x\neq 0,$$

we obtain the one-dimensional Schrödinger equation with potential $q(x)$.

Our problem is the following: If $\lambda=\cos\theta$, with $-\pi\leq\theta\leq\pi$, then Eq. (I.1) has the solutions $\phi_\pm$ which for $|n|\to\infty$ behave so that

$$\begin{aligned} n\to+\infty,\quad \phi_+(\theta,n) &\to S_{++}(\theta)e^{in\theta},\\ \phi_-(\theta,n) &\to e^{-in\theta}+S_{-+}(\theta)e^{+in\theta},\\ n\to-\infty,\quad \phi_+(\theta,n) &\to e^{in\theta}+S_{+-}(\theta)e^{-in\theta},\\ \phi_-(\theta,n) &\to S_{--}(\theta)e^{-in\theta}. \end{aligned} \tag{I.3}$$

The question is to determine $g(n)$ given the 2×2 matrix S. As in other inverse scattering problems we will find that the solution is unique provided we are given the positions of the bound states[2] and their normalization constants.

In Sec. II some needed properties of solutions of Eq. (I.1) are summarized. The explicit solution to the problem is then obtained in Sec. III. An example is described in Sec. IV and the continuous (Schrödinger) limit is given in Sec. V.

II. SOME PROPERTIES

From Eq. (I.1) it follows that if $\phi^{(1)}$, $\phi^{(2)}$ are two solutions then the Wronskian

$$W[\phi^{(1)},\phi^{(2)}] = \tfrac{1}{2}\{\phi^{(1)}(\lambda,n+1)\,\phi^{(2)}(\lambda,n)-\phi^{(1)}(\lambda,n)\,\phi^{(2)}(\lambda,n+1)\}$$

is independent of n. Then from the asymptotic forms given by Eq. (I.3) we conclude that (for $\lambda=\cos\theta$) $S_{++}=S_{--}$ and the matrix S is unitary, i.e.,

$$|S_{++}|^2+|S_{+-}|^2=1 \quad\text{and}\quad S_{++}S_{+-}^*+S_{+-}S_{--}^*=0.$$

Two functions related to $\phi_\pm$ are defined so that

$$f_\pm \xrightarrow[n\to\pm\infty]{} e^{\pm in\theta}.$$

Clearly

$$\phi_+=S_{++}f_+$$

and

$$\phi_-=S_{--}f_-. \tag{II.1}$$

More generally we define $f_\pm$ so: Let $z=e^{i\theta}$, then for *all* z we define $f_\pm(z,n)$ so that

$$\lim_{n\to\pm\infty} f_\pm(z,n)=z^{\pm n}. \tag{II.2}$$

Bound states [i.e. square summable solutions of Eq. (I.1)] clearly correspond to those values z_i where $|z_i|<1$ and $f_+(z,n)\sim f_-(z_i,n)$. From Green's type identities one readily concludes that z_i are real, simple and occur in $\pm$ pairs. Comparing with the asymptotic behavior given by Eq. (I.3) we see these z_i are poles of all elements of S. Further[3]

$$f_+(z_i,n)=\alpha_i f_-(z_i,n), \tag{II.3}$$

where

$$\alpha_i=\lim_{z\to z_i}\frac{S_{+-}(z)}{S_{++}(z)}=\lim_{z\to z_i}\frac{S_{--}(z)}{S_{-+}(z)}. \tag{II.4}$$

Again from a Green's type identity we conclude that

$$\sum_{n=-\infty}^{\infty} g(n)f_+^2(z_i,n)=\frac{1}{M_i^2}\equiv-\alpha_i z_i\left[\frac{d}{dz}\frac{1}{S_{++}}\right]_{z=z_i}, \tag{II.5}$$

and

$$\sum_{n=-\infty}^{\infty} g(n)f_-^2(z_i,n)=\frac{1}{N_i^2}=-\frac{z_i}{\alpha_i}\left[\frac{d}{dz}\frac{1}{S_{++}}\right]_{z=z_i}. \tag{II.6}$$

Finally we note from Eq. (I.1) written in the forms

$$f_+(z,n-1)=(z+1/z)g(n)f_+(z,n)-f_+(z,n+1)$$

and

$$f_-(z,n+1)=(z+1/z)g(z)f_-(z,n)-f_-(z,n-1) \tag{II.7}$$

that we have representations of the form

$$f_+(z,n) = \sum_{m=n}^{\infty} K_+(n,m)\, z^m,$$

and

$$f_-(z,n) = \sum_{m=-\infty}^{n} K_-(m,m)\, z^{-m}, \tag{II. 8}$$

where $K_\pm$ are independent of z. Moreover,

$$\lim_{n\to\infty} K_+(n,m) = \delta(n,m) = \lim_{n\to-\infty} K_-(n,m) \tag{II. 9}$$

and

$$\begin{aligned} g(n) &= K_+(n-1,n-1)/K_+(n,n) \\ &= K_-(n+1,n+1)/K_-(n,n). \end{aligned} \tag{II. 10}$$

III. SOLUTION

Given the matrix S for z on the unit circle, the z_i and the M_i^2 (or N_i^2) we now want to construct $g(n)$.

Notice that for $|z| = 1$, f_+ and f_+^* are two linearly independent solutions of Eq. (I. 1). Hence we can write

$$\phi_- = S_{--} f_- = A f_+ + B f_+^*, \qquad |z| = 1.$$

From the constancy of the Wronskian and the asymptotic behaviors we obtain $B = 1, A = S_{-+}$.

$$\therefore S_{--} f_- = f_+^* + S_{-+} f_+, \qquad |z| = 1. \tag{III. 1}$$

Let us insert the representations of Eq. (II. 8). Then

$$\sum_{m=n}^{\infty} K_+(n,m) z^{-m} + S_{-+} \sum_{m=n}^{\infty} K_+(n,m) z^m = S_{--} \sum_{m=-\infty}^{n} K_-(n,m) z^{-m}, \qquad |z| = 1. \tag{III. 2}$$

Let us multiply this equation by $(2\pi i)^{-1} z^{l-1}$ and integrate around the unit circle. Since

$$(2\pi i)^{-1} \oint z^{l-m-1}\, dz = \delta(l,m),$$

the first term on the left becomes $K_+(n,l)$.

If we define $\gamma_s(m,l)$ by

$$\gamma_+^s(m,l) = \frac{1}{2\pi i} \oint S_{-+} z^{m+l-1}\, dz, \tag{III. 3}$$

the second term is

$$\sum_{m=n}^{\infty} K_+(n,m) \gamma_+^s(m,l).$$

For the right-hand side we need

$$I(m,l) = (2\pi i)^{-1} \oint z^{l-m-1} S_{--}\, dz. \tag{III. 4}$$

Since S_{--} is analytic except for simple poles at the bound states z_i we can evaluate this by residues. Thus

$$I = I_0 + I_{\text{b.s.}},$$

where I_0 is the contribution of the possible pole at $z = 0$ and $I_{\text{b.s.}}$ are the bound state contribution.

We find

$$I_0(m,l) = S_{--}(0)\, \delta(l,m),$$

while

$$I_{\text{b.s.}}(m,l) = \sum_i z_i^{l-m-1} \left\{ \left[\frac{d}{dz} \frac{1}{S_{++}} \right]_{z=z_i} \right\}^{-1}.$$

Hence Eq. (III. 2) is transformed to

$$\begin{aligned} &\sum_{m=n}^{\infty} K_+(n,m)\, \delta(m,l) + \sum_{m=n}^{\infty} K_+(n,m) \gamma_s(m,l) \\ &\quad = S_{--}(0) \sum_{m=-\infty}^{n} K_-(n,m)\, \delta(l,m) \\ &\qquad + \sum_i z_i^l \left(\frac{d}{dz} \frac{1}{S_{--}} \Big|_{z=z_i} \right)^{-1} \sum_{m=-\infty}^{n} K_-(n,m) z_i^{-m}, \end{aligned} \tag{III. 5}$$

but

$$\begin{aligned} \sum_{m=-\infty}^{n} K_-(n,m) z_i^{-m} &= f_-(z_i,n) = \frac{1}{\alpha_i} f_+(z_i,n) \\ &= \frac{1}{\alpha_i} \sum_{m=n}^{\infty} K_+(n,m) z_i^m. \end{aligned}$$

Thus the second term on the right of Eq. (III. 5) is

$$-\sum z_i^l M_i^2 \sum_{m=n}^{\infty} K_+(n,m) z_i^m.$$

Transcribing this term to the left of Eq. (III. 5) we find for $l > n$ the Marchenko-type equation

$$K_+(n,l) + \sum_{m=n}^{\infty} K_+(n,m) \gamma_+(m,l) = 0, \tag{III. 6}$$

where

$$\gamma_+(m,l) = \sum_i M_i^2 z_i^{l+m} + (2\pi i)^{-1} \oint S_{-+} z^{m+l-1}\, dz. \tag{III. 7}$$

Let us remark that in the present context, where S_{-+} is analytic except for simple poles at the bound states z_i, this formula has a simple interpretation, namely:

$$\gamma_+(m.l) = (2\pi i)^{-1} \oint' S_{-+} z^{m+l-1}\, dz. \tag{III. 8}$$

(Here $\oint'$ means merely omit bound state contributions.)

For $l = n$ we get the condition

$$K_+(n,n) + \sum_{m=n}^{\infty} K_+(n,m) \gamma_+(m,n) = S_{--}(0) K_-(n,n). \tag{III. 9}$$

The Eqs. (III. 6, 9) can be simplified if we define

$$\kappa_+(n,l) = K_+(n,l)/K_+(n,n).$$

Then they are

$$\kappa_+(n,l) + \gamma_+(n,l) + \sum_{m=n+1}^{\infty} \kappa_+(n,m) \gamma_+(m,l) = 0, \qquad l > n \tag{III. 10}$$

and

$$\begin{aligned} 1 + \gamma_+(n,n) + \sum_{m=n+1}^{\infty} &\kappa_+(n,m) \gamma_+(m,n) \\ &= S_{--}(0) K_-(n,n)/K_+(n,n). \end{aligned} \tag{III. 11}$$

To have a determinate set of equations, we still must eliminate $K_-(n,n)$. This is accomplished so: From the definitions we have

$$W[f_+,f_-] = (z - z^{-1})/2S_{--}(z) \tag{III. 12}$$

while, using our representations of Eq. (II. 8), we have

$$\begin{aligned} W[f_+,f_-] &\equiv \tfrac{1}{2}\{f_+(z,n+1) f_-(z,n) - f_+(z,n) f_-(z,n+1)\} \\ &= \tfrac{1}{2} \sum_{m=n+1}^{\infty} \sum_{m'=-\infty}^{n} K_+(n+1,m) K_-(n,m') z^{m-m'} \\ &\quad - \sum_{m=n}^{\infty} \sum_{m'=-\infty}^{n+1} K_+(n,m) K_-(n+1,m') z^{m-m'}. \end{aligned} \tag{III. 13}$$

Evaluating $(2\pi i)^{-1} \oint W[f_+,f_-]\, dz$ with the two expressions of Eq. (III. 12 and 13) yields the identity

$1/S_{--}(0) = K_+(n,n)K_-(n+1,n+1)$ (independent of n).

Thus

$$S_{--}(0)K_-(n,n) = 1/K_-(n-1,n-1)$$

and Eq. (III. 11) is

$$1 + \gamma_+(n,n) + \sum_{m=n+1}^{\infty} \kappa_+(n,m)\gamma_+(m,n) = 1/K_+(n,n)K_+n-1,n-1). \quad \text{(III. 14)}$$

The procedure to determine $g(n)$ is then the following: From S_{-+}, M_i^2, and z_i we calculate γ_+. The κ_+ are found by solving Eq. (III. 10). Then $K_+(n,n)$ is determined from Eq. (III. 14) with the boundary condition

$$\lim_{n\to\infty} K_+(n,n) = 1.$$

The $g(n)$ are given by Eq. (II. 10).

Before considering an example it is convenient to note that we could just as well have worked with the K_-. The analogous results are

Let

$$\gamma_-(m,l) = \sum_i N_i^2 z_i^{-(l+m)} + (2\pi i)^{-1} \oint S_{+-} z^{-l-m-1}\, dz, \quad \text{(III. 15)}$$

$$\kappa_-(n,l) = K_-(n,l)/K_-(n,n)$$

Then our equations to determine $g(n)$ are

$$\kappa_-(n,l) + \gamma_-(n,l) + \sum_{m=-\infty}^{n-1} \kappa_-(n,m)\gamma_-(m,l) = 0, \quad l < n, \quad \text{(III. 16)}$$

$$1 + \gamma_-(n,n) + \sum_{m=-\infty}^{n-1} \kappa_-(n,m)\gamma_-(m,n) = 1/K_-(n+1,n+1)K_-(n,n), \quad \text{(III. 17)}$$

$$\lim_{n\to-\infty} K_-(n,n) = 1 \quad \text{and} \quad g(n) = K_-(n+1,n+1)/K_-(n,n). \quad \text{(III. 18)}$$

IV. AN EXAMPLE

Suppose no bound states and

$$S_{+-}(\theta) = S_{-+}(\theta) = \{e^{2i\theta} - 1/[(2-g)e^{2i\theta} - g]\} - 1. \quad \text{(IV. 1)}$$

(The condition that there be no bound states is $g > 1$.)

Then

$$\gamma_+(m,l) = (2\pi i)^{-1} \oint S_{-+} z^{m+l-1}\, dz$$

and

$$\gamma_+(m,l) = 0, \quad m + l > 0.$$

Then

$$\kappa_+(n,l) = 0, \quad l + n > 0,$$

and we have for $n > 0$

$$1 = 1/K_+(n,n)K_+(n-1,n-1).$$

We conclude [since $K_+(\infty,\infty) = 1$] that

$$K_+(n,n) = 1, \quad n \geq 0.$$

When $n = 0$ we have as the only relevant nonzero $\gamma_+(m,l)$ that for $m = l = 0$. Then

$$\gamma_+(0,0) = S_{-+}(0) = (1/g) - 1.$$

Equation (III. 14) becomes

$$1/g = 1/K_+(0,0)K_+(-1,-1);$$
$$\therefore K_+(-1,-1) = g.$$

Then we have

$$g(n) = 1, \quad n \geq 1,$$

and

$$g(0) = K_+(-1,-1)/K_+(0,0) = g/1 = g. \quad \text{(IV. 2)}$$

When $n = -1$, we need in addition to

$$\gamma_+(+1,-1) = \gamma_+(-1,1) = \gamma_+(0,0) = 1/g - 1$$

the quantity

$$\gamma_+(-1,-1) = 1/g.$$

Equation (III. 10) then gives

$$\kappa_+(-1,1) = -\gamma_+(-1,1)$$

and Eq. (III. 14) gives

$$1 + \gamma_+(-1,-1) - \gamma_+(1,-1)^2 = 1/K_+(-1,-1)K_+(-2,-2) = 1/g^2.$$

Thus

$$K_+(-2,-2) = g$$

and

$$g(-1) = 1.$$

Proceeding in this way, we would obtain

$$K_+(-n,-n) = g, \quad n \leq 1,$$

and

$$g(n) = 1, \quad n < -1.$$

An alternate, more efficient calculation, is to use Eqs. (III. 16, 17, 18) to show this result.

V. THE CONTINUOUS LIMIT

Let us consider the "+" form of our equations in the limit $\Delta \to 0$ – so that Eq. (I. 1) passes over into the one-dimensional Schrödinger. For simplicity omit bound states. If we put $z = e^{i\theta}$, the expression for γ_+ becomes

$$\gamma_+(m,l) = (1/2\pi) \int_{-\pi}^{\pi} S_{-+}(\theta) e^{i(m+l)\theta}\, d\theta. \quad \text{(V. 1)}$$

From

$$\cos\theta = \lambda = 1 - E\Delta^2,$$

we obtain

$$\theta = \pm\, \Delta k, \quad \text{where } k = \sqrt{2E}\,.$$

Thus

$$\gamma_+(m,l) = \Delta\gamma_+'(m,l),$$

$$\gamma_+' = (1/2\pi) \int_{-\infty}^{\infty} S_{-+}(k) e^{i(m\Delta+l\Delta)k}\, dk.$$

Let

$$x = m\Delta, \quad y = l\Delta.$$

Then Eq. (III. 10) becomes

$$\kappa_+(x,y) + \Delta\gamma_+'(x,y) + \Delta \sum_{m=n+1}^{\infty} \kappa_+(x,z)\gamma_+'(z,y) = 0, \quad y > x.$$

Clearly $\kappa_+ = \Delta\kappa_+'$ and for small Δ this becomes

$$\kappa_+'(x,y) + \gamma_+'(x,y) + \int_x^\infty \kappa_+'(x,t)\gamma_+'(t,y) = 0, \quad y > x. \tag{V.2}$$

The condition of Eq. (III. 14) becomes when the limit of Eq. (V. 2) for $y \to x$ is inserted becomes

$$1 - \Delta\kappa_+'(x,x) = \frac{1}{K_+(n,n)K_+(n-1,n-1)}. \tag{V.3}$$

In principle we should express $K_+(n-1,n-1)$ in the form

$$K_+(n-1,n-1) \approx K_+(x,x) - \Delta \frac{d}{dx} K_+(x,x)$$

before passing to the limit $\Delta \to 0$. However, we will find $K_+(x,x) = 1 + O(\Delta)$. Hence to an adequate approximation the right side of (V. 3) can be written as $K_+^{-2}(x,x)$. Then

$$K(x,y) \cong 1 + (\Delta/2)\kappa_+'(x,x).$$

For the potential $q(x)$ we use our formulas

$$g(n) = e^{q(x)\Delta^2} \cong 1 + q(x)\Delta^2 = K_+(n-1,n-1)/K_+(n,n)$$
$$\approx 1 - \frac{\Delta^2}{2} \frac{d}{dx} \kappa_+'(x,x),$$

i.e.,

$$q(x) = -\tfrac{1}{2} \frac{d}{dx} \kappa_+'(x,x). \tag{V.4}$$

We note also that the integral representation for f_+ becomes

$$f_+(k,x) = e^{ikx} + \int_x^\infty \kappa_+'(x,y) e^{iky}\, dy.$$

VI. CONCLUSION

It has been found that while the initial formulation of the inverse scattering problem is somewhat different from that for a spherically symmetric problem,[1] the resulting equations for the solution are almost identical. One difference is that the condition of Eq. (III. 14) involves $1/K(n,n)K(n-1,n-1)$ instead of $1/K^2(n,n)$. However, after passage to the continuous limit even this difference disappears.

*Support in part by the U.S. Air Force under Grant No. 722187.

[1] K. M. Case and M. Kac, J. Math. Phys. **14**, 594 (1973). K. M. Case, J. Math. Phys. **14**, 916 (1973); see also I. Kay and H. E. Moses, Nuovo Cimento **10**, 3 (1956); and L. D. Faddeev, Tr. Mat. Inst. Steklov 73, 314 (1964) [English transl., Am. Math. Soc. Transl., Ser. 2 **65**, 139 (1964)].

[2] By "bound states" we mean square summable solutions of Eq. (I.1). An explicit definition of the "normalization constants" is given in Sec. II.

[3] Clearly conditions need to be prescribed so that the element of S can be continued into the unit circle. However, with our present strong limitations the elements of S are all rational functions of z.

Local observations of geodesics in the extended Kerr manifold

Richard H. St. John* and J. D. Finley III

Department of Physics and Astronomy, The University of New Mexico, Albuquerque, New Mexico 87131
(Received 1 August 1973)

Geodesics along the axis of symmetry in Carter's extension of the Kerr metric are divided into two types by the sign of the constant of the motion associated with the timelike Killing vector, and it is shown that this also divides them as to their place of origin on the manifold, which contains infinitely many copies of two different spaces which are flat at $r = \pm\infty$. It is shown that geodesics cannot cross from one space to the other, but that a trajectory with properly applied acceleration can cross over.

Gravitational collapse of a material body into a black hole has been the center of much study recently because of its possible explanation of various astronomical energy sources that have been observed. As a first attempt to understand the features of gravitational collapse, knowledge of what might happen to matter falling into an already existing black hole is certainly important. The solution to the field equations found by Kerr almost certainly describes the space exterior to a rotating black hole; however, no interior metric has yet been found. The actual causal and topological properties of the space inside the first horizon may differ from that predicted by the Kerr metric, but we consider it a useful first approximation. We use Carter's[1,2] analytic extension, which is geodesically complete, of Kerr's original metric. Generally, the geodesics in the four-dimensional case are quite complicated, so, for simplification, only geodesics on the axis of symmetry will be considered at this time.

In Boyer and Lindquist's[3] "Schwarzschild-like" coordinates, hereafter referred to as B&L coordinates, the Kerr metric can be written as[4]

$$ds^2 = \rho^2(dr^2/\Delta + d\theta^2) + (r^2 + a^2)\sin^2\theta\, d\phi^2 - dt^2 + (2mr/\rho^2)(dt - a\sin^2\theta d\phi)^2; \quad (1)$$

when restricted to the axis of symmetry Eq. (1) becomes

$$ds^2 = \rho^2/\Delta\, dr^2 - \Delta/\rho^2\, dt^2, \quad (2)$$

where $\rho^2 = r^2 + a^2\cos^2\theta$, a is the angular momentum per unit mass, $\Delta = r^2 + a^2 - 2mr$, and m is the mass. Carter's extension of the axis of symmetry is created from the repeated use of two null metrics. We define one coordinate system (r,u) with null metric

$$ds^2 = 2dr\,du - \Delta/\rho^2 du^2 \quad (3a)$$

and another similar coordinate system (r,w) with null metric

$$ds^2 = 2dr\,dw - \Delta/\rho^2 dw^2, \quad (3b)$$

where

$$u = \tfrac{1}{2}F(r) + t, \qquad w = \tfrac{1}{2}F(r) - t,$$

$$dF/dr = 2\rho^2/\Delta,$$

such that $F(r) = 2r + K_+^{-1}\ln|r - r_+| + K_-^{-1}\ln|r - r_-|$,

$$K_\pm = \tfrac{1}{2}(r_\pm^2 + a^2)^{-1}(r_\pm - r_\mp),$$

while $r_\pm = m \pm (m^2 - a^2)^{1/2}$ are the roots of $\Delta = 0$.

We note that the function $F(r)$ is separately monotonic in each of the three regions

$$\begin{aligned} &\text{I: } r_+ < r < \infty, \\ &\text{II: } r_- < r < r_+, \\ &\text{III: } -\infty < r < r_-. \end{aligned} \quad (5)$$

Each of these coordinate systems is analytic and extensible to a manifold larger than the one upon which the B&L coordinates were defined. Where these two manifolds overlap, one may introduce full null coordinates (w,u) with the metric

$$ds^2 = \Delta/\rho^2 du\,dw. \quad (6)$$

This overlap region will be one of the three regions in Eq. (5); therefore, given u, w and a region one may uniquely determine r. We may then introduce, following Carter, a new coordinate system (ξ,ψ) by

$$\pm h(u) = \tan(\psi + \xi), \qquad \pm h(w) = \tan(\psi - \xi), \quad (7)$$

where $h(z)$ must be a first-order monotone increasing function such that $h(z) = O(e^{-K_\pm z})$ as $z \to \mp\infty$. The complete manifold will then consist of an infinite sequence of (r,u) patches labeled $(-,m)$ and, superimposed on this, a similar sequence of (r,w) patches labeled $(n,-)$ running perpendicularly to the (r,u) sequence. Labeling each intersection by (n,m), the manifold consists of those intersections where $|n-m| \le 1$. If $n = m$ is odd [even] then it is a II [$\overline{\text{II}}$] region; if n is even (odd) and $< (>)\, m$, then it is a I (I′) region; if n is even (odd) and $> (<) m$ then it is a III (III′) region. The choice in sign in the definition of ξ and ψ is determined by which of the regions I, I′, II, etc., is under consideration. Given an (n,m), the sign is $+h(u)$ $[-h(u)]$ for m odd [even], and equivalently for n with $\pm h(w)$.[5]

From the null metric and the definition of w, the equations of motion are

$$\begin{aligned} r' &= \pm(E^2 - \Delta/\rho^2)^{1/2}, \\ u' &= \rho^2(r' + E)/\Delta, \\ w' &= \rho^2(r' - E)/\Delta, \end{aligned} \quad (8)$$

where E is the constant of the motion associated with the timelike Killing vector, in B&L coordinates, and the prime denotes the total derivative with respect to τ, the proper time. The geodesics in terms of ξ and ψ are not as useful as those in null coordinates, but the sign of ψ' can be used to divide the geodesics into four classes. A short calculation shows that

$$\operatorname{sgn}\psi' = \operatorname{sgn}(\pm E), \quad (9a)$$

if the region is of type I or III; $+E$ is used for unprimed regions and $-E$ for primed regions. For a region of type II

$$\operatorname{sgn}\psi' = \operatorname{sgn}(\pm r'), \tag{9b}$$

$+r'$ if it is barred and $-r'$ if it is unbarred.

The four classes of geodesics which may be obtained from Eqs. (9) are distinguished by their starting region, primed or unprimed, and the sign of ψ. When two particles are in the same region with their respective constants, E, equal in magnitude but opposite in sign, then they are on the same geodesic; the difference is the direction of increasing proper time: toward greater ψ for one and lesser ψ for the other. So for a particle starting in a I region, we need look only at those with increasing ψ. Geodesics in a primed region with decreasing ψ will violate causality when they intersect with unprimed geodesics with increasing ψ. Therefore only those geodesics with $E > 0$ in an unprimed region and $E < 0$ in a primed region will be discussed further. The constant of the motion E associated with the timelike Killing vector, is the total energy per unit mass, as measured by a stationary observer at infinity, in the unprimed region I. However, in the primed region I', the B&L coordinate t tends to minus infinity as one approaches the outer horizon. Therefore, we must insist that a stationary observer at infinity in I' has $d\hat{t}/dt = -1$, so that his proper time agrees in direction with that of all other geodesic coordinate systems, in particular, those which eventually enter II. So in the I' region, the constant E (which is negative for geodesics in I') is the *negative* of the energy per unit mass of a particle as measured by a stationary observer at infinity. This leaves two essentially different geodesics, one starting in a primed region and one in an unprimed region, both with $\psi' > 0$. It is not possible for a *geodesic* to start as one type and end up as another. As an example, we consider a particle which would begin at I and go through II into a III' region. A geodesic in a I region with $\psi' > 0$ has $E > 0$ and, to approach r_+, must have $r' < 0$. Since the perihelion point, $r' = 0$, cannot be in the II region, r' remains negative while in the II region, so u' remains finite; but, to cross from II to III', u' must go to infinity. A similar argument holds for the return trip, with $r' > 0$ and the coordinate w.

It is, however, possible for a particle in II (or $\overline{\text{II}}$) to be accelerated onto a geodesic which will carry it into the opposite type of region from whence it came; e.g., from I to III'. To see this in detail we may consider a particle with $\tilde{E} < 0$, on a geodesic from I' to III', with 4-velocity $(\tilde{r}', \tilde{u}')$, and another particle with $E > 0$, on a geodesic from I to III, with 4-velocity (r', u'); the two paths will cross in the II region. We examine the momentum and energy of the primed geodesic as measured by an observer on the other. We transform to geodesic coordinates by first transforming to a locally nonrotating reference frame (LNRF)[6] from B&L coordinates, and then to the geodesic coordinates by a pure Lorentz transformation, giving us

$$\begin{pmatrix} d\bar{r} \\ d\tau \end{pmatrix} = \begin{pmatrix} u' & -r' \\ -u' & E \end{pmatrix} \begin{pmatrix} dr \\ du \end{pmatrix}, \tag{10}$$

where $\bar{r}$ and τ are the local geodesic coordinates. (We insist that $d\bar{r}/dr > 0$; i.e., they align their axes in the same way.)

The 4-momentum of the primed particle as measured by our observer, as their world lines cross in the II region, is then

$$\mathfrak{p}^1 = E\tilde{u}' - \tilde{E}u' = \rho^2(E\tilde{r}' - \tilde{E}r')/\Delta, \tag{11a}$$

$$\mathfrak{p}^4 = \rho^2(E\tilde{E} - \tilde{r}'r')/\Delta > 0, \tag{11b}$$

$$\text{with } v = \mathfrak{p}^1/\mathfrak{p}^4 = (E\tilde{r}' - \tilde{E}r')/(E\tilde{E} - \tilde{r}'r'). \tag{11c}$$

(These results hold also for the intersection of geodesics in a $\overline{\text{II}}$ region. The difference is in the change of sign of r' when moving from a III to a I region, which changes the sign of $\mathfrak{p}^1$ with respect to the previous case, but leaves $\mathfrak{p}^4$ with the same sign.)

Since $\tilde{E} < 0$, $v > 0$ in II and $v < 0$ in $\overline{\text{II}}$, while $|v| \leq 1$, so that by imparting this velocity to some previously comoving object our observer may put the object onto a primed geodesic. As should be expected, $|v|$ approaches 1 at both horizons. Additionally, $|v|$ is a minimum at $r = |a|$. It is therefore clear that there is some kind of "access" to this "other space," even though of course the observer going there will never return.

*Partially supported by an NSF Traineeship.

[1]B. Carter, Phys. Rev. 141, 1242 (1966).
[2]B. Carter, Phys. Rev. 174, 1559 (1968).
[3]R. Boyer and R. Lindquist, J. Math. Phys. 8, 265 (1967). In this reference, a is the negative of the angular momentum per unit mass.
[4]We use units such that $c = 1 = G$. Furthermore, we only consider the case $|a| \leq m$, since $|a| > m$ makes the problem trivial.
[5]A more complete description, and sketch, of the manifold will be found in Ref. 1.
[6]A LNRF has $d\hat{r}^2 - d\hat{t}^2 = ds^2$. See J. M. Bardeen, Astrophys. J. 178, 347 (1972).

Multiple time scales and the ϕ^4 model of quantum field theory*

B. B. Varga and S. Ø. Aks

University of Illinois, Chicago, Illinois 60680
(Received 13 July 1973)

Multiple time scales perturbation theory is applied to the weakly nonlinear ϕ^4 quantum field theory model. The multiple time scales perturbation equations are solved to lowest order, leading to the removal of secular and quasisecular terms from the standard perturbative solution. This removal occurs in a manner similar to that developed in a previous quasisecular perturbative approach which focused on small energy denominators. The multiple time scales approach provides a better rationale for the quasisecular perturbation theory, as well as providing a systematic method which can be extended to higher orders in the coupling constant. It leads to the natural introduction of a first-order renormalized Hamiltonian, which is a well-defined self-adjoint operator on a certain Hilbert space of physical states. This renormalized Hamiltonian is a direct sum of Schrödinger Hamiltonians on N-particle subspaces, which describe the interactions of pairs of particles via a nonlocal potential.

1. INTRODUCTION

The primary purpose of this work is to investigate the method of multiple time scales[1] in the context of quantum field theory. In recent years general perturbative techniques have been developed for uniformizing perturbation expansions.[2] The general uniformizing method which is called the method of extension[2] introduces an extension of the domain of the independent variable, which is the time variable (t) in this work. The variable t is replaced by a large number of formally independent variables $t_0, t_1, t_2, \cdots$, which at the end of the calculation are taken to depend upon t, thus defining the "physical line"[2] in the extended domain. Physicists have long used the idea of time scales intuitively, for example to separate strong interaction phenomena operating on a time scale of 10^{-23} sec from weak interactions operating on a time scale of 10^{-10} sec. The introduction of multiple variables achieves this separation in a systematic manner.

The secondary purpose of this paper is to place a previously developed perturbation theory into the general and systematic context of the multiple time scales method. The perturbation theory was called quasisecular perturbation theory[3] and it differs from standard perturbation theory through the treatment of those terms which in the standard approach involve small or zero energy denominators. Such terms give rise to secular behavior[4] in the standard approach; the terms are not periodic and blow up at large times. It has been shown that such terms are physically associated with persistent effects which occur over long times and lead to mass renormalization and the binding together of particles.[3]

In the quasisecular perturbative approach the secular terms (vanishing energy denominators) and quasisecular terms (small energy denominators) were modified by a heuristic technique of mass and amplitude renormalization. The presence of small denominators is in conflict with the basic rationale of perturbation theory, that higher-order corrections be small. In quasisecular perturbation theory the conflict was resolved. Nevertheless, the method adopted for resolving this problem had a somewhat arbitrary character since it was based upon techniques which have proven useful in the classical theory of nonlinear oscillations for periodic systems with a small number of degress of freedom.[5]

In this paper we examine a particular case of the method of extension known as the method of multiple linear time scales which has been successfully applied to systems with a large number of degrees of freedom.[1] In this case the physical line in the extended domain is given by $t_0 = t, t_1 = \lambda t, t_2 = \lambda^2 t$ and so on, where λ is a small parameter in the model problem. The technique has been applied to nonperiodic processes such as the irreversible approach to equilibrium in gases.[1] For certain aspects of kinetic theory the linear time scales approach proves inadequate in removing all singularities and a more general approach using multiple time and space scales has to be introduced.[6] However, in this paper we restrict ourselves to the linear time scales approach since it is sufficiently general to handle the secularities of interest.[3]

The model treated is the ϕ^4 model of quantum field theory[7] with the equation of motion

$$(\Box + m^2)\phi(t,x) = \lambda : \phi(t,x)^3 :, \tag{1.1}$$

where $\Box$ is the d'Alembert operator $\partial^2/\partial t^2 - \nabla^2$ and : : denotes normal ordering. With the assumption of periodic boundary conditions for a finite, cubical box, the formal Hamiltonian associated with Eq. (1.1) is

$$H = H_0 + \lambda H_I, \tag{1.2}$$

where

$$H_0 = \sum_l \omega_l \alpha_l^* \alpha_l \tag{1.3}$$

in the notation of Sec. 2 (α_l^* creates a bare particle of momentum l and energy ω_l). The interaction Hamiltonian is

$$H_I = -(16L)^{-1} \sum_{lkqr} \delta_{l+k,q+r} (\omega_l \omega_k \omega_q \omega_r)^{-1/2} : (\alpha_l^* + \alpha_{-l}) \times (\alpha_k^* + \alpha_{-k})(\alpha_q + \alpha_{-q}^*)(\alpha_r + \alpha_{-r}^*) : . \tag{1.4}$$

In Sec. 2 Eq. (1.1) is transformed into a more convenient form using the Fourier amplitudes α_l. The assumption of the linear time scales method is that a solution can be found in the form

$$\phi(t,x) = \phi^{(0)}(t, \lambda t, \lambda^2 t, \cdots; x) + \lambda \phi^{(1)}(t, \lambda t, \lambda^2 t, \cdots; x) + \cdots, \tag{1.5}$$

where $t, \lambda t, \lambda^2 t, \cdots$ are treated as independent variables $t_0, t_1, t_2, \cdots$ during the calculations, and at the end identified with $t, \lambda t, \lambda^2 t, \ldots$, thus defining the physical line in the extended domain.[2] The approximation of

Eq. (1.5) is made uniform on the successive time scales t_0, t_1, t_2 during the stages of the calculation by eliminating any secular or quasisecular terms.

The calculation is carried out to first order in λ in Secs. 3 and 4. It is found that the results of quasisecular perturbation theory[3] are recovered in this order. In the course of the calculation it is found to be natural to introduce a first-order renormalized Hamiltonian which is a well-defined self-adjoint operator on the usual Fock space. In addition it is a particularly simple operator in that it commutes with the particle number operator and thus can be written as the direct sum of Schrödinger type Hamiltonians on N particle subspaces ($N = 0, 1, 2, \cdots$). The dynamical information carried by the first order field of Eq. (1.5) is shown to be unitarily equivalent to the dynamics of the renormalized Hamiltonian in Sec. 4. However, the unitary transformation does not remain in Fock space in the cases of two and three space dimensions.[8] The connection with the strange representations of the canonical commutation relations has been discussed previously.[8]

The above results bear some similarities to the renormalization group approach of Wilson[9] in which multiple space scales are introduced to describe larger and larger clusters of "partons". An effective Lagrangian is obtained which describes the interaction of "particles" considered as clusters of partons. However in the Wilson approach the multiple scales are not related via a small parameter and the method is not based on a perturbative expansion, as it is a strong coupling method.

2. THE COUPLED NONLINEAR FIELD AMPLITUDES

The model chosen has been described previously[3,8]; it is the ϕ^4 model of quantum field theory.[7] The equation of motion for the case of one space dimension is written as

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} + m^2\right)\phi = \lambda : \phi^3 : \tag{2.1}$$

and the real field ϕ is studied on the interval $-\frac{1}{2}L \leq x \leq \frac{1}{2}L$ with periodic boundary conditions ($\hbar = c = 1$). The results obtained are valid for two and three space dimensions as well, with the appropriate natural notational changes. It is convenient however to express most equations in the one-dimensional form to avoid vector subscripts.

By using the Fourier decomposition

$$\phi(t,x) = L^{-1/2}\sum_l a_l(t)e^{ilx}, \tag{2.2}$$

where $l = 2\pi n L^{-1}$ and $n = 0, \pm 1, \pm 2, \cdots$, the field equations transform into an infinite array of nonlinearly coupled equations[3]:

$$\ddot{a}_l + \omega_l^2 a_l = \lambda L^{-1}\sum_{kpr} :a_k a_p a_r: \delta_{l,k+p+r}, \tag{2.3}$$

where $\omega_l^2 = m^2 + l^2$, $\ddot{a}_l = d^2a_l/dt^2$, and δ denotes the Kronecker function. It is possible to transform these equations into a set of first-order differential equations via the substitution

$$A_l(t) = (\omega_l/2)^{1/2}[a_l(t) + i\omega_l^{-1}\dot{a}_l(t)], \tag{2.4}$$

which leads to the equations

$$\dot{A}_l + i\omega_l A_l = (i\lambda/4L)\sum_{kpr}\delta_{l,k+p+r}(\omega_l\omega_k\omega_p\omega_r)^{-1/2} \times : (A_k + A_{-k}^*)(A_p + A_{-p}^*)(A_r + A_{-r}^*): \tag{2.5}$$

via the use of the reality conditions $\phi(t,x) = \phi^*(t,x)$ and $\dot{\phi}(t,x) = \dot{\phi}^*(t,x)$, which imply $a_l(t)^* = a_{-l}(t)$ and $\dot{a}_l(t)^* = \dot{a}_{-l}(t)$. Here * indicates a formal algebraic adjoint operation which is abstractly defined as an antiliner involution.

The initial conditions for Eq. (2.1) at $t = 0$ are taken to be the canonical commutation relations

$$[\phi(0,x), \dot{\phi}(0,y)] = i\delta(x-y), \tag{2.6a}$$

$$[\phi(0,x), \phi(0,y)] = [\dot{\phi}(0,x), \dot{\phi}(0,y)] = 0, \tag{2.6b}$$

which lead to the equivalent commutation relations

$$[A_l(0), A_k^*(0)] = \delta_{l,k}, \tag{2.7a}$$

$$[A_k(0), A_l(0)] = [A_k(0)^*, A_l(0)^*] = 0. \tag{2.7b}$$

These are the initial conditions for Eq. (2.5). As the final transformation it is convenient to extract the unperturbed time dependence, $\alpha_l(t) = A_l(t)e^{i\omega_l t}$, because then

$$\dot{\alpha}_l(t) = (i\lambda/4L)\sum_{kpr}\delta_{l,k+p+r}(\omega_l\omega_k\omega_p\omega_r)^{-1/2}e^{i\omega_l t} \times : (e^{-i\omega_k t}\alpha_k + e^{i\omega_k t}\alpha_{-k}^*)(e^{-i\omega_p t}\alpha_p + e^{i\omega_p t}\alpha_{-p}^*) \times (e^{-i\omega_r t}\alpha_r + e^{i\omega_r t}\alpha_{-r}^*): \tag{2.8}$$

so that $\dot{\alpha}_l = 0$ if $\lambda = 0$.

The essence of the linear multiple time scales method is the assumption that a solution of the system of Eqs. (2.8) can be found in the form[1]

$$\alpha_l(t) = \alpha_l^{(0)}(t_0, t_1, t_2, \cdots) + \lambda\alpha_l^{(1)}(t_0, t_1, t_2, \cdots) + \cdots, \tag{2.9}$$

where $t_0 = t$ and the new variables are given by $t_1 = \lambda t$, $t_2 = \lambda^2 t$, and so on. It is shown in the next section how the extra freedom which arises from the introduction of the new variables (method of extension) can be used to eliminate secular and quasisecular terms from the standard perturbative solution of Eq. (2.8). The variable t_0 describes processes occurring on a fast time scale whose unit interval is m^{-1}, whereas t_1, t_2, etc. describe processes occurring on progressively slower time scales involving the cumulative persistent effects of the interactions in higher order. Strictly speaking the slower time scales should be related to t_0 by the dimensionless coupling constant (which is $|\lambda|m^{-2}$ in the case of one space dimension); however, no difficulties are caused by the simpler assumptions $t_1 = \lambda t$, etc.

The first term $\alpha_l^{(0)}$ is chosen to satisfy the initial conditions of Eq. (2.7) and the subsequent terms are taken to vanish at $t = 0$ [e.g., $\alpha_l^{(1)}(0, 0, \cdots) = 0$].

3. PERTURBATIVE EQUATIONS

The first two equations of the perturbative hierarchy obtained by substituting Eq. (2.9) into Eq. (2.8) are

$$\frac{\partial}{\partial t_0}\alpha_l^{(0)}(t_0, t_1, \cdots) = 0, \tag{3.1}$$

$$\frac{\partial}{\partial t_0}\alpha_l^{(1)}(t_0, t_1, \cdots) + \frac{\partial}{\partial t_1}\alpha_l^{(0)}(t_0, t_1, \cdots) = (i/4L)\sum_{kpr}\delta_{l,k+p+r} \times (\omega_l\omega_k\omega_p\omega_r)^{-1/2}e^{i\omega_l t_0} : (e^{-i\omega_k t_0}\alpha_k^{(0)} + e^{i\omega_k t_0}\alpha_{-k}^{(0)*}) \times (e^{-i\omega_p t_0}\alpha_p^{(0)} + e^{i\omega_p t_0}\alpha_{-p}^{(0)*})(e^{-i\omega_r t_0}\alpha_r^{(0)} + e^{i\omega_r t_0}\alpha_{-r}^{(0)*}): \tag{3.2}$$

Equation (3.1) implies that $\alpha_l^{(0)}$ depends only on the slow variables t_1, t_2, etc. Among the many terms on the right-hand side of Eq. (3.2) there are some which do not vary with t_0 or which vary very slowly. For example, if $\omega_l + \omega_k - \omega_p - \omega_r = 0$, then the term $\alpha_{-k}^{(0)*}\alpha_p^{(0)}\alpha_r^{(0)}$ has a coefficient which is constant in t_0. This leads to secular growth for the corresponding contribution to $\alpha_l^{(1)}$ in the standard perturbation theory, [i.e., a contribution like $(it/4L)\delta_{l,k+p+r}\alpha_{-k}^{(0)*}\alpha_p^{(0)}\alpha_r^{(0)}(\omega_l\omega_k\omega_p\omega_r)^{-1/2}$], which becomes large for long times, violating the perturbation theory rationale that successive terms should be uniformly small. On the other hand, if $E = \omega_l + \omega_k - \omega_p - \omega_r$ is of order $(\lambda m^{-2})m$ (where λm^{-2} is the dimensionless coupling constant for one space dimension), then the corresponding contribution to $\alpha_l^{(1)}$ in standard perturbation theory involves the small energy denominator E^{-1} (of order $m\lambda^{-1}$) and this part of $\lambda\alpha_l^{(1)}$ is of zeroth order in λ. This violates the expectation that all zeroth-order terms be contained in $\alpha_l^{(0)}$.

For the consistency of the perturbative expansion of Eq. (2.9) we demand that $\lambda\alpha_l^{(1)}$ be uniformly smaller (for all times) than $\alpha_l^{(0)}$ by one power of the small parameter λm^{-2}. This consistency condition determines $(\partial/\partial t_1)\alpha_l^{(0)}$ in Eq. (3.2) which must be chosen to cancel out all potentially secular and quasisecular terms which appear on the right-hand side of Eq. (3.2). The potentially quasisecular terms are the ones which oscillate at frequencies e^{-iEt_0} such that $|E| < |\lambda| m^{-1}$.

Assuming that the above cancellation has occurred, we are left with the equation

$$\begin{aligned}\frac{\partial}{\partial t_0}\alpha_l^{(1)} = (i/4L)\sum_{kpr}{}' \delta_{l,k+p+r}(\omega_l\omega_k\omega_p\omega_r)^{-1/2} \\ \times [e^{i(\omega_l+\omega_k+\omega_p+\omega_r)t_0}\alpha_{-k}^{(0)*}\alpha_{-p}^{(0)*}\alpha_{-r}^{(0)*} \\ + 3e^{i(\omega_l+\omega_k+\omega_p-\omega_r)t_0}\alpha_{-k}^{(0)*}\alpha_{-p}^{(0)*}\alpha_r^{(0)} \\ + 3e^{i(\omega_l+\omega_k-\omega_p-\omega_r)t_0}\alpha_{-k}^{(0)*}\alpha_p^{(0)}\alpha_r^{(0)} \\ + e^{i(\omega_l-\omega_k-\omega_p-\omega_r)t_0}\alpha_k^{(0)}\alpha_p^{(0)}\alpha_r^{(0)}],\end{aligned} \tag{3.3}$$

where $\sum'$ indicates that the sum is restricted in such a way that all the terms on the right-hand side of Eq. (3.3) oscillate at frequencies e^{-iEt_0} satisfying $|E| > |\lambda| m^{-1}$. The solution of Eq. (3.3) gives the behavior of $\alpha_l^{(1)}$ on the fast time scale characterized by t_0. This part of the problem can be solved explicitly in a straightforward manner because $\alpha_k^{(0)}$ on the right-hand side of Eq. (3.3) is constant on the fast time scale t_0.

It should be noted that the consistency condition which requires the determination of $\alpha_l^{(0)}(t_1)$ is where the real difficulty lies. The cancellation condition for potentially quasisecular terms is formulated in Eq. (4.1) in which the frequencies e^{-iEt_0} appearing on the right hand side satisfy the condition $|E| < |\lambda| m^{-1}$. Equation (4.1) is difficult to solve because $\alpha_l^{(0)}$ on the right-hand side also varies with t_1 and therefore the equations form a complicated nonlinear system which does not admit an exact solution.

The solution of Eq. (4.1) can be regarded as the solution of a Heisenberg equation of motion for a new Hamiltonian, which is referred to as the renormalized Hamiltonian. In the following this terminology is justified. A prime advantage of such an approach is that it is possible to focus upon a solution valid in certain small subspaces of Hilbert space (e.g., one and two particle subspaces) rather than finding the global solution of Eq. (4.1) for the entire Hilbert space.

4. THE RENORMALIZED HAMILTONIAN

The only potentially quasisecular terms in Eq. (3.2) are those of the form $\alpha^{(0)*}\alpha^{(0)}\alpha^{(0)}$ since the only way the sum of four $\pm\omega_k$'s can be near zero is when two have plus signs and two have minus signs. This leads to the simpler looking consistency condition

$$\begin{aligned}\frac{\partial}{\partial t_1}\alpha_l^{(0)}(t_1) = (3i/4L)\sum_{kpr} D_{lkpr}(\omega_l\omega_k\omega_p\omega_r)^{-1/2} \\ \times e^{i(\omega_l+\omega_k-\omega_p-\omega_r)t}\alpha_k^{(0)*}(t_1)\alpha_p^{(0)}(t_1)\alpha_r^{(0)}(t_1).\end{aligned} \tag{4.1}$$

The quasisecular restriction on the sum in Eq. (4.1) has been met by introducing the function D_{lkpr} which is 1 if both $l + k = p + r$ and $|\omega_l + \omega_k - \omega_p - \omega_r| \le |\lambda| m^{-1}$ are satisfied, and 0 otherwise. Since the frequency in the exponential is of order λ, it may be written as $e^{i\lambda^{-1}(\omega_l+\omega_k-\omega_p-\omega_r)t_1}$ which has a zeroth-order frequency, and thus Eq. (4.1) involves only the slow time scale t_1, so that the compatibility condition[2]

$$\frac{\partial}{\partial t_0}\frac{\partial\alpha_l^{(0)}}{\partial t_1} = \frac{\partial}{\partial t_1}\frac{\partial\alpha_l^{(0)}}{\partial t_0} = 0$$

is satisfied. If the substitutions $t_1 = \lambda t$ and $b_l(t) e^{-i\omega_l t}\alpha_l^{(0)}(\lambda t)$ are made in Eq. (4.1), the result is seen to be the system of Heisenberg equations for a new Hamiltonian

$$\begin{aligned}H_R = \sum_l \omega_l b_l^* b_l - (3\lambda/8L)\sum_{lkpr} D_{lkpr}(\omega_l\omega_k\omega_p\omega_r)^{-1/2} \\ \times b_l^* b_k^* b_p b_r\end{aligned} \tag{4.2}$$

provided that the commutation relations

$$[b_l(t), b_k(t)] = [b_l^*(t), b_k^*(t)] = 0, \tag{4.3a}$$

$$[b_l(t), b_k^*(t)] = \delta_{l,k} \tag{4.3b}$$

are valid at all times. From Eq. (2.7) they are seen to be valid at the initial time. A simple calculation based on Eq. (4.1) which is given in Appendix A shows that all the higher order time derivatives of the commutators in Eq. (4.3) vanish at the initial time. The assumption that the commutation relations are valid at all times is clearly consistent since the time evolution operator associated with the Hermitian operator H_R of Eq. (4.2) is unitary $[b_l(t) = e^{iH_R t}b_l(0)e^{-iH_R t}]$. This implies the operators $\alpha_l^{(0)}(t_1)$ satisfy the canonical commutation relations at all times.

The Hamiltonian H_R has several pleasant features. First it commutes with the number operator

$$N = \sum_l b_l^* b_l .$$

Such a result suggests the Hamiltonian might be defined on a Hilbert space containing a unique (normalized) vacuum state $|0\rangle$ with the property $N|0\rangle = 0$, which implies $b_l|0\rangle = 0$ for all l. The other states in the Hilbert space are in the closure of the subspace of states on the form $\mathcal{P}(b_{l_1}^*, \cdots, b_{l_n}^*)|0\rangle$, where $\mathcal{P}$ is a polynomial in n variables. This Hilbert space is just the usual Fock space which is constructed via the field operators at the initial time. The Hamiltonian leaves invariant subspaces of a definite particle number, and thus it is a direct sum of Hamiltonians over the Hilbert spaces of definite particle number N, $N = 0$ (no particle state or vacuum), $N = 1$, $N = 2$, $\cdots$ and so on, which lead to standard Schrödinger equations on each such

subspace. The commutation rules imply Bose statistics for the particles. It is remarkable that the consistency condition leads to a system of Schrödinger equations.

The Hamiltonian on the N-particle subspace may be written as

$$H_R^{(N)} = \sum_{i=1}^{N} \omega(p_i) + \sum_{i<j=1}^{N} V_{ij}, \tag{4.4}$$

where in the momentum representation for the symmetric wave functions $\Psi = \Psi(p_1, p_2, \ldots, p_N)$, the kinetic energy $\omega(p_i) = (p_i^2 + m^2)^{1/2}$ is multiplicative and V_{ij} is a non-local potential acting between the pair of particles labeled by (i,j). For example

$$\begin{aligned}(V_{12}\Psi)(p_1, \ldots, p_N) &= -(3\lambda/4L) \sum_{k_1 k_2} D_{p_1 p_2 k_1 k_2} (\omega_{p_1}\omega_{p_2}\omega_{k_1}\omega_{k_2})^{-1/2} \\ &\times \Psi(k_1, k_2, p_3, \ldots, p_N).\end{aligned} \tag{4.5}$$

Evidently V_{12} is symmetric. In Appendix B it is proved that V_{12} is a bounded operator whose bound is independent of L, the size of the box. It follows that the total interaction potential in Eq. (4.4) is a bounded, symmetric operator and therefore $H_R^{(N)}$ is a self-adjoint operator whose domain is the same as the domain of the kinetic energy

$$T^{(N)} = \sum_{i=1}^{N} \omega(p_i).$$

The one-particle states of the Hamiltonian $H_R^{(1)}$ lie at the unperturbed energy $\omega(p) = (p^2 + m^2)^{1/2}$. The two particle states undergo scattering due to the interaction V_{12}. For $\lambda > 0$ the potential V_{12} has negative matrix elements and corresponds to an attractive interaction. In previous work it was shown that a bound two-particle state occurs for arbitrarily weak $\lambda > 0$ for the cases of one and two space dimensions, but not for the case of three.[10]

5. THE FIRST-ORDER HEISENBERG FIELD

The first-order solution of the field equation (2.1) requires finding $\alpha_l^{(1)}$ as the solution of Eq. (3.3) on the fast time scale t_0. The solution which satisfies the prescribed zero initial condition of Sec. 2 is

$$\begin{aligned}\alpha_l^{(1)} &= (4L)^{-1} \sum_{kpr} \delta_{l+k,p+r} (\omega_l\omega_k\omega_p\omega_r)^{-1/2} \\ &\times \left[\frac{e^{i(\omega_l+\omega_k+\omega_p+\omega_r)t_0} - 1}{(\omega_l + \omega_k + \omega_p + \omega_r)} \alpha_k^{(0)*}\alpha_{-p}^{(0)*}\alpha_{-r}^{(0)*} \right. \\ &+ 3\frac{e^{i(\omega_l+\omega_k+\omega_p-\omega_r)t_0} - 1}{(\omega_l + \omega_k + \omega_p - \omega_r)} \alpha_k^{(0)*}\alpha_{-p}^{(0)*}\alpha_r^{(0)} \\ &+ 3(1 - D_{lkpr})\frac{e^{i(\omega_l+\omega_k-\omega_p-\omega_r)t_0} - 1}{(\omega_l + \omega_k - \omega_p - \omega_r)} \alpha_k^{(0)*}\alpha_p^{(0)}\alpha_r^{(0)} \\ &\left. + \frac{e^{i(\omega_l-\omega_k-\omega_p-\omega_r)t_0} - 1}{(\omega_l - \omega_k - \omega_p - \omega_r)} \alpha_{-k}^{(0)}\alpha_p^{(0)}\alpha_r^{(0)} \right] + \gamma_l(t_1).\end{aligned} \tag{5.1}$$

Note that $\alpha_l^{(1)}$ depends on t_0 as well as on t_1 through the t_1 dependence of $\alpha^{(0)}$ and $\alpha^{(0)*}$. The factor $1 - D_{lkpr}$ in the third term takes care of the nonsecular restriction on the sum in Eq. (3.3); it prevents the denominator of this term from becoming smaller than $|\lambda| m^{-1}$. This is the only potentially dangerous denominator in Eq. (5.1). The last term $\gamma_l(t_1)$ is determined by eliminating quasisecular terms from the third member of the hierarchy which was started in Eq. (3.1). The left-hand side of the third member of the hierarchy involves

$$\frac{\partial}{\partial t_0}\alpha^{(2)} + \frac{\partial}{\partial t_1}\alpha^{(1)} + \frac{\partial}{\partial t_2}\alpha^{(0)}$$

and therefore

$$\frac{\partial}{\partial t_1}\alpha^{(1)}$$

is determined by the consistency condition to eliminate quasisecular terms on the fast (t_0) time scale, whereas

$$\frac{\partial}{\partial t_2}\alpha^{(0)}$$

is chosen to eliminate quasisecular terms on the slow (t_1) time scale. The initial condition on $\gamma_l(t_1)$ is $\gamma_l(0) = 0$.

The complete first-order Fourier amplitude of the field $\phi(t,x)$ is

$$a_l(t) = (2\omega_l)^{-1/2}[(\alpha_l^{(0)} + \lambda\alpha_l^{(1)})e^{-i\omega_l t} + (\alpha_{-l}^{(0)*} + \lambda\alpha_{-l}^{(1)*})e^{i\omega_l t}]. \tag{5.2}$$

In Eq. (5.2) the independent variable appears as $\alpha_l^{(0)} = \alpha_l^{(0)}(\lambda t)$ and $\alpha_l^{(1)} = \alpha_l^{(1)}(t, \lambda t)$. The expression Eq. (5.2) is a solution of Eq. (2.1) to terms of $O(\lambda^2)$ uniformly over the time interval $|t| < m|\lambda|^{-1}$ provided that γ_l in Eq. (5.1) has been chosen to eliminate quasisecular terms which appear in second order, as discussed above. We defer the details of such a calculation to a later paper.

In previous work a heuristic method was devised for obtaining dynamical information from the first-order Heisenberg field.[10] The method involved extracting all terms in Eq. (5.2) which oscillate like $e^{-i\omega_l t}$ or close to it. The result can be written as

$$a_l(t) = (2\omega_l)^{-1/2}[U_l e^{-i\omega_l t} + U_{-l}^* e^{i\omega_l t} + W_l(t)], \tag{5.3}$$

where $W_l(t)$ has the property that its Fourier transform vanishes in intervals of width $2|\lambda| m^{-1}$ centered at $\pm\omega_l$. Then

$$U_l = \beta_l + (3\lambda/8\omega_l L)\sum_{rkp} (\omega_l\omega_k\omega_p\omega_r)^{-1/2} D_{lrkp}\, \alpha_r^{(0)*}\alpha_k^{(0)}\alpha_p^{(0)} \tag{5.4}$$

and β_l comes from the t_0 independent terms in Eq. (5.1),

$$\begin{aligned}\beta_l(t_1) &= \alpha_l^{(0)} - (\lambda/4L)\sum_{kpr} \delta_{l+k,p+r}(\omega_l\omega_k\omega_p\omega_r)^{-1/2} \\ &\times [(\omega_l + \omega_k + \omega_p + \omega_r)^{-1}\alpha_k^{(0)*}\alpha_{-p}^{(0)*}\alpha_{-r}^{(0)*} \\ &+ 3(\omega_l + \omega_k + \omega_p - \omega_r)^{-1}\alpha_k^{(0)*}\alpha_{-p}^{(0)*}\alpha_r^{(0)} \\ &+ 3(1 - D_{lkpr})(\omega_l + \omega_k - \omega_p - \omega_r)^{-1}\alpha_k^{(0)*}\alpha_p^{(0)}\alpha_r^{(0)} \\ &+ (\omega_l - \omega_k - \omega_p - \omega_r)^{-1}\alpha_{-k}^{(0)}\alpha_p^{(0)}\alpha_r^{(0)}] + \lambda\gamma_l(t_1).\end{aligned} \tag{5.5}$$

In Eq. (5.4) the last term arises from $\alpha_{-l}^{(1)*}$ which has an oscillation close to $e^{-i\omega_l t_0}$ when the conjugate of the second term in Eq. (5.1) satisfies $|\omega_k + \omega_p - \omega_r - \omega_l| \leq |\lambda| m^{-1}$ (i.e., $D_{lrkp} = 1$). Approximating such frequencies by $e^{-i\omega_l t_0}$ is valid on the fast time scale $|t| < m^{-1}$ on which $\gamma_l(t_1)$ can be replaced by zero. Actually since our solution is accurate to first order in λ, the last term in Eq. (5.4) can be dropped as well since the D function restricts the summation and introduces a factor proportional to λ.

Strictly speaking Eqs. (5.4) and (5.5) are valid only if the t_1 dependence of $\alpha^{(0)}$ is neglected so that the frequency support of terms in Eq. (5.3) is given by the t_0 dependence. This approximation is quite accurate because the frequency shifts implied by the t_1 dependence

are given by H_R and are quite small. Detailed calculations[10] give frequency shifts proportional to λ^2 for one space dimension, and much smaller for two and three space dimensions.

A Hilbert space of physical states was constructed in previous work.[10] The vacuum state is taken to be the (unique) state Φ_0 annihilated by all the positive frequency amplitudes β_l, and its energy is taken as zero. One particle states of momentum k and energy ω_k are of the form $\beta_k^*\Phi_0$. On the subspace of two-particle states generated by the vectors $\beta_l^*\beta_k^*\Phi_0$ the Hamiltonian satisfies[10]

$$H\beta_l^*\beta_k^*\Phi_0 = (\omega_l + \omega_k)\beta_l^*\beta_k^*\Phi_0 - (3\lambda/4L)\sum_{qr}(\omega_l\omega_k\omega_q\omega_r)^{-1/2}D_{lkqr}\beta_r^*\beta_q^*\Phi_0. \quad (5.6)$$

By comparison with Eqs. (4.4) and (4.5) this result is seen to be identical with the action of the renormalized Hamiltonian $H_R^{(2)}$ on its two-particle subspace. This suggests there exists an isomorphism between the dynamical information contained in the first-order (renormalized) field of Eq. (5.2) and the renormalized Hamiltonian of Sec. 2.

Such an isomorphism is established generally via a formally unitary clothing transformation.[8] We may write

$$\beta_l(0) = e^{\lambda S}\alpha_l e^{-\lambda S} \quad (5.7)$$

where both sides are taken at $t = 0$, and the transformation is evaluated by expanding, $\beta_l = \alpha_l + \lambda[S, \alpha_l] + O(\lambda^2)$. The formally anti-Hermitian operator S is given by[8]

$$\begin{aligned} S = {} & (4L)^{-1}\sum_{kpqr}\delta_{k,p+q+r}(\omega_k\omega_p\omega_q\omega_r)^{-1/2} \\ & \times [(\omega_k - \omega_p - \omega_q - \omega_r)^{-1}(\alpha_k^*\alpha_p\alpha_q\alpha_r - \alpha_r^*\alpha_q^*\alpha_p^*\alpha_k) \\ & + \tfrac{1}{4}(\omega_k + \omega_p + \omega_q + \omega_r)^{-1}(\alpha_{-k}^*\alpha_p^*\alpha_q^*\alpha_r^* - \alpha_r\alpha_q\alpha_p\alpha_{-k})] \\ & + 3(8L)^{-1}\sum_{kpqr}\delta_{k+p,q+r}(1 - D_{kpqr})(\omega_k\omega_p\omega_q\omega_r)^{-1/2} \\ & \times (\omega_k + \omega_p - \omega_q - \omega_r)^{-1}\alpha_k^*\alpha_p^*\alpha_q\alpha_r, \end{aligned} \quad (5.8)$$

where the operators α, α^* are all at $t = 0$. If we define the time dependent β by the equation $\beta_l(t)e^{-i\omega_l t} = e^{iHt}\beta_l(0)e^{-iHt}$, where e^{-iHt} is the unitary time evolution operator associated with the (formal) Hamiltonian H then

$$\beta_l(t)e^{-i\omega_l t} = e^{\lambda S}e^{iH't}\alpha_l e^{-iH't}e^{-\lambda S}, \quad (5.9)$$

where

$$H' = e^{-\lambda S}He^{\lambda S} = H + \lambda[H_0, S] + O(\lambda^2). \quad (5.10)$$

From the relation $[H_0, \alpha_l] = -\omega_l\alpha_l$ (which follows from the commutation relations at the initial time) we see that

$$H + \lambda[H_0, S] = H_R = H' + O(\lambda^2), \quad (5.11)$$

where H_R is the expression of Eq. (4.2) with b_l replaced by $\alpha_l(0) = \alpha_l^{(0)}(0)$. Substituting Eq. (5.10) in Eq. (5.9), we get to within terms of order λ^2

$$\beta_l(t) = e^{\lambda S}\alpha_l^{(0)}(\lambda t)e^{-\lambda S} \quad (5.12)$$

and thus $\beta_l(t)e^{-i\omega_l t}$ [which is a solution of the Heisenberg equation in terms of the original Hamiltonian H of Eq. (1.2)] is related by a formal unitary transformation to $\alpha_l^{(0)}(\lambda t)e^{-i\omega_l t}$, which is the solution of the Heisenberg equation in terms of the renormalized Hamiltonian H_R.

Since we can interpret $\beta_l(t)$ as the renormalized (positive frequency) field amplitude which annihilates a "physical particle",[10] we see that the renormalized Hamiltonian H_R describes the interaction of the physical particles in the physical Hilbert space (through the clothing transformation).

The properties of the clothing transformation have been investigated previously.[8] In the case of one space dimension it is a unitary transformation which acts within the usual Fock space. In the cases of two and three space dimensions it is interpreted as an "improper" unitary transformation[11] (i.e., its domain is Fock space and its range a new Hilbert space, the space of physical states, which is outside the Fock space).

ACKNOWLEDGMENT

The authors would like to acknowledge helpful discussions with Professor G. V. Ramanathan.

APPENDIX A

In Sec. 4 an operator $b_l(t)$ is defined which satisfies the equation of motion

$$\dot{b}_l(t) = -i\omega_l b_l(t) + (3i\lambda/4L)\sum_{kpr}D_{lkpr} \times (\omega_l\omega_k\omega_p\omega_r)^{-1/2}b_k^*(t)b_p(t)b_r(t). \quad (A1)$$

The commutation relations of Eq. (2.7) imply at the initial time ($t = 0$) that

$$[b_l(0), b_k(0)] = [b_l^*(0), b_k^*(0)] = 0, \quad (A2a)$$

$$[b_l(0), b_k^*(0)] = \delta_{lk}. \quad (A2b)$$

We show that all the time derivatives of these commutators must vanish at the initial time. The relations Eq. (A2) and H_R as defined by Eq. (4.2) imply that

$$\dot{b}_l = i[H_R, b_l], \quad (A3a)$$

$$\dot{b}_l^* = i[H_R, b_l^*] \quad (A3b)$$

holds at $t = 0$, and by Leibniz' rule for derivatives of products and for commutators of products,

$$\frac{d}{dt}\mathcal{P}(b, b^*) = i[H_R, \mathcal{P}], \quad (A4)$$

holds at $t = 0$, where $\mathcal{P}$ is any polynomial in the various $b = b^*$.

The time derivative of the commutator of Eq. (A2b) for instance is given by

$$\left[\frac{d}{dt}b_l, b_k^*\right] + \left[b_l, \frac{d}{dt}b_k^*\right] = i[[H_R, b_l], b_k^*] + i[b_l, [H_R, b_k^*]] = i[H_R, [b_l, b_k^*]], \quad (A5)$$

where the last equality follows from the Jacobi identity. From Eq. (A2b) the right-hand side vanishes at $t = 0$. In a similar manner from Eq. (A4) at the initial time we have

$$\frac{d^2}{dt^2}[b_l, b_k^*] = i^2[H_R, [H_R, [b_l, b_k^*]]] \quad (A6)$$

which again vanishes. A proof by induction shows all higher derivatives vanish at $t = 0$. A similar proof applies to the other commutators in Eq. (A2).

These considerations suggest strongly that the basic

commutation rules of Eq. (A2) are valid at all times. Of course, this is not a strict proof as there are nonanalytic functions like e^{-1/t^2} which vanishes at $t = 0$ and all of whose derivatives vanish at $t = 0$.

APPENDIX B

The nonlocal potential acting between particles 1 and 2 was given in Sec. 4 in the N particle momentum representation by Eq. (4. 5). From the Cauchy–Schwarz inequality it follows that

$$|(V_{12}\Psi)(p_1,\ldots,p_N)|^2 \leq (3\lambda/4L)^2 \left(\sum_{k_1k_2} \omega_{k_1}^{-1}\omega_{k_2}^{-1} D_{p_1p_2k_1k_2}\right) \times \left(\sum_{k_1k_2} \omega_{p_1}^{-1}\omega_{p_2}^{-1} D_{p_1p_2k_1k_2} |\Psi(k_1,k_2,p_3,\ldots,p_N)|^2\right). \tag{B1}$$

We consider the case of three space dimensions, where the boundedness is the most difficult to establish. Then

$$\sum_{k_1k_2} \omega_{k_1}^{-1}\omega_{k_2}^{-1} D_{p_1p_2k_1k_2} \leq BL, \tag{B2}$$

where B is some (volume independent) constant, because for p_1, p_2 fixed the momentum conserving Kronecker delta restricts the sum to one variable, and the number of terms in the sum is bounded by $BL\omega_k^2$ (where L is the volume of the box), since in three dimensions the density of states goes as $k^2 \approx \omega_k^2$ at high energies. Thus the squared norm of the vector $V_{12}\Psi$ satisfies

$$\|V_{12}\Psi\|^2 \leq L^{-1}B' \times \sum_{p_1\cdots p_N} \sum_{k_1k_2} \omega_{p_1}^{-1}\omega_{p_2}^{-1} D_{p_1p_2k_1k_2} |\Psi(k_1,k_2,p_3,\ldots,p_N)|^2 \tag{B3}$$

where B' is another constant. Doing the sum over p_1, p_2 first in Eq. (B3), we obtain via Eq. (B2)

$$\|V_{12}\Psi\|^2 \leq B'' \sum_{k_1k_2} \sum_{p_3\cdots p_N} |\Psi(k_1,k_2,p_3,\ldots,p_N)|^2 = B''\|\Psi\|^2, \tag{B4}$$

where B'' is another volume-independent constant. Hence V_{12} is a bounded operator.

It is amusing to note that V_{12} is not a bounded operator in the case of four space dimensions.

*Work supported in part by National Science Foundation Grants NSF GP-8734 and 21052.

[1] E. Frieman, J. Math. Phys. 4, 410 (1963). G. Sandri, Ann. Phys. (N.Y.) 24, 332 (1963).
[2] G. Sandri, Nuovo Cimento 36, 67 (1965).
[3] S. Aks, J. Sienicki, and B. Varga, Phys. Rev. D 6, 520 (1972).
[4] S. Aks, Fortschr. Phys. 15, 661 (1967). S. Aks and R. Carhart, Nuovo Cimento A 64, 798 (1969).
[5] N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Hindustani Press, Delhi, 1961).
[6] G. V. Ramanathan and G. Sandri, J. Math. Phys. 10, 1763 (1969).
[7] A. M. Jaffe, Rev. Mod. Phys. 41, 576 (1969).
[8] B. Varga and S. Aks, Phys. Rev. D 6, 529 (1972).
[9] K. Wilson, Phys. Rev. 140, B 445 (1965); Phys. Rev. D 2, 1438 (1970); and Phys. Rev. D 6, 419 (1972).
[10] S. Aks and B. Varga, Phys. Rev. D 6, 2773 (1972).
[11] G. Barton, *Introduction to Advanced Field Theory* (Interscience, New York, 1963), p. 126.

Reduction on the Lorentz subgroup of UIR's of the Poincaré group induced by a semisimple little group

E. Angelopoulos

Institut Henri Poincaré, 11 rue Pierre et Marie Curie, Paris V°, France
(Received 15 February 1972)

All UIR's of the Poincaré group corresponding to nonzero mass are reduced on a Lorentz subgroup by means of a unique formalism. The maximal differential domain of each UIR is proved to be a nuclear space. The completeness relations (generalized matrix elements) are established between the energy-orbital angular momentum—total angular momentum and one of its components—basis, and the relativistic—orbital—total angular momentum and one of the components of t.a.m. basis.

INTRODUCTION

The unitary irreducible representations (UIR's) of the universal covering of the Poincaré group, $\overline{\mathcal{P}}$, are of great interest. Since Wigner,[1] Bargman,[2] and Shirokov[3] have given the exhaustive list and explicit realizations of them, many authors have given other realizations of these UIR's or their tensor products, by a change of basis which diagonalizes a set of commuting operators corresponding to a definite choice of physical observables.[4-6]

Such problems often take the general form of the decomposition of a representation on a subgroup, the operators to diagonalize being closely related to such a subgroup, and their spectra to the UIR's of that subgroup.

Our purpose is to reduce UIR's of $\overline{\mathcal{P}}$, induced either by $SU(2)$ or $SU(I,I)$, on a Lorentz subgroup. This problem was first treated by Joos[4] for the case of the so-called physical representations. However, his method can not easily be transposed to UIR's with negative squared mass, and it is not very rigorous mathematically. Moreover, the UIR's corresponding to negative mass square, besides their interest from the physical point of view (they appear as irreducible components in tensor products decomposition) present many analogies with the positive-mass-square ones and can, therefore, be studied simultaneously.

The method used here keeps a unique formalism for positive and negative masses, trying to put in evidence the structural similarities of $SU(2)$ and $SU(1,1)$; an outline of it figures in a previous article[7] treating only the zero spin case (the simplest) for both little groups.

Since the Lorentz group is noncompact, the operators to be diagonalized are unbounded and their eigenvectors are outside the Hilbert space of the representation. To solve the reduction problem, one must introduce a Guelfand triplet of test functions and distributions. Then, provided the corresponding spaces are nuclear (and in our case they are), one can effect the reduction in the form of completeness relations defined by a measure on the spectrum of the diagonalized operators.

In Sec. I the problem is stated and the different notations introduced; a discrete parameter $\chi = \pm 1$ distinguishes the two little groups, and the expression of U^S, the UIR to be reduced, is given by means of the Wigner operator τ_p. In Sec. IB we put in evidence some properties of the Lorentz Lie Algebra $\mathfrak{l}$ and its enveloping algebra $E(\mathfrak{l})$, such as ladder operators for the eigenvalues of the angular momentum, and the discrete and continuous parameters of the spectra of Casimirs of $\mathfrak{l}$. The reduction of U^S on the maximal compact subgroup $SU(2)$ is effected in Sec. II and in Sec. III the relevant infinitesimal operators are given. In Sec. IV we prove that the maximal differential domain of U^S is a nuclear space.

The discrete reduction of U^S is effected in Sec. V, by using the properties established in Sec. IB; in Sec. VI the diagonalization is effected along the continuous spectrum; this step is carried out by introducing derivations of noninteger order for some series of UIR's. The completeness relations are established in Sec. VIII. In Sec. IX the principal results are outlined and remarks on open or similar problems are made. In the Appendix are given the explicit expression as well as the renormalization of the eigenfunctions of the Casimirs.

I. STATEMENT OF THE PROBLEM AND DIFFERENT NOTATIONS

A. Parametrization of the UIR's of $\overline{\mathcal{P}}$

Timelike and spacelike UIR's of the universal covering group $\overline{\mathcal{P}}$ of the Poincaré group $\mathcal{P}$ are well known.[1] They are induced representations, characterized by an orbit and a UIR of the little group. The orbit is, for timelike representations, an upper or lower sheet of hyperboloid in the momentum space, denoted Ω^+_M or Ω^-_M; for spacelike ones, the orbit is an one-sheet hyperboloid, Ω_{iM}. The little groups are $SU(2)$ and $SU(1,1)$, respectively.

To obtain a unifying formalism for the two cases, one must have a convenient parametrization for the orbits. We shall thus put, for $(p_0, \mathbf{p}) \in \Omega^{\epsilon}_{\sqrt{\chi}M}$ $(M > 0)$:

$$\begin{aligned} p_1 &= \tfrac{1}{2}\epsilon M(e^x + \chi e^{-x}), \\ p_1 + ip_2 &= \tfrac{1}{2}\epsilon M(e^x - \chi e^{-x})\sin\varphi\, e^{i\theta}, \\ p_3 &= \tfrac{1}{2}\epsilon M(e^x - \chi e^{-x})\cos\varphi, \end{aligned} \tag{I.1}$$

where the polar angles $\varphi \in]0, \pi[$ and $\theta \in]0, 2\pi[$ parametrize the unit sphere S_2. The discrete parameter χ takes the values $\chi = +1$ in the $SU(2)$ case, in which x is a real positive number and $\epsilon = \pm 1$ (the choice of ϵ distinguishing the two sheets $\Omega^{\pm}_M$), and $\chi = -1$ in the $SU(1,1)$ case, where x is a real number and ϵ can be taken equal to 1. We shall write $x \in \mathbf{R}_\chi$ (with $\mathbf{R}_{+1} = \mathbf{R}^+ : \mathbf{R}_{-1} = \mathbf{R}$) for convenience.

Remark 1: The above parametrization covers an open dense set of each of the Ω's; this is sufficient, since we shall deal with square integrable functions over the orbits.

Remark 2: In the case $\chi = 1$, the apex of $\Omega^{\pm}_M$ becomes a singular point of the parametrizing space $\mathbf{R}^+ \times S_2$; therefore, one expects to find technical difficulties in the form of boundary conditions.

To express the invariant measure on an orbit, we introduce the functions

$$\rho(p) = M^{-1/2}(p_1^2 + p_2^2 + p_3^2)^{1/2} = \tfrac{1}{2}(e^x - \chi e^{-x}), \quad \text{(I. 2)}$$

$$\sigma(p) = \epsilon M^{-1} p_0 = \tfrac{1}{2}(e^x + \chi e^{-x}), \quad \text{(I. 2')}$$

and we find, modulo a multiplicative constant:

$$d\mu(p) = \rho^2\, dx(4\pi)^{-1} \sin\varphi\, d\varphi\, d\theta = \rho^2 dx du, \quad \text{(I. 3)}$$

where du denotes the invariant measure on the sphere $S_2 = U(1)\backslash SU(2)$.

Now, let S be the inducing UIR of the little group and $\mathcal{H}^S$ its carrier (Hilbert) space; let $p \to \tau_p$ be a section which assigns to every p the element of $SL(2,\mathbf{C})$ defined by

$$\tau_p = \begin{pmatrix} \cos(\varphi/2)e^{(x+i\theta)/2} & \sin(\varphi/2)e^{(x-i\theta)/2} \\ -\sin(\varphi/2)e^{(-x+i\theta)/2} & \cos(\varphi/2)e^{(-x-i\theta)/2} \end{pmatrix} = \begin{pmatrix} e^{x/2} & 0 \\ 0 & e^{-x/2} \end{pmatrix} \cdot u(\varphi,\theta), \quad \text{(I. 4)}$$

p being given by (I. 1). Then, every element $\Lambda \in SL(2,\mathbf{C})$ can be written $\Lambda = \gamma \cdot \tau_p$ for some p and some γ which belongs to the little group, the decomposition being unique except for a null measured set.

The UIR of $\overline{\mathcal{P}}$ induced by S can then be written in a canonical way, for $(a,\Lambda) \in \overline{\mathcal{P}}$:

$$U^S(a,\Lambda)f(p) = \rho(p)\cdot[\rho(p\cdot\Lambda)]^{-1} \exp(ip_\mu a^\mu)\cdot S(\tau_p \Lambda \tau_{p\cdot\Lambda})\cdot f(p\cdot\Lambda), \quad \text{(I. 5)}$$

where f is a square-integrable function on $\mathbf{R}_x \times S_2$, with values in $\mathcal{H}^S$; the Hilbert space of the representation being $\mathcal{H} = \mathcal{H}^S \otimes L^2(\mathbf{R}_x \times S_2; dxdu)$. One obtains the restriction of U^S on $SL(2,\mathbf{C})$ by putting $a = 0$ in (I. 5).

To close this survey, we shall briefly list the inducing UIR's, by giving the expression of the Lie algebra operators of the little group G_χ (standing for $SU(2)$ or $SU(1,1)$ according to values of χ).

The Lie Algebra $\mathfrak{I}_\chi$ has three generators, J_1^χ, J_2^χ, J_3, verifying the commutation relations

$$[J_1^\chi, J_2^\chi] = \chi J_3;\ [J_2^\chi, J_3] = J_1^\chi;\ [J_3^\chi, J_2^\chi] = J_2^\chi, \quad \text{(I. 6)}$$

and one Casimir operator, $C^\chi = J_3 + \chi[(J_1^\chi)^2 + (J_2^\chi)^2]$.

TABLE I.

χ	Series	Range s	Range n
1	A_+	$1, 2, 3, \ldots$	$-s, -s+1, \ldots, s-1, s$
1	B_+	$\frac{1}{2}, \frac{3}{2}, \ldots$	
1	O_+	0	0
-1	A_-	$-\frac{1}{2}+it$; $t \geq 0$	$\ldots, -1, 0, 1, \ldots$
-1	A'_-	$-\frac{1}{2}+t$; $0 < t < \frac{1}{2}$	
-1	B_-	$-\frac{1}{2}+it$; $t > 0$	$\ldots, -\frac{1}{2}, \frac{1}{2}, \ldots$
-1	C'	$-\frac{1}{2}, -1, -\frac{3}{2}, -2, \ldots$	$-s, -s+1, \ldots$
-1	C''		$\ldots, s-1, s$
-1	O_-	0	0

The expression of these operators for a UIR are, on a convenient basis $|s,n\rangle$:

$$\mathbf{C}^\chi\ |s,n\rangle = -s(s+1)\,|s,n\rangle,$$

$$J_3\ |s,n\rangle = in\,|s,n\rangle, \quad \text{(I. 7)}$$

$$(J_1^\chi \pm iJ_2^\chi)\,|s,n\rangle = \pm\tfrac{1}{2}\sqrt{\chi(s\mp n)(s\pm n+1)}\cdot|s,n\pm1\rangle.$$

Corresponding ranges for s and n are given in Table I.

B. On the representations of the Lorentz Lie algebra I; generalization of ladder operators

I is generated by six infinitesimal generators, which verify the commutation relations:

$$[J_i, J_j] = [K_j, K_i] = \epsilon_{ijk} J_k, \quad [J_i, K_j] = \epsilon_{ijk} K_k. \quad \text{(I. 8)}$$

The maximal compact subalgebra, $\mathfrak{su}(2)$, is generated by the J's, its Casimir being $Q = J_1^2 + J_2^2 + J_3^2$.

The Casimir operators of the whole algebra are

$$C_1 = \Sigma\, K_i J_i, \quad C_2 = \Sigma(K_i^2 - J_i^2).$$

Since C_1, C_2, Q, J_3 form a maximal set of commuting operators for the enveloping algebra $E(\mathrm{I})$, every algebraic representation of $E(\mathrm{I})$ can be written on a basis of simultaneous eigenvectors of these operators. Putting in evidence those of Q and J_3, we have

$$Q\ |j,m\rangle = -j(j+1)\,|j,m\rangle,$$

$$J_3\ |j,m\rangle = im\,|j,m\rangle.$$

We want to introduce ladder operators $Q^\pm$ for the eigenvalues of Q, in a similar way as $J_1 \pm iJ_2$ are defined for those of J_3 [see (I. 7)].

The expression of $Q^\pm$ shall be j-dependent, since Q's eigenvalues are quadratic on j. Thus $Q^\pm$ shall not be operators of $E(\mathrm{I})$ on a strict sense; one can however choose them to coincide with such operators, when restricted on a j-dependent subspace.

In fact, let

$$L_j^\pm = (j + \tfrac{1}{2} \pm \tfrac{1}{2})^2 K_3 \mp (j + \tfrac{1}{2} \pm \tfrac{1}{2})(K_1 J_2 - K_2 J_1) + J_3 \cdot C_1. \quad \text{(I. 9)}$$

One can easily check that

$$QL_j^\pm\,|\,j,m\rangle = -(j\pm1)(j\pm1+1)\,L_j^\pm\,|\,j,m\rangle. \quad \text{(I. 10)}$$

Then let $Q^\pm$ be defined by

$$Q^+\,|\,j,m\rangle = (2j+3)^{1/2}(2j+1)^{-1/2}(j+1+m)^{-1/2} \times (j+1-m)^{-1/2}\,L_j^+\,|\,j,m\rangle,$$

$$Q^-\,|\,j,m\rangle = (2j-1)^{1/2}(2j+1)^{-1/2}(j+m)^{-1/2}(j-m)^{-1/2} \times L_j^-\,|\,j,m\rangle, \quad \text{if } j \neq m^2$$

$$Q^-\,|\,j,\pm j\rangle = 0. \quad \text{(I. 11)}$$

The following relations then hold (we drop the non-relevant index m)

$$[C_1, Q^\pm] = [C_2, Q^\pm] = [J_3, Q^\pm] = 0, \quad \text{(I. 12)}$$

$$Q^+Q^-\,|\,j\rangle = [C_1^2 + j^2 C_2 - (j-1)j^2(j+1)]\,|\,j\rangle, \quad \text{(I. 13)}$$

$$Q^-Q^+\,|\,j\rangle = [C_1^2 + (j+1)^2 C_2 - j(j+1)^2(j+2)]\,|\,j\rangle. \quad \text{(I. 13')}$$

Now, if one imposes to the representation to be integrable over the subalgebra $\mathfrak{su}(2)$, $j+m$ and $j-m$ must be positive integers, and $j>0$; let then $j_0=\min j$. Then $Q^-\mid j_0,m\rangle=0$, and (I.13) gives

$$[C_1^2+j_0^2C_2-j_0^2(j_0^2-1)]\mid j_0\rangle=0. \tag{I.14}$$

Considering the invariant subspace generated by $(Q^+)^k\cdot\mid j_0\rangle$, one can define on it a new operator D (resp: D^2), if $j_0\neq 0$ (resp: if $j_0=0$), given by $iD=j_0^{-1}C_1$ (resp. $D^2=C_2+1$). Combining these definitions, one can write

$$C_1=ij_0D,\qquad C_2=j_0^2+D^2-1. \tag{I.15}$$

So that (I.14) holds (keeping in mind that only D^2 is defined for $j_0=0$). Writing $\mid j_0;j\rangle=(Q^+)^{j-j_0}\mid j_0\rangle$, Eqs. (I.13) become in the j_0 subspace:

$$Q^+Q^-\mid j_0;j\rangle=(j^2-j_0^2)(D^2-j^2)\mid j_0;j\rangle, \tag{I.16}$$

$$Q^-Q^+\mid j_0;j\rangle=((j+1)^2-j_0^2)(D^2-(j+1)^2)\mid j_0;j\rangle. \tag{I.16'}$$

Combining these results, one gets the following expression for K_3 in this subspace:

$$\begin{aligned}K_3\mid j_0;j,m\rangle&=\left(\frac{(j+1)^2-m^2}{(2j+1)(2j+3)}\right)^{1/2}\mid j_0;j+1,m\rangle\\&+mj_0(j^2+j)^{-1}D\mid j_0;j,m\rangle\\&+\left(\frac{j^2-m^2}{(2j-1)(2j+1)}\right)^{1/2}(j^2-j_0^2)(D^2-j^2)\mid j_0;j-1,m\rangle.\end{aligned} \tag{I.17}$$

Since $\mathfrak{l}$ is generated by K_3 and $\mathfrak{su}(2)$, D (or D^2 if $j_0=0$) is enough to characterize the representation in the j_0 subspace. To obtain a representation algebraically irreducible, D must be a scalar, and the inverse is also true, except for $D^2=(j_0+k)$ with k nonnegative integer.

We recall that the representations obtained by differentiation of UIR's of $SL(2,\mathbf{C})$ are given by $D=i\xi$, $2j_0$ integer (or $D^2=-\xi^2$, $j_0=0$) with $\xi\in\mathbf{R}$, for the principal series; and by $D=\xi$, $j_0=0$, with $0<\xi<1$ for the supplementary series.[8]

II. REDUCTION ON THE MAXIMAL COMPACT SUBGROUP $SU(2)$

Using (1.5) we see that, if $\Lambda\in SU(2)$, the matrix $\gamma=\tau_{p\cdot\Lambda}\Lambda\tau_p^{-1}$ is a diagonal unitary matrix, regardless to the value of χ; also σ and ρ are left invariant.

But we can impose to S to be diagonal on element of the form $\exp(tJ_3)$, i.e:

$$S\begin{pmatrix}\beta&0\\0&\bar\beta\end{pmatrix}\mid s,n\rangle=\beta^{2n}\mid S,n\rangle \text{ for } \beta\bar\beta=1,\beta\in\mathbf{C}. \tag{II.1}$$

Putting

$$f(p)=\sum_n F_n(x,u)\mid s,n\rangle,$$

we obtain for $\Lambda\in SU(2)$

$$V^S(\Lambda)f(p)=\sum_n\beta^{2n}F_n(x,u')\mid s,n\rangle, \tag{II.2}$$

where

$$u\cdot\Lambda=\beta\cdot u'=\begin{pmatrix}\beta&0\\0&\bar\beta\end{pmatrix}u'.$$

One can decompose F_n in terms of the unitary characters $Y_n^{j\,m}$ of $SU(2)$, since u can also be considered as an element of $SU(2)$. We recall that

$$Y_n^{j\,m}(\beta\cdot u\cdot\beta')=\beta^{2n}\beta'^{2m}Y_n^{j\,m}(u)$$

and one can write

$$F_n(x,u)=F_n(x,\beta\cdot u)=\sum_{j,m,k}\beta^{-2k}F_{k,n}^{j,m}(x)\cdot Y_k^{j\,m}(\beta\cdot u) \tag{II.3}$$

since F_n does not depend upon the choice of a representative in the class $u\in U(1)\backslash SU(2)$. For that reason the function $\beta^{2n}F_n(x,u')=\beta^{2n}F_n(x,\beta^{-1}\cdot u\Lambda)$ must depend on the class of u alone, hence, putting $\Lambda=1$, we obtain $n=k$.

In view of these results, we can identify $\mathcal{H}$ to a subspace of $L^n(\mathbf{R}_\chi\times SU(2);dxdw)$, by putting

$$w=\begin{pmatrix}e^{i\psi/2}&0\\0&e^{-i\psi/2}\end{pmatrix}u$$

and

$$f(p)\to F(x,w)=\sum_n e^{in\psi}F_n(x,u), \tag{II.4}$$

p being parametrized by x and u, and range n being restricted by the representation S.

We can thus write for $\Lambda\in SU(2)$:

$$\begin{aligned}U^S(\Lambda)F(x,w)&=F(x,w\Lambda)\\&=\sum_{j,m,n}\left(\sum_k Y^j(\Lambda)_m^k F_n^{j\,k}(x)\right)\cdot Y_n^{j\,m}(w)\end{aligned} \tag{II.5}$$

and we see that, for n,x, and j fixed, $F_n^{j\,m}(x)$ behaves as a $2j+1$ line vector under the action of the UIR of $SU(2)$ characterized by j.

Since $j\pm n$ must be positive integers, there are the following restrictions on range j:

Series $0_\pm, A_\pm, A'_\pm$: $j=0,1,\dots$,

Series $B_\pm$: $j=\frac12,\frac32,\dots$,

Series C', C'' : $j=-s,-s+1,\dots$,

and $\mathcal{H}$ decomposes to the infinite direct sum (in the Hilbert space sense)

$$\mathcal{H}=\bigoplus_{j\pm n\in\mathbf{N}}(H_n^j(x)\otimes\mathbf{C}^{2j+1}), \tag{II.6}$$

with

$$F_n^{j\,m}(x)\in H_n^j(x)=L^2(\mathbf{R}_\chi;dx).$$

The infinitesimal operators of $SU(2)$ are given by (I.7) (with $\chi=1$) by changing s to j and n to m.

We shall also give their differential expression on functions of the form $e^{in\psi}\sum_{j,m}F^{j\,m}Y_n^{j\,m}(u)$, where $u\in U(1)\backslash SU(2);u(\varphi,\theta)$ is given by (I.4), and one has

$$\begin{aligned}J_\pm&=-\frac{1}{\sqrt2}(J_1\mp iJ_2)=e^{\pm i\theta}\left(\mp i\frac{\partial}{\partial\varphi}+\cot\varphi\frac{\partial}{\partial\theta}-in(\sin\varphi)^{-1}\right),\\J_3&=\frac{\partial}{\partial\theta},\\Q&=\frac{\partial^2}{\partial\varphi^2}+\cot\varphi\frac{\partial}{\partial\varphi}+(\sin\varphi)^{-2}\left(\frac{\partial^2}{\partial\theta^2}-2ni\cos\varphi\frac{\partial}{\partial\theta}-n^2\right).\end{aligned} \tag{II.7}$$

Using these relations, writing $in=\partial/\partial\psi$, and choosing convenient phase factors for $Y_n^{j\,m}$ one can establish the following relations:

$$\cos\varphi\ Y_n^{j,m} = \alpha(j,m,n)Y_n^{j-1,m} + [mn/(j^2+j)]Y_n^{j,m} + \alpha(j+1,m,n)Y_n^{j+1,m}, \quad \text{(II. 8)}$$

$$[\cos\varphi + \sin\varphi(\partial/\partial\varphi)]Y_n^{j,m} = -j\cdot\alpha(j,m,n)Y_n^{j-1,m} + (j+1)\alpha(j+1,m,n)Y_n^{j+1,m}, \quad \text{(II. 9)}$$

$$e^{\pm i\psi}\sin\varphi\ Y_n^{j,m} = \mp\sqrt{(j-1\mp n)/(j\pm n)}\ \alpha(j,m,n)Y_{n\pm1}^{j-1,m} - [sn/j(j+1)]\sqrt{(j\mp n)(j\pm n+1)} \times Y_{n\pm1}^{j,m} \pm \sqrt{(j+2\pm n)/(j+1\mp n)} \times \alpha(j+1,m,n)Y_{n\pm1}^{j+1,m}, \quad \text{(II. 10)}$$

where

$$\alpha(j,m,n) = \sqrt{(j^2-m^2)(j^2-n^2)}/j\sqrt{4j^2-1}.$$

III. THE INFINITESIMAL GENERATORS

The action of noncompact one-parameter groups on the orbits being quite complicated, we shall make use of the infinitesimal generators' expressions.

Rotations are given by (II. 7) and translations by $P_\mu = ip_\mu$, defined by (I. 1). We shall put in evidence only K_3 among boosts.

Straightforward calculus gives for K_3:

$$K_3 \sum_{j,m,n} F(x)_n^{jm}\ Y_n^{jm}(w) = \sum_{j,m,n} \frac{d}{dx} F(x)_n^{jm} \cos\varphi\, Y_n^{jm}(w) - \sum_{j,m,n} \sigma\rho^{-1} F(x)_n^{jm}(\cos\varphi + \sin\varphi\ \partial/\partial\varphi)Y_n^{jm}(w) + \sum_{j,m,n} \rho^{-1}F(x)_n^{jm}\sin\varphi\ dS \begin{bmatrix} 0 & \chi\frac{1}{2}e^{i\psi} \\ -\frac{1}{2}e^{-i\psi} & 0 \end{bmatrix}_n^k Y_k^{jm}(w), \quad \text{(III. 1)}$$

where σ,ρ are defined in (I. 2) and $w = e^{1/2\ i\psi\cdot u} \subset SU(2)$. Writing

$$\begin{pmatrix} 0 & \chi\frac{1}{2}e^{i\psi} \\ -\frac{1}{2}e^{-i\psi} & 0 \end{pmatrix} = e^{i\psi/2\cdot J_1^\chi\cdot e^{-i\psi/2}}$$

and using (I. 7) and (II. 8)–(II. 10), one obtains

$$K_3 \sum_{j,m,n} F_n^{jm}Y_n^{jm} = \sum_{j,m,n}\alpha(j,m,n)\left[\left(\frac{d}{dx}+j\sigma\rho^{-1}\right)F_n^{jm} - \tfrac{1}{2}\rho^{-1}\sum_\pm\left(\frac{(s\mp n)(s\pm n+1)(j\pm n+1)}{\chi(j\mp n)}\right)^{1/2}F_{n\pm1}^{j,m}\right]Y_n^{j-1,m} + \sum_{j,m,n}\frac{m}{j(j+1)}\left[n\frac{d}{dx}F_n^{jm} + \tfrac{1}{2}\rho^{-1}\sum_\pm^{\pm}\sqrt{\chi(s\mp n)(s\pm n+1)(j\mp n)(j\pm n+1)}\ F_{n\pm1}^{j,m}\right]Y_n^{j,m} + \sum_{j,m,n}\alpha(j+1,m,n)\left[\left(\frac{d}{dx}-(j+1)\sigma\rho^{-1}\right)F_n^{jm} + \tfrac{1}{2}\rho^{-1}\sum_\pm\left(\frac{(s\pm n)(s\pm n+1)(j\mp n)}{\chi(j\pm n+1)}\right)^{1/2}F_{n\pm1}^{j,m}\right]Y_n^{j+1,m} \quad \text{(III. 2)}$$

From this expression, we can get the ones of the Casimir and ladder operators according to results of Sec. I. B. One obtains, dropping the index m:

$$(Q^-F^j)_n = \sqrt{j^2-n^2}\left(\frac{d}{dx}+j\sigma\rho^{-1}\right)F_n^j - \tfrac{1}{2}\rho^{-1}\sum_\pm\sqrt{\chi(s\mp n)(s\pm n+1)(j\pm n)(j\pm n+1)} \times F_{n\pm1}^j \in H_n^{j-1}(x), \quad \text{(III. 3)}$$

$$(Q^+F^j)_n = \sqrt{(j+1)^2-n^2}\left(\frac{d}{dx}-(j+1)\sigma\rho^{-1}\right)F_n^j + \tfrac{1}{2}\rho^{-1}\sum_\pm\sqrt{\chi(s\mp n)(s\pm n+1)(j\mp n)(j\mp n+1)} \times F_{n\pm1}^j \in H_n^{j+1}(x), \quad \text{(III. 4)}$$

$$(-iC_1F^j)_n = n\frac{d}{dx}F_n^j + \tfrac{1}{2}\rho^{-1}\sum_\pm\pm\sqrt{\chi(s\mp n)(s\pm n+1)(j\mp n)(j\pm n+1)} \times F_{n\pm1}^j \in H_n^j(x), \quad \text{(III. 5)}$$

TABLE II.

Series	Range j_0
$O_\pm$, A_+	$0,1,\ldots,s$
B_+	$\frac{1}{2},\frac{3}{2},\ldots,s$
A_-, A'_-	$0,1,\ldots$
B_-	$\frac{1}{2},\frac{3}{2},\ldots$
C', C''	$-s,-s+1,\ldots$

$$(C_2F^j)_n = \left[\frac{d^2}{dx^2} - \chi(j^2+j+s^2+s-2n^2)\rho^{-2}+n^2-1\right] \times F_n^j + \sigma\rho^{-2}\sum_\pm\sqrt{\chi(s\mp n)(s\pm n+1)(j\mp n)(j\pm n+1)} \times F_{n\pm1}^j \in H_n^j(x). \quad \text{(III. 6)}$$

Calling H the direct sum $\oplus_{j,n} H_n^j(x)$, we see that it splits into two complementary subspaces, $H = H_+ \oplus H_-$ defined by $H_\pm = \{FsF_n^j = \pm F_{-n}^j\}$ (except for series C' and C'', in which n takes only positive or only negative values). It is immediate to check that $H_\pm$ are conserved by $Q^\pm, C_2$ and interverted by C_1.

Calling H^j the direct sum $\oplus H_n^j(x)$, we see that Q^- has nonzero kernel on every H^j if range n is unbounded (i.e., $\chi = -1$, except for the O-representation); Q^- has nonzero kernel on H^j for $j \leqslant s$ and it has kernel zero on H^j for $j > s$, if range n is bounded (i.e., $\chi = 1$ and the O-representation).

We put $\mathcal{K}^j = \mathrm{Ker}\,Q_- \cap H^j$ and $\mathcal{K}_\pm^j = \mathcal{K}^j \cap H_\pm$. We remark that if $\mathcal{K}^j$ is nonzero, it contains complex-valued functions in case $j = 0$ or for series C', C'' and two-component- complex-valued functions in other cases (hence the splitting in $\mathcal{K}_\pm^j$).

Following the pattern of Sec. I. B, we see that the subspace K^j of H generated by the action of Q^+ on $\mathcal{K}^j$ is invariant. Mere addition of components proves

$$H_j^j = \bigoplus_{j_0\leqslant j}(Q^+)^{j-j_0}\mathcal{K}^{j_0}, \text{ hence } H = \bigoplus_{j_0} K^{j_0}.$$

Range j_0 is the same as range $|n|$ and is given in Table II.

However, before carrying on the reduction, we shall deal with the maximal domain of definition of the infinitesimal operators, the calculation up to now being just a formal one.

IV. THE MAXIMAL DIFFERENTIAL DOMAIN. NUCLEARITY

Let $\mathcal{E}$ be the maximal subspace of $\mathcal{K}$, on which every polynomial expression over the generators of the Poincaré Lie algebra $\mathfrak{r}$ is defined. $\mathcal{E}$ can also be considered as the common domain of definition of all powers of the Nelson–Laplace operator $\Delta = -\Sigma K_i^2 - \Sigma J_i^2 - \Sigma P_\mu^2$.[9]

Vectors of $\mathcal{E}$ must be, first of all, C^∞ functions of the spherical and noncompact variables. Effecting the Fourier transform over the compact variables we have further conditions on the scalar functions $F_n^{jm}(x)$ (besides the fact that summation over j,m,n, of their norms must give a finite sum): they are required by the expressions of translations and boosts.

Translations impose faster decrease than any power of $\cosh x$, for infinite x. Boosts impose the same conditions on the derivatives, and these are the only conditions at infinity for the coordinate functions, since the expressions $\sigma\rho^{-1}$ and ρ^{-1} which enter in (III. 3) to (III. 6) are bounded for, say, $|x| \geq 1$. These conditions can be summarized:

$$\lim_{|x|\to\infty} (\cosh x)^k (d/dx)^q F_n^{jm}(x) = 0 \quad \text{for } k,q \in \mathbf{N}. \tag{IV.1}$$

In case $\chi = -1$ these are the only conditions on F_n^{jm}; but for $\chi = +1$ there are boundary conditions for $x = 0$, since $\rho = \sinh x$ and quantities like ρ^{-1} and $\sigma\rho^{-1}$ are unbounded in a neighborhood of the origin. It is a quite tedious job to give a global expression of these boundary conditions under the actual expression of vectors: therefore they shall be established in the expression one gets after effecting the j_0 splitting. However, we can already remark that $L^2(\mathbf{R}^+)$, in which lies F_n^{jm}, is canonically isomorphic to the subspace of even or odd functions (indifferently) of $L^2(\mathbf{R})$. The domain of definition of F_n^{jn} can be extended to the whole real line in a compatible way with the expression of Q_+, Q_-, C_1 (which change parity for $\pm x$) and C_2 (which keeps it). A convenient extension is thus

$$F_n^{j,m}(-x) \pm F_n^{j,m}(-x) = (-1)^{s-j}(\pm P(s))(F_n^{j,m}(x) \pm F_n^{j,m}(+x)),$$

the "initial parity" $P(s) = \pm 1$ to be fixed by the boundary conditions. Expecting them to be of the form "all even (or odd) and a finite number of odd (or even) derivatives vanish at the origin," which shall be proved later, one can identify the space $\mathcal{E}_n^j(x)$ in which belongs F_n^{jm} to a closed infinite dimensional subspace of the space $\mathcal{E}(x)$ defined by (IV. 1).

The question which now arises is whether $\mathcal{E}$ is a nuclear space. We recall that a subspace $\mathcal{E}$ of a Hilbert space is nuclear if it is the projective limit (intersection) of a decreasing sequence of Hilbert spaces H_k, such that the canonical imbedding of H_k in $H_{k'}$ $(k > k')$ is, for any k,k', a Hilbert–Schmidt operator.

In the present case

$$\mathcal{E} = \bigcap_{k=0}^{\infty} \mathrm{Dom}(\Delta^k).$$

Since $\Delta + 1$ is a strictly positive, essentially self-adjoint operator, it has negative powers which are bounded self-adjoint operators. If one of them is of Hilbert–Schmidt class, say $(1+\Delta)^{-k}$, one can construct the following decreasing sequence, of which $\mathcal{E}$ is the projective limit:

$$H_k = \{\phi \in \mathcal{K}; (\phi \mid (1+\Delta)^{2k} \phi) < \infty\}.$$

The existence of a traceful negative power is thus a criterion for nuclearity. In fact we shall prove it for a negative power of an e.s.a. operator Δ_1, majorated by $\Delta + 1$.

Let $|j,m,n,k\rangle$ be a Hilbert basis of $\mathcal{K}$, the new label k denoting eigenfunctions of the harmonic oscillator $-d^2 + x^2$, corresponding to the eigenvalue $2k+1$ $(k = 0, 1, \ldots)$. For fixed j,m,n, these vectors are also a basis of $\mathcal{E}(x)$, and they form a convenient basis also for $\chi = 1$, since an operator in $\mathcal{E}(x)$ is traceful independently of the basis chosen, and if it is traceful in $\mathcal{E}(x)$ it also has a trace in its subspace $\mathcal{E}_n^j(x)$.

It is easy to establish the inequality between positive operators: $\Delta + 1 \geq \Delta_1 = j^2 + x^2 - d^2$. One obtains then

$$\langle j,m,n,k \mid \Delta_1^{-1} \mid j,m,n,k\rangle = (j^2 + 2k + 1)^{-1}$$

and

$$\mathrm{Tr}(\Delta_1^{-4}) = \sum_j \sum_k \sum_{m,n} (j^2 + 2k + 1)^{-4} \leq \sum_{j,k} (2j+1)^2 (j^2 + 2k + 1)^{-4} < \infty.$$

Since $(\Delta + 1)^{-4} < \Delta_1^{-4}$, $(\Delta + 1)^{-4}$ is of Hilbert–Schmidt class and $\mathcal{E}$ is a nuclear space.

Remark: The existence of a trace is due to the presence of translations; if one drops the term $p_0^2 + \mathbf{p}^2$, considering thus the representation of $\bar{\mathfrak{L}}$ obtained by restriction of V^s, the corresponding differential domain is not nuclear. This fact is not surprising, since this representation is highly reducible.

The space $\mathcal{E}$ and its antidual $\mathcal{E}'$ are reflexive Montel spaces. The nuclear spectral theorem applies to the triplet $\mathcal{E} \subset \mathcal{K} \subset \mathcal{E}'$; the Casimir operators C_1 and C_2 have a complete system of generalized eigenvectors in $\mathcal{E}$[1]. There exists a spectral measure for the decomposition of the scalar product of vectors of $\mathcal{E}$ along the eigenvalues of C_1, C_2; and the restriction of U^S can be expressed as a direct integral of UIR's of $SL(2,\mathbf{C})$ by means of the completeness relations thus obtained.

V. THE DISCRETE REDUCTION

We propose here to define an operator T on $\mathcal{E}$, such that T shall be an isometric mapping from $\mathcal{E}$ to $T\mathcal{E}$, and apply the results of Sec. I. B to the representation of $T^{-1}dU^S T$ thus obtained. We shall proceed in several steps in the definition of T.

To obtain a convenient basis for the reduction, we introduce the operator T_j^{j-k} from H^j to H^{j-k}:

$$T_j^{j-k} F_n^{j,m} = \{(j^2 - n^2) \ldots [(j-k+1)^2 - n^2]\}^{1/2} \rho^k F_n^{j,m}. \tag{V.1}$$

T_j^{j-k} commutes with Q_-, hence $T_j^{j-k}\,\mathcal{K}_j \subset \mathcal{K}_{j-k}$.

Moreover, if $F \in \mathcal{K}_j$, one has

$$D_{j-k}\, T_j^{j-k} F = T_j^{j-k}\, D_j F \text{ for } j > k \tag{V.2}$$

(if $j = k$, the same relation holds for D^2, D being not defined in $\mathcal{K}_0$). According to the value of $\min|n|$, which can be $0, 1$, or $-s$, according to the series, we then define the operator T_{j_0} on $\mathcal{K}_{j_0}$:

Case I ($\min|n| = 0$):

$T_0 = 1$ if $j_0 = 0$, and, if $j_0 \neq 0$:

$$T_{j_0}: F_n^{j_0,m} \in \mathcal{K}_{j_0} \longmapsto \begin{pmatrix} \Phi_0^{j_0,m} \\ \Phi_1^{j_0,m} \end{pmatrix} = \begin{pmatrix} T_{j_0}^0 F_0^{j_0,m} \\ \sqrt{(s^2+s)/2\chi}\ T_{j_0}^1 (F_1^{j_0,m} - F_{-1}^{j_0,m}) \end{pmatrix} = (TF)_{j_0}^{j_0,m}.$$

Case II ($\min|n| = \frac{1}{2}$):

$$T_{j_0}: F_n^{j_0,m} \in \mathcal{K}_{j_0} \longmapsto \begin{pmatrix} \Phi_{1/2}^{j_0,m} \\ \Phi_{-1/2}^{j_0,m} \end{pmatrix} = \begin{pmatrix} T_{j_0}^{1/2}(F_{1/2}^{j_0,m} + F_{-1/2}^{j_0,m}) \\ T_{j_0}^{1/2}(F_{1/2}^{j_0,m} - F_{-1/2}^{j_0,m}) \end{pmatrix} = (TF)_{j_0}^{j_0,m}.$$

Case III ($\min|n| = -s$):

$$T_{j_0}: F_n^{j_0,m} \in \mathcal{K}_{j_0} \longmapsto \Phi_{\pm s}^{j_0,m} = T_{j_0}^{-s} F_{\pm s}^{j_0,m} = (TF)_{j_0}^{j_0,m}.$$

Next, we want to extend T_{j_0} over the whole $\mathcal{E}$. Introducing the notation

$$F = (F_{j_0;n}^{j,m}) = (Q_+^{j-j_0} F_n^{j_0,m})$$

we put in evidence the discrete splitting of $\mathcal{E}$ (and, by closure, H) along values of j_0. Then let T' be the one-to-one mapping

$$Q_+^{j-j_0} F_n^{j_0,m} \xrightarrow{T'} F_n^{j_0,m}$$

(the index m being irrelevant). We define then $T = T_{j_0} T'$. and we shall write

$$TF = \Phi = (\Phi_{j_0;\alpha}^{j,m}), \quad \alpha = 0, 1 \text{ or } \alpha = \pm\tfrac{1}{2} \text{ or } \alpha = \pm s. \tag{V.3}$$

Now we want to introduce a scalar product in $T\mathcal{E}$, such that T will be an isometric mapping, and $T\mathcal{E}$ a pre-Hilbert nuclear space. Let first $F^j \in \mathcal{K}_j$; from (III. 3)–(III. 6) one has

$$[\chi(s \mp n)(s \pm n + 1)(j \pm n)(j \pm n + 1)]^{1/2} F_{n\pm1}^j = (j^2 - n^2)^{1/2} \rho \left(\frac{d}{dx} + j\sigma\rho^{-1} \pm n\sigma\rho^{-1} \pm D_j\right) F_n^j, \tag{V.4}$$

$$\left[\left(\frac{d}{dx} + j\sigma\rho^{-1}\right)^2 - \chi s(s+1)\rho^{-2} - n^2 - 2n\sigma\rho^{-1}D - D^2\right] \times F_n^j = 0. \tag{V.4'}$$

Using these relations, and after a somewhat tedious calculation, one finds

$$\begin{aligned} &\tfrac{1}{2} j(2j-1)\chi(s+j)(s-j+1) \sum_n \| F_n^j \|^2 \\ &= \sum_n (j^2 - n^2)[(j-1)^2 \| \rho F_n^j \|^2 + \tfrac{1}{2} \| \rho D F_n^j \|^2 \\ &\quad - \tfrac{1}{2} \langle \rho D^2 F_n^j \mid \rho F_n^j \rangle] \\ &= (j-1)^2 \sum_n \| T_j^{j-1} F_n^j \|^2 + \tfrac{1}{2} \sum_n \| D T_j^{j-1} F_n^j \|^2 \\ &\quad - \tfrac{1}{2} \sum_n \langle D^2 T_j^{j-1} F_n^j \mid T_j^{j-1} F_n^j \rangle. \end{aligned}$$

Having in mind that only D^2 is defined on $\mathcal{K}_0$, one finds by carrying on the induction and using $D^* = -D$,

$$\begin{aligned} &\sum_n \| F_n^j \|^2 \\ &= c(j,s,\chi) \sum_\alpha \int [((j-1)^2 - D^2) \ldots (\alpha^2 - D^2) \Phi_\alpha] \cdot \Phi_\alpha \, dx \end{aligned}$$

with the following notations:

Case I (for $j \neq 0$), $\alpha = 0, 1$:

$$c(j,s,1) = 2^{2j-1}(2j)!^{-1}(s+j)!^{-1}(s-j)!,$$

$$c(j,s,-1) = 2^{2j-1}(2j)!^{-1}(s+j)!^{-1}(j-s-1)!^{-1}s!(-s-1)!.$$

Case II ($\alpha = \pm\frac{1}{2}$):

$$c(j,s,1) = 2^{2j-1}(2j)!^{-1}(s+j!)^{-1}(s-j)! \cdot (s+\tfrac{1}{2}),$$

$$c(j,s,-1) = 2^{2j-1}(2j)!^{-1}(s+j!)^{-1}(j-s-1!)^{-1} \times (s+\tfrac{1}{2})!(-s-\tfrac{1}{2})!.$$

Case III: $c(j,s,-1) = 4^{j+s}(2j)!\,(s+j)!^{-1}(j-s-1)!^{-1} \times (-2s)!(-2s-1)!,$

$\alpha = s$, or $\alpha = -s$ (no summation).

Next, we consider $Q_+^{j-j_0}\mathcal{K}_{j_0}$. Since $Q_+ + Q_-^* = 0$, using (I. 16) one obtains

$$\begin{aligned} &\sum_n \| Q_+^{j-j_0} F_n^{j_0} \|^2 \\ &= c'(j,j_0) \sum_n \int [(j^2 - D^2) \ldots ((j_0+1)^2 - D^2) F_n^{j_0}] F_n^{j_0} \, dx, \end{aligned} \tag{V.5}$$

with

$$c'(j,j_0) = (2j_0)!^{-1}(j_0+j)!(j-j_0)!.$$

We define now the scalar product in $T\mathcal{E}$ by

$$\langle \phi \mid \psi \rangle = \sum_{j_0,j,m,\alpha} \int \overline{\psi_{j_0,\alpha}^{j,m}}(x) \cdot E(s,j_0,j,\alpha^2;D_s^2)\, \phi_{j_0,\alpha}^{j,m}(x)\, dx, \tag{V.6a}$$

where E is the positive e.s.a. operator:

$$E = c(j_0,s,\chi)c'(j,j_0)((j^2 - D_s^2)\ldots((j_0+1)^2 - D_s^2) \times ((j_0-1)^2 - D_s^2)\ldots(|\alpha|^2 - D_s^2)), \tag{V.6b}$$

and $D_s = TDT^{-1}$. T is thus defined as an isometric operator. Let $\overline{T}$ be its closure on the completed space $(\overline{T\mathcal{E}})$; the representation $\overline{T}U^s\overline{T}^{-1}$ is unitarily equivalent to U^s, and the new expression of the Casimir operators depends only on s (and j_0). One obtains, for Cases I, II, III:

$$D_s \cdot \begin{pmatrix} \Phi_0 \\ \Phi_1 \end{pmatrix} = \begin{pmatrix} \Phi_1 \\ (d^2/dx^2 - \chi s(s+1)\rho^{-2})\,\Phi_0 \end{pmatrix} \quad \text{if } j_0 \neq 0 \tag{V.7a}$$

and, if $j_0 = 0$, $D_s^2 = (d^2/dx^2) - \chi s(s+1)\rho^{-2}$,

$$(D_s\Phi)_{\pm1/2} = [(d/dx) \pm \sqrt{\chi(s+\tfrac{1}{2})^2}\rho^{-1}]\Phi_{\pm1/2}, \tag{V.7b}$$

$$D_s\Phi_{\pm s} = \mp (d/dx)\,\Phi_{\pm s}. \tag{V.7c}$$

The boundary conditions for $\chi = 1$ are now easy to establish for vectors of $T\mathcal{E}$. In Case I one must have, for $x \to 0$, $\Phi_\alpha \sim \rho^{s+1}$, or $\Phi_\alpha \sim \rho^{-s}$; this last solution is not acceptable, since it gives no square integrable functions (for $s = 0$ it is $Q_+\Phi_0$ which is not square-integrable). In Case II one must have $\Phi_{1/2} \sim \rho^{s+1/2}$, hence $\Phi_{-1/2} \sim \rho^{s+3/2}$. The boundary conditions can thus be written in the form

$$\Phi_\alpha(x) = \rho^{s+k(\alpha)}\,\varphi_\alpha(\sigma), \tag{V.8}$$

where φ_α is a C^∞ function of $\sigma = \cosh x$ for $1 \leqslant \sigma < \infty$ and

$$k(0) = k(1) = 1; k(\tfrac{1}{2}) = \tfrac{1}{2}; k(-\tfrac{1}{2}) = \tfrac{3}{2}.$$

iD_s, as defined in (V.7) is an e.s.a. operator; we shall put $T\mathcal{E}_s = \bigoplus_{j_0,j,m} T\mathcal{E}_s(x)$ its domain.

VI. DIAGONALIZATION OF D_S AND D_S^2

Rather than find an explicit expression for the generalized eigenvectors (which lie in $T\mathcal{E}'$), our purpose is to find in each case an operator K_s, such that

$$D_s^2 K_s = K_s(d^2/dx^2). \tag{VI.1}$$

In series $C', C'', O, K_s = 1$; the eigenvectors of D_s (for series C', C'') are $V_\lambda = e^{i\lambda x}$, each eigenvalue $i\lambda$ having multiplicity one; those of D_s^2 (for series O_-) are $V_\lambda^{(1)} = i\sin\lambda x$ and $V_\lambda^{(2)} = \cos\lambda x$ ($\lambda \geqslant 0$), each eigenvalue $-\lambda^2$ occurring twice. For what remains, we shall distinguish $\chi = 1$ and $\chi = -1$.

A. $\chi = 1$

We consider explicitly the values $s = 0$ and $s = \frac{1}{2}$, and write again the original expressions of Q_+ and D:

$$s = 0: (Q_+F)^j = j\left(\frac{d}{dx} - j\sigma\rho^{-1}\right)F^{j-1} = j\cdot\rho^{j+1}\frac{d}{d\sigma}\,\rho^{-j}F^{j-1},$$

$$(D^2F)^j = \left[\frac{d^2}{dx^2} - j(j+1)\rho^{-2}\right]F^j,$$

hence

$$\left[\frac{d^2}{dx^2} - j(j+1)\rho^{-2}\right]\cdot\rho^{j+1}\left(\frac{d}{d\sigma}\right)^j\rho^{-1} = \rho^{j+1}\left(\frac{d}{d\sigma}\right)^j\rho^{-1}\frac{d^2}{dx^2}, \tag{VI.2}$$

since Q_+ and D^2 commute.

For $s = \frac{1}{2}$ we have

$$(Q_+F)^j_{\pm 1/2} = (j^2 - \tfrac{1}{4})^{1/2}\left[\left(\frac{d}{dx} - j\sigma\rho^{-1}\right)F^{j-1}_{\pm 1/2} + \tfrac{1}{2}\rho^{-1}F^{j-1}_{\mp 1/2}\right],$$

$$(DF)^j_{\pm 1/2} = \pm\frac{d}{dx}\,F^j_{\pm 1/2} \mp (j+\tfrac{1}{2})\rho^{-1}F^j_{\mp 1/2}.$$

Writing that D and Q_+ commute and effecting symmetric and skew-symmetric combinations, we obtain

$$\left(\frac{d}{dx} \pm (j+\tfrac{1}{2})\rho^{-1}\right)\left(\frac{d}{dx} - j\sigma\rho^{-1} \mp \tfrac{1}{2}\rho^{-1}\right) = \left(\frac{d}{dx} - j\sigma\rho^{-1} \pm \tfrac{1}{2}\rho^{-1}\right)\left(\frac{d}{dx} \pm (j-\tfrac{1}{2})\rho^{-1}\right).$$

Since $\frac{1}{2}\rho^{-1} = A^{-1}\,dA/dx$, with $A = (\tanh\frac{1}{2}x)$, induction on j gives

$$\left(\frac{d}{dx} \pm (j+\tfrac{1}{2})\rho^{-1}\right)\cdot A^{\pm 1}\rho^{j+1}\left(\frac{d}{d\sigma}\right)^{j\pm 1/2}\rho^{-1/2}A^{\mp 1} = A^{\mp 1}\rho^{j+1}\left(\frac{d}{d\sigma}\right)^{j+1/2}\rho^{-1/2}A^{\pm 1}\frac{d}{dx}. \tag{VI.3}$$

According to these results, we put

$$K_s = \rho^{s+1}(d/d\sigma)^s\,\rho^{-1} \quad \text{for } j_0 = 0, \tag{VI.4a}$$

$$K_s = \rho^{s+1}(d/d\sigma)^s\begin{pmatrix}\rho^{-1} & 0\\ 0 & d/d\sigma\end{pmatrix} \quad \text{for } j_0 \neq 0, j_0 \text{ integer}, \tag{VI.4b}$$

$$K_s = \rho^{s+1}\begin{pmatrix}A(d/d\sigma)^{s+1/2}A^{-1} & \\ 0 & A^{-1}(d/d\sigma)^{s+1/2}A\end{pmatrix}\rho^{-1/2} \quad \text{for } j_0 \text{ half-integer}. \tag{VI.4c}$$

Relation (VI.1) is true for K_s defined in (VI.4); moreover, we have (for $j_0 \neq 0$):

$$D_sK_s = K_s\begin{pmatrix}0 & d/dx\\ d/dx & 0\end{pmatrix}. \tag{VI.1'}$$

In this expression, it is easy to show the spectrum and the eigenvectors V_λ of D_s (or D_s^2 if $j_0 = 0$).

We shall put

$$V_\lambda = K_s C_\lambda. \tag{VI.5}$$

The boundary conditions (V.8) impose to $(C_\lambda)_\alpha$ to be an even function of x for $\alpha = 1, \frac{1}{2}$ and an odd one for $\alpha = 0, -\frac{1}{2}$.

We obtain thus:

$$j_0 \neq 0: D_sV_\lambda = D_sK_s\begin{pmatrix}i\sin\lambda x\\ \cos\lambda x\end{pmatrix} = i\lambda V_\lambda, \quad -\infty < \lambda < \infty,$$

$$j_0 = 0: D_s^2V_\lambda = D_s^2\cdot i\sin\lambda x = -\lambda^2 V_\lambda, \quad 0 \leqslant \lambda < \infty.$$

Finally, if K_s^* is the adjoint of K_s, the C_λ are eigenvectors of the self-adjoint operator $K_s^*EK_s$ [E being given in (V.6b)]. One can substitute $-D^2$ by λ^2 in the expression of E; as for $(K_s^*K_s)$, substitution in (V.5) for $s = 0$ and $\frac{1}{2}$ gives

$$(K_s^*K_s)^\alpha_\alpha = \left(s^2 - \frac{d^2}{dx^2}\right)\cdots\left((1-|\alpha|)^2 - \frac{d^2}{dx^2}\right). \tag{VI.6}$$

Hence, for integer s

$$K_s^*EK_s\cdot C_\lambda = c(j_0,s,1)\cdot c'(j,j_0)\cdot(j^2+j^2)\ldots\lambda^2\cdot(s^2+\lambda^2)\ldots(1+\lambda^2)\cdot(j_0^2+\lambda^2)^{-1}C_\lambda = \omega(j,j_0,s,\lambda)\cdot C_\lambda,$$

and for half-integer s

$$K_s^*EK_s\cdot C_\lambda = c(j_0,s,1)\,c'(j,j_0)(j^2+\lambda^2)\ldots(\tfrac{1}{4}+\lambda^2)\times(s^2+\lambda^2)\ldots(\tfrac{1}{4}+\lambda^2)(j_0^2+\lambda^2)^{-1}C_\lambda, = \omega(j,j_0,s,\lambda)\,C_\lambda.$$

B. $\chi = -1$

In order to find K_s verifying (VI.1) and (VI.1'), we shall introduce derivations of noninteger order. Recalling that now $\rho = \cosh x$; $\sigma = \sinh x$, we introduce the Fourier transform on the variable σ:

$$\mathfrak{F}\Phi(x) = \int_{-\infty}^{\infty} e^{i\sigma y}\,\Phi(x)d\sigma. \tag{VI.7}$$

Noticing that $\mathcal{E}(x) = \mathcal{S}_o$ (the Schwartz space of rapidly decreasing C^∞ functions of σ), we see that $\mathfrak{F}\cdot\mathcal{E}(x) = \mathcal{S}_y$ and $\mathfrak{F}\mathcal{E}'(x) = S'_y$. Since $d\sigma = \rho dx$, we also obtain

$$\mathfrak{F}^* = \rho\mathfrak{F}^{-1}, \quad \mathfrak{F}^{*-1} = \mathfrak{F}\cdot\rho^{-1}, \qquad \frac{d}{d\sigma} = \rho^{-1}\frac{d}{dx} = -i\mathfrak{F}^{-1}y\mathfrak{F}. \tag{VI.7'}$$

We now introduce the following operators, in analogy to the case $\chi = 1$:

$$K_s^0 = \rho^{s+1}\,\mathcal{F}^{-1}\,|y|^s\,\mathcal{F}\rho^{-1},$$

$$K_s^1 = -\,i\rho^{s+1}\mathcal{F}^{-1}\,|y|^s\cdot y\,\mathcal{F} = K_s^0\,\frac{d}{dx}\,,$$

$$K_s^{\pm} = B(x)^{\pm 1}\rho^{s+1}\,\mathcal{F}^{-1}\,|y|^{s+1/2}\,\mathcal{F}\rho^{-1/2}B(x)^{\mp 1},$$

where $B(x) = \sqrt{\rho}\,(\cosh\tfrac{1}{2}x + i\sinh\tfrac{1}{2}x) = \overline{B(x)}^{-1}$ is a function verifying

$$dB/dx = \tfrac{1}{2}\,i\rho^{-1}B = \tfrac{1}{2}\,i(1+i\sigma)\rho^{-2}\overline{B}\,.$$

By straightforward computation one can establish:

$$\left(\frac{d^2}{dx^2} + s(s+1)\rho^{-2}\right)K_s^{\alpha} = K_s^{\alpha}\,\frac{d^2}{dx^2} \qquad (\alpha = 0, 1), \qquad \text{(VI. 8a)}$$

$$\left(\frac{d}{dx} \pm (s+\tfrac{1}{2})\,i\rho^{-1}\right)K_s^{\pm} = K_s^{\mp}\,\frac{d}{dx}\,. \qquad \text{(VI. 8b)}$$

We can define now K_s satisfying (VI. 1)–(VI. 1′) by

$$K_s = K_s^0 \qquad \text{if } j_0 = 0, \qquad \text{(VI. 9a)}$$

$$K_s = \begin{pmatrix} K^0 & 0 \\ 0 & K_s^1 \end{pmatrix} \qquad \text{if } j_0 \neq 0,\ j_0 \text{ integer}, \qquad \text{(VI. 9b)}$$

$$K_s = \begin{pmatrix} K_s^- & 0 \\ 0 & K_s^+ \end{pmatrix} \qquad \text{if } j_0 \text{ is half-integer}. \qquad \text{(VI. 9c)}$$

These formulas are exactly the same as those of $\chi = +1$; since no boundary conditions exist, we find two eigenvectors for each eigenvalue:

$$j_0 \neq 0: \qquad V_\lambda^{(1)} = K_s\begin{pmatrix} i\sin\lambda x \\ \cos\lambda x \end{pmatrix}, \qquad V_\lambda^{(2)} = K_s\begin{pmatrix} \cos\lambda x \\ i\sin\lambda x \end{pmatrix}, \qquad -\infty < \lambda < \infty,$$

$$j_0 = 0: \qquad V_\lambda^{(1)} = K_s\ \ i\sin\lambda x\,, \qquad V_\lambda^{(2)} = K_s\cos\lambda x, \qquad 0 \leq \lambda < \infty.$$

Up to now these results are formal: we still have to check that K_s is a "good" operator, that is, show that V_λ^η is in $\mathcal{E}'(x)$. That means that the integral

$$\int_{-\infty}^{\infty} K_s^{\alpha}(e^{i\lambda x})\cdot\overline{E\Phi(x)}dx$$

converges for any $\Phi \in \mathcal{E}(x)$, E being defined in (V. 6b).

To prove this, we first notice that E conserves $\mathcal{E}(x)$, so E can be omitted. The same applies to multiplication by ρ^s (any s) and $B^{\pm 1}$. It is then sufficient to show that

$$\left|\int \Phi(x)(\mathcal{F}^{-1}\,|y|^{s+\alpha}\,\mathcal{F}\rho^{\alpha-1}e^{i\lambda x})d\sigma\right| < \infty\,, \qquad \alpha = 0, \tfrac{1}{2}, 1,$$

or, equivalently, for $\psi \in \mathcal{S}y$:

$$\left|\int \psi(y)\cdot|y|^{s+\alpha}(\mathcal{F}\rho^{\alpha-1}e^{i\lambda x})dy\right| < \infty\,.$$

Since $\rho^{\alpha-1}e^{i\lambda x}$ is a bounded C^∞ function (and majorated by ρ^{-1} for $\alpha = 0$), its Fourier transform is in $\mathcal{S}'y$ (and for $\alpha = 0$ it is a fast decreasing, continuous function of y, hence bounded). Therefore the integral converges for $y \to \pm\infty$; for $y \to 0$ it still converges, since $||y|^{s+\alpha}| \leq 1$ (for $\alpha = \tfrac{1}{2}, 1$ and $-1 \leq y \leq 1$), and since for $\alpha = 0, \psi$ and $\mathcal{F}\rho^{-1}e^{i\lambda x}$ are bounded and

$$\int_{-1}^{1} |y|^{\mathrm{Res}}\,dy$$

converges.

These results are still valid when substituting K_s^{*-1} to K_s. One can write $K_s = K_s^{*-1}K_s^*K_s$; since d^2/dx^2 and $K_s^*K_s$ commute, $K_s^{*-1}C_\lambda^{(\eta)}$ is a linear combination of $V_\lambda^{(1)}$ and $V_\lambda^{(2)}$.

In fact we have a stronger result:

Proposition: The self-adjoint operator $K_s^*EK_s$ is diagonal on the set of eigenvectors $\{C_\lambda^{(\eta)}\}$ [with the notation $C_\lambda^{(\eta)} = K_s^{-1}V_\lambda^{(\eta)}$].

Proof: E being a polynomial on D_s^2, we have

$$E_\alpha^\alpha\,V_{\lambda;\,\alpha}^{(\eta)} = E_\alpha^\alpha(-\lambda^2)\,V_{\lambda,\,\alpha}^{(\eta)}$$

and from (V. 6) we see that $E_{1/2}^{1/2}(-\lambda^2) = E_{-1/2}^{-1/2}(-\lambda^2)$; $E_0^0(-\lambda^2) = \lambda^2 E_1^1(-\lambda^2)$.

Now, for j_0 half-integer, we see that $K_s^*K_s = 1$ and the proposition is proved. For j_0 integer we first remark that $K_s^{1*}K_s^1 = -(d/dx)K_s^{0*}K_s^0(d/dx)$, and that $K_s^{0*}K_s^0$ commutes with d^2/dx^2; thus $K_s^{0*}K_s^0e^{i\lambda x}$ is a linear combination of $e^{i\lambda x}$ and $e^{-i\lambda x}$. Moreover, one easily sees that $K_s^{0*}K_s^0$ transforms odd functions to odd ones and even to even ones. One can thus write

$$K_s^{0*}K_s^0\ \sin\lambda x = \omega_1(\lambda, s)\cdot\sin\lambda x\,,$$

$$K_s^{0*}K_s^0\ \cos\lambda x = \omega_2(\lambda, s)\cdot\cos\lambda x\,,$$

hence, for j_0 integer:

$$K_s^*EK_s\cdot C_\lambda^{(\eta)} = E_0^0(-\lambda^0)\cdot\omega_\eta(\lambda, s)\,C_\lambda^{(\eta)}\,,$$

ω_η being of course a positive function.

Remark 1: The choice of K_s is of course not unique, as one can see, e.g., by substituting $|y|^s\,\mathrm{sgn}y$ to $|y|^s$. Because of multiplicity two, the different choices for K_s, do not forcingly give the same set of eigenvectors for $K_s^*EK_s$ (and for D_s). However, since it is not our purpose to calculate explicitly the eigenvectors of D_s by means of K_s, we shall not pursue this point. An explicit expression of the eigenvectors figures in the Appendix.

Remark 2: In fact, the above results, including the convergence of integrals, are valid for complex λ.

VII. THE COMPLETENESS RELATIONS

According to the nuclear spectral theorem, the scalar product of vectors of $\mathcal{E}$ decomposes according to the formula

$$\langle\phi\,|\,\psi\rangle = \sum_l \int \langle\phi\,|\,V^l\rangle\,\overline{\langle\psi\,|\,V^l\rangle}\,d\mu\,, \qquad \text{(VII. 1)}$$

where $V^l \in \mathcal{E}'$ are the generalized eigenvectors of C_1 and C_2; $d\mu$ is a positive measure on their spectrum and l denotes symbolically multiplicity for each set of eigenvalues. In the present case l stands for j, m; and, as shown above, for a "pure multiplicity" index $\eta = 1, 2$ in series A_-, A'_-, B_- and O_-.

Our object is to write explicitly (VII. 1); from the results of the discrete reduction we have

$$\langle\phi\,|\,\psi\rangle = \sum_{\substack{j_0\neq 0 \\ |m|\leq j}}\ \sum_{j\geq j_0}\ \sum_{\alpha}\langle\phi_{j_0;\,\alpha}^{j,\,m}\,|\,\psi_{j_0,\,\alpha}^{j,\,m}\rangle + \sum_{\substack{j\geq 0=j_0 \\ |m|\leq j}}\langle\phi_{0;\,0}^{j,\,m}\,|\,\psi_{0;\,0}^{j,\,m}\rangle\,, \qquad \text{(VII. 2)}$$

with the notation of (V. 3) and (V. 6a).

Dealing with complex-valued functions which belong to $\mathcal{S}(x)$, we have the usual Fourier formula

$$(\varphi \mid \psi) = \int_{-\infty}^{\infty} \varphi(x)\overline{\psi(x)}dx = \int_{-\infty}^{\infty} (\varphi \mid e^{i\lambda x})(\overline{\psi \mid e^{i\lambda x}})\, d_\lambda .$$

If φ and ψ are both odd (resp. even), $e^{i\lambda x}$ can be replaced by $i \sin \lambda x$ (resp. $\cos \lambda x$) and the integration domains for x and λ by $[0, \infty]$ (we neglect multiplicative constants).

Treating first the case $\chi = -1$, where the whole real line is the domain, we have for series C', C'': $C_\lambda = V_\lambda = e^{i\lambda x}$ and for series O_-, A_-, A'_-, B_- :

$$j_0 = 0: \begin{cases} e^{i\lambda x} = C_\lambda^{(1)} + C_\lambda^{(2)} & \text{if } \lambda > 0, \\ e^{i\lambda x} = -C_{-\lambda}^{(1)} + C_{-\lambda}^{(2)} & \text{if } \lambda < 0, \end{cases}$$

$$j_0 \neq 0. \begin{cases} \begin{pmatrix} e^{i\lambda x} \\ 0 \end{pmatrix} = \tfrac{1}{2}(C_\lambda^{(1)} + C_\lambda^{(2)} - C_{-\lambda}^{(1)} + C_{-\lambda}^{(2)}), \\ \begin{pmatrix} 0 \\ e^{i\lambda x} \end{pmatrix} = \tfrac{1}{2}(C_\lambda^{(1)} + C_\lambda^{(2)} + C_{-\lambda}^{(1)} - C_{-\lambda}^{(2)}). \end{cases}$$

Combining adequately these relations we see that, if φ, ψ are, according to the case, one- or two-component functions, we obtain (neglecting constant factors)

$$(\varphi \mid \psi) = \sum_\alpha (\varphi_\alpha \mid \psi_\alpha) = \sum_\eta \int_{-\infty}^{+\infty} (\varphi \mid C_\lambda^\eta)(\overline{\psi \mid C_\lambda^\eta})\, d_\lambda \quad \text{if } j_0 \neq 0,$$

$$(\varphi \mid \psi) = \sum_\eta \int_0^\infty (\varphi \mid C_\lambda^\eta)(\overline{\psi \mid C_\lambda^\eta})\, d_\lambda \quad \text{if } j_0 = 0.$$

Putting now

$$\varphi = K^* E \Phi_{j_0}^{j,m}, \qquad \psi = K^{-1} \Psi_{j_0}^{j,m},$$

we obtain

$$\begin{aligned} \langle \Phi_{j_0}^{j,m} \mid \Psi_{j_0}^{j,m} \rangle &= (\varphi \mid \psi) \\ &= \sum_\eta \int (E\Phi_{j_0}^{j,m} \mid KC_\lambda^\eta)(\overline{E\Psi_{j_0}^{j,m} \mid E^{-1}K^{*-1}C_\lambda^\eta})\, d_\lambda \\ &= \sum_\eta \int \langle \Phi_{j_0}^{j,m} \mid V_\lambda^\eta \rangle \langle \overline{\Psi_{j_0}^{j,m} \mid K(K^*EK)^{-1}C_\lambda^\eta} \rangle\, d_\lambda \\ &= \sum_\eta \int \langle \Phi_{j_0}^{j,m} \mid V_\lambda^\eta \rangle \langle \overline{\Psi_{j_0}^{j,m} \mid V_\lambda^\eta} \rangle\, d\mu(\lambda), \end{aligned}$$

with $d\mu(\lambda) = [\omega_\eta(s,\lambda) \cdot E(-\lambda^2)_\alpha^\alpha]^{-1}$, where $\alpha = 0, \frac{1}{2}, \pm s$ according to the case, and $\omega_\eta(s,\lambda) = 1$ for series O_-, B_-, C', C''. This result together with (VII. 2) establishes the completeness relations for $\chi = -1$.

For $\chi = +1$, we first notice the relations

$$j_0 = 0: \quad i \sin \lambda x = C_\lambda ,$$

$$j_0 \neq 0: \begin{cases} \begin{pmatrix} i \sin \lambda x \\ 0 \end{pmatrix} = \tfrac{1}{2}(C_\lambda - C_{-\lambda}), \\ \begin{pmatrix} 0 \\ \cos \lambda x \end{pmatrix} = \tfrac{1}{2}(C_\lambda + C_{-\lambda}). \end{cases}$$

The Fourier inversion on odd functions can thus be written

$$(\varphi \mid \psi) = \int_0^\infty \varphi(x)\overline{\psi(x)}dx = \int_0^\infty (\varphi \mid C_2)(\overline{\psi \mid C_\lambda})\, d_\lambda$$

and on two-component functions such that φ_0 is odd and φ_1 even (again neglecting a factor $\frac{1}{2}$):

$$\begin{aligned} (\varphi \mid \psi) &= \sum_\alpha (\varphi_\alpha \mid \varphi_\alpha) \\ &= \sum_\alpha \int_0^\infty (\varphi \mid C_\lambda - (-1)^\alpha C_{-\lambda})(\overline{\psi \mid C_\lambda - (-1)^\alpha C_{-\lambda}})\, d_\lambda \\ &= \int_{-\infty}^{\infty} (\varphi \mid C_\lambda)(\psi \mid C_\lambda)\, d\lambda . \end{aligned}$$

Now, one easily verifies that $(K^m E \Phi_{j_0}^{j,m})_\alpha$ is odd if $\alpha = 0$ or $-\frac{1}{2}$ and even if $\alpha = 1$ or $\frac{1}{2}$; the same holds of course for $(K^{-1}\Phi_{j_0}^{j,m})_\alpha$ since K^*EK conserves parity. We can thus write[0]

$$\begin{aligned} \langle \phi_{j_0}^{j,m} \mid \psi_{j_0}^{j,m} \rangle &= (K^*E\Phi_{j_0}^{j,m} \mid K^{-1}\Psi_{j_0}^{j,m}) \\ &= \int_{\mathrm{Ror}\ \mathbf{R}_+} (K^*E\Phi_{j_0}^{j,m} \mid C_\lambda)(\overline{K^{-1}\Psi_{j_0}^{j,m} \mid C_\lambda})\, d\lambda \\ &= \int_{\mathrm{Ror}\ \mathbf{R}_+} \langle \phi_{j_0}^{j,m} \mid V_\lambda \rangle \langle \overline{\Psi_{j_0}^{j,m} \mid V_\lambda} \rangle\, d\mu(\lambda), \end{aligned}$$

with

$$d\mu(\lambda) = [\omega(j, j_0, s, \lambda)]^{-1} d\lambda .$$

VIII. RESULTS AND REMARKS

(1) The diagonalization of the Casimir operators and the completeness relations provide the final results of the reduction of U^S into irreducible components. The couple of numbers $(j_0 > 0, \lambda)$ or $(j_0 = 0, \lambda^2)$ and, eventually, the multiplicity index, determine an invariant subspace which can be given a prehilbert structure and completed to a hilbert space (see Appendix). The representation of $\mathfrak{l}$ on this space can be integrated to a representation of the principal series of $SL(2, \mathbf{C})$. Since range $\lambda = \mathbf{R}$ for $j_0 \neq 0$ (and range $\lambda = [0, \infty[$, for $j_0 = 0$), in all cases, the only restriction on the spectrum of UIR's obtained is given by j_0. Table III gives these results:

TABLE III.

Series of S	Range j_0	Multiplicity
A_+, B_+, O_+	$s, s-1, \ldots, \frac{1}{2}$ or 0	1
O_-	0	2
A_-, A'_-	$0, 1, \ldots$	2
B_-	$\frac{1}{2}, \frac{3}{2}, \ldots$	2
C', C''	$-s, -s+1, \ldots$	1

(2) In Ref. 4, Joos obtains the same results for the $SU(2)$ case. Moreover, he gives the "mixed scalar products" between vectors of U^S and those of its irreducible components, which do not figure explicitly here; however the spectral measure calculated in Sec. VII and the normalization of the eigenvectors given in the Appendix, provide equivalent results.

The main difference in the technique used, besides topological considerations which are absent from Joos's article, is that the discrete splitting of U^S is effected here after passing to the Lie algebra. This method allows us to keep a unique formalism along all steps of the reduction, while the one of Ref. 4 does not: it makes use of a finite dimensional representation of $SL(2, \mathbf{C})$ obtained by complexification of S to get the discrete splitting before differentiation; in the $SU(1,1)$ case the complexification gives a local, nonintegrable representation, that is why differentiation precedes the discrete splitting.

(3) It must be pointed out that the unified formalism works very well, in many steps of the reduction, and

especially, in establishing the nuclearity property. It is interesting to notice that the differential domain of the UIR's of $\mathcal{P}$ induced by E_2 is not nuclear,[10] while the one of UIR's induced by semisimple little groups is nuclear. Although this fact is not sufficient to make a conjecture, it is a hint for further research on nuclearity properties of induced UIR's.[11]

(4) The nuclearity property can easily be shown to be valid for some other classes of representations of $\overline{\mathcal{P}}$. If the inducing representation S is a nonunitary, irreducible representation of $SU(1,1)$ (or a finite direct sum of such representations) on a Hilbert space, U^S is nonunitary, but its maximal differential domain is a nuclear space. This is due to the fact that the proof of Sec. IV does not depend on the value of S; and such representations are obtained by allowing $s(s+1)$ to be any complex number. The same remark holds for $SU(2)$, provided one takes adequate topological restrictions (which we do not specify here) since one has to deal with nonintegrable local representations.

(5) The boundary conditions established in Sec. V for $SU(2)$ eliminate one of the two eigenvectors of d^2/dx^2, and the multiplicity of the irreducible components is one. If the test functions Φ belong to a smaller space $\tilde{\mathcal{E}} \subset \mathcal{E}$, such that Φ and all its derivatives vanish at the origin, both eigenvectors of d^2/dx^2 belong to the dual $\tilde{\mathcal{E}}'$. The space $\tilde{\mathcal{E}}$, dense in $\mathcal{H}$ is invariant under the Lie algebra $\mathfrak{r}$, but not under the group action: in fact, the apex of the hyperboloid Ω, which is just an ordinary point, plays a singular role in the definition of $\tilde{\mathcal{E}}$. The local representation $d\tilde{U}$, defined by restriction of $d\tilde{U}$ on $\tilde{\mathcal{E}}$, is not equivalent to dU^S, although its algebraic expression is the same. This example shows the importance of topological considerations when speaking of Lie algebra representations, and the rigor needed to define notions like equivalence or irreducibility: dU^S, which is the differential of an irreducible representation U^S of $\mathcal{P}$ has (at least!) one nontrivial invariant subspace on which another representation is defined by restriction!

(6) Since $\mathcal{P}$ is a contraction of the de Sitter groups, it would be interesting to compare the reduction of UIR's of both groups on $SL(2\mathbf{C})$. In fact, examining our results and those of $SO_0(4,1)$,[12] one sees that spectra and multiplicities are the same for UIR's obtained one from the other by contraction or deformation [taking account of the fact that two UIR's of $\mathcal{P}$ with opposite masses and little group $SU(2)$ are needed to obtain by deformation UIR of $SO_0(4,1)$].

ACKNOWLEDGMENTS

The author is indebted to Professor M. Flato and D. Sternheimer for helpful discussions and suggestions and pertinent criticism.

APPENDIX

1. Explicit expression of the eigenfunctions of D_S, $D_S{}^2$

We first consider the case $j_0 = 0$; the case $j_0 \neq 0$ integer can be trivially deduced from $j_0 = 0$, since the two components of an eigenfunction are proportional with ratio $i\lambda$, and they are solutions of the same differential equation as for $j_0 = 0$.

This equation is

$$((d^2/dx^2) - \chi s(s+1)\rho^{-2} + \lambda^2)\, V^{(\eta)}_{\lambda,s} = 0, \tag{A1}$$

with $V^{(1)}_{\lambda,s} = iK_s \sin\lambda x$; $V^{(2)}_{\lambda,s} = K_s \cos\lambda x$.

We recall that for $\chi = 1$ only $V^{(1)}_{\lambda,s}$ is admitted.

Let $V_{\lambda,s} = \rho^{s+1} f_{\lambda,s}$ be a solution of (A1). Then $f_{\lambda,s}$ is a solution of

$$((\sigma(d/d\sigma) + s + 1)^2 - \chi(d^2/d\sigma^2) + \lambda^2) f_{\lambda,s} = 0. \tag{A1'}$$

This equation admits power series of σ or σ^{-1} for solutions the interval of convergence being $\sigma^2 < 1$ or $\sigma^{-2} < 1$ in each case. The inequalities become large for $\chi = -1$, since the numeric series is alternate in this case.

We consider first the expansion for $\sigma > 1$ (in fact the only relevant one for $\chi = 1$, since in this case $\sigma = \cosh x$); let

$$f_{\lambda,s} = \sum_{n=0}^{\infty} \alpha_n(\lambda, s)\, \sigma^{-s-1-i\lambda-2n}. \tag{A2}$$

The coefficients α_n must verify

$$4n(n + i\lambda)\alpha_n - \chi(2n + s + i\lambda)(2n + s - 1 + i\lambda)\alpha_{n-1} = 0, \tag{A3}$$

hence

$$(s + i\lambda)!\, n!\, (n + i\lambda)!\, \alpha_n = (4\chi)^{-n}(2n + s + i\lambda)!\, i\lambda!\, \alpha_0. \tag{A3'}$$

Identifying $f_{\lambda,0}$ to $\rho^{\pm 1} e^{-i\lambda x}$ one finds $\alpha_0(\lambda, 0) = 2^{-i\lambda}$; we shall put $\alpha_0(\lambda, s) = c.\, 2^{-i\lambda}$, c being a constant which may depend on s.

In the domain $\sigma = \sinh x \leq -1$ (for $\chi = -1$), we define $f_{\lambda,s}$ by

$$f_{\lambda,s}(x) = f_{-\lambda,s}(-x).$$

One can then show that in the domain $\sigma^2 > 1$ one has, up to a constant factor:

$$V^{(1)}_{\lambda,s} = \tfrac{1}{2}\rho^{s+1}(f_{-\lambda,s} - f_{\lambda,s});\quad V^{(2)}_{\lambda,s} = \tfrac{1}{2}\rho^{s+1}(f_{-\lambda,s} + f_{\lambda,s}).$$

Next, we give the expansion in terms of powers of σ, valid for the domain $\sigma^2 = (\sinh x)^2 \leq 1$ (and $\chi = -1$):

$$f^{(\eta)}_{\lambda,s} = \sum_k \beta^{(\eta)}_k(\lambda, s)\, \sigma^k. \tag{A4}$$

The coefficients β_k must verify

$$((k + s + 1)^2 + \lambda^2)\beta_k + (k + 2)(k + 1)\beta_{k+2} = 0. \tag{A5}$$

Imposing that k be odd (even) if η is odd (even), one has, up to a constant factor, on the domain $\sigma^2 \leq 1$, $V^{(\eta)}_{\lambda,s} = \rho^{s+1} f^{(\eta)}_{\lambda,s}$.

The solutions of (A1) for $\chi = 1$ can be put into a polynomial form. The function $\psi_{\lambda,s}$

$$\psi_{\lambda,s}(x) = e^{-i\lambda x} \sum_{k=0}^{s} a_k(\lambda, s)\, (\coth x)^k \tag{A6}$$

is a solution of (A1) if the coefficients a_k verify

$$(s - k)(s + k + 1)a_k - 2i\lambda(k + 1)a_{k+1} + (k + 1)(k + 2)a_{k+2} = 0. \tag{A7}$$

The acceptable solution $V^{(1)}$ is then the antisymmetric combination $\psi_\lambda - \psi_{-\lambda}$ (the initial coefficient, a_s, being the same.

For $\chi = -1$, (A7) is still valid if one replaces $\coth x$ by $\tanh x$; however, one obtains no more a polynomial, but an infinite expansion, since s is no more an integer.

Let us now consider the case j_0 half-integer. One has to solve the equations

$$\left(\frac{d}{dx} \pm \sqrt{\chi}\ (s + \tfrac{1}{2})\rho^{-1}\right) V^{\pm}_{\lambda,s} = i\lambda\ V^{\mp}_{\lambda,s}\ . \qquad \text{(A8)}$$

Let $\mathfrak{A}$ be the 2×2 matrix

$$\mathfrak{A} = \rho^{s+1} \begin{pmatrix} g_\chi^{-1} & -g_\chi^{-1} \\ g_\chi & g_\chi \end{pmatrix},$$

where $g_\chi^2 = (\sigma - \sqrt{\chi})\rho^{-1}$ and $dg_\chi/dx = \frac{1}{2}\sqrt{\chi}\,\rho^{-1} g_\chi$.

Putting $f_{\lambda,s} = \mathfrak{A}^{-1} V_{\lambda,s}$, we obtain two new equations

$$\pm(\sigma(d/d\sigma) + s \pm 1) f^{\pm}_{\lambda,s} \mp \sqrt{\chi}(d/d\sigma) f^{\mp}_{\lambda,s} = i\lambda f^{\pm}_{\lambda,s}. \qquad \text{(A9)}$$

Squaring the differential operator of (A9) and summing, one obtains

$$((\sigma(d/d\sigma) + s + \tfrac{3}{2})^2 - \chi(d^2/d\sigma^2)) f^{\pm}_{\lambda,s} = (i\lambda \pm \tfrac{1}{2})^2 f^{\pm}_{\lambda,s}. \qquad \text{(A10)}$$

One can thus transpose to the functions found for $j_0 = 0$, the formulas being valid for complex λ and s, in general. The four eigenfunctions obtained by (A10) are associated in two pairs by means of (A9); one obtains

$$f_{\lambda,s} = f_{\lambda \mp (1/2)i,\, s+1/2} \quad \text{or} \quad f^{\pm}_{\lambda,s} = f_{-\lambda \pm i\,1/2,\, s+1/2}\ .$$

Using the notations of the previous paragraph one can write:

$$\chi = 1: \quad V^{(1)}_{\lambda,s} = \mathfrak{A} \begin{pmatrix} f_{-\lambda+i/2,\,s+1/2} + f_{\lambda-i/2,\,s+1/2} \\ f_{-\lambda-i/2,\,s+1/2} + f_{\lambda+i/2,\,s+1/2} \end{pmatrix},$$

$$\chi = -1: \quad V^{(\eta)}_{\lambda,s} = \mathfrak{A} \begin{pmatrix} f^{(\eta)}_{-\lambda+i/2,\,s+1/2} - i f^{(\eta')}_{-\lambda+i/2,\,s+1/2} \\ f^{(\eta)}_{-\lambda-i/2,\,s+1/2} - i f^{(\eta')}_{-\lambda-i/2,\,s+1/2} \end{pmatrix},$$

with $\eta, \eta' = 1, 2$ and $\eta \neq \eta'$.

2. Normalization of the Hilbert basis in each irreducible component of U^S

The eigenvectors $V^{(\eta);\,j,m}_{\lambda;\,j_0}$, defined by $\langle \Phi \mid V^{(\eta);\,j,m}_{\lambda;\,j_0} \rangle = \langle d^{j,m}_{j_0} \mid V^{(\eta)}_{\lambda;\,j_0} \rangle$, generate an invariant subspace E^{η}_{λ,j_0} for the extension of dU^S on $\mathcal{E}$, when η, λ, and j_0 are fixed. E^{η}_{λ,j_0} can be given a pre-Hilbert structure and then be completed to a Hilbert space, if one defines a scalar product by means of the orthogonal basis $V^{j,m}_{\lambda,j_0}$. The scalar product must be such that the infinitesimal generators be skew-adjoint.

We shall put (dropping the multiplicity index)

$$\langle V^{j,m}_{\lambda,j_0} \mid V^{j',m'}_{\lambda',j_0} \rangle = \delta_{jj'}\,\delta_{mm'}\,\delta_{\lambda\lambda'}\,\delta_{j_0 j_0'} \cdot \Omega(j). \qquad \text{(A11)}$$

The norm Ω does not depend on m, since the compact generators are still given by (I. 7).

But Ω depends on j: from the expression of the scalar product in $\mathcal{E}$ in (V. 6) one gets

$$Q_+: \quad V^{j,m}_{\lambda,j_0} \to V^{j+1,\,m}_{\lambda,j_0}\ ,$$

$$Q_-: \quad V^{j,m}_{\lambda,j_0} \to (j^2 - j_0^2)(j^2 + \lambda^2)\, V^{j-1,\,m}_{\lambda,j_0}\ ,$$

and K_3 has the expression (I. 17), where $D = i\lambda$. Since $Q_+ + Q_-^* = 0$, the norm Ω must verify

$$\Omega(j) = (j - j_0)(j + j_0)(j - i\lambda)(j + i\lambda)\Omega(j-1), \qquad \text{(A12)}$$

or

$$\Omega(j) = (j - j_0)!(j + j_0)!(j - i\lambda)!(j + i\lambda)!\, C, \qquad \text{(A12')}$$

where C is a positive constant.

To obtain an orthonormal Hilbert basis (and a more symmetric expression for Q_+, Q_-), one must change from $V^{j,m}_{\lambda,j_0}$ to $V^{j,m}_{\lambda,j_0}\,\Omega(j)^{-1/2}$.

[1] E. Wigner, Ann. Math. **40**, 149 (1939).
[2] V. Bargmann, Ann. Math. **59**, 1 (1954).
[3] Y. Shirokov, Zh. Eksp. Teor. Fiz. **33**, 861, 1196, 1208 (1957); **34**, 717 (1958); **36**, 879 (1959) [Sov. Phys.-JETP **6**, 664, 918, 929 (1958); **7**, 493 (1959); **9**, 620 (1961)].
[4] H. Joos, Fortschr. Phys. **10**, 65 (1962).
[5] J. M. Levy-Leblond, Nuovo Cimento **39**, 967 (1965).
[6] J. Lomont and H. Moses, J. Math. Phys. **5**, 294 (1964); J. Math. Phys. **5**, 1438 (1964).
[7] E. Angelopoulos, Ann. Inst. H. Poincaré, XV, 4, 303 (1971).
[8] M. Nay-Mark, Transl. Ann. Math. Soc., Ser. 2 **6**, 379 (1900).
[9] E. Nelson, Ann. Math. **70**, 572 (1959).
[10] B. Nagel, Seminar Notes, Collège de France (April 1970).
[11] In fact, the technique used here allows a conjecture for semidirect products $S \cdot T$, such that S be semisimple, T Abelian, and such that there exists a symmetric bilinear form on T which is an invariant, as in the case $SO(p, q) \cdot Tp + q$: Then for almost every UIR, except for a set of Plancherel measure zero, the maximal differential domain should be nuclear space.
[12] S. Strom, Inst. Theor. Phys., Göteborg, Preprints 68-1 and 68-12.

The general exact solution of the equation of geodesics of the general exact homogeneous plane gravitational wave

Dieter W. Ebner

Lehrstuhl für Theoretische Physik der Universität Konstanz, 775 Konstanz, Germany
(Received 14 August 1973)

The general exact solution of the equation of geodesics for the general exact plane homogeneous gravitational wave is given by solving the vacuum field equations for the line element $ds^2 = A(u)dx^2 + B(u)dy^2 + 2C(u)dx\,dy + dz^2 - dt^2$, with $u = (z-t)/2^{1/2}$ of these waves. The subclass of the exact plane sandwich waves is given explicitly by solving exactly the flatness condition for that line element. The relative energy and momentum transfer on test particles by the sandwich wave is discussed.

1. INTRODUCTION

In this paper we give the solution of a classical problem, i.e., the general exact solution of the equations of geodesics of all plane gravitational waves.

The metric of the plane gravitational wave propagating in the direction of the z axis is usually given in the form (Ref. 1, Gl. 4.1.17)

$$ds^2 = dx^2 + dy^2 + 2\,du\,dv + 2H(x,y,u)\,du^2, \tag{1.1a}$$

with

$$u = (1/\sqrt{2})(z-t), \quad v = (1/\sqrt{2})(z+t), \quad dz^2 - dt^2 = 2\,du\,dv, \tag{1.1b}$$

whereby the vacuum field equations are reduced to

$$\frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2} = 0. \tag{1.1c}$$

The subclass of the *homogeneous* plane waves is given by the following special solution of (1.1c):

$$H = \tfrac{1}{2}A(u)\,[\cos\theta(u)\cdot(x^2-y^2) - \sin\theta(u)\,2xy], \tag{1.1d}$$

where $A(u)$ and $\theta(u)$ are arbitrary real-valued functions. In case of

$$A(u) = 0 \tag{1.2}$$

the space–time is flat, and vice versa. Thus choosing $A(u) = 0$ outside the range $u_1 \leq u \leq u_2$, we get a so-called sandwich wave.

The coordinates of (1.1) have the disadvantage that in them the equation of geodesics cannot be solved in general form. However, as we shall show, there exists another system of coordinates for the homogeneous plane waves in which we can solve the equation of geodesics exactly. Thus we are able to calculate the relative energy-momentum transfer on test particles by a strong gravitational plane wave. This is of interest because the gravitational waves of realistic sources are asymptotically homogeneous plane waves in small regions.

2. THE HOMOGENEOUS PLANE WAVE

In the coordinates above-mentioned which we denote also by x,y,z,t, the line element for the homogeneous plane wave takes the form[2]

$$ds^2 = A(u)dx^2 + B(u)dy^2 + 2C(u)dx\,dy + 2du\,dv, \tag{2.1}$$

with u and v defined as in (1.1b):

$$u = (1/\sqrt{2})(z-t), \quad v = (1/\sqrt{2})(z+t), \quad 2du\,dv = dz^2 - dt^2. \tag{2.1a}$$

The formulas transforming (1.1) into (2.1) is given by Ref. 1.

Numbering of the coordinates in the sequence x,y,u,v, the only nonvanishing component of the Ricci tensor is comp. (4.1).

$$R_{33} = (1/4D^2)(2DD'' - D'^2 - 2D\Delta) \tag{2.2}$$

using the abbreviation[4]

$$D = AB - C^2, \tag{2.3a}$$

$$\Delta = A'B' - C'^2, \tag{2.3b}$$

where the prime means differentiation with respect to u. Thus the vacuum field equations result in the vacuum condition

$$2DD'' - D'^2 - 2D\Delta = 0. \tag{2.4}$$

In order that (2.1) is a normal-hyperbolic Riemannian space of signature $(+++-)$ the quadratic form

$$A(u)\,dx^2 + B(u)\,dy^2 + 2C(u)\,dx\,dy$$

must be positive definite, being equivalent to

$$A(u) > 0, \quad B(u) > 0, \quad D(u) > 0. \tag{2.5}$$

The equations (2.1), (2.4), and (2.5) represent the general homogeneous plane vacuum wave propagating in z direction.

In the next step we satisfy the condition (2.4), by which only two of the functions $A(u), B(u)$, and $C(u)$ are independent from one another. Therefore we look for an appropriate independent pair of variables for the general vacuum solutions. As such a pair we choose $C(u)$ and $D(u)$, with which we get from (2.3)

$$AB = D + C^2, \tag{2.6a}$$

$$A'B' = \Delta + C'^2, \tag{2.6b}$$

wherein Δ is determined with respect to (2.4) by

$$\Delta = D'' - \tfrac{1}{2}(D'^2/D). \tag{2.4'}$$

The quotient of Eqs. (2.6b) and (2.6a) gives

$$\frac{A'B'}{AB} = \frac{\Delta + C'^2}{D + C^2} =: \vartheta(u). \tag{2.7}$$

Furthermore, the logarithmic differentiation of (2.6a) results in

$$\frac{A'}{A} + \frac{B'}{B} = \frac{(D + C^{2\prime})}{D + C^2} =: \eta(u). \tag{2.8}$$

Herein the function $\vartheta(u)$ and $\eta(u)$, in view of (2.4'), are determined by the knowledge of C and D only. Solving the quadratic equation system (2.7), (2.8) for A'/A and B'/B, we find

$$A'/A = \tfrac{1}{2}(\eta \pm \sqrt{\eta^2 - 4\vartheta}), \tag{2.9a}$$

$$B'/B = \tfrac{1}{2}(\eta \mp \sqrt{\eta^2 - 4\vartheta}). \tag{2.9b}$$

Integration of (2.9) gives the result

$$A(u) = A_0 \exp \tfrac{1}{2} \int_{u_0}^{u} (\eta \pm \sqrt{\eta - 4\vartheta})\, du, \tag{2.10a}$$

$$B(u) = B_0 \exp \tfrac{1}{2} \int_{u_0}^{u} (\eta \mp \sqrt{\eta^2 - 4\vartheta})\, du. \tag{2.10b}$$

Because of the condition (2.5), the constants of integration $A_0 = A(u_0)$ and $B_0 = B(u_0)$ must be positive and can be normalized to 1 by an appropriate gauge of the coordinates x and y.[5] Furthermore we can restrict ourselves to the upper sign of the root in the integrands of (2.10), because the choice of the lower sign means a permutation of A and B, which results only in an interchange of the coordinates x and y

Thus after specification of the functions $C(u)$ and $D(u)$ with $D(u) > 0$ [compare (2.5)] the metric (2.1) is determined uniquely according to (2.10).[6] Certainly, the metric must be real. This restricts the free choice of the functions $C(u)$ and $D(u)$, because the radicands in (2.10) have to be positive. With respect to (2.4), (2.7), and (2.8) this results in the subsidiary condition:

$$3D'^2D - 4D^2C'^2 - 4D^2D'' + 2C^2D'^2 - 4DC^2D'' \geq 0. \tag{2.11}$$

3. INTEGRATION OF THE EQUATION OF GEODESICS

We use the equation of geodesics in form[7]

$$\frac{d}{ds}(g_{o\lambda}u^o) = \tfrac{1}{2} u^\mu u^\nu \frac{\partial g_{\mu\nu}}{\partial x^\lambda}, \qquad u^o = \frac{dx^o}{ds}. \tag{3.1}$$

In case of the metric (2.1) this reads[8]:

$$\frac{d}{ds}(Au^1 + Cu^2) = 0 \rightarrow Au^1 + Cu^2 = \gamma = \text{const}, \tag{3.2a}$$

$$\frac{d}{ds}(Bu^2 + Cu^1) = 0 \rightarrow Bu^2 + Cu^1 = \delta = \text{const}, \tag{3.2b}$$

$$\frac{du^3}{ds} = 0 \rightarrow u^3 = \alpha = \text{const}, \tag{3.2c}$$

$$\frac{du^4}{ds} = \tfrac{1}{2}[(u^1)^2 A' + (u^2)^2 B' + 2C' u^1 u^2]. \tag{3.2d}$$

Solving (3.2a, b) with respect to u^1 and u^2, we get

$$u^1 = D^{-1}(\gamma B - \delta C), \tag{3.3a}$$

$$u^2 = D^{-1}(\delta A - \gamma C), \tag{3.3b}$$

$$u^3 = \alpha. \tag{3.3c}$$

In the case of $\alpha \neq 0$, instead of integrating (3.2d), we determine u^4 by the normalization condition

$$u^\mu u_\mu = \epsilon,$$

where $\epsilon = -1, 0, 1$ for the time-like, light-like, and space-like geodesics, respectively. Thus we find

$$u^4 = (2\alpha)^{-1}[\epsilon - A(u^1)^2 - B(u^2)^2 - 2Cu^1u^2]. \tag{3.3d}$$

With the help of (3.3c) the further integration of (3.3) yields ($\alpha \neq 0$)

$$x = \alpha^{-1} \int_{u_0}^{u} D^{-1}(\gamma B - \delta C)\, du + x_0, \tag{3.4a}$$

$$y = \alpha^{-1} \int_{u_0}^{u} D^{-1}(\delta A - \gamma C)\, du + y_0, \tag{3.4b}$$

$$u = \alpha s + u_0, \tag{3.4c}$$

$$v = \alpha^{-1} \int_{u_0}^{u} u^4\, du + v_0, \tag{3.4d}$$

with u^4 given by (3.3d).

In the case of $\alpha = 0$, the integration of (3.3a, b, c) are

$$x = x_0 + u_0^1 \cdot s, \tag{3.5a}$$

$$y = y_0 + u_0^2 \cdot s, \tag{3.5b}$$

$$u = u_0, \tag{3.5c}$$

whereas double integration of (3.2d) gives

$$v = v_0 + u_0^4 \cdot s + \tfrac{1}{4} s^2 [A'_0(u_0^1)^2 + B'_0(u_0^2)^2 + 2C'_0 u_0^1 u_0^2]. \tag{3.5d}$$

With the use of (2.5), it is easy to show that the geodesics (3.5) are spacelike or lightlike only.

The entirety of the timelike geodesics, used in the following, are obtained by setting $\epsilon = -1$ in (3.3d).

4. THE SOLUTION OF THE FLATNESS CONDITION

We deduce the general form of the functions $A(u), B(u), C(u)$ of the metric (2.1) representing flat space-time in the region $u_0 \leq u \leq u_1$.

The nonvanishing Christoffel symbols of (2.1) are[8]

$$\begin{aligned}
\Gamma^1_{13} &= \Gamma^1_{31} = (2D)^{-1}(BA' - CC'),\\
\Gamma^1_{23} &= \Gamma^1_{32} = (2D)^{-1}(BC' - CB'),\\
\Gamma^2_{13} &= \Gamma^2_{31} = (2D)^{-1}(AC' - CA'),\\
\Gamma^2_{23} &= \Gamma^2_{32} = (2D)^{-1}(AB' - CC'),\\
\Gamma^4_{11} &= -\tfrac{1}{2}A',\\
\Gamma^4_{22} &= -\tfrac{1}{2}B',\\
\Gamma^4_{12} &= \Gamma^4_{21} = -\tfrac{1}{2}C',
\end{aligned} \tag{4.1}$$

and the nonvanishing independent components of the Riemann tensor of (2.1) are[8]

$$R_{1313} = -4D^3[2DA'' - D'A' + \Delta A], \tag{4.2a}$$

$$R_{2323} = -4D^3[2DB'' - D'B' + \Delta B], \tag{4.2b}$$

$$R_{1313} = -4D^3[2DC'' - D'C' + \Delta C]. \tag{4.2c}$$

The necessary and sufficient condition for (2.1) to represent flat space-time in $u_0 \leq u \leq u_1$ is given according to (4.2) by the following set of equations (flatness condition).

$$2DA'' - D'A' + \Delta A = 0, \tag{4.3a}$$

$$2DB'' - D'B' + \Delta B = 0, \tag{4.3b}$$

$$2DC'' - D'C' + \Delta C = 0. \tag{4.3c}$$

As will be shown in the Appendix, the general solution of (4.3) is given by

$$A(u) = A_0 + A_1(u-u_0) + \tfrac{1}{2}A_2(u-u_0)^2, \quad (4.4a)$$

$$B(u) = B_0 + B_1(u-u_0) + \tfrac{1}{2}B_2(u-u_0)^2, \quad (4.4b)$$

$$C(u) = C_0 + C_1(u-u_0) + \tfrac{1}{2}C_2(u-u_0)^2, \quad (4.4c)$$

where the real constants $A_0, B_0, C_0, A_1, B_1, C_1, A_2, B_2, C_2$ are restricted by (2.5) and by the conditions

$$A_2 = (2D_0)^{-1}(D_1A_1 - \Delta_0A_0), \quad (4.5a)$$

$$B_2 = (2D_0)^{-1}(D_1B_1 - \Delta_0B_0), \quad (4.5b)$$

$$C_2 = (2D_0)^{-1}(D_1C_1 - \Delta_0C_0), \quad (4.5c)$$

with the abbreviations $D_0 := A_0B_0 - C_0^2, D_1 := A_0B_1 + A_1B_0 - 2C_0C_1$, and $\Delta_0 = A_1B_1 - C_1^2$. For the independent functions D and C, (comp. Sec. 2) we find from (4.4) (4.5)

$$C(u) = C_0 + C_1(u-u_0) + \tfrac{1}{2}(2D_0)^{-1}D_1C_1(u-u_0)^2, \quad (4.6)$$

$$D(u) = D_0 + D_1(u-u_0) + \left(\frac{1}{4D_0}D_1^2 + \tfrac{1}{2}\Delta_0\right)(u-u_0)^2 + \frac{1}{4D_0}D_1\Delta_0(u-u_0)^3 + \frac{1}{16D_0}\Delta_0^2(u-u_0)^4. \quad (4.7)$$

It is easy to show by the general method of Eq. (2.10) that (4.6) and (4.7) lead back to (4.4), (4.5) with the constants

$$A_0 > 0 \text{ arbitrary}, \quad B_0 = \frac{D_0 + C_0^2}{A_0}, \quad (4.8a)$$

$$A_1 = \frac{A_0(D_1 + 2C_0C_1 + \kappa_0)}{2(D_0 + C_0^2)}, \quad (4.8b)$$

$$B_1 = \frac{B_0(D_1 + 2C_0C_1 - \kappa_0)}{2(D_0 + C_0^2)}, \quad (4.8c)$$

$$A_2 = \frac{D_1A_1 - \Delta_0A_0}{2D_0}, \quad (4.8d)$$

$$B_2 = \frac{D_1B_1 - \Delta_0B_0}{2D_0} \quad (4.8e)$$

$$C_2 = \frac{D_1C_1 - \Delta_0C_0}{2D_0} \quad (4.8f)$$

with

$$\kappa_0 = (D_1^2 + 4D_1C_0C_1 - 4D_0C_1^2 - 4D_0\Delta_0 - 4\Delta_0C_0^2)^{1/2}. \quad (4.8g)$$

The reality condition (2.11) for (4.6) (4.7) reads

$$D_1^2 + 4D_1C_0C_1 - 4D_0C_1^2 - 4D_0\Delta_0 - 4\Delta_0C_0^2 \geq 0. \quad (4.9)$$

Thus the 6 constants $D_0 > 0, D_1, \Delta_0, C_0, C_1, A_0 > 0$ fulfilling (4.9) can be chosen arbitrarily and the functions $A(u), B(u), C(u)$ are uniquely determined according to (4.3) and (4.7). The result (4.7) is suitable to construct sandwich waves as will be done in Sec. 5.

5. ACCELERATION OF TEST PARTICLES BY SANDWICH WAVES

In this paragraph we calculate the acceleration effect of a plane gravitational wave on test particles. We choose a so called sandwich wave which is different from flat space-time only in a finite interval $u_0 \leq u \leq u_1$. Since we have a gravitational wave without source, only changes of the *relative* velocities of two test particles, i.e., relative momentum and kinetic energy, have physical significance.

We obtain a sandwich wave if we choose the twice continuously differentiable independent functions $D(u)$, $C(u)$ such that outside the range $u_0 \leq u \leq u_1$ they have the form (4.6), (4.7) whereas in the range $u_0 \leq u \leq u_1$ they are arbitrary fulfilling only the conditions (2.5), (2.11) and the conditions of differentiability at the limits u_0 and u_1 (arbitrary wave profiles). Without loss of generality we can assume that in one of the two flat space-time regions, e.g., for $u \geq u_1$ we have[9]

$$D(u) = 1, \quad C(u) = 0, \quad A(u) = 1, \quad B(u) = 1. \quad (5.1)$$

In this paper, for reason of simplicity, we restrict ourselves to the case where the two particles are at rest relative to one another before the wave has arrived at them.[10]

Without loss of generality, we can write the geodesics of the particles [putting $\gamma = \delta = 0, \alpha = -1/\sqrt{2}$ in (3.3) and (3.4)] as:[9]

1. particle:

$$x = 0, \quad y = 0, \quad u = -(1/\sqrt{2})s + u_0, \quad v = (1/\sqrt{2})s + v_0; \quad (5.2.1)$$

2. particle:

$$x = x_0, \quad y = y_0, \quad u = -(1/\sqrt{2})\tau + u_0, \quad v = (1/\sqrt{2})\tau + u_0 + v_0. \quad (5.2.2)$$

Equation (5.2) is valid for $-\infty < u < +\infty$. In both cases the coordinate velocities of both particles are

$$u^1 = u^2 = 0, \quad u^3 = -1/\sqrt{2}, \quad u^4 = 1/\sqrt{2} \quad (5.3)$$

independent of s and τ. For $u \geq u_1$ this means that the two particles are at rest relative to one another. For $u \leq u_1$ and in particular for $u \leq u_0$, the comparison of the velocities of the two particles must be performed by a parallel transport using the Christoffel symbols (4.1). This gives the following for the coordinate velocity of the first particle at the position of the second one after the wave has swept over them ($u = u_0$):

1. particle:

$$u_1^1 = \frac{1}{\sqrt{2}}[\Gamma^1_{31}\cdot x_0 + \Gamma^1_{32}\cdot y_0], \quad u_1^2 = \frac{1}{\sqrt{2}}[\Gamma^2_{31}\cdot x_0 + \Gamma^2_{32}\cdot y_0], \quad u_1^3 = -1/\sqrt{2}, \quad u_1^4 = (1/\sqrt{2})(1 + A(u_1^1)^2 + B(u_1^2)^2 + 2Cu_1^1u_1^2), \quad (5.3.1)$$

while the coordinate velocity of the second particle at the same position is given by (5.3)

2. particle:

$$u_2^1 = u_2^2 = 0, \quad u_2^3 = -1/\sqrt{2}, \quad u_2^4 = 1/\sqrt{2}. \quad (5.3.2)$$

We separate u_1^μ into two components

$$u_1^\mu = (\Delta E/m_0)u_2^\mu + \mathfrak{u}_1^\mu \tag{5.4}$$

determined by the condition $\mathfrak{u}_1^\mu u_{2\mu} = 0$. Then $\Delta E = -m_0 u_1^\mu u_{2\mu}$ is the relativistic kinetic energy (including the rest-energy) of the first particle judged by the observer u_2^μ (second particle). (m_0 = mass of the first particle). In the same way we separate the (light-like) ray vector k^μ of the wave, given by

$$k^1 = k^2 = k^3 = 0, \quad k^4 = 1 \tag{5.5}$$

into two components:

$$k^\mu = (1/\sqrt{2})\cdot u_2^\mu + \kappa^\mu . \tag{5.6}$$

With respect to $\kappa^\mu u_{2\mu} = 0, \kappa^\mu$ is the vector of the direction of propagation of the wave judged by the observer u_2^μ. According to this we separate

$$\mathfrak{u}_1^\mu = \mathfrak{u}_\perp^\mu + \mathfrak{u}_\parallel^\mu$$

with $\mathfrak{u}_\parallel^\mu \parallel \kappa^\mu$ and $\mathfrak{u}_\perp^\mu \kappa_\mu = 0$. Then $\Delta p_\parallel = m_0\sqrt{\mathfrak{u}_\parallel^\mu \mathfrak{u}_{\parallel\mu}}$ is the momentum of the first particle in the direction of propagation of the wave and $\Delta p_\perp = m_0\sqrt{\mathfrak{u}_\perp \mathfrak{u}_{\perp\mu}}$ is the momentum of the first particle perpendicular to the direction of propagation of the wave, both judged by the observer u_2^μ.

The calculation of these quantities according to (5.3) yields

$$\Delta E = m_0(1 + F_0), \tag{5.7}$$

$$\Delta p_\parallel = m_0 F_0 , \tag{5.8}$$

$$\Delta p_\perp = m_0\sqrt{2F_0}, \tag{5.9}$$

with

$$F_0 = \tfrac{1}{8}[A_2 x_0^2 + B_2 y_0^2 + 2C_2 x_0 y_0], \tag{5.10}$$

where A_2, B_2, C_2 are given by (4.8). According to this, the transfer of energy and momentum by the gravitational wave is determined by the "constants of the gravitational wave" alone (and not by integrals over the profile of the gravitational wave). x_0, y_0 are the components of the transverse distance vector of the two particles before the wave has swept over them, whereas the longitudinal distance (in the direction of propagation of the wave) is irrelevant.[11]

For special orientations of the particles and for special sandwich waves, the relative energy and momentum transfer vanishes if $F_0 = 0$ is fulfilled. Two cases should be discussed explicitly:

First, if the distance vector of the particles has the direction of the propagation of the wave ($x_0 = y_0 = 0$) the energy and momentum transfer vanishes for any sandwich wave. Second, if for the constants of the wave $A_2 = B_2 = C_2 = 0$ is valid, the energy momentum transfer vanishes also for any orientation of the particles.

ACKNOWLEDGMENT

I thank Professor H. Dehnen for valuable discussions during all phases of the work.

APPENDIX A

That (4.4) with (4.5) is a solution of (4.3) and therefore a *sufficient* condition for (2.1) to represent flat space-time can be checked by direct calculation.

To show that (4.4) with (4.5) is also a *necessary* condition, we assume (2.1) to represent flat space-time. Expressed in a Lorentz coordinate system the equation for two infinitesimally neighboring geodesics read

$$x^\mu = x_0^\mu + u_0^\mu s \quad \text{and} \quad x^\mu = x_0^\mu + dx_0^\mu + (u_0^\mu + du_0^\mu)\tau, \tag{A1}$$

where the subscript null refers to the initial values.

The relative distance of these geodesics judged by the observer u_0^μ is given by

$$(dr(s))^2 = [(dx_0^\mu dx_{0\mu}) + (dx_0^\mu u_{0\mu})^2 + 2(dx_{0\mu} Du_0^\mu)s + (Du_0^\mu Du_{0\mu})s^2, \tag{A2}$$

wherein

$$Du_0^\mu = du_0^\mu + (\Gamma_{\rho\sigma}^\mu)_0 u_0^\rho dx_0^\sigma . \tag{A3}$$

Equation (4.12) is exact in s but in dx_0^μ, dy_0^μ only correct up to the second order. Furthermore, (A2) with (A3) is independent of the choice of coordinates and is valid in any curvilinear coordinate system of flat space-time. In particular (A2) must be valid for the two neighboring time-like geodesics given by (3.3), (3.4) setting $\gamma = \delta = 0, \alpha = 1$:

$$x = 0, \quad y = 0, \quad u = u_0 + s, \quad v = v_0 - \tfrac{1}{2}s, \tag{A4a}$$

$$x = dx_0, \quad y = dy_0, \quad u = u_0 + \tau, \quad v = v_0 - \tfrac{1}{2}\tau. \tag{A4b}$$

In both cases we have

$$u^\mu(s) = w^\mu(\tau) = u_0^\mu = (0, 0, 1, -\tfrac{1}{2}), \tag{A5}$$

i.e., $$du_0^\mu = 0, \tag{A6}$$

and

$$dx_0^1 = dx_0, \quad dx_0^2 = dy_0, \quad dx_0^3 = 0, \quad dx_0^4 = 0; \tag{A7}$$

with (4.1), (A5), (A6), (A7) Eq. (A3) reads

$$Du_0^1 = (2D_0)^{-1}(B_0A_0' - C_0C_0')dx_0 + (2D_0)^{-1}(B_0C_0' - C_0B_0)dy_0,$$

$$Du_0^2 = (2D_0)^{-1}(A_0C_0' - C_0A_0')dx_0 + (2D_0)^{-1}(A_0B_0' - C_0C_0')dy_0,$$

$$Du_0^3 = Du_0^4 = 0, \tag{A8}$$

and (A2) results in

$$(dr(s))^2 = (dx_0^\mu dx_{0\mu}) + 2(dx_{0\mu}Du_0^\mu)s + (Du_0^\mu Du_{0\mu})s^2, \tag{A9}$$

with Du_0^μ given by (A8).

On the other hand, $(dr(s))^2$ can be calculated directly. Because $dx^\mu = dx_0^\mu$ is joining the two geodesics and is orthogonal to u_0^μ at any point s, we find

$$(dr(s))^2 = g_{\mu\nu}dx^\mu dx^\nu = A(u)dx_0^2 + B(u)dy_0^2 + 2C(u)dx_0 dy_0. \tag{A10}$$

Comparing (A9) with (A10) we find that $A(u), B(u), C(u)$ are quadratic functions in $s = u - u_0$. Thus we have arrived at (4.4).

To derive (4.5), we put (4.4) into (4.3). Comparing coefficients in $(u - u_0)$, we get the set of equations

$$-X_1[A_1B_2 + A_2B_1 - 2C_1C_2] + X_2[A_1B_1 - C_1^2] + 2X_0[A_2B_2 - C_2^2] = 0, \tag{A11a}$$

$$X_2[A_0B_1 + A_1B_0 - 2C_0C_1] - X_1[A_0B_2 + A_2B_0 + A_1B_1 - C_1^2 - 2C_0C_2] = 0, \tag{A11b}$$

$$2X_2[A_0B_0-C_0^2]-X_1[A_0B_1+A_1B_0-2C_0C_1]+X_0[A_1B_1-C_1^2]=0, \quad \text{(A11c)}$$

where X stands for $A, B,$ or C.

Under the supposition (4.4) these 9 equations (A11) are equivalent to (4.3). By using the abbreviations

$$\begin{aligned} D_0&=A_0B_0-C_0^2,\\ \Delta_0&=A_1B_1-C_1^2,\\ \Box_0&=A_2B_2-C_2^2,\\ D_1&=A_0B_1+A_1B_0-2C_0C_1,\\ Q_0&=A_1B_2+A_2B_1-2C_1C_2,\\ R_0&=A_0B_2+A_2B_0-2C_0C_2, \end{aligned} \quad \text{(A12)}$$

(A11) reads

$$X_2\Delta_0-X_1Q_0+2X_0\Box_0=0, \quad \text{(A13)}$$

$$X_2D_1-X_1(R_0+\Delta_0)+X_0Q_0=0, \quad \text{(A14)}$$

$$2X_2D_0-X_1D_1+X_0\Delta_0=0. \quad \text{(A15)}$$

Since $D>0$, (A15) yields (4.5), i.e.,

$$X_2=(2D_0)^{-1}(X_1D_1-X_0\Delta_0), \quad \text{(A16)}$$

whereas (A13), (A14) is then satisfied identically.

APPENDIX B

In this Appendix we show that there exist coordinates $(\bar{x},\bar{y},\bar{u},\bar{v})$ preserving the general form (2.1) for the metric of the plane gravitational wave, i.e.,

$$\begin{aligned} ds^2&=A(u)\,dx^2+B(u)\,dy^2+2C(u)\,dxdy+2dudv,\\ &=\bar{A}(\bar{u})\,d\bar{x}^2+\bar{B}(\bar{u})\,d\bar{y}^2+2\bar{C}(\bar{u})\,d\bar{x}d\bar{y}+2d\bar{u}d\bar{v}, \end{aligned} \quad \text{(B1)}$$

with the property that for any prescribed point $P_0=(x_0,y_0,u_0,v_0)$ lying in the flat space–time region $u\geq u_1$ (i.e., $u_0\geq u_1$) and for any prescribed normalized time-like vector situated at P_0 we have

$$P_0=(0,0,\bar{u}_0,\bar{v}_0), \quad \text{(B2)}$$

$$\mathrm{u}_0=(0,0,-1/\sqrt{2},1/\sqrt{2}), \quad \text{(B3)}$$

$$\bar{A}(\bar{u})=\bar{B}(\bar{u})=1,\bar{C}(\bar{u})=0 \text{ for } \bar{u}\geq u_0. \quad \text{(B4)}$$

The metric (2.1) has the Killing vectors

$$\begin{aligned} \underset{1}{\xi}{}^{\mu}&=\delta^{\mu}_1, \quad \underset{2}{\xi}{}^{\mu}=\delta^{\mu}_2, \quad \chi^{\mu}=\delta^{\mu}_4,\\ \underset{1}{\eta}{}^{\mu}&=\delta^{\mu}_1\int_{u_0}^{u}\frac{B}{D}\,du-\delta^{\mu}_2\int_{u_0}^{u}\frac{C}{D}\,du-\delta^{\mu}_4 x,\\ \underset{2}{\eta}{}^{\mu}&=-\delta^{\mu}_1\int_{u_0}^{u}\frac{C}{D}\,du+\delta^{\mu}_2\int_{u_0}^{u}\frac{A}{D}\,du-\delta^{\mu}_4 y, \end{aligned} \quad \text{(B5)}$$

as can be checked by using the Killing equations

$$g_{\mu\nu,\rho}\xi^{\rho}+g_{\mu\rho}\xi^{\rho}{}_{,\nu}+g_{\nu\rho}\xi^{\rho}{}_{,\mu}=0. \quad \text{(B6)}$$

Then

$$\underset{i}{\varphi}{}^{\mu}=a_{i1}\underset{1}{\xi}{}^{\mu}+a_{i2}\underset{2}{\xi}{}^{\mu}+\kappa_i\chi^{\mu}+\mu_{i1}\underset{1}{\eta}{}^{\mu}+\mu_{i2}\underset{2}{\eta}{}^{\mu}, \quad i=1,2, \quad \underset{3}{\varphi}{}^{\mu}=\kappa_0\chi^{\mu} \quad \text{(B7)}$$

with constants $a_{ij},\mu_{ij},\kappa_i,\kappa_0$ are also Killing vectors.

$\underset{3}{\varphi}{}^{\mu}$ is lightlike and orthogonal to $\underset{1}{\varphi}{}^{\mu}$ and $\underset{2}{\varphi}{}^{\mu}$.

The vectors $\underset{1}{\xi}{}^{\mu},\underset{2}{\xi}{}^{\mu},\chi^{\mu},\underset{1}{\eta}{}^{\mu},\underset{2}{\eta}{}^{\mu}$ have vanishing Lie deviations among one another with exception of

$$[\underset{i}{\xi},\underset{i}{\eta}]=:\underset{i}{\xi}{}^{\rho}\underset{i}{\eta}{}^{\nu}{}_{,\rho}-\underset{i}{\eta}{}^{\rho}\underset{i}{\xi}{}^{\nu}{}_{,\rho}=-\chi^{\nu}. \quad \text{(B8)}$$

Therefore, $\underset{i}{\varphi}{}^{\mu};i=1,2,3$ have also vanishing Lie deviations if we require

$$\mu_{11}a_{21}+\mu_{12}a_{22}=a_{11}\mu_{21}+a_{12}\mu_{22}. \quad \text{(B9)}$$

Besides (B9), we choose the constants a_{ij},μ_{ij},κ_i such that we have at P_0:

$$u_0^{\mu}\underset{i}{\varphi}{}_{\mu}=0, \quad i=1,2, \quad u_0^{\mu}\underset{3}{\varphi}{}_{\mu}=-1/\sqrt{2} \quad \text{(B10)}$$

and such that

$$\sigma_{ij}(x,y,u)=\underset{i}{\varphi}{}^{\mu}\underset{j}{\varphi}{}_{\mu}$$

satisfy at P_0:

$$\sigma_{11}=\sigma_{22}=1, \quad \sigma_{12}=0, \quad \sigma'_{11}=\sigma'_{22}=\sigma'_{12}=0. \quad \text{(B11)}$$

($'$ means partial derivative with respect to u).

Equations (B10) and (B11) are satisfied if we require:

$$\kappa_i=-(u_0^3)^{-1}[A_0u_0^1a_{i1}+B_0u_0^2a_{i2}+C_0u_0^1a_{i2}+C_0u_0^2a_{i1}], \quad i=1,2. \quad \text{(B12)}$$

(Note: $u_0^3\neq 0$ since u_0^{μ} is time-like.) $\kappa_0=-(\sqrt{2}\,u_0^3)^{-1}$ (B13)

$$A_0a_{i1}a_{j1}+B_0a_{i2}a_{j2}+C_0a_{i1}a_{j2}+C_0a_{j1}a_{i2}=\delta_{ij}, \quad i,j=1,2 \quad \text{(B14)}$$

$$A_1a_{i1}a_{j1}+B_1a_{i2}a_{j2}+C_1a_{i1}a_{j2}+C_1a_{j1}a_{i2}+a_{i1}\mu_{j1}+a_{j1}\mu_{i1}+a_{i2}\mu_{j2}+a_{j2}\mu_{i2}=0, \quad i,j=1,2. \quad \text{(B15)}$$

Equations (B9), (B14), (B15) can be solved, e.g., with

$$\begin{aligned} &a_{11}=(A_0)^{-1/2}, \quad a_{12}=0, \quad a_{21}=-C_0(A_0D_0)^{-1/2}, \quad a_{22}=A_0^{1/2}D_0^{-1/2},\\ &\mu_{11}=-A_1A_0^{-1/2}, \quad \mu_{12}=\tfrac{1}{4}C_0A_1A_0^{-3/2}-\tfrac{1}{2}C_1A_0^{-1/2}-\tfrac{1}{4}C_0A_1A_0^{-1/2},\\ &\mu_{21}=\tfrac{3}{4}A_1C_0(A_0D_0)^{-1/2}-\tfrac{1}{2}C_1A_0^{1/2}D_0^{-1/2}-\tfrac{1}{4}C_0A_1A_0^{1/2}D_0^{-1/2},\\ &\mu_{22}=-\tfrac{1}{2}A_1C_0^2A_0^{-3/2}D_0^{-1/2}-\tfrac{1}{2}B_1A_0^{1/2}D_0^{-1/2}+\tfrac{1}{2}C_0C_1(A_0D_0)^{-1/2}+\tfrac{3}{4}C_0^2A_1(A_0)^{-3/2}D_0^{-1/2}-\tfrac{1}{4}C_0^2A_1(A_0D_0)^{-1/2}. \end{aligned}$$

The Killing vectors $\underset{i}{\varphi}{}^{\mu}$, $i=1,2,3$, are linearly independent (as vector fields) because they are linearly independent as P_0. Because they have vanishing Lie deviations among one another, we can choose a system of coordinates $(\bar{\bar{x}},\bar{\bar{y}},\bar{\bar{u}},\bar{\bar{v}})$ such that[13]

$$\underset{1}{\varphi}{}^{\mu}=\delta^{\mu}_1, \quad \underset{2}{\varphi}{}^{\mu}=\delta^{\mu}_2, \quad \underset{3}{\varphi}{}^{\mu}=\delta^{\mu}_4. \quad \text{(B16)}$$

From Killing's equation (B6) there follows

$$\bar{\bar{g}}_{\mu\nu}=\bar{\bar{g}}_{\mu\nu}(\bar{\bar{u}}). \quad \text{(B17)}$$

From the orthogonality of φ^μ_1 and φ^μ_2 to φ^μ_3 and from the time-like character of φ^μ_3 there follows:

$$\bar{\bar{g}}_{14} = \bar{\bar{g}}_{24} = \bar{\bar{g}}_{44} = 0. \tag{B18}$$

Now, we introduce coordinates $x^\mu = (\bar{x}, \bar{y}, \bar{u}, \bar{v})$ by

$$\bar{\bar{x}}^\mu = \bar{x}^\mu + f^\mu(\bar{u}).$$

With (B17), (B18) this leads to ($\cdot$ means differentiation with respect to $\bar{u}$)

$$\bar{g}_{\mu\nu} = \bar{\bar{g}}_{\mu\nu}, \quad \mu, \nu = 1, 2, 4, \tag{B19}$$

$$\dot{f}^3(\bar{u}) = (\bar{\bar{g}}_{34})^{-1}(\bar{g}_{34} - \bar{\bar{g}}_{34}), \tag{B20}$$

$$\bar{\bar{g}}_{11}\dot{f}^1(\bar{u}) + \bar{\bar{g}}_{21}\dot{f}^2(\bar{u}) = \bar{g}_{31} - \bar{\bar{g}}_{31}(1 + \dot{f}^3(\bar{u})), \tag{B21a}$$

$$\bar{\bar{g}}_{12}\dot{f}^1(\bar{u}) + \bar{\bar{g}}_{22}\dot{f}^2(\bar{u}) = \bar{g}_{32} - \bar{\bar{g}}_{32}(1 + \dot{f}^3(u)), \tag{B21b}$$

$$4\dot{f}^4(\bar{u}) = \bar{g}_{33} - \bar{\bar{g}}_{33} - 2\sum_{i=1}^{3} \bar{\bar{g}}_{3i}\dot{f}^i(\bar{u}) - \sum_{i,j=1}^{3} \bar{\bar{g}}_{ij}\dot{f}^i(\bar{u})\dot{f}^j(\bar{u}). \tag{B22}$$

From (B18) we get

$$\bar{\bar{g}}_{34} \neq 0, \quad \bar{\bar{g}}_{11}\bar{\bar{g}}_{22} - (\bar{\bar{g}}_{12})^2 \neq 0.$$

Therefore (B20), (B22) can be solved such flat

$$\bar{g}_{34} = 1, \quad \bar{g}_{3i} = 0, \quad i = 1, 2, 3, \tag{B23}$$

and P_0 has the coordinates $(0, 0, \bar{u}_0, \bar{v}_0)$.

From (B19), (B23) there follows (B1). Equation (B16) is also valid in the coordinates $(\bar{x}, \bar{y}, \bar{u}, \bar{v})$. Thus from (B11) we obtain

$$\bar{A}_0 = \bar{B}_0 = 1, \quad \bar{C}_0 = \bar{C}_1 = \bar{A}_1 = \bar{B}_1 = 0 \tag{B24}$$

at P_0. Since P_0 lies in a region of flat space-time, (4.4), (4.5) is valid. From this we conclude (B4). From (B13) there follows (B3).

[1] P. Jordan, John Ehlers, and W. Kundt, Abh. Math.-Naturwiss. K. L. Akad. Wiss. Lit. Mainz 2, 21 (1960), see Eq. (4.5.1).

[2] The form (1.1a) of the plane wave could give rise to the erroneous conclusion that the exact homogeneous plane wave is not as exactly plane as e.g. a plane electromagnetic wave (comp. Ref. 3, p. 526) since $H(x, y, u)$ depends explicitly on the transversal coordinates x and y. [If $H(x, y, u)$ is independent of x and y. (1.1a) is flat space-time.] However the existence of the form (2.1) for the exact plane wave disproves this proposition. On the other hand the metric (1.1a) has the advantage that because of (1.1c) it fulfills a linear superposition principle, whereas, in view of (2.4), (2.1) does not.

[3] H. Bondi, F. Pirani and I. Robinson, Proc. R. Soc. Lond. A 251, 519 (1959).

[4] Note that $D = -g$, where g is the determinant of $g_{\mu\nu}$.

[5] In (2.1) we can assume without loss of generality that for some $u = u_0$ we have $A(u_0) = B(u_0) = 1$ which implies $D(u_0) = 1 - C(u_0)^2$ [comp. (2.3a)]

[6] The functions $C(u)$ and $D(u)$ must be twice continuously differentiable in view of (2.4).

[7] The 4-velocity u^σ should not be confused with the (index-free) coordinate u.

[8] We use the following ordering of the coordinates: $(x^1, x^2, x^3, x^4) = (x, y, u, v)$.

[9] See Appendix B.

[10] A general discussion of the influence of sandwich waves on test particles with arbitrary initial conditions as well as a comparison with linearized theory will be the subject of a forthcoming paper.

[11] Equations (5.7), (5.8), (5.9) agrees with the result (2-5.58) of Ehlers-Kundt (Ref. 12) with $|g'|^2 = 4F_0$. The calculation of (2-5.58) was performed in the coordinates of (1.1a). Therefore, (2-5.58) is no explicit formula since g' can only be calculated by solving the differential equation (2-5.5.7), a problem which cannot be reduced to an integration, generally. On the other hand, our method has lead to the explicit result (5.10). Furthermore, we remark that the result (2-5.58) is by no means general. As in our calculation, it was supposed that initially both particles have no relative velocity.

[12] Z. Ehlers and W. Kundt, "Exact Solution of the Gravitational Field Equations", *Gravitation: An Introduction to Current Research*, edited by Louis Witten (Wiley, New York, 1962), Chap. II.

[13] See, e. g., L. P. Eisenhart, Ref. 14, p. 48. Compare Ref. 1, p. 100 for a similar treatment.

[14] L. P. Eisenhart, *Continuous Groups of Transformations* (Princeton U.P., Princeton, N.J., 1933).

Lie algebra of the hypergeometric functions

E. Kyriakopoulos

Nuclear Research Center "Democritos", Aghia Paraskevi Attikis, Athens, Greece

*Theory Division, CERN, Geneva, Switzerland**

(Received 17 April 1973; revised manuscript received 3 August 1973)

It is shown that six operators forming the Lie algebra D_2 can be introduced in order to treat the hypergeometric functions by Lie theory techniques. The relation of these operators to the three operators of the Lie algebra $sl(2)$, which were used previously in treating the hypergeometric functions by Lie theory techniques, is discussed.

I. INTRODUCTION

In the study of special functions the techniques of the theory of analytic functions are usually applied. It is possible however to derive relations involving these functions, for example generating functions, additions theorems, etc by replacing partially the analytic methods by group theoretical methods.[1-6] The new techniques are based on the representations of Lie algebras by generalized Lie derivatives and the multiplier representations of local Lie groups.

Because of its importance in mathematical physics the hypergeometric functions $F(a, b; c; w)$ have attracted a great deal of interest, and several generating functions have been derived by Lie theory methods by Weisner,[1] and by Miller,[5] using a Lie algebra with three operators.

In the present work we consider six operators which obey the commutation relations of the generators of an $SO(3, 1)$ group. They give after complexification the algebra D_2 in Cartan's notation. Six new operators are derived from these, which transform the functions $F(a, b; c; w)t^m$ among themselves, and therefore can be used in treating the hypergeometric functions by Lie theory techniques. In Sec. III the six generators of the $SO(3, 1)$ of Sec. II are transformed to six new operators, three of which forming an $sl(2)$ algebra, are identical with those considered by Miller[5] in connection with the hypergeometric functions. The other three operators however do not transform the functions $F(a, b; c; w)t^m$, which are basis of the $sl(2)$ algebra, among themselves, and therefore they are not useful in the Lie theory treatment of the hypergeometric functions. Finally in Sec. IV we show that the six operators of Sec. II can be used to treat the incomplete beta functions by Lie theory techniques.

II. LIE ALGEBRA

In this section a six-dimensional Lie algebra will be given, which transforms the hypergeometric functions among themselves. Consider the operators

$$L'_{12} = M_3 = i\frac{\partial}{\partial\varphi}$$

$$L'_{23} = M_1 = -i\left(\frac{\cos\varphi\cos\vartheta}{\sin\vartheta}\frac{\partial}{\partial\varphi} + \sin\varphi\frac{\partial}{\partial\vartheta} - i\sigma\frac{\cos\varphi}{\sin\vartheta}\right),$$

$$L'_{31} = M_2 = -i\left(\frac{\sin\varphi\cos\vartheta}{\sin\vartheta}\frac{\partial}{\partial\varphi} - \cos\varphi\frac{\partial}{\partial\vartheta} - i\sigma\frac{\sin\varphi}{\sin\vartheta}\right),$$

$$L'_{41} = N_1 = -i\left(\frac{\sin\varphi}{\sin\vartheta}\frac{\partial}{\partial\varphi} - \cos\varphi\cos\vartheta\frac{\partial}{\partial\vartheta} - i\sigma\frac{\sin\varphi\cos\vartheta}{\sin\vartheta} - k\cos\varphi\sin\vartheta\right),$$

$$L'_{42} = N_2 = -i\left(\frac{\cos\varphi}{\sin\vartheta}\frac{\partial}{\partial\varphi} - \sin\varphi\cos\vartheta\frac{\partial}{\partial\vartheta} + i\sigma\frac{\cos\varphi\cos\vartheta}{\sin\vartheta} - k\sin\varphi\sin\vartheta\right),$$

$$L'_{43} = N_3 = -i\left(\sin\vartheta\frac{\partial}{\partial\vartheta} - k\cos\vartheta\right), \tag{2.1}$$

which satisfy the commutation relations

$$[L'_{\mu\nu}, L'_{\rho\lambda}] = (g_{\mu\lambda}L'_{\nu\rho} + g_{\nu\rho}L'_{\mu\lambda} - g_{\mu\rho}L'_{\nu\lambda} - g_{\nu\lambda}L'_{\mu\rho}),$$
$$g_{11} = g_{22} = g_{33} = -g_{44} = 1, \tag{2.2}$$

i.e., they are generators of an $SO(3, 1)$ group. The Casimir operators are

$$\mathbf{M}^2 - \mathbf{N}^2 = \sigma^2 + (k+1)^2 - 1, \quad \mathbf{M}\cdot\mathbf{N} = -i\sigma(k+1). \tag{2.3}$$

In Eqs (2.1) σ and k are free parameters. Therefore the operators (2.1) generate all $SO(3, 1)$ representations.

Let $(\cos\vartheta)^k R'$ be the functions on which the operators $L'_{\mu\nu}$ act. We "extract" the factor $(\cos\vartheta)^k$ from the Hilbert space and define new operators $L''_{\mu\nu}$ such that

$$L'_{\mu\nu}(\cos\vartheta)^k R' = (\cos\vartheta)^k L''_{\mu\nu} R'. \tag{2.4}$$

The operators $L'_{\mu\nu}$ and $L''_{\mu\nu}$ satisfy the same commutation relations. We introduce new variables x and y by the relations

$$\sin\vartheta = i\tan x, \quad \varphi = iy, \tag{2.5}$$

and then we consider the transformation

$$w = [\cos(x/2)]^2, \quad t = \{1 - [\cos(x/2)]^{-2}\}^{1/2} e^y \tag{2.6}$$

Also we extract as before from the Hilbert space the factor $[w(w-1)]^{\sigma/2}$ and define new operators $L_{\mu\nu}$ such that

$$L''_{\mu\nu}[w(w-1)]^{\sigma/2} R = [w(w-1)]^{\sigma/2} L_{\mu\nu} R. \tag{2.7}$$

In the variables w and t we get

$$L_{12} = t\frac{\partial}{\partial t},$$

$$L_{14} = \frac{j}{2}\left[t\left(w\frac{\partial}{\partial w} - t\frac{\partial}{\partial t} + \sigma\right) + \frac{1}{t}\left((w-1)\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} + \sigma\right)\right],$$

$$L_{24} = \tfrac{1}{2}\left[-t\left(w\frac{\partial}{\partial w} - t\frac{\partial}{\partial t} + \sigma\right) + \frac{1}{t}\left((w-1)\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} + \sigma\right)\right],$$

$$L_{34} = i\left(2w(w-1)\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} + 2(\sigma - k)w + k - \sigma\right),$$

$$L_{32} = \frac{[w(t^2-1)+1](2w-1)}{2t}\frac{\partial}{\partial w} + \frac{t^2-1}{2}\frac{\partial}{\partial t} - k\frac{w(t^2-1)+1}{t} + \sigma\frac{(t^2-1)(2w-1)}{2t},$$

$$L_{13} = i\left(\frac{[w(t^2+1)-1](2w-1)}{2t}\frac{\partial}{\partial w} + \frac{t^2+1}{2}\frac{\partial}{\partial t} - k\frac{w(t^2+1)-1}{t} + \sigma\frac{(t^2+1)(2w-1)}{2t}\right). \quad (2.8)$$

We define the operators $J_3, J^{\pm}$ by

$$L_{12} = J_3, L_{14} = -(i/2)(J^+ + J^-), \quad L_{24} = \tfrac{1}{2}(J^+ - J^-). \quad (2.9)$$

From (2.8) and (2.9) we get

$$J_3 = t\frac{\partial}{\partial t},$$

$$J^+ = -t\left(w\frac{\partial}{\partial w} - t\frac{\partial}{\partial t} + \sigma\right),$$

$$J^- = -\frac{1}{t}\left((w-1)\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} + \sigma\right), \quad (2.10)$$

which satisfy the commutation relations

$$[J_3, J^{\pm}] = \pm J^{\pm}, \quad [J^+, J^-] = 2J_3. \quad (2.11)$$

Consider the $SO(2,1)$ subalgebra with basis L_{12}, L_{24}, and L_{41}. The Casimir operator C is

$$C = L_{12}^2 - L_{14}^2 - L_{24}^2 = J_3^2 + J^+J^- - J_3. \quad (2.12)$$

We find

$$C = w(w-1)\frac{\partial^2}{\partial w^2} + t\frac{\partial^2}{\partial w \partial t} + (\sigma+1)(2w-1)\frac{\partial}{\partial w} + \sigma(\sigma+1). \quad (2.13)$$

Since L_{12}, L_{24}, and L_{41} generate an $SO(2,1)$ group we must have

$$[C - u(u+1)]R(w,t) = 0, \quad (2.14)$$

where $R(w,t)$ are the basis of the representations. If we write

$$R(w,t) = \Phi_m(w)t^m, \quad (2.15)$$

Eq. (2.14) gives

$$\left(w(1-w)\frac{d^2}{dw^2} + [\sigma - m + 1 - 2(\sigma+1)w]\frac{d}{dw} - (\sigma-u)(\sigma+u+1)\right)\Phi_m(w) = 0. \quad (2.16)$$

This is the hypergeometric equation. If $\sigma - m + 1 \neq 0, -1, -2, \cdots$, its solution which is regular at $w = 0$ is

$$\Phi_m(w) = F(\sigma-u, \sigma+u+1; \sigma-m+1; w). \quad (2.17)$$

Its second solution which is not regular at $w = 0$ is

$$\Phi_m(w) = [w^{m-\sigma}/(m-\sigma)]F(m-u, m+u+1; m-\sigma+1; w), \quad (2.18)$$

where the denominator $m - \sigma$ has been inserted for convenience. Since $[C, J_3] = [C, J^{\pm}] = 0$ the operators J_3, J^+, and J^- transform the functions $R(w,t)$ among themselves. If we define $|u,m\rangle$ by

$$|u,m\rangle = F(\sigma-u, \sigma+u+1; \sigma-m+1; w)t^m, \quad (2.19)$$

we find from the explicit expressions of the operators J_3, J^+, and J^-

$$J_3|u,m\rangle = m|u,m\rangle,$$
$$J^+|u,m\rangle = (m-\sigma)|u,m+1\rangle,$$
$$J^-|u,m\rangle = -\frac{(u+m)(u-m+1)}{\sigma-m+1}|u,m-1\rangle. \quad (2.20)$$

We also find that the operators L_{34}, L_{13}, and L_{23} transform the functions $|u,m\rangle$ among themselves. To simplify the notation let us define the operators $H^{\pm}$ by

$$H^{\pm} = L_{23} \pm iL_{13}. \quad (2.21)$$

Then we get

$$H^+|u,m\rangle = \frac{(\sigma-m)\sigma(k+1)}{u(u+1)}|u,m+1\rangle - \frac{(\sigma-m)(u+\sigma+1)(k-u)}{(u+1)(2u+1)}|u+1,m+1\rangle + \frac{(\sigma-m)(u-\sigma)(u+k+1)}{u(2u+1)}|u-1,m+1\rangle,$$

$$H^-|u,m\rangle = \frac{(u+m)(u-m+1)\sigma(k+1)}{u(u+1)(\sigma-m+1)}|u,m-1\rangle + \frac{(u+\sigma+1)(u-m+1)(k-u)(u-m+2)}{(u+1)(2u+1)(\sigma-m+1)}|u+1,m-1\rangle - \frac{(u+m)(u-\sigma)(u+k+1)(u+m-1)}{u(2u+1)(\sigma-m+1)} \times |u-1,m-1\rangle,$$

$$L_{34}|u,m\rangle = i\left(\frac{m\sigma(k+1)}{u(u+1)}|u,m\rangle + \frac{(u+\sigma+1)(u-m+1)(k-u)}{(u+1)(2u+1)}|u+1,m\rangle + \frac{(u-\sigma)(u+k+1)(u+m)}{u(2u+1)}|u-1,m\rangle\right). \quad (2.22)$$

Equations (2.20) and (2.22) tell us that the functions $|u,m\rangle$ form a basis of the representation of the operators $L_{\mu\nu}$, which give after complexification the algebra D_2. Therefore a six-dimensional Lie algebra was introduced, which can be used to treat the hypergeometric functions by Lie theory techniques.

III. EXTENSION OF MILLER'S ALGEBRA

The Lie theory treatment of the hypergeometric functions has been extensively discussed by Miller.[5] His approach is based on the Lie algebra $sl(2)$ with basis J^+, J^-, and J_3 satisfying the commutation relations (2.11). His operators can also be derived from the $SO(3,1)$ generators of Eqs (2.1). To do that we introduce the operators $L_{\mu\nu}$ as in (2.4), we make the transformation (2.5), we make the transformation

$$w = [\cos(x/2)]^{-2}, \quad t = \{1 - [\cos(x/2)]^{-2}\}^{1/2}e^y \quad (3.1)$$

and we extract from the Hilbert space the factor $w^{-u}(1-w)^{\sigma/2}$ defining new operators $N_{\mu\nu}$ such that

$$L''_{\mu\nu} w^{-u}(1-w)^{\sigma/2} R'' = w^{-u}(1-w)^{\sigma/2} N_{\mu\nu} R''. \tag{3.2}$$

The operators $N_{\mu\nu}$ become in the variables w and t

$$N_{12} = t\frac{\partial}{\partial t},$$

$$N_{14} = -\frac{i}{2}\left(t(w\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} - u) + \frac{1}{t}[w(1-w)\frac{\partial}{\partial w} - t\frac{\partial}{\partial t} + (u-\sigma)w - u]\right),$$

$$N_{24} = \tfrac{1}{2}\left(t(w\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} - u) - \frac{1}{t}[w(1-w)\frac{\partial}{\partial w} - t\frac{\partial}{\partial t} + (u-\sigma)w - u]\right),$$

$$N_{34} = i\left(2(w-1)\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} + 2(u-k)\frac{1}{w} - 2u + k + \sigma\right),$$

$$N_{23} = \frac{(t^2-1+w)(2-w)}{2t}\frac{\partial}{\partial w} - \frac{t^2-1}{2}\frac{\partial}{\partial t} - u\frac{(t^2-1+w)(2-w)}{2tw} + k\frac{t^2-1+w}{tw} + \sigma\frac{2-w}{2t},$$

$$N_{13} = i\left(-\frac{(t^2+1-w)(2-w)}{2t}\frac{\partial}{\partial w} + \frac{t^2+1}{2}\frac{\partial}{\partial t} + u\frac{(t^2+1-w)(2-w)}{2tw} - k\frac{t^2+1-w}{tw} + \sigma\frac{2-w}{2t}\right). \tag{3.3}$$

Writing $N_{12} = J'_3, N_{14} = -(i/2)(J'^{+} + J'^{-}), N_{24} = \frac{1}{2}(J'^{+} - J'^{-})$ we get

$$J'_3 = t\frac{\partial}{\partial t},$$

$$J'^{+} = t\left(w\frac{\partial}{\partial w} + t\frac{\partial}{\partial t} - u\right),$$

$$J'^{-} = \frac{1}{t}\left(w(1-w)\frac{\partial}{\partial w} - t\frac{\partial}{\partial t} + (u-\sigma)w - u\right), \tag{3.4}$$

which are the three operators used by Miller. Their connection with the hypergeometric functions is easily seen if we consider the equation

$$[N_{12}^2 - N_{14}^2 - N_{24}^2 - u(u+1)]f(w,t) = \left(w^2(1-w)\frac{\partial^2}{\partial w^2} - w^2 t\frac{\partial^2}{\partial w \partial t} + w[-2u + w(2u-\sigma-1)]\frac{\partial}{\partial w} + wt(u-\sigma)\frac{\partial}{\partial t} - wu(u-\sigma)\right) f(w,t) = 0 \tag{3.5}$$

and the expansion

$$f(w,t) = \sum_m h_m(w) t^m. \tag{3.6}$$

We find that $h_m(w)$ must be a solution of the hypergeometric equation

$$\left(w(1-w)\frac{d^2}{dw^2} + [-2u - w(m-2u+\sigma+1)]\frac{d}{dw} - (\sigma-u)(m-u)\right) h_m(w) = 0. \tag{3.7}$$

Therefore if $2u$ is not an integer the function $h_m(w)$ is a linear combination of the solutions $F(\sigma-u, m-u; -2u; w)$ and $[w^{2u+1}/(2u+1)]F(\sigma+u+1, m+u+1; 2u+2; w)$. We see that a treatment of the hypergeometric functions by methods of Lie theory can be based on the operators[5] N_{12}, N_{14}, and N_{24}.

In the present case the operators N_{34}, N_{23}, and N_{13} cannot be used to treat the hypergeometric functions by Lie theory techniques. To show that consider the functions

$$|u,m\} = w^{-u}F(\sigma-u, m-u; -2u; w)t^m. \tag{3.8}$$

From the series expansion of the hypergeometric functions we find

$$(1-w)\frac{d}{dw}F(a,b;c;w) = \frac{ab}{c}F(a,b;c;w) - \frac{ab(c-a)(c-b)}{c^2(c+1)}wF(a+1,b+1;c+2;w), \tag{3.9}$$

$$\left(1-\frac{2}{w}\right)F(a,b;c;w) = \frac{(c-2a)(c-2b)}{c(c-2)}F(a,b;c;w) - \frac{2ab(c-a)(c-b)}{c^2(c^2-1)}wF(a+1,b+1;c+2;w) - \frac{2}{w}F(a-1,b-1;c-2;w). \tag{3.10}$$

Then we get from (3.3), (3.9), and (3.10)

$$N_{34}|u,m\} = i\left(\frac{m\sigma(k-u+1) - u^2(u+1)}{u(u+1)}|u,m\} - \frac{(\sigma^2-u^2)(m^2-u^2)(k+1)}{2u^2(4u^2-1)}|u-1,m\} + 2(2u-k)|u+1,m\}\right). \tag{3.11}$$

Therefore the operator N_{34} transforms among themselves the functions $|u,m\}$ of Eq. (3.8) and not the functions $F(\sigma-u, m-u; -2u; w)t^m$. Using Eq. (3.11) and the commutation relations of the operators $N_{\mu\nu}$, we find again that the operators N_{23} and N_{13} do not transform the functions $F(\sigma-u, m-u; -2u; w)t^m$ among themselves. Therefore, contrary to the case of the operators $L_{\mu\nu}$, the extended Miller's algebra formed by the operators $N_{\mu\nu}$, $\mu, \nu = 1, \ldots, 4$, cannot be used to treat the hypergeometric functions by Lie theory techniques. If, however, we restrict ourselves to the subset N_{12}, N_{14}, and N_{24}, several nice relations involving the hypergeometric functions can be derived.[5]

IV. INCOMPLETE BETA FUNCTIONS

So far we have considered only the solutions of the hypergeometric equations (2.16) and (3.7) which are regular at $w = 0$. In this section we shall consider their solutions which are not regular at $w = 0$, i.e., the functions

$$I_{u,m}(w,t) = [w^{m-\sigma}/(w-\sigma)]F(m-u, m+u+1; m-\sigma+1; w)t^m \tag{4.1}$$

in connection with the operators $L_{\mu\nu}$ of Eqs (2.8), and the functions

$$I'_{u,m}(w,t) = w^{2u+1}/(2u+1)]F(\sigma+u+1, m+u+1; 2u+2; w)t^m \tag{4.2}$$

in connection with the operators $N_{\mu\nu}$ of Eqs. (3.3).

The incomplete beta function $B_w(p,q)$ is given by[7]

$$B_w(p,q) = \int_0^w d\xi\ \xi^{p-1}(1-\xi)^{q-1} = (w^p/p)F(p, 1-q; p+1; w). \tag{4.3}$$

Therefore, if $\sigma = u$, we get

$$I_{u,m}(w,t) = B_w(m-u,-m-u)t^m, \tag{4.4}$$

$$I'_{u,m}(w,t) = B_w(2u+1,-m-u)t^m. \tag{4.5}$$

It is obvious that the operators L_{12}, L_{14}, L_{24} transform the functions $I_{u,m}(w,t)$ among themselves. More specifically we have

$$\begin{aligned} J_3 I_{u,m} &= m I_{u,m}, \\ J^+ I_{u,m} &= -\frac{(m-u)(m+u+1)}{m-\sigma} I_{u,m+1}, \\ J^- I_{u,m} &= (m-\sigma-1) I_{u,m-1}. \end{aligned} \tag{4.6}$$

Also we find

$$\begin{aligned} L_{34} I_{u,m} = i \Bigg(& \frac{m\sigma(k+1)}{u(u+1)} I_{u,m} \\ & + \frac{(u+m+1)(u-\sigma+1)(k-u)}{(u+1)(2u+1)} I_{u+1,m} \\ & + \frac{(u-m)(u+k+1)(u+\sigma)}{u(2u+1)} I_{u-1,m} \Bigg). \end{aligned} \tag{4.7}$$

From Eqs. (4.6), (4.7), and the commutation relations of the operators $L_{\mu\nu}$ we easily find that the operators L_{13} and L_{23} also transform the functions $I_{u,m}$ among themselves. Therefore we see that the whole set of generators $L_{\mu\nu}$ transforms the functions $I_{u,m}(w,t)$, which for $\sigma = u$ are the product of an incomplete beta function times t^m, among themselves. Using this algebra one can treat the incomplete beta functions by Lie theory techniques. For $w = 1$ the incomplete beta functions become complete. The amplitudes in the Veneziano model[8] are expressed in terms of complete beta functions.

Again it is obvious that the operators J'_3, J'^+, and J'^- transform the functions $I'_{u,m}(w,t)$ among themselves. However the other three operators N_{13}, N_{23}, and N_{34} do not have this property.

* Present address.

[1]L. Weisner, Pac. J. Math. 5, 1033 (1955).
[2]L. Weisner, Can. J. Math. 11, 141, 148 (1959).
[3]W. Miller, Mem. Am. Math. Soc. 50, (1964).
[4]W. Miller, Commun. Pure Appl. Math. 17, 527 (1964); 18, 493 (1964); 18, 679 (1965); 19, 125 (1966).
[5]W. Miller, *Lie Theory and Special Functions* (Academic, New York, 1968).
[6]B. Kaufman, J. Math. Phys. 7, 447 (1966).
[7]Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, p. 87.
[8]G. Veneziano, Nuovo Cimento A 57, 190 (1968).

Fuzzy observables in quantum mechanics*

S. Twareque Ali

Department of Physics and Astronomy, The University of Rochester, Rochester, New York 14627

Gérard G. Emch

Departments of Mathematics, and of Physics and Astronomy, The University of Rochester, Rochester, New York 14627

(Received 26 July 1973)

The formalism of covariant conditional expectations is described as leading to an operational definition of generalized observables in quantum mechanics, wide enough to account for the fuzziness inherent in actual measurement processes, relative to a multidimensional physical continuum. As an application, a position operator for the photon is defined and its intrinsic fuzziness is discussed.

1. INTRODUCTION

The aim of this paper is to present a theoretical analysis of measurements where residual imprecisions are allowed. The fact that actual measurements are always imprecise is well-known and led Poincaré[1] to distinguish carefully the "mathematical continuum" from the "physical continuum." In the mathematical continuum the notion of identity satisfies the usual transitivity condition, i.e., for three points A, B, and C of the mathematical continuum:

$$A = B, B = C \quad \text{implies} \quad A = C. \tag{1.1}$$

By contrast, this property cannot be assumed for the notion of "indistinguishability" in the physical continuum attached to the raw data of experiments. Indeed, for any two elements E_1 and E_n of the physical continuum, there exists a sequence $\{E_1, E_2, \ldots, E_{n-1}, E_n\}$ of consecutive elements such that, for $i = 1, 2, \ldots, n-1, E_i$ is experimentally indistinguishable from E_{i+1}, although E_1 and E_n are distinguishable from one another; we have

$$E_1 \equiv E_2,\ E_2 \equiv E_3, \ldots,\ E_{n-1} \equiv E_n, \text{ and } E_1 \not\equiv E_n. \tag{1.2}$$

To pass from the physical continuum to the mathematical continuum requires an idealization, namely that infinitely precise measurements are in principle, if not in fact, attainable. In an effort to avoid this perhaps dubious idealization, Zeeman[2] suggested a consideration of what has come to be called "tolerance spaces" or[3] "fuzzy geometry."

Aside from the fact that actual measurements are always imprecise there are at least two, perhaps even more compelling, reasons why a theory of fuzzy observations should be developed. First, in the usual formalism of quantum mechanics—linked as it is to the notion of a mathematical continuum, rather than a physical continuum—one cannot even formulate an unbiased test of whether there exists an elementary length in nature. Second, the usual formalism leads to some serious difficulties when one wishes to give a theoretical account of position correlation experiments with photons. The modifications, that the existence of an elementary length would imply, in the algebraic formulation of quantum mechanics have recently been explored by Jordan.[4] The possible existence of observables which do not admit dispersion-free states led him to give up the power-associative law:

$$(A \circ A) \circ (A \circ A) = A \circ (A \circ (A \circ A)), \tag{1.3}$$

in general, for an arbitrary observable A of the system. The problem of the position operator for the photon has been the subject of several papers in the last ten years. Wightman[5] showed that a position operator in the usual sense cannot be defined for the photon, a relativistic elementary particle of mass zero and spin one. Jauch and Piron,[6] and Amrein[7] suggested a genuine mathematical generalization of the concept of localizability; they constructed as an archetype of their theory a localization operator for the photon. We shall see that the position observable for the photon appears as an intrinsically "fuzzy observable," and that this term can be given a physically motivated and mathematically precise meaning, using the formalism suggested by Davies and Lewis.[8]

Our paper is organized in the following manner: In Sec. 2 we explore in some detail the essential features of a simple example, and indicate how fuzzy observables can be acommodated within a natural extension of the formalism of quantum mechanics. In Sec. 3 we recall the definitions and a theorem pertinent to "operationally defined observables"; these encompass fuzzy observables as a genuine generalization of the usual concept of (fuzzy-free) observables. In Sec. 4 we study the restrictions imposed upon the observables by the requirement of covariance under certain symmetry groups. The reader's attention is called to our operator T_x, defined in this section. Finally, in Sec. 5, we apply these results to two particular situations, the second one leading to an operationally acceptable definition of a covariant "fuzzy position operator" for the photon.

2. PRELIMINARY EXAMPLES OF FUZZY MEASUREMENTS

To motivate our discussion, consider a massive, spinless particle, constrained to move on an infinite line $\mathbb{R}$, and assume that we are interested in making a measurement of its position.

Let $L^2(\mathbb{R})$ be the space of square-integrable, complex valued functions on $\mathbb{R}$, and Q the usual position operator defined by

$$(Q\psi)(x) = x\psi(x), \tag{2.1}$$

for all x in $\mathbb{R}$ and all ψ in $\mathfrak{D}(Q)$, the dense domain in $L^2(\mathbb{R})$ of the self-adjoint operator Q. We have then, by the spectral theorem:[9]

$$Q = \int_{\mathbb{R}} x\, P(dx), \tag{2.2}$$

where, for every Borel set E in $\mathbb{R}$, the projector $P(E)$ is defined by

$$(P(E)\psi)(x) = \chi_E(x)\psi(x), \tag{2.3}$$

for all ψ in $L^2(\mathbb{R})$, χ_E being the characteristic function of E, i.e.,

$$\chi_E(x) = 1, \quad \text{if } x \in E,$$
$$= 0, \quad \text{otherwise.} \tag{2.4}$$

According to the usual rules of quantum mechanics, the expectation value $\langle \Psi; Q \rangle$ of Q, when our system is in the state Ψ associated to a normalized vector ψ in $\mathfrak{D}(Q)$ is given by

$$\langle \Psi; Q \rangle = (Q\psi, \psi), \tag{2.5}$$

which we can rewrite as

$$\langle \Psi; Q \rangle = \int_{\mathbb{R}} x\, p_\psi(dx), \tag{2.6}$$

where $p_\psi(E)$, defined by

$$p_\psi(E) = (P(E)\psi, \psi), \tag{2.7}$$

is naturally interpreted as the probability that the particle be found in E when the system is in the state Ψ.

Clearly, the above interpretation requires that E be sharply defined. Suppose now that our measuring apparatus has a finite resolution Δ, i.e., it cannot distinguish between two points which are separated by a distance less than Δ. This inaccuracy will be reflected in the location of the midpoint of the set E, for example, and consequently, we shall not obtain the quantities $p_\psi(E)$ by our experimental procedure any longer. Rather, we shall obtain average quantities $p_\psi^\Delta(E)$, over some appropriate distribution.

Let x_0 be the coordinate of the midpoint of the interval E. Then, experimentally x_0 is uncertain by the amount Δ, so that an experimental determination of the set E would in general lead to a set $E_{x'-x_0}$ with midpoint x' located somewhere in the interval $[x_0 - \Delta/2, x_0 + \Delta/2]$. (Here, by E_x we mean the translate of the set E by the distance x.) Better still, may we assume that the observed midpoint x' of $E_{x'-x_0}$ is distributed on the real line $\mathbb{R}$ according to the probability density $x \mapsto f_{x_0}^\Delta(x)$, centered at x_0 and having a standard deviation Δ. We shall assume $f_{x_0}^\Delta$ to be symmetric around x_0. Let $x \mapsto f^\Delta(x)$ be the corresponding distribution around the origin, i.e.,

$$f_{x_0}^\Delta(x) = f^\Delta(x - x_0).$$

Then, $x \mapsto f^\Delta(x)$ satisfies

(I) $f^\Delta(x) \geq 0$, for all x in $\mathbb{R}$;

(II) $f^\Delta(x) = f^\Delta(-x)$, for all x in $\mathbb{R}$;

(III) $\int_{\mathbb{R}} f^\Delta(x)\, dx = 1$;

(IV) $\int_{\mathbb{R}} x^2 f^\Delta(x)\, dx = \Delta^2$.

Besides, we should also assume that f^Δ be strongly peaked around the origin, if our determination of the midpoint x' of $E_{x'-x_0}$ is at all to be meaningful. Actually we shall require that

(V) $\int_{\mathbb{R}} x^{2n} f^\Delta(x)\, dx \leq O(\Delta^{2n})$, for all integers $n \geq 0$,

a condition which is clearly satisfied, for instance, for the Gaussian distribution

$$f^\Delta(x) = (\Delta\sqrt{2\pi})^{-1} \exp[-\tfrac{1}{2}(x/\Delta)^2],$$

and for the flat distribution

$$f^\Delta(x) = (1/\Delta')\, \chi_{[-1/2\Delta', 1/2\Delta']},$$

with $\Delta = \Delta'/2\sqrt{3}$.

Assuming such a distribution f^Δ, we may write

$$p_\psi^\Delta(E) = \int_{\mathbb{R}} f_{x_0}^\Delta(x')\, p_\psi(E_{x'-x_0})\, dx'$$
$$= \int_{\mathbb{R}} f^\Delta(x')(P(E_{x'})\psi, \psi)\, dx'.$$

Using Eq. (2.4) it is then easy to see that

$$p_\psi^\Delta(E) = \int_{\mathbb{R}} \chi_E * f^\Delta(x) \mid \psi(x) \mid^2 dx$$
$$= (a^\Delta(E)\psi, \psi), \tag{2.8}$$

where the $*$ denotes convolution, defined by

$$g * f(x) = \int_{\mathbb{R}} g(y) f(x - y)\, dy,$$

and $a^\Delta(E)$ is the bounded, positive, linear operator on $\mathfrak{K}$, defined as:

$$(a^\Delta(E)\psi)(x) = \chi_E * f^\Delta(x)\psi(x), \tag{2.9}$$

for all ψ in $\mathfrak{K}[\equiv L^2(\mathbb{R})]$.

Equation (2.8) makes it clear that if our position measurements are not infinitely precise, we obtain, not the projection operators $P(E)$, but rather the more general positive operators $a^\Delta(E)$. It is easily checked that the $a^\Delta(E)$'s satisfy all the properties of being a positive operator valued (POV) measure,[10] viz.,

(I) $a^\Delta(\phi) = 0$, ϕ being the null set,
$a^\Delta(\mathbb{R}) = I$;

(II) $a^\Delta(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} a^\Delta(E_i)$, for disjoint sets E_i

(strong convergence being meant).

Thus, as in Eq. (2.2), we may now construct the operationally determined position observable Q^Δ:

$$Q^\Delta = \int_{\mathbb{R}} x\, a^\Delta(dx). \tag{2.10}$$

This operator acts on a vector ψ in $\mathfrak{K}$ in the following manner:

$$(Q^\Delta\psi)(x) = x^\Delta\psi(x), \tag{2.11}$$

where, as expected, x^Δ is the average

$$x^\Delta = \int_{\mathbb{R}} x' f_x^\Delta(x')\, dx'$$

of x with respect to the probability density distribution f_x^Δ. Since we took f_x^Δ to be symmetric about x, x^Δ is x itself, so that in fact,

$$(Q^\Delta\psi)(x) = x\psi(x). \tag{2.12}$$

Since the probability distribution $p(E)$ has been replaced by $p^\Delta(E)$, an operationally defined function A^Δ of the position will in general be different from the corresponding fuzzy-free observable A. To make this precise let $\mathfrak{A} = \{Q\}''$ be the von Neumann algebra generated by the projection operators $P(E)$ defined in Eq. (2.3). Then, as is well known,[11] $\mathfrak{A}$ consists of operators A of the form

$$(A\psi)(x) = A(x)\psi(x), \tag{2.13}$$

where, $A: x \in \mathbb{R} \mapsto A(x) \in \mathbb{C}$, is a bounded Borel function on $\mathbb{R}$. $\mathfrak{A}$ is thus isomorphic to the $*$-algebra $L^\infty(\mathbb{R})$ of all bounded measurable functions on $\mathbb{R}$. An arbitrary nor-

mal state ρ on A can then be identified with a positive function p in $L^1(\mathbb{R})$ via

$$\langle \rho; A \rangle = \int_{\mathbb{R}} A(x) p(dx) = \int_{\mathbb{R}} A(x) p(x)\, dx. \tag{2.14}$$

In line with what has been said above, we associate to every normal state ρ on $\mathfrak{A}$ the fuzzy state ρ^Δ, defined by the distribution

$$p^\Delta = p * f^\Delta. \tag{2.15}$$

Correspondingly we can associate to every observable A in $\mathfrak{A}$ a fuzzy observable A^Δ, defined operationally by

$$\langle \rho; A^\Delta \rangle \equiv \langle \rho^\Delta; A \rangle, \tag{2.16}$$

for all normal states on $\mathfrak{A}$. Clearly,

$$A^\Delta(x) = A * f^\Delta(x) \quad \text{in } L^\infty(\mathbb{R}) \tag{2.17}$$

and thus:

$$A^\Delta = \int_{\mathbb{R}} x\, a^\Delta(dx) \quad \text{on } L^2(\mathbb{R}). \tag{2.18}$$

It is important for the sequel to notice that if one repeats the averaging procedure, one gets

$$[(A^\Delta)^\Delta](x) = (A * f^\Delta) * f^\Delta(x) \tag{2.19}$$

and the corresponding dispersion increases, namely to $\sqrt{2}\Delta$

For illustrative purposes, we formally extend (2.17) to define, for all positive integers m,

$$Q^{(m)}(x) \equiv (Q^m)^\Delta(x) = \sum_{n=0}^{m} \binom{m}{n} (-1)^n x^{m-n} \int_{\mathbb{R}} y^n f^\Delta(y)\, dy, \tag{2.20}$$

so that

$$Q^{(0)}(x) = 1,$$
$$Q^{(1)}(x) = x,$$
$$Q^{(2)}(x) = x^2 + \Delta^2$$

and, up to first order in Δ^2, for all $m \geq 2$,

$$Q^{(m)}(x) \simeq Q^m(x) + \tfrac{1}{2} m(m-1)\Delta^2 x^{m-2}.$$

From this the well-known error law:

$$\delta Q^m(x) \equiv [[Q^{(2m)} - (Q^{(m)})^2]/(Q^{(m)})^2]^{1/2} \simeq m(\Delta/x) = m\delta Q(x)$$

follows immediately, indicating that we are indeed on the right track.

The relation with Jordan's idea[4] is obtained upon equipping the real vector space $\mathfrak{A}^\Delta$ generated by our $Q^{(m)}$'s with the product law:

$$Q^{(m)} \circ Q^{(n)} = \sum_{(2\rho \leq m,n)}^{\rho} (2\rho)! \binom{m}{2\rho}\binom{n}{2\rho} Q^{(m+n-4\rho)} \Delta^{4\rho}. \tag{2.21}$$

We notice that with this product $\mathfrak{A}^\Delta$ does indeed become a power algebra, i.e., in particular

$$Q^{(1)} \circ Q^{(n)} = Q^{(n+1)}, \tag{2.22}$$

for all integers $n \geq 0$, but that it fails to satisfy the power associativity condition; we have

$$(Q \circ Q) \circ (Q \circ Q) - Q \circ (Q \circ (Q \circ Q)) = 2\Delta^4, \tag{2.23}$$

which vanishes [compare with Eq. (1.3)] only in the limit $\Delta \to 0$ of infinite precision. In this limit $Q^{(m)}$ reduces to Q^m (and more generally A^Δ reduces to A) and the product $\circ$ reduces to the ordinary product. We therefore obtain the expected result that, in the limit $\Delta \to 0$, $\mathfrak{A}^\Delta$ with its algebraic structure tends to the algebra generated by the ordinary powers of Q, and equipped with its usual power structure.

It might also be proper to notice at this point that Jordan[4] linked the failure of the power associative law to the fact that there do not exist dispersion free states on the algebra of observables of interest (here $\mathfrak{A}^\Delta$). We have here indeed that the admissible states ρ^Δ have a dispersion $\delta\rho^\Delta$ which is always bounded below by Δ.

One might now wonder whether the generalization *from* the projection valued measure $P(E)$, associated to $\mathfrak{A}$, to the positive operator valued measure $a^\Delta(E)$, associated to $\mathfrak{A}^\Delta$, can be axiomatized properly. This is precisely where the axiomatic approach to quantum probability, of Davies and Lewis,[8] can be used, and we shall come back to this in the next section. What is specific to the example studied here is that both the approximate observable $E \mapsto a^\Delta(E)$ (i.e., Q^Δ) and the exact observable $E \mapsto P(E)$ (i.e., Q) exist for the system. There are, however, physical situations where only an approximate operator $E \mapsto a^\Delta(E)$ can be defined. This happens, for example, in the case of a photon, for which a position operator is definable only as long as measurements are made with a finite precision. We shall construct such an operator in Sec. 5.

We ought to point out in passing that, in virtue of a theorem of Naimark, on the embedding of POV measures into projection valued (PV) measures,[12] it is possible to find an enlarged Hilbert space $\tilde{\mathcal{H}}$, a projection operator $\mathbb{P}$ and a PV measure $E \mapsto \tilde{P}(E)$ on R, extending $E \mapsto a^\Delta(E)$ in the following manner:

(I) $\mathcal{H} = \mathbb{P}\tilde{\mathcal{H}}$,

(II) $a^\Delta(E) = \mathbb{P}\tilde{P}(E)\mathbb{P}$.

Further, the space $\tilde{\mathcal{H}}$ can be chosen minimal in the sense that it is generated by elements of the sort $\tilde{P}(E)\psi$, where ψ is in $\mathcal{H}$. It is probably not always possible to give a physical meaning to the extended space $\tilde{\mathcal{H}}$. We shall construct, however in Sec. 5, a projection operator $\mathbb{P}$ for the photon system, following Jauch and Piron,[6] which will be seen to correspond to a gauge condition. In our present example, $\tilde{\mathcal{H}}$ is the direct integral space:

$$\tilde{\mathcal{H}} = \int_{\oplus} \mathcal{H}_{f_x^\Delta}\, dx, \tag{2.24}$$

where $\mathcal{H}_{f_x^\Delta}$ is the Hilbert space of all complex valued functions on $\mathbb{R}$ which are square integrable with respect to the measure μ_x, where $\mu_x(dy) = f_x^\Delta(y)\, dy$. An element $\tilde{\psi}$ in $\tilde{\mathcal{H}}$ is now a function of two real variables x and y such that $\tilde{\psi}(x) \in \mathcal{H}_{f_x^\Delta}$ and $\tilde{\psi}_y(x) \in \mathbb{C}$ (complex numbers). The scalar product in $\tilde{\mathcal{H}}$ is given as

$$(\psi, \chi)_{\tilde{\mathcal{H}}} = \int_{\mathbb{R}} (\psi(x), \chi(x))_{\mathcal{H}_{f_x^\Delta}}\, dx, \tag{2.25}$$

where $(\ ,\ )_{\mathcal{H}_{f_x^\Delta}}$ is the scalar product in $\mathcal{H}_{f_x^\Delta}$:

$$(\tilde{\psi}(x), \tilde{\chi}(x))_{\mathcal{H}_{f_x^\Delta}} = \int_{\mathbb{R}} \overline{\tilde{\psi}_y(x)}\, \tilde{\chi}_y(x) f_x^\Delta(y)\, dy. \tag{2.26}$$

The operators $\tilde{P}(E)$ and $\mathbb{P}$ are then

$$(\tilde{P}(E)\tilde{\psi})_y(x) = \chi_E(y)\tilde{\psi}_y(x), \tag{2.27}$$

and

$$(\mathbb{P}\tilde{\psi})_y(x) = (\mathbb{1}, \tilde{\psi}(x)_{\mathcal{H}_{f_{\hat{x}}^{\Delta}}} \mathbb{1}(y), \tag{2.28}$$

where $\mathbb{1}$ is the unit vector in $\mathcal{H}_{f_{\hat{x}}^{\Delta}}$ which takes the value $\mathbf{1}(y) = 1$, for all y in $\mathbf{R}$.

3. OBSERVABLES AND MEASUREMENT

We shall make now a systematic study of generalized observables, particular examples of which were mentioned in the last section. We shall extract in this section the notion of a generalized (operationally defined) observable, starting from a consideration of the "collapse of the wave packet" during a process of measurement. Much of the material of this section is known. For the sake of completeness however, we shall review it here in a manner which will prepare the way for our discussions of the next two sections.

Let us begin by introducing von Neumann's expression[13] for "collapse of the wave packet." Suppose that we wish to measure an observable A in a state of the physical system given by the vector ψ in the Hilbert space $\mathcal{H}$ (assumed separable). Let ρ_ψ denote the normalized density matrix $\psi \otimes \bar{\psi}$ corresponding to ψ. Suppose that A, as a self-adjoint operator on $\mathcal{H}$, has a purely discrete spectrum: $\{\lambda_i\}_{i\in\mathfrak{I}^+}$, $(\mathfrak{I}^+ = \{1, 2, 3, \ldots, \text{dimension of } \mathcal{H}\})$; let $\{\xi_i\}_{i\in\mathfrak{I}^+}$ be the corresponding eigenvectors and $\{P_i\}_{i\in\mathfrak{I}^+}$ the one dimensional projectors $\{\xi_i \otimes \bar{\xi}_i\}$. Then, as a result of the measurement, the state ρ_ψ changes into the new state $\rho^{\#}$ in the manner:

$$\rho_\psi \to \rho^{\#} = \sum_{i\in\mathfrak{I}^+} |(\xi_i, \psi)|^2 P_i = \sum_{i\in\mathfrak{I}^+} P_i \rho_\psi P_i . \tag{3.1}$$

The quantity

$$\begin{aligned} |(\xi_i, \psi)|^2 &= (P_i\psi, \psi) = \operatorname{tr}[P_i \rho_\psi] \\ &= \langle \rho_\psi ; P_i \rangle = p_i \end{aligned} \tag{3.2}$$

represents the probability of obtaining the eigenvalue λ_i as a result of the measurement. Clearly we have

$$\langle \rho_\psi ; B \rangle = \langle \rho^{\#} ; B \rangle , \tag{3.3}$$

for all B of the form $B = \sum_i b_i P_i$, as is indeed required for a consistent interpretation of $\rho^{\#}$ as the state of the system after the measurement of the observable $A = \sum_i \lambda_i P_i$ with which such B's commute. If our measuring apparatus is such that it suppresses all eigenstates of A whose corresponding eigenvalues do not lie in a given set E, the state ρ_ψ goes over, as a result of the measurement, into the "collapsed state" ρ^c:

$$\rho^c = \sum_{\lambda_i \in E} p_i P_i \tag{3.4}$$

where the p_i's are the probabilities defined in Eq. (3.2). If we measure the observable A in the arbitrary (perhaps mixed) state ρ, we again obtain the collapse expression (3.4) with the p_i's now given by

$$p_i = \operatorname{tr}[P_i \rho] \tag{3.5}$$

Equations (3.4) and (3.5) define von Neumann's expression for the collapse of the wave packet during the course of a measurement. Let us introduce the notation of Davies and Lewis[8] and write

$$\mathcal{E}(E, \rho) = \sum_{\lambda_i \in E} P_i \rho P_i . \tag{3.6}$$

It is this quantity $\mathcal{E}(E, \rho)$ which is basic to our discussion, and the one which shall be generalized below.

If $p_\rho(E)$ is the probability of finding an eigenvalue of A in the set E, when the system measured is initially in the state ρ, we have [on taking the trace of both sides of Eq. (3.6)]

$$p_\rho(E) = \operatorname{tr}[\mathcal{E}(E, \rho)] = \operatorname{tr}[P(E)\rho], \tag{3.7}$$

where

$$P(E) = \sum_{\lambda_i \in E} P_i$$

is the spectral projector of A corresponding to the set E. Clearly, Eq. (3.7) is the analogue of Eq. (2.7) for the general state ρ. Thus, a study of the quantity $\mathcal{E}(E, \rho)$ allows us to determine the observable A, once again via the spectral theorem:

$$A = \sum_i \lambda_i P_i = \int_{-\infty}^{\infty} \lambda P(d\lambda).$$

Next, we observe that for a fixed ρ the quantity $\mathcal{E}(E, \rho)$ satisfies all the properties of a vector-valued measure, taking values in the Banach space $\mathcal{T}(\mathcal{H})$ of all trace class operators on $\mathcal{H}$ (under the trace norm), viz.,

(I) $\rho \in \mathcal{T}(\mathcal{H})^+ \Rightarrow \mathcal{E}(E, \rho) \in \mathcal{T}(\mathcal{H})^+$,

(II) $\mathcal{E}(\phi, \rho) = 0$,

(III) $\mathcal{E}(\bigcup_{i\in\mathfrak{I}^+} E_i, \rho) = \sum_{i\in\mathfrak{I}^+} \mathcal{E}(E_i, \rho)$

(weak convergence in $\mathcal{T}(\mathcal{H})$ being meant).

In addition, $\mathcal{E}(E, \rho)$ satisfies the further condition

$$\operatorname{tr}[\mathcal{E}(\mathbf{R}, \rho)] = \operatorname{tr}\rho. \tag{3.8}$$

It is now clear in what sense the notion of a collapse ought to be generalized in order that the probabilities $p_\rho(E)$ may refer to observables with POV measures, as opposed to those with only PV measures. Following Davies and Lewis,[8] this generalization is provided through Definitions 1 and 2 below. From now on we shall denote by X a locally compact space, and by $K(X)$ the set of all continuous, complex valued functions on X with compact supports. [$K(X)$ is equipped with the standard inductive topology.[14]] We also note that a vector valued (Borel) measure $\mathbf{m}$ on X may be defined either as a σ-additive set function, on the Borel sets of X, which assumes values in some Banach space F, or, equivalently, as a continuous linear map

$$\mathbf{m}: K(X) \to F.$$

We shall use here this second definition of a measure.

Let $\mathcal{H}$ be a separable Hilbert space, $\mathfrak{L}(\mathcal{H})$ the set of all bounded operators on $\mathcal{H}$, and $\mathfrak{L}(\mathcal{H})^+$ the positive elements in $\mathfrak{L}(\mathcal{H})$.

Definition 1: A generalized observable (a, X), on a separable Hilbert space $\mathcal{H}$, is a normalized, positive operator valued measure $a: K(X) \to \mathfrak{L}(\mathcal{H})$, defined on some locally compact space X, with values in $\mathfrak{L}(\mathcal{H})$. X is called the value space of the observable (a, X).

The normalization of a: $K(X) \to \mathfrak{L}(\mathcal{H})$ is to be understood in the following sense: If f_n is a monotone sequence of elements in $K(X)$ tending pointwise to 1, then $a(f_n) \nearrow \mathrm{I}$, the identity operator on $\mathcal{H}$. The positivity of the measure a implies that $a(f) \in \mathfrak{L}(\mathcal{H})^+$ for $f \in K(X)^+$.

In this definition if we take, in particular, $\mathbf{R}$ for X and projection valued measures for a, we retrieve the standard observables of quantum mechanics.

Definition 2: A measurement $\mathcal{E}$ on X is a continuous bilinear map:

$$\mathcal{E}: K(X) \times \mathcal{T}(\mathcal{H}) \to \mathcal{T}(\mathcal{H}),$$

which satisfies

(I) $f \in K(X)^+$, $\rho \in \mathcal{T}(\mathcal{H})^+ \Rightarrow \mathcal{E}(f,\rho) \in \mathcal{T}(\mathcal{H})^+$,

(II) if f_n is a sequence in $K(X)$ such that $f_n \nearrow 1$, pointwise, then $\mathrm{tr}\,[\mathcal{E}(f_n,\rho)] \nearrow \mathrm{tr}\,\rho, \rho \in \mathcal{T}(\mathcal{H})^+$.

We have thus a generalized notion for the collapse through that of a measurement. We shall take the point of view that the quantities $\mathrm{tr}\,[\mathcal{E}(f,\rho)]$ for positive f and ρ define the experimentally observed probabilities when an observable is measured in the state ρ. The fact that we have indeed achieved a legitimate generalization is borne out by the following theorem (due to Davies and Lewis, c.f., Ref. 8).

Theorem: To every measurement $\mathcal{E}$ on X there corresponds a unique generalized observable (a,X) having X as its value space, and given by the formula

$$\mathrm{tr}\,[\mathcal{E}(f,\rho)] = \mathrm{tr}\,[a(f)\rho], \tag{3.9}$$

for all $f \in K(X)$, $\rho \in \mathcal{T}(\mathcal{H})$. Conversely, every generalized observable (a,X), with value space X, is so determined by at least one measurement on X.

Comparing Eq. (3.9) with Eq. (2.8) we immediately see how the above definition of a measurement allows us to construct the generalized observable $E \to a^{\Delta}(E)$ through an analysis of the collapse process. In the next two sections we shall indicate what experimental situations might bring about such a collapse. We also note that Definition 2, of a measurement, is general enough to include observables which need not have purely discrete spectra. This was not possible in von Neumann's definition of the collapse given in Eq. (3.4), since for an observable with a continuous spectrum the projectors P_i do not exist. The price we have to pay to achieve this generalization is that, unlike von Neumann's collapse expression, ours is not necessarily repeatable, i.e., $\mathcal{E}(f,\rho)$ is not necessarily the same as $\mathcal{E}(f, \mathcal{E}(f,\rho))$; compare, for instance, with Eq. (2.19).

4. ANALYSIS OF THE COLLAPSE EXPRESSION

This section is devoted to a discussion of a mathematical characterization of a measurement $\mathcal{E}$, when a certain group restriction is present. This characterization will help us to write down a collapse expression specific to a large class of laboratory measurement processes. In this way we shall be able to construct in the next section a covariant, "fuzzy" position operator for the photon, starting from localization measurements.

In the position measurement of the Schrödinger particle discussed in Sec. 2, consider again the probabilities $p_{\psi}^{\Delta}(E)$. If both the state vector ψ and the interval E are translated by an amount x, to $x[\psi]$ and E_x, respectively, we expect the new probability $p_{x[\psi]}^{\Delta}(E_x)$ to be the same as the old probability $p_{\psi}^{\Delta}(E)$. It is this sort of group invariance that we wish to build into our theory. If we assume unitary implementability in $\mathcal{H}$ of this translational symmetry, we notice that the measurement $\mathcal{E}$ ought to satisfy:

$$\mathcal{E}(E_x,\rho) = U_x \mathcal{E}(E, U_x^* \rho U_x) U_x^*$$

for translations through x, where U_x is the unitary operator on $\mathcal{H}$ implementing the transformation due to $x \in \mathbf{R}$.

Since in many interesting physical measurement situations—of position or momentum, for example—the relevant symmetry groups G are semidirect products of the form

$$G = H \circledS X, \tag{4.1}$$

we shall, without further ado, specialize to group symmetries such as these. We shall then speak about measurements of observables whose value spaces are the Abelian groups X themselves. Thus, in our previous example, G is just the one-dimensional translation group T, which is isomorphic to $\mathbf{R}$ itself. Similarly, for a position measurement in three-dimensional space, G is the Euclidean group $\mathbf{E}^3 = SU(2) \circledS T^3$ of all three-dimensional proper rigid rotations and translations, and T^3 is isomorphic to $\mathbf{R}^3$—the value space of the position operator.

Let us denote by $g[\cdot], g[\cdot]$, and $[\cdot]g$ the action of the symmetry $g \in G$ on $\mathcal{T}(\mathcal{H}), K(X)$ and X, respectively, i.e.,

$$g[f](x) = f([x]g), \quad \text{for all } x \in X \tag{4.2}$$

and, if these symmetries are unitarily implemented $(g \mapsto U_g)$,

$$g[\rho] = U_g^* \rho U_g . \tag{4.3}$$

In general, for an arbitrary measurement $\mathcal{E}$ on X, we shall assume the group covariance under the automorphisms of $G = H \circledS X$, in the state space $\mathcal{T}(\mathcal{H})$, to be of the form:

$$\mathcal{E}(g[f],\rho) = g^{-1}[\mathcal{E}(f, g[\rho])], \tag{4.4}$$

for all $g \in G$.

An immediate consequence of Eq. (4.4) is that the observable (a,X) determined by $\mathcal{E}$ satisfies the relation

$$a(g[f]) = U_g^* a(f) U_g \tag{4.5}$$

for all $g \in G, f \in K(X)$. Equation (4.5) is referred to as the generalized imprimitivity condition and (a,X) and $g \mapsto U_g$ are said to form a generalized system of imprimitivity. If $f \mapsto a(f)$ were a projection valued measure, then Eq. (4.5) would define a projective system of imprimitivity in the sense of Mackey.[15]

A measurement $\mathcal{E}$ on X which is subject to a covariance condition of the type given in Eq. (4.4), under a group G as in (4.1) is characterized[16] by a positive operator valued function

$$T: X \to \mathcal{L}(\mathcal{T}(\mathcal{H}), \mathcal{T}(\mathcal{H}))^+, \tag{4.6}$$

where $\mathcal{L}(\mathcal{T}(\mathcal{H}), \mathcal{T}(\mathcal{H}))^+$ is the space of all positive linear maps of $\mathcal{T}(\mathcal{H})$, the space all density matrices, into itself. Further, T is measurable with respect to the Haar measure μ on X, in the sense that for each $\rho \in \mathcal{T}(\mathcal{H})$ and $A \in \mathcal{L}(\mathcal{H})$, the numerical function $x \mapsto \mathrm{tr}\,[T_x(\rho)A]$ is measurable. Indeed, one has the relation

$$\mathcal{E}(f,\rho) = \int_X f(x)\, T_x(\rho)\, \mu(dx), \tag{4.7}$$

weak convergence in $\mathcal{T}(\mathcal{H})$ of the integral being meant. T_x also has the properties:

$$T_{x[g]}(\rho) = U_g^* T_x(U_g \rho U_g^*) U_g \tag{4.8}$$

and

$$\mathrm{tr}\left[\int_X T_x(\rho)\,\mu(dx)\right] = \mathrm{tr}\,\rho, \tag{4.9}$$

for all ρ in $\mathcal{T}(\mathcal{K})^+$.

The dual $\mathcal{E}^*(\cdot, A)$ of $\mathcal{E}(\cdot, \rho)$ defined by

$$\mathrm{tr}\left[\mathcal{E}^*(f, A)\rho\right] = \mathrm{tr}\left[A\mathcal{E}(f, \rho)\right],$$

for all A in $\mathcal{L}(\mathcal{K})$, all f in $K(X)$ and all ρ in $\mathcal{T}(\mathcal{K})$, will be referred to as the "covariant conditional expectation" of A, relative to the measurement process $\mathcal{E}$ satisfying Eq. (4.4).

The representation of $\mathcal{E}(f, \rho)$ through the operator function $x \mapsto T_x$, as given in Eq. (4.7) is very general, and contains several particular results obtained earlier by Davies.[17] At each point $x \in X$, the operator T_x describes the change that the state ρ undergoes due to the measurement being performed there. From our earlier association of projection valued measures with infinitely precise measurements, we intuitively expect therefore, that for $x \mapsto T_x$ to generate a projection valued measure, the operator T_x ought to change the state ρ only locally in an infinitesimally small neighborhood of x.

Mathematically we can demonstrate this as follows: Let us assume that the Hilbert space $\mathcal{K}$, in the problem, is of the form $\mathcal{L}^2_{\mathcal{K}}(X, \mu) = \{\psi: X \to \mathcal{K} \mid \int_X \|\psi(x)\|^2_{\mathcal{K}}$ $\mu(dx) < \infty\}$, where $\mathcal{K}$ is some Hilbert space, with norm denoted by $\|\cdots\|_{\mathcal{K}}$. Such a choice for $\mathcal{K}$ is always possible if (a, X) in Eq. (4.5) is projection valued—in virtue of the imprimitivity theorem of Mackey.[15] Let us denote by $\mathcal{L}^1_{\mathcal{T}(\mathcal{K})}(X, \mu)$ the space of all functions

$$r: X \to \mathcal{T}(\mathcal{K})$$

for which

$$\int_X \|r(x)\|_{\mathcal{T}(\mathcal{K})}\,\mu(dx) < \infty,$$

$\|\cdots\|_{\mathcal{T}(\mathcal{K})}$ being the norm in $\mathcal{T}(\mathcal{K})$, and let λ be the linear map:

$$\lambda: \mathcal{T}(\mathcal{K}) \to L^1_{\mathcal{T}(\mathcal{K})}(X, \mu),$$

which on elements in $\mathcal{T}(\mathcal{K})$ of the form $\psi \otimes \bar{\psi} (\psi \in \mathcal{K})$ is defined by

$$\{\lambda(\psi \otimes \bar{\psi})\}(x) = \psi(x) \otimes \overline{\psi(x)}.$$

For $\rho \in \mathcal{T}(\mathcal{K})^+$, the function $x \mapsto \|(\lambda\rho)(x)\|_{\mathcal{T}(\mathcal{K})} = \mathrm{tr}_{\mathcal{K}}$ $[(\lambda\rho)(x)]$ actually gives the probability density distribution of the wave packet represented by ρ. Thus, in our example of the particle on a line, having wavefunction $\psi, x \mapsto \|(\lambda\rho)(x)\|$ is indeed the usual probability function $x \mapsto |\psi(x)|^2$. It has been proved in Ref. 16 that the operators T_x generate an observable with a PV measure if and only if every T_x has a representation of the type:

$$T_x(\rho) = \int_X \mathbf{T}_x(dx')(\lambda\rho)(x') \tag{4.10}$$

where $\mathbf{T}_x$ is a positive vector valued measure on X with values in $\mathcal{L}(\mathcal{T}(\mathcal{K}), \mathcal{T}(\mathcal{K}))$ [= Banach space of bounded linear maps from $\mathcal{T}(\mathcal{K})$ to $\mathcal{T}(\mathcal{K})$], whose support is concentrated at the point x. Thus, for a measurement to have a PV observable, the collapse at each point x ought to affect the probability distribution at that point alone—hence the necessity for it to be infinitely precise.

5. APPLICATIONS OF THE GENERAL FORMALISM

We saw at the end of the preceding section, that the generalized observables defined in Sec. 3, reduce to the usual observables of quantum mechanics only when the measurements to which they correspond are of infinite precision. In this section we illustrate first the general meaning of our operator valued function T in terms of the fuzzy observables defined in Sec. 2. We next analyze the problem of position measurements for the photon.

Going back to our original example of a particle on a line, we may now construct the operator function $x \mapsto T_x$ for it, corresponding to measurements accurate only up to a length Δ'. Suppose that the system is originally in the state ρ, and suppose that we have a counter which, when introduced at the point x, measures the total intensity of the particle beam in the region $\Delta' x$ surrounding x. Suppose further, that after the measurement at x the counter gives out a beam consisting of a flat pulse over the region $\Delta' x$. Then the measured intensity in $\Delta' x$ is

$$\int_{\Delta' x} (\lambda\rho)(x')\,dx' = \mathrm{tr}\left[P_{\Delta' x}\rho\right],$$

where $P_{\Delta' x}$ is the projection operator $P(\Delta' x)$, corresponding to multiplication by the characteristic function $\chi_{\Delta' x}$ of the set $\Delta' x$. Hence the observed probability density at x is $(1/\Delta')\,\mathrm{tr}\left[P_{\Delta' x}\rho\right]$. After the measurement, the outgoing state at x is $(1/\Delta')\,\chi_{\Delta' x} \otimes \bar{\chi}_{\Delta' x}$. Thus

$$T_x(\rho)\left[1/(\Delta')^2\right]\mathrm{tr}\left[P_{\Delta' x}\rho\right]\chi_{\Delta' x} \otimes \bar{\chi}_{\Delta' x}. \tag{5.1}$$

It is now straightforward to verify that this T_x defines a measurement $\mathcal{E}$, through Eq. (4.7), with the correct group covariance properties, and further, that it also yields the observable $E \mapsto a^{\Delta}(E)$ defined in Eq. (2.9), where now f^{Δ} is the function:

$$f^{\Delta}(x) = (1/\Delta')\,\chi_{[-\Delta'/2, \Delta'/2]}(x)$$

centered at the origin, and having standard deviation

$$\Delta = \Delta'/2\sqrt{3}.$$

It is clear that taking f^{Δ} to be any other function would amount to introducing a certain bias in the counter over its sensitive volume.

Using Eq. (4.7) we shall next construct an approximate position operator for the photon. The Hilbert space appropriate to the photon system is defined as follows.[6]

Let $\tilde{\mathcal{K}} = L^2_{C^3}(\mathbf{R}^3, d^3\mathbf{x})$ denote the Hilbert space of all complex 3-vector valued, square integrable functions on $\mathbf{R}^3$. Denote by $\mathcal{K} = L^2_{C^3_{\mathbf{A}}}(\mathbf{R}^3, d^3\mathbf{x})_0$ the (closed) subspace of those vectors $\mathbf{A}$ in $\tilde{\mathcal{K}}$ which satisfy the divergence condition.

$$(\nabla \cdot \mathbf{A})(\mathbf{x}) = 0. \tag{5.2}$$

It is this space $\mathcal{K}$ which describes the photon. Let $\mathbb{P}$ denote the projection operator which, acting on $\tilde{\mathcal{K}}$, projects out the subspace $\mathcal{K}$:

$$\mathcal{K} = \mathbb{P}\tilde{\mathcal{K}}$$

so that, in virtue of Eq. (5.2), $\mathbb{P}$ corresponds to the Coulomb gauge condition. On $\tilde{\mathcal{K}}$ we may define the projection operators $\tilde{P}(E)$ in the manner:

$$(\tilde{P}(E)\tilde{A})(x) = \chi_E(\mathbf{x})\,\tilde{\mathbf{A}}(\mathbf{x}), \tag{5.3}$$

for all $\tilde{\mathbf{A}} \in \tilde{\mathcal{K}}$. However, these operators cannot be defined in $\mathcal{K}$, because even if $\mathbf{A}$ satisfies Eq. (5.2), $\chi_E\mathbf{A}$ need not do so any longer. Hence no position operator, in the usual sense, may be defined for the photon, and consequently, the quantity $\mathrm{tr}\left[\tilde{P}(E)\rho\right]$, for a state ρ of the

electromagnetic field, does not define the total probability of finding a photon in the region E.

It was suggested by Jauch and Piron (Ref. 6) that the photon be considered a weakly localizable system, i.e., a system for which it is not the operators $\tilde{P}(E)$ which define localization in the regions E but rather the new set of projection operators $F(E)$ given by

$$F(E) = \tilde{P}(E) \cap \mathbb{P} \tag{5.4}$$

and defined on $\mathfrak{K}$.

[$F(E)$ is the operator which projects onto the intersection of the ranges of $\mathbb{P}$ and $\tilde{P}(E)$.]
The operators $F(E)$ do not, however, form a PV measure. But they do satisfy the conditions:

(I) $F(\phi) = 0, \quad F(\mathbb{R}^3) = I,$

(II) $F(E_1) \perp F(E_2), \quad$ if $E_1 \cap E_2 = \phi,$

(III) $F(E_1 \cap E_2) = F(E_1) \cap F(E_2),$

and the imprimitivity condition

$$F(E_g - 1) = U_g F(E) U_g^*, \tag{5.5}$$

under the automorphisms of the Euclidean group $\mathbb{E}^3$. Thus, $\mathrm{tr}[F(E)\rho]$ is assumed to define the total field intensity over the region E.

Using this definition for the localization of photons, we shall write down a collapse expression for a position measurement, and hence derive an approximate position operator for the photon. Suppose that we introduce at the point $\mathbf{x}$ a probe (such as an antenna) into the electromagnetic field, and let this probe be sensitive to the field intensity measured over a small volume $\Delta\mathbf{x}$ surrounding $\mathbf{x}$. Suppose that initially the field is in a state ρ, and that the state at $\mathbf{x}$ after the measurement is given by the normalized density matrix $\rho_{\mathbf{x}}^{\Delta}$. We shall assume that $\rho_{\mathbf{x}}^{\Delta} = U_{\mathbf{x}}^* \rho_0^{\Delta} U_{\mathbf{x}}$, where $\mathbf{x} \mapsto U_{\mathbf{x}}$ is a unitary representation in $\mathfrak{K}$ of the three-dimensional translation group T^3, and ρ_0^{Δ} is the outgoing state after a measurement at the origin. The total field intensity in $\Delta\mathbf{x}$ is $\mathrm{tr}[F_{\Delta\mathbf{x}}\rho]$, where $F_{\Delta\mathbf{x}} \equiv F(\Delta\mathbf{x})$. However, the number density of photons observed by this measurement is not $(1/\Delta)\,\mathrm{tr}[F_{\Delta\mathbf{x}}\rho]$ but rather $\mathrm{tr}[\mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2} F_{\Delta\mathbf{x}} C^{\Delta}(\mathbb{R}^3)^{-1/2}\rho]$, where $\mathcal{C}^{\Delta}(\mathbb{R}^3)$ is the positive operator

$$\mathcal{C}^{\Delta}(\mathbb{R}^3) = \int_{\mathbb{R}^3} F_{\Delta\mathbf{x}}\, d^3\mathbf{x}. \tag{5.6}$$

The reason for this is that $(1/\Delta)\int_{\mathbb{R}^3} \mathrm{tr}[F_{\Delta\mathbf{x}}\rho] \neq \mathrm{tr}\rho$, for all $\rho \in \mathcal{T}(\mathfrak{K})^+$, and $\mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2}$ is needed as a normalizing factor. The operator $T_{\mathbf{x}}$ can again be written as in Eq. (5.1). Thus,

$$T_{\mathbf{x}}(\rho) = \mathrm{tr}[\mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2} F_{\Delta\mathbf{x}}\, \mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2}\rho]\rho_{\mathbf{x}}^{\Delta}, \tag{5.7}$$

so that the corresponding position operator $E \mapsto a^{\Delta}(E)$ is

$$a^{\Delta}(E) = \mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2} \int_E F_{\Delta\mathbf{x}}\, d^3\mathbf{x}\, \mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2} \tag{5.8}$$

(the integral converging ultraweakly in $\mathfrak{K}$).

Equation (5.7) may be interpreted by saying that the interaction of the electromagnetic field with the probe renormalizes the state ρ into the new state $\rho' = \mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2}\rho\mathcal{C}^{\Delta}(\mathbb{R}^3)^{-1/2}$, and it is the intensity at $\mathbf{x}$ of this state ρ' that the probe measures. As a final remark, we emphasize once again that we have been able to define a position operator for the photon only because we were dealing with fuzzy measurements of the field intensity. An infinitely precise measurement of intensity in any region E ought to lead to the quantity $\mathrm{tr}[F(E)\rho]$, and hence no position observable would be definable then. Our result also clears up the apparent anomaly between the experimental situation, where photon space-correlation measurements do implicitly assume the existence of a position operator for it, while no such operator could be predicted theoretically so far. We shall report elsewhere the relationship between our position operator and the observed photon space correlations.

ACKNOWLEDGMENTS

In the course of the investigation reported in this paper, we benefited from discussions with Professor W. F. Eberlein, Mrs. B. Whitten-Wolfe, and Dr. J. C. Wolfe.

* Based in part on the Ph.D. dissertation submitted by one of us (S.T.A.) to the Physics Department of the University of Rochester.
[1]H. Poincaré, *La science et l'hypothèse* (Flammarion, Paris, 1903); the same ideas were pursued again in his subsequent book, *La valeur de la science* (Flammarion, Paris, 1904).
[2]C. Zeeman, in *Topology of 3-manifolds*, edited by K. Fort (Prentice-Hall, Englewood Cliffs, N.J., 1961), p. 240.
[3]T. Poston, "Fuzzy Geometry". Ph.D. thesis, University of Warwick (1971), unpublished; a colorful introduction may be found in Manifold **10**, 25 (1971).
[4]P. Jordan, Commun. Math Phys. **9**, 279 (1968); **11**, 293 (1969).
[5]A. Wightman, Rev. Mod. Phys. **34**, 845 (1962).
[6]J. M. Jauch and C. Piron, Helv. Phys. Acta **40**, 559 (1967).
[7]W. O. Amrein, Helv. Phys. Acta **42**, 149 (1969).
[8]E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970).
[9]See, for instance, N. I. Akheizer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1966), Vol. 2.
[10]See, for instance, R. Biels, *Topics in Operator Theory* (University of Chicago Press, Chicago, 1971).
[11]J. Dixmier, *Les algèbres d'operateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1957).
[12]F. Riesz and B. Sz.-Nagy, *Leçons d'analyse fonctionnelle* (Gauthier-Villars, Paris, 1955).
[13]J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1955).
[14]See, for instance, J. Dieudonné, *Treatise on Analysis* (Academic, New York, 1970).
[15]G. W. Mackey, Proc. Natl. Acad. Sci. USA, **35**, 537 (1949).
[16]S. T. Ali and G. G. Emch (to be published).
[17]E. B. Davies, J. Funct. Anal. **6**, 318 (1970).

Bose–Einstein condensation in the presence of impurities. II

M. Kac

Rockefeller University, New York, New York 10021*

J. M. Luttinger

Columbia University,† New York, New York
(Received 27 August 1973)

An alternative discussion, based on function space integration, is given of results given in our preceding paper.

1. INTRODUCTION AND PRELIMINARIES

Consider a cubical container V of side L and imagine M spheres $S_k(\delta)$ of radius δ placed "at random" in V (i.e., the center of the spheres are chosen independently and each center with uniform distribution in V).

For a given configuration C of the spheres we consider the eigenvalue problem

$$(\hbar^2/2m)\nabla^2\psi + E^{(C)}\psi = 0, \tag{1.1}$$

with ψ satisfying periodic boundary conditions on V and vanishing on the surface of each sphere, i.e.,

$$\psi = 0 \quad \text{on } \partial S_k(\delta), \quad k = 1, 2, \ldots, M, \tag{1.2}$$

(∂S, denotes as usual, the boundary of S).

We wish to study the statistical properties of the sums

$$Q^{(C)}(L, \beta t) = \frac{1}{L^3}\sum_{k=1}^{\infty} e^{-t\beta E_k^{(C)}} \tag{1.3}$$

in the "thermodynamic limit", i.e.,

$$L \to \infty, \quad M \to \infty, \quad M/L^3 = \nu, \tag{1.4}$$

where $\beta = 1/kT$ and t a dimensionless parameter introduced for the sake of convenience and future use.

Setting

$$\psi(\mathbf{r}) = \phi(\mathbf{r}/\lambda), \tag{1.5}$$

where

$$\lambda^2 = \hbar^2/2mkT \tag{1.6}$$

is the square of the de Broglie wavelength (we use a slight modification differing by a numerical factor from the usual definition), we are led to an entirely dimensionless formulation of our problem.

In fact,

$$Q^{(C)}(L, \beta t) = \frac{1}{(L/\lambda)^3}\frac{1}{\lambda^3}\sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)}), \tag{1.7}$$

where $\epsilon_k^{(C)}$ are the eigenvalues of the Schrödinger equation

$$\tfrac{1}{2}\nabla^2\phi + \epsilon^{(C)}\phi = 0, \tag{1.8}$$

with ϕ satisfying periodic boundary conditions on $\overline{V} = V/\lambda^3$, i.e., the cube of (dimensionless) side L/λ and vanishing on the boundaries of the spheres $S_k(\delta/\lambda)$, i.e.,

$$\phi = 0 \quad \text{on } \partial S_k(\delta/\lambda), \quad k = 1, 2, \ldots, M. \tag{1.9}$$

We shall set

$$\overline{L} = L/\lambda, \quad \overline{\delta} = \delta/\lambda \tag{1.10}$$

and keep in mind that in taking the thermodynamic limit we have

$$M/\overline{L}^3 = \nu\lambda^3. \tag{1.11}$$

2. CALCULATION OF AUXILIARY QUANTITIES IN TERMS OF FUNCTION SPACE INTEGRALS

Let $\mathbf{r}_{\overline{L}}(\tau)$ [$\mathbf{r}_{\overline{L}}(0) = 0$, where 0 is an arbitrarily chosen origin within $\overline{V}$] denote a Brownian motion path ($0 \le \tau < \infty$) modified by the requirement that when the ordinary (unrestricted) Brownian path $\mathbf{r}(\tau)$ [$\mathbf{r}(0) = 0$] hits the boundary of V, it is continued in accordance with the periodic boundary condition. {In one dimension $x_{\overline{L}}(\tau)$ could be written as $x(\tau) \pmod{\overline{L}}$, i.e., $x_{\overline{L}}(\tau) = x(\tau) - \overline{L}[\bar{x}(\tau)/\overline{L}]$, where $[a]$ denotes, as usual, the greatest integer not exceeding a.} In our dimensionless presentation the diffusion constant association with the Brownian motion is $\frac{1}{2}$.

The basis of our calculation is the formula

$$\frac{1}{\overline{L}^3}\sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)}) = \frac{1}{\overline{L}^3}\frac{1}{(\sqrt{2\pi t})^3}\int_V d\mathbf{r}\ \mathrm{Prob}\{\mathbf{r} + \mathbf{r}_{\overline{L}}(\tau) \notin S_k(\overline{\delta}/\lambda), \quad k = 1, 2, \ldots, M, \ 0 \le \tau \le t \,|\, \mathbf{r}_{\overline{L}}(t) = 0\} \tag{2.1}$$

the probability inside the integral $\int d\mathbf{r}$ being the *conditional* probability conditioned on the path $\mathbf{r} + \mathbf{r}_{\overline{L}}(\tau)$ terminating at $\mathbf{r}$ at time t.

Strictly speaking formula (2.1) does not appear in the literature, but it is only a minor extension to restricted Brownian motion $\mathbf{r}_{\overline{L}}(\tau)$ of well-known formulas concerning ordinary (unrestricted) Brownian motion.

[If one discretizes the problem so that Brownian motion becomes a simple random walk and the Laplacian in the Schrödinger equation is replaced by the appropriate difference operator, the analogue of (2.1) becomes nearly obvious.]

We now show how one can calculate the average

$$\frac{1}{\overline{L}^3}\left\langle \sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)}) \right\rangle_{(C)}$$

over all configurations (C).

It is clear that we need the average

$$\langle \mathrm{Prob}\{\mathbf{r} + \mathbf{r}_{\overline{L}}(\tau) \notin S_k(\delta/\lambda), \quad k = 1, 2, \ldots, M, \quad 0 \le \tau \le t \,|\, \mathbf{r}_{\overline{L}}(t) = 0\rangle \tag{2.2}$$

and the simple trick is to note that we are dealing here with a *double* average (one over the Brownian paths and the other over the positions of the spheres) and to invert the order of averaging.

We therefore fix a path $\mathbf{r}_{\bar{L}}(\tau)$, $0 \le \tau \le t$, terminating at the origin and perform the average over the configurations (C).

If with *each* point of the path $\mathbf{r} + \mathbf{r}_{\bar{L}}(\tau)$, $0 \le \tau \le t$, $\mathbf{r}_{\bar{L}}(t) = 0$ as a center we draw a sphere of radius δ/λ and consider the (set theoretical) union of all these spheres we obtain a sausage-like set

$$W_{\delta/\lambda}(\mathbf{r} + \mathbf{r}_{\bar{L}}(\tau), \quad 0 \le \tau \le t, \mathbf{r}_{\bar{L}}(t) = 0), \tag{2.3}$$

and it is clear that our fixed path will not enter any of the spheres $S_k(\delta/\lambda)$ if their centers lie outside the set (2.3).

The probability that this should happen is clearly

$$\left(1 - \frac{|W_{\delta/\lambda}(\mathbf{r} + \mathbf{r}_{\bar{L}}(\tau), 0 \le \tau \le t, \mathbf{r}_{\bar{L}}(t) = 0)|}{\bar{L}^3}\right)^M, \tag{2.4}$$

where $|W_{\delta/\lambda}|$ denotes the three-dimensional Lebesgue measure (volume) of the set (2.3).

To calculate (2.2), we must still average over the paths, and it is here that integration in function spaces comes in.

Denoting the integral over the paths by the usual symbol $E\{\ \}$ (mathematical expectation) we have

$$\frac{1}{\bar{L}^3}\left\langle \sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)})\right\rangle_{(C)} = \frac{1}{(\sqrt{2\pi t})^3}\frac{1}{\bar{L}^3}\int_{\bar{V}} d\mathbf{r}\, E\left\{\left(1 - \frac{|W_{\delta/\lambda}(\mathbf{r} + \mathbf{r}_L(\tau), 0 \le \tau \le t, \mathbf{r}_L(t) = 0)|}{\bar{L}^3}\right)^M\right\}. \tag{2.5}$$

This can be simplified if we imagine that the spheres $S_k(\delta/\lambda)$ conform to the periodic boundary conditions, i.e., if the center of a sphere falls sufficiently close to the boundary of $\bar{V}$ the part of the sphere which would be outside of $\bar{V}$ is placed inside in accordance with the assumed periodicity. With this convention $|W_{\delta/\lambda}|$ becomes independent of $\mathbf{r}$ and therefore formula (2.5) assumes the form

$$\frac{1}{\bar{L}^3}\left\langle \sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)})\right\rangle_{(C)} = \frac{1}{(\sqrt{2\pi t})^3} E\left\{\left(1 - \frac{|W_{\delta/\lambda}(\mathbf{r}_{\bar{L}}(\tau), 0 \le \tau \le t)|}{\bar{L}^3}\right)^M \Big| \mathbf{r}_{\bar{L}}(t) = 0\right\}, \tag{2.6}$$

where the notation makes it clear that the expectation is the conditional expectation, the condition being that $\mathbf{r}_{\bar{L}}(t) = 0$, i.e., the path returns at time t to the starting point.

In the thermodynamic limit recalling that

$$M/\bar{L}^3 = \nu\lambda^3$$

we obtain

$$\lim \frac{1}{\bar{L}^3}\left\langle \sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)})\right\rangle_{(C)} = \frac{1}{(\sqrt{2\pi t})^3} E\exp(-\nu\lambda^3 |W_{\delta/\lambda}(\mathbf{r}(\tau), 0 \le \tau \le t)|\,\Big|\, \mathbf{r}(t) = 0), \tag{2.7}$$

and it should be noted that since $\bar{L} \to \infty$ we end up with the ordinary Brownian motion $\mathbf{r}(\tau)$.

The method of this section permits one also to calculate easily the second moment of the left-hand side of (2.1) and we leave it to the reader to verify that

$$\left\langle\left(\frac{1}{\bar{L}^3}\sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)})\right)^2\right\rangle_{(C)} = \frac{1}{\bar{L}^6}\frac{1}{(2\pi t)^3}\int_V\int_V d\mathbf{r}_1 d\mathbf{r}_2 \times E\left\{\left(1 - \frac{|W_{\delta/\lambda}(\mathbf{r}_1 + \mathbf{r}_{\bar{L}}^{(1)}(\tau), 0 \le \tau \le t) \cup W_{\delta/\lambda}(\mathbf{r}_2 + \mathbf{r}_{\bar{L}}^{(2)}(\tau), 0 \le \tau \le t)|}{\bar{L}^3}\right)^M \Big| r_{\bar{L}}^{(1)}(t) = 0, r_{\bar{L}}^{(2)}(t) = 0\right\},$$

where $r_{\bar{L}}^{(1)}(\tau)$ and $r_{\bar{L}}^{(2)}(\tau)$ are two *independent* Brownian motions and the symbol $\cup$ denotes as usual the union of the two sets.

It should be clear that unless $\mathbf{r}_1$ and $\mathbf{r}_2$ are close to each other the sets $W_{\delta/\lambda}(\mathbf{r}_1)$ and $W_{\delta/\lambda}(\mathbf{r}_2)$ will be non-overlapping and that, therefore, we shall have in the thermodynamic limit

$$\lim\left\langle\left(\frac{1}{\bar{L}^3}\sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)})\right)^2\right\rangle_{(C)} = \left(\lim\frac{1}{\bar{L}^3}\left\langle\sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)})\right\rangle_{(C)}\right)^2. \tag{2.8}$$

The result expressed by formula (2.8) is important because it shows that the quantities

$$\bar{Q}^{(C)}(\bar{L}; t) = \frac{1}{\bar{L}^3}\sum_{k=1}^{\infty} \exp(-t\epsilon_k^{(C)}) \tag{2.9}$$

which (for $t = 1, 2, 3, \cdots$) enter the formulas of pressure and density of the ideal Bose gas in a container with randomly distributed spherical impurities *do not fluctuate* (in the thermodynamic limit).

3. BOSE-EINSTEIN CONDENSATION ONCE AGAIN

If we consider N free Bosons are in V with spherical impurities $S_k(\delta)$ we have for a (fixed) fugacity ζ

$$\rho = \frac{N}{L^3} = \sum_{l=1}^{\infty} \zeta^l Q^{(C)}(L; \beta l) = \left(\frac{1}{\lambda\sqrt{2\pi}}\right)^3 \sum_{l=1}^{\infty} \frac{\zeta^l}{l^{3/2}} \bar{Q}^{(C)}(\bar{L}; l). \tag{3.1}$$

Since as we have seen the quantities $\bar{Q}^{(C)}(\bar{L}; l)$ do not fluctuate in the thermodynamic limit, we replace (in that limit) the $\bar{Q}^{(C)}$'s by their averages and we obtain

$$\rho = \left(\frac{1}{\lambda\sqrt{2\pi}}\right)^3 \sum_{l=1}^{\infty} \frac{\zeta^l}{l^{3/2}} E\{\exp[-\nu\lambda^3 |W_{\delta/\lambda}(\mathbf{r}(\tau), 0 \le \tau \le l)|]\,|\,\mathbf{r}(l) = 0\} \tag{3.2}$$

as long, of course, as the series converges, i.e., ζ is sufficiently small.

The question now arises what is the radius of convergence of the series (3.2). In our paper[1] we have shown that it is still equal to unity by showing that $\epsilon_1^{(C)}$ approaches zero in probability. (We actually claimed to have proved somewhat more, i.e., $\epsilon_1^{(C)} \to 0$ with probability one, but although this claim is not justified, the stronger statement is not needed.)

We shall now give an alternate proof based on formula (2.1).

With $\mathbf{r}$ as center, consider the sphere $S(\mathbf{r}; a)$ of radius a and note that

$$\begin{aligned}&\langle \text{Prob}\{\mathbf{r} + \mathbf{r}_{\bar{L}}(\tau) \notin S_k(\delta/\lambda),\\ &\quad k = 1, 2, \ldots, M, 0 \le \tau \le t \,|\, \mathbf{r}_{\bar{L}}(t) = 0\}_{(C)}\\ &\ge \text{Prob}\{\mathbf{r} + \mathbf{r}_{\bar{L}}(\tau) \in S(\mathbf{r}; a), 0 \le \tau \le t \,|\, \mathbf{r}_{\bar{L}}(t) = 0\}\\ &\quad \times \text{Prob}\{\text{all of the centers of } S_k(\delta/\lambda) \text{ lie outside}\\ &\quad\quad S(\mathbf{r}; a + (\delta/\lambda))\}. \qquad (3.3)\end{aligned}$$

The second of the probabilities on the right-hand side of (3.3) is clearly

$$\left(1 - \frac{\frac{4}{3}\pi(a + (\delta/\lambda))^3}{\bar{L}^3}\right)^M$$

and hence in the thermodynamic limit we obtain

$$\begin{aligned}&\lim\langle \text{Prob}\{\mathbf{r} + \mathbf{r}_{\bar{L}}(\tau) \notin S_k(\delta/\lambda),\\ &\quad k = 1, 2, \ldots, M, 0 \le \tau \le t) \,|\, \mathbf{r}(t) = 0)\}\\ &\ge \text{Prob}\{\mathbf{r} + \mathbf{r}(\tau) \in S(\mathbf{r}; a), 0 \le \tau \le t \,|\, \mathbf{r}(0) = 0\}\\ &\quad \times \exp\{-\tfrac{4}{3}\pi\nu\lambda^3(a + (\delta/\lambda))^3\}. \qquad (3.4)\end{aligned}$$

The probability appearing on the right-hand side of (3.4) is related to the classical first passage problem for Brownian motion and is given by the formula

$$\begin{aligned}&\text{Prob}\{\mathbf{r} + \mathbf{r}(\tau) \in S(\mathbf{r}; a), 0 \le \tau \le t \,|\, \mathbf{r}(0) = 0\}\\ &\qquad = (\sqrt{2\pi t})^3 \sum_{n=1}^{\infty} e^{-\lambda_n^{(a)} t} \phi_n^{(a)2}(0), \qquad (3.5)\end{aligned}$$

where $\lambda_n(a)$ and $\phi_n^{(a)}(\boldsymbol{\rho})$ are the eigenvalues and normalized eigenfunctions corresponding to the eigenvalue problem

$$\tfrac{1}{2}\nabla^2\phi + \lambda\phi = 0, \quad \phi = 0 \quad \text{on } \partial S(0; a). \qquad (3.6)$$

It is easily seen that

$$\lambda_n(a) = \lambda_n(1)/a^2, \quad \phi_n^{(a)2}(0) = \phi_n^{(1)2}(0)/a^3, \qquad (3.7)$$

and hence by (3.4) and (3.5) we have

$$\begin{aligned}&\lim\langle \text{Prob}\{\mathbf{r} + \mathbf{r}_L(\tau) \notin S_k(\delta/\lambda), k\\ &\quad = 1, 2, \ldots, M, 0 \le \tau \le t \,|\, \mathbf{r}(t) = 0\}\rangle_{(C)}\\ &\ge (2\pi t/a^2)^{3/2} \exp[-(\lambda_1(1)t/a^2 + \tfrac{4}{3}\pi\nu\lambda^3(a + \delta/\lambda)^3].\end{aligned} \qquad (3.8)$$

Setting

$$a = t^{1/5} \qquad (3.9)$$

and going back to (3.8), we obtain that for $t \to \infty$

$$E\{\exp(-\nu\lambda^3 |W_{\delta/\lambda}(\mathbf{r}(\tau), 0 \le \tau \le t)| \,|\, \mathbf{r}(t) = 0\} \ge \exp(-Ct^{3/5}), \qquad (3.10)$$

where C depends on $\nu\lambda^3$ (but not on δ/λ).

Formula (3.10) implies easily that the radius of convergence of the series (3.2) is equal to unity.

We suspect that actually

$$\lim_{t\to\infty} \frac{1}{t^{3/5}} \log E\{\exp[-\nu\lambda^3 |W_{\delta/\lambda}(\mathbf{r}(\tau), 0 \le \tau \le t)| \,|\, \mathbf{r}(t) = 0\} = A \ne 0, \qquad (3.11)$$

but we have no proof except in one dimension (with $^3/_5$ replaced by $^1/_3$) where the problem is particularly simple since $W_{\delta/\lambda}(x(\tau), 0 \le \tau \le t)$ is simply δ/λ plus the *range* of the Brownian motion.

Heuristic support for conjecture (3.11) has been given by Lifschitz.[2]

Finally, we should state once again that even though convergence of the series (3.2) for $\zeta = 1$ heralds an onset of a condensation, we are unable to prove that the condensation is of the Bose–Einstein type, i.e., that we have simply a macroscopic occupation of only one lowest state.

What is needed is a proof that in probability

$$\lim \bar{L}^3(\epsilon_2^{(C)} - \epsilon_1^{(C)}) = \infty$$

or, more precisely,

$$\lim \text{Prob}\{\bar{L}^3(\epsilon_2^{(C)} - \epsilon^{(C)}) < \alpha\} = 0,$$

for *every* α.

4. THE SHIFT IN CRITICAL DENSITY DUE TO IMPURITIES

We obtain the critical density ρ_c for the Bose gas in the presence of impurities by setting $\zeta = 1$ in (3.2).

Thus

$$\begin{aligned}\rho_c = &\left(\frac{1}{\lambda\sqrt{2\pi}}\right)^3 \sum_{l=1}^{\infty} \frac{1}{l^{3/2}}\\ &E\{\exp[-\nu\lambda^3 |W_{\delta/\lambda}(\mathbf{r}(\tau), 0 \le \tau \le l)]| \,\mathbf{r}(l) = 0\}, \qquad (4.1)\end{aligned}$$

while the usual critical density $\rho_c^{(0)}$ is given by the well-known formula

$$\rho_c^{(0)} = \left(\frac{1}{\lambda\sqrt{2\pi}}\right)^3 \sum_{l=1}^{\infty} \frac{1}{l^{3/2}}. \qquad (4.2)$$

In one limiting case it is easy to derive a convenient formula for the correction to $\rho_c^{(0)}$ due to impurities.

The limiting case in question is

$$\delta/\lambda \ll 1, \quad (\nu\lambda^3)(\delta/\lambda)^2 = \nu\lambda^2\delta = \omega, \qquad (4.3)$$

with ω being of order 1.

By simple dimensional analysis, it can be shown that the statistical properties of

$$|W_{\delta/\lambda}(\mathbf{r}(\tau)), 0 \le \tau \le l|$$

are the same as those of

$$(\delta/\lambda)^3 |W_1(\mathbf{r}(\tau), 0 \le \tau < l/(\delta/\lambda)^2)|$$

and therefore

$$E\{\exp[-\nu\lambda^3|(W_{\delta/\lambda}(\mathbf{r}(\tau)), 0 \le \tau \le l)|] \,|\, \mathbf{r}(l) = 0\}$$
$$= E\{\exp[-\nu\lambda^3|(W_1(\mathbf{r}(\tau)), 0 \le \tau \le l/(\delta/\lambda)^2)|]|\mathbf{r}(l/(\delta/\lambda)^2) = 0. \quad (4.4)$$

Without the restriction $\mathbf{r}(t) = 0$ it is known[3] that

$$\lim_{t\to\infty}[|W_a(\mathbf{r}(\tau)), 0 \le \tau \le t|/t] = 2\pi a \quad (4.5)$$

in probability, and it is not difficult to show that (4.5) also holds with the restriction $\mathbf{r}(t) = 0$.

It thus follows that in the limit $\delta/\lambda \to 0$, $\nu\lambda^2\delta = \omega$

$$\lim E\{\exp[-\nu\delta^3|W_1(\mathbf{r}(\tau), 0 \le \tau \le l/(\delta/\lambda)^2)|]|\mathbf{r}(l/(\delta/\lambda)^2) = 0\} = \exp(-2\pi\omega l) \quad (4.6)$$

and therefore in this limit

$$\rho_c = \left(\frac{1}{\lambda\sqrt{2\pi}}\right)^3 \sum_{l=1}^{\infty} \frac{\exp(-2\pi\omega l)}{l^{3/2}}. \quad (4.7)$$

If, in addition, ω is small we have (asymptotically)

$$\rho_c = \rho_c^{(0)} - \frac{2}{\sqrt{\pi}} \frac{1}{\lambda^3} (\nu\lambda^2\delta)^{1/2}. \quad (4.8)$$

It is clear that in the limit considered in this section $\nu\delta^3 = \omega(\delta/\lambda)^2 \ll 1$ and hence one is justified in neglecting the effects of overlapping of the impurities.

* Supported in part by AFOSR Grant 72-2187.
† Supported in part by the National Science Foundation.
[1]M. Kac and J. M. Luttinger, J. Math. Phys. 14, 1626 (1973)
[2]I. M. Lifschitz, Usp. Fiz. Nauk 83, 617 (1964).
[3]F. Spitzer, *Principles of Random Walk* (Van Nostrand, Princeton, N.J., 1964), see esp. p. 40.

A note on lattice sums in two dimensions

I. J. Zucker

Department of Physics, University of Surrey, Guildford, England
(Received 10 September 1973)

A slight simplification of Glasser's approach for obtaining lattice sums in two dimensions is suggested. The result for the triangular lattice is given.

In two recent publications Glasser[1,2] has given some powerful methods for obtaining sums of the form

$$S = \sum_{m,n=1}^{\infty} f(m,n), \tag{1}$$

where $f(m,n)$ is the reciprocal of some power of a linear or quadratic equation. He mentions that the one result existing in the literature is

$$\sum_{m,n=1}^{\infty} (m^2+n^2)^{-s} = \zeta(s)\beta(s) - \zeta(2s). \tag{2}$$

This is not quite the case. Fletcher *et al*[3] besides giving (2) list several others such as the two-dimensional Madelung-type sums

$$\sum_{m,n=-\infty}^{\infty}{}' (-1)^{m+n}(m^2+n^2)^{-s} = -4\eta(s)\beta(s),$$
$$\eta(s) = (1-2^{1-s})\zeta(s). \tag{3}$$

The $'$ excludes $m=n=0$. Further they give the result for the triangular lattice, namely,

$$\sum_{m,n=-\infty}^{\infty}{}' (m^2+mn+n^2)^{-s} = 6\zeta(s)g(s),$$
$$g(s) = \sum_{n=0}^{\infty} (3n+1)^{-s} - (3n+2)^{-s}. \tag{4}$$

A slight simplification of Glasser's approach is suggested here. We take as an example

$$\sum_{m,n=-\infty}^{\infty} (m^2+n^2)^{-s} = \frac{1}{\Gamma(s)} \int_0^{\infty} t^{s-1} \left(\sum\sum{}' e^{-(m^2+n^2)}\right) dt. \tag{5}$$

This can be written

$$\frac{1}{\Gamma(s)} \int_0^{\infty} t^{s-1}[\theta_3^2(0,e^{-t}) - 1]\,dt. \tag{6}$$

Now by using the Jacobi identity

$$\theta_3^2(0,q) \equiv 1 + 4\sum_{n=1}^{\infty} \frac{q^n}{1+q^{2n}} \equiv 1 + 4\sum_{n=1}^{\infty}\sum_{m=0}^{\infty} (-1)^m q^{n(2m+1)} \tag{7}$$

(6) becomes

$$\frac{4}{\Gamma(s)} \sum_{n=1}^{\infty}\sum_{m=0}^{\infty} (-1)^m \int_0^{\infty} t^{s-1} e^{-n(2m+1)t}\,dt$$
$$= 4\sum_{n=1}^{\infty}\sum_{m=0}^{\infty} \frac{(-1)^m}{n^s(2m+1)^s} = 4\zeta(s)\beta(s). \tag{8}$$

Similarly the Madelung sum in two dimensions may be written

$$\sum_{m,n=-\infty}^{\infty}{}' (-1)^{m+n}(m^2+n^2)^{-s}$$
$$= \frac{1}{\Gamma(s)} \int_0^{\infty} t^{s-1}[\theta_4^2(0,e^{-t}) - 1]\,dt \tag{9}$$

and then use of the Jacobi identities for θ_4^2 will give the result $-4\eta(s)\beta(s)$. The extensions to d dimensions are obvious, e.g.,

$$\sum_{-\infty}^{\infty}{}' \cdots \sum_{-\infty}^{\infty}{}' (m_1^2 + m_2^2 \cdots m_d^2)^{-s}$$
$$= \frac{1}{\Gamma(s)} \int^{\infty} t^{s-1}[\theta_3^d(0,e^{-t}) - 1]\,dt \tag{10}$$

but identities for odd powers of θ functions are not known. However the linear case in d dimensions can be put in reasonably closed form. Thus

$$\sum_{m_1=1}^{\infty} \cdots \sum_{m_d=1}^{\infty} (m_1 + m_2 + m_3 \cdots m_d)^{-s}$$
$$= \sum_{m=1}^{\infty} \frac{(d+m-2)!}{(d-1)!\,(m-1)!} \frac{1}{(m+d-1)^s}. \tag{11}$$

ACKNOWLEDGMENT

The author wishes to thank Dr. G. S. Joyce of King's College, London for interesting discussions.

[1] M. L. Glasser, J. Math. Phys. **14**, 409 (1973).
[2] M. L. Glasser, J. Math. Phys. **14**, 701 (1973).
[3] A. Fletcher, J. C. P. Miller, L. Rosenhead, and L. J. Comrie, *An Index of Mathematical Tables* (Blackwell, London, 1962), Vol. I, p. 95.

The evaluation of lattice sums. III. Phase modulated sums

M. L. Glasser

Battelle Memorial Institute, Columbus, Ohio 43201
(Received 29 August 1973)

Two-dimensional lattice sums of the form $\Sigma'\exp(i\mathbf{k}\cdot\mathbf{s})k^{-2n}$ are evaluated exactly in terms of Jacobian theta functions.

Electronic structure studies relating to chain polymers, crystal surfaces, and solids frequently involve the computation of sums having the form

$$\Phi = \sum_{k\neq 0} \frac{e^{i\mathbf{k}\cdot\mathbf{S}}}{k^s} \tag{1}$$

extending over the (nonzero) lattice vectors of a one-, two-, or three dimensional array.[1] Along with progress in such studies interest in these sums is increasing and various *ad hoc* procedures have been advanced for their numerical evaluation. However, these algorithms tend to be expensive in terms of computer time and it is generally acknowledged that more expeditious approximation schemes are desirable. The purpose of this note is to show that in a few cases "exact" analytic expressions for these sums can be obtained; it is hoped that our examination of these results will reveal enough of the mathematical nature of these sums to suggest alternative numerical procedures for their rapid evaluation. In addition, the results are interesting in their own right and, in the two-dimensional case, may have direct applications in surface science.

In one dimension these sums reduce immediately to well-known Fourier series so we proceed to the two-dimensional case. For simplicity, we assume that the vectors $\mathbf{k}$ are reciprocal lattice vectors for a rectangular lattice

$$\begin{aligned}&\mathbf{k} = 2\pi(u/a, v/b),\\ &u, v = 0, \pm 1, \pm 2, \cdots \quad ;\end{aligned} \tag{2}$$

since there is no loss in generality in assuming that $\mathbf{S}$ is a vector in the unit cell, we take

$$\mathbf{S} = (ax, by), \qquad 0 \le x, y < 1. \tag{3}$$

We then have

$$\Phi = \left(\frac{a}{2\pi}\right)^s \sum_{u,v\neq 0,0} \frac{e^{2\pi i(ux+vy)}}{[u^2 + (a/b)^2 v^2]^{s/2}} \tag{4}$$

and by using the identity

$$\Gamma(s)b^{-s} = \int_0^\infty dt \;\; t^{s-1}e^{-bt} \tag{5}$$

we write

$$\begin{aligned}&\Gamma(s)\Phi = (a/2\pi)^s \int_0^\infty dt \;\; t^{1/2s-1}[T(t) - 1],\\ &T(t) = \sum_{u,v} \exp\{-t[u^2 + (a/b)^2 v^2] + 2\pi i(ux + vy)\}\end{aligned} \tag{6}$$

where the sum extends over all values of u, v. By applying Jacobi's transformation[2] from the theory of theta functions we have

$$\sum_u \exp(-tu^2 + 2\pi iux) = (\pi/t)^{1/2} \sum_u \exp[-(\pi^2/t)(u + x)^2] \tag{7}$$

and therefore, after separating out the term v = 0 from the double sum in (6),

$$T(t) - 1 = \sum_{u\neq 0} e^{-tu^2 + 2\pi iux} + (\pi/t)^{1/2} \sum_{\substack{u\\ v\neq 0}} e^{-(\pi^2/t)(u+x)^2} e^{-(a^2v^2t/b^2) + 2\pi ivy}. \tag{8}$$

Consequently,

$$\Phi = \left(\frac{a}{2\pi}\right)^s \left\{\sum_{u\neq 0} \frac{e^{2\pi iux}}{u^s} + \frac{\pi^{1/2}}{\Gamma(s)} \sum_{\substack{u\\ v\neq 0}} e^{2\pi ivy} \int_0^\infty dt\, t^{(s-3)/2}\, e^{-\pi^2(u+x)/t} e^{-a^2v^2t/b^2}\right\}. \tag{9}$$

However,

$$\int_0^\infty dt\, t^{s-1}e^{-pt}e^{-q/t} = 2(q/p)^{s/2}K_s[2(pq)^{1/2}], \tag{10}$$

where K_s denotes, as usual, the modified Bessel function; so we obtain

$$\Phi = \left(\frac{a}{2\pi}\right)^s \left[\sum_{u\neq 0} \frac{e^{2\pi iux}}{u^s} + \frac{2\pi^{1/2}}{\Gamma(s)} \sum_{\substack{u\\ v\neq 0}} \left(\frac{\pi b|u+x|}{|v|}\right)^{(s-1)/2} \times K_{(s-1)/2}\left(\frac{2\pi a}{b}|v||u+x|\right) e^{2\pi ivy}\right]. \tag{11}$$

This appears to be as far as one can go in general. However, if s is an even integer $K_{(s-1)/2}(z)$ reduces to $z^{-1/2}e^{-z}$ multiplied by a polynomial and the double sum can be performed. We illustrate this by the case $s = 2$ where we have

$$K_{1/2}(z) = (\pi/2z)^{1/2}e^{-z} \tag{12}$$

and

$$\Phi = \left(\frac{a}{2\pi}\right)^2 \left[\sum_{u\neq 0} \frac{e^{2\pi iux}}{u^2} + \left(\frac{\pi b}{a}\right) \sum_{\substack{u\\ v\neq 0}} \frac{1}{|v|}\, e^{-2\pi(a/b)|v||u+x| + 2\pi ivy}\right]. \tag{13}$$

Out of the double sum we separate the term $u = 0$ and then express all the sums as a sum over positive values of u and v. By introducing the quantities

$$q = e^{-\pi(a/b)}, \qquad \alpha = y + i(a/b)x, \tag{14}$$

Φ can be expressed simply as follows:

$$\Phi = \frac{a^2}{2\pi^2} \sum_{u=1} \frac{\cos 2\pi ux}{u^2} + \frac{ab}{4\pi} \sum_{v=1}^\infty \frac{1}{v}\left(e^{2\pi i\alpha v} + \text{c.c.}\right) + \frac{ab}{4\pi} \sum_{u=1}^\infty \sum_{v=1}^\infty \frac{1}{v} q^{2uv}\left(e^{2\pi i\alpha v} + e^{-2\pi i\alpha v} + \text{(c.c.)}\right), \tag{15}$$

where c.c. denotes complex conjugate. The v sums are elementary since

$$\sum_{v=1}^\infty v^{-1}x^v = -\ln(1 - x) \tag{16}$$

and

$$\Phi = \frac{a^2}{2\pi^2}\sum_{u=1}^{\infty}\frac{\cos 2\pi u x}{u^2} - \frac{ab}{2\pi}|1 - e^{2\pi i\alpha}| - \frac{ab}{2\pi}\sum_{u=1}^{\infty}\ln|(1-q^{2u}e^{2\pi i\alpha})(1-q^{2u}e^{-2\pi i\alpha})|. \quad (17)$$

Inserting the value for the first sum, which is a well-known Fourier series,[3] and performing an elementary manipulation we have

$$\Phi = \frac{a^2}{2}(x^2 - x + \tfrac{1}{6}) - \frac{ab}{2\pi}\ln(2e^{-\pi(a/b)x}|\sin\pi\alpha|) - \frac{ab}{2\pi}\ln\left|\prod_{u=1}^{\infty}(1 - 2q^{2u}\cos 2\pi\alpha + q^{4u})\right|. \quad (18)$$

However[4]

$$\prod_{u=1}^{\infty}(1 - 2q^{2u}\cos 2z + q^{4u}) = 2^{1/3}\csc z\; q^{-1/6}\frac{\theta_1(z,q)}{[\theta_1'(0,q)]^{1/3}}, \quad (19)$$

where θ_1 denotes a Jacobian theta function. Hence, after some simple algebra we find the result

$$\Phi = \frac{a^2}{2}x^2 - \frac{2ab}{3\pi}\ln 2 - \frac{ab}{2\pi}\ln\left|\frac{\theta_1(\pi\alpha,q)}{[\theta_1'(0,q)]^{1/3}}\right| \quad (20)$$

That the result (20) is jointly symmetric in a, b and x, y, as is manifest in the definition of Φ, follows from Jacobi's transformation.[2] Alternatively, the above derivation can be used to furnish a new proof of this important identity. Equation (20) can also be expressed in terms of the function θ_3 which is computationally somewhat simpler than θ_1. Because the theta functions are generally approximated well by two or three terms of their series representations, (20) has obvious computational value. By comparing (20) with the results of Refs. 5 and 6 a number of new interesting mathematical results can be obtained relating to the exact evaluation of theta functions of special arguments which the reader is invited to work out.

Although the above calculation is susceptible of some generalization [for example, denominators of the form $(u^2/a^2 + uv/c + v^2/b^2)^s$ can be treated], extention to other values of s or to three dimensions does not appear to be possible. However, for the denominator $(u/a)^2 + (v/b)^2 + w^2$, the sum over u and v can be reduced to an exponentially converging single sum which may be of value in three dimensions; since our concern has been with exact results, we do not explore these avenues here.

ACKNOWLEDGMENT

The author thanks the Aspen Center for Physics for its hospitality while this work was carried out.

[1]See, e.g., F. E. Harris and H. J. Monkhorst, Phys. Rev. B 2, 4400 (1970).
[2]E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1958), 4th ed., Sect. 21.51.
[3]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, (Academic, New York, 1965), 4th ed., Formula 1.443(3).
[4]Reference 2, Secs. 21.3 and 21.42.
[5]M. L. Glasser, J. Math Phys. 14, 409 (1973).
[6]M. L. Glasser, J. Math. Phys. 14, 701 (1973).

The polaron without cutoffs in two space dimensions

Alan D. Sloan

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 25 September 1973)

Hamiltonians for the polaron of fixed total momentum are defined using momentum cutoffs. A renormalized Hamiltonian of fixed total momentum is defined in two space dimensions by proving the strong convergence of the resolvents of the cutoff Hamiltonians. The Hamiltonian for the physical polaron is defined as the direct integral of the fixed momentum Hamiltonians.

1. INTRODUCTION

The model considered in this paper is that of a spinless relativistic electron interacting with relativistic phonons. The one-particle momentum state space for a d-dimensional spinless electron is $L^2(R^d)$ and this is also the state space for the phonon, or quatum of lattice vibration. The Hamiltonian for the electron is

$$H_e = \int_{R^d} E_{m_e}(p) b^*(p) b(p) dp, \tag{1.1}$$

where $b^*(p)$ and $b(p)$ are the electron creation and annihilation operators of momentum p, respectively, and where

$$E_{m_e}(p) = (|p|^2 + m_e^2)^{1/2} \tag{1.2}$$

is the relativistic kinetic energy function for the electron of mass m_e. H_e acts on antisymmetric Fock space over $L^2(R^d)$, A.

The free phonon Hamiltonian is

$$K = \int_{R^d} w(k) a^*(k) a(k) dk, \tag{1.3}$$

where $a^*(k)$ and $a(k)$ are the phonon creation and annihilation operators of momentum k, respectively, and where

$$w(k) = (|k|^2 + \mu^2)^{1/2} \tag{1.4}$$

is the kinetic energy function for the phonon of mass μ. K acts on symmetric Fock space $\mathfrak{F}$.

The state space for arbitrarily many electrons and phonons is $A \otimes \mathfrak{F}$. On this space we define the interaction Hamiltonian as

$$H_I = \int_{R^n} \int_{R^n} w(k)^{-1/2} \times [b^*(p + k) b(p) a(k) + b^*(p - k) b(p) a^*(k)] dk dp. \tag{1.5}$$

This is a modified Yukawa-type interaction in which the pair creation and annihilation terms have been dropped. The total Hamiltonian

$$H = H_e \otimes I + I \otimes K + \lambda H_I \tag{1.6}$$

also acts on $A \otimes \mathfrak{F}$. λ is a real number and is called the coupling constant. An important property of H is that it leaves the number of electrons invariant. Consequently we may consider the case in which a certain fixed number of electrons are present.

If no electrons are present then H is just $I \times K$ and the dynamics is understood. The bare and physical vacuums coincide. There is no wavefunction renormalization.

In this paper the case of one electron is considered. If, in this case, $\lambda = 0$, then the electrons and phonons do not interact. Again the situation is well understood. Thus for the rest of the paper we assume $\lambda \neq 0$. When $d > 1$, $w^{-1/2}$ is not in $L^2(R^d)$. Consequently, H is only defined on the zero vector.

To overcome this problem we first replace $w^{-1/2}$ in H_I by a function f in $L^2(R^d)$. That is, we introduce a momentum cutoff. Let $H_I(f)$ and $H(f)$ denote the modified interaction and total Hamiltonians, respectively. Modulo technical conditions on f, it is known that $H(f)$ is a self-adjoint operator on $L^2(R^d) \otimes \mathfrak{F}$, which is bounded below. It is also known that there exists a zero total momentum rest state for each fixed f. That is, if we write the state space as a direct integral of infinitesimal subspaces of fixed total momentum,

$$L^2(R^d) \otimes \mathfrak{F} = \int_{R^d} H_p dp, \tag{1.7}$$

then the Hamiltonian also decomposes as

$$H(f) = \int H_p(f) dp, \tag{1.8}$$

where each $\mathcal{K}_p$ is a copy of $\mathfrak{F}$ and each $H_p(f)$ is a self-adjoint operator on $\mathcal{K}_p$. It is known that $H_0(f)$ is bounded below and that the bottom of the spectrum of $H_0(f)$ is an eigenvalue. A corresponding eigenvector is called a zero total momentum rest state. Let $\Psi(f)$ denote such a rest state.

To remove cutoffs we specialize to the case of two space dimensions, $d = 2$. Using a sequence $\{f_n\}$ of cutoff functions that agree with $w^{-1/2}$ on sets that increase to R^2 as $n \to \infty$ we obtain corresponding sequences of rest states, $\{\Psi(f_n)\}$ and operators $\{H(f_n)\}$ and $\{H_p(f_n)\}$. It is shown that $\{\Psi(f_n)\}$ lies in a norm compact subset of $\mathcal{K}_0 = \mathfrak{F}$, so that some subsequence converges as we remove cutoffs:

$$\lim_{n_i \to \infty} \Psi(f_{n_i}) = \Psi_\infty. \tag{1.9}$$

Thus there is no infinite field strength renormalization.

After adjusting the electron mass so that the lowest point in the spectrum of $H(f_n)$ is zero we find a subsequence Λ, of the sequence of indices for which (1.9) is true, such that for every p in R^2 and $r > 0$ the resolvents $\{(H_p(f_n) + r)^{-1} : n \in \Lambda\}$ converge strongly to the resolvent $H_p(\infty)$ of a densely defined, self-adjoint, positive definite operator on $\mathcal{K}_p$. Since

$$(H(f_n) + r)^{-1} = \int (H_p(f_n) + r)^{-1} dp$$

it will also follow that there is a self-adjoint operator, $H(\infty)$, on $L^2(R^2) \otimes \mathfrak{F}$, such that $\{(H(f_n) + r)^{-1}; n \in \Lambda\}$ converges strongly to $(H(\infty) + r)^{-1}$ as $n \to \infty$ for $r > 0$. We show that inf (spectrum $H(\infty)$) = 0. The contraction semigroups and unitary groups of $H(f_n)$ also are shown to converge to those of $H(\infty)$.

$H(\infty)$ is the one polaron Hamiltonian without cutoffs in two space dimensions. $H(\infty)$ is defined on $L^2(R^2) \otimes \mathfrak{F}$.

Thus, it is not necessary to change Hilbert spaces when we remove cutoffs.

This model is related to the one Nelson considered.[1] He investigated the case of a nonrelativistic free electron Hamiltonian in three space dimensions and was able to remove cutoffs without changing Hilbert spaces. Nelson's method is probably applicable to constructing a total Hamiltonian in Fock space for the model considered here.

Another model which does not exhibit vacuum polarization was considered by Eckmann.[2] He considered a three-dimensional model with the same singularity structure as the two-dimensional relativistic polaron. By expanding the resolvent in a Born series he was able to obtain norm convergence of the resolvents whereas we are only able to obtain strong convergence. Frohlich,[3] has shown that Eckmann's techniques apply to the polaron model considered here.

Work on Nelson's model and Eckmann's model has been extended in several interesting directions. Cannon,[4] constructed the basic field theoretic objects for Nelson's model while Frohlich,[5] obtained field theoretic properties for Eckmann's model. Albeverio[6,7] has worked on the scattering problem for Eckmann's model. Frohlich,[8] has considered the infrared problem resulting when the bosons have zero mass.

The techniques we employ do not involve dressing transformations, perturbation theory, mass gaps, or resolvent expansions and so are distinct from those previously mentioned. The point of view taken here is that one should be able to construct a total Hamiltonian once one has a physical ground state. In realistic models this is done via the Gel'fand–Naimark–Segal construction but for the simple model considered here this is not necessary. Thus while our results are not as extensive as those already cited our methods hopefully capture the spirit of realistic quantum field theory. Gross has taken this viewpoint in Ref. 9.

He has considered the case of relativistic free electron energy and three or more space dimensions. Let $\mathfrak{A}_0$ be the set of all finite linear combinations of operators on $\mathfrak{F}$ of the form $\exp[iR(g)]$, where g is in $L^2(R^d)$ and $R(g)$ is given by (2.10). Gross chooses a C^*-algebra $\mathfrak{A}$ containing $\mathfrak{A}_0$ and finds a representation σ of $\mathfrak{A}$ on a Hilbert space $\mathfrak{K}$, and also a nonnegative self-adjoint operator H on $\mathfrak{K}$, and a sequence of cutoff functions $\{f_n\}$ such that $[H(f_n)(U \otimes A\Psi(f_n)), v \otimes B\Psi(f_n)]$ converges to $[H(U \otimes \sigma(A)\Psi\infty, v \otimes \sigma(B)\Psi\infty)]$ as n goes to ∞ in some subsequence of $\{1, 2, \ldots\}$, and where A, B are in $\mathfrak{A}_0$ and U and v are in $C_c(R^d)$. In two dimensions it will be shown that one may take $\mathfrak{K} = \mathfrak{F}$ and that for every h in $D(H(\infty))$, there is a $\{h_n\}$ with $h_n \in D(H(f_n))$ and $H(f_n)h_n \to H(\infty)h$.

The results presented here have been announced in Ref. 10.

2. A CUTOFF MODEL WITH FIXED TOTAL MOMENTUM

A. Hamiltonians

Our purpose here is to define the $H_p(f)$ of (1.8). We follow Ref. 9 in making the definitions of this section.

Let $\mathfrak{F}$ denote symmetric Fock space over $\mathfrak{F}_1 = L^2(R^d, dk)$, where dk denotes Lebesgue measure on d-dimensional Euclidean space, R^d. In other words $\mathfrak{F}$ is the Hilbert space direct sum of $\mathfrak{F}_n$ = the n-fold symmetric tensor produce of $\mathfrak{F}_1$ and $\mathfrak{F}_0$ is the complex numbers: $\mathfrak{F} = \sum_{n=0}^{\infty} \mathfrak{F}_n$. We will identify $\mathfrak{F}_n$ with the functions in $L^2(R^{d\cdot n}, dk_1 \ldots dk_n)$ which are symmetric: $\mathfrak{F}_n = \{g \in L^2(R^{d\cdot n})$: $g(k_1, \ldots k_n) = g(k_{p(1)}, \ldots, k_{p(n)})$ for almost every $(k_1, \ldots, k_n)$ in $R^{d\cdot n}$ and every permutation p of $\{1, 2, \ldots, n\}\}$.

By $\mathcal{S}^n$ we mean the Schwartz space of rapidly decreasing, infinitely differentiable complex functions on $(R^d)^n$. Let $\mathcal{E}$ denote the vector subspace of $\mathfrak{F}$ consisting of all $g = \sum_{n=0}^{\infty} g_n$ such that g_n is in $\mathcal{S}^n$ and $g_n = 0$ for all sufficiently large n. By $\mathfrak{F}^0$ we mean $\{\sum_{n=0}^{\infty} g_n \in \mathfrak{F}$: $g_n \in \mathfrak{F}_n$ and $g_n = 0$ for all sufficiently large $n\}$.

For k in R^d we define the annihilation operator $a(k)$ as follows. $D(a(k)) = \mathcal{E}$. If g is in $\mathcal{E} \cap \mathfrak{F}_n$ then $a(k)g_n$ is in $\mathcal{E} \cap \mathfrak{F}_{n-1}$ and is given by

$$(a(k)g_n)(k_1, \ldots, k_{n-1}) = n^{1/2} g_n(k, k_1, \ldots, k_{n-1}). \tag{2.1}$$

$a(k)$ is zero on $\mathfrak{F}_0$. $a(k)$ is then extended linearly. By $a^*(k_1) \ldots a^*(k_j)a(k_{j+1}) \ldots a(k_n)$ we mean the bilinear form on $\mathcal{E} \times \mathcal{E}$:

$$g \times h \to (a(k_{j+1}) \ldots a(k_n)g, a(k_j) \ldots a(k_1)h).$$

If v is in $L^2(R^{d\cdot n})$, then the bilinear form

$$g \times h \to \int v(k_1, \ldots, k_n)(a(k_{j+1}) \ldots a(k_n)g, a(k_g) \ldots a(k_1)h) \times dk_1 \ldots dk_n$$

defined on $\mathcal{E} \times \mathcal{E}$ is actually the bilinear form of a closed operator on $\mathfrak{F}$ with core $\mathcal{E}$ (Ref. 9, p. 9). We denote the operator by

$$\int v(k_1, \ldots, k_n)a^*(k_1) \ldots a(k_n)dk_1 \ldots dk_n. \tag{2.2}$$

For any self-adjoint operator, T, on $\mathfrak{F}_1$, let $\gamma(T)$ denote its quantization, (Ref. 11, p. 223) $\gamma(T)$ is self-adjoint and on $\mathfrak{F}_n$ is the closure of

$$T \otimes I \otimes \ldots \otimes I + I \otimes T \otimes I \otimes \ldots \otimes I + \ldots + I \otimes \ldots \otimes I \otimes T.$$

When T is multiplication by a nonnegative, measurable function h, (in this case we will write $T = M_h$), then we will say $\gamma(T) = \int h(k)b^*(k)b(k)dk$, since for f and g in $\mathcal{E} \cap D(\gamma(T))$ we have

$$(\gamma(T)f, g) = \int h(k)(a(k)f, a(k)g)dk. \tag{2.3}$$

The *mass* of the phonon is a positive number μ, which is fixed throughout this paper. The relativistic free phonon Hamiltonian, $K = \gamma(M_w)$ is given in (1.3) and w is given in (1.4).

For each measurable subset $B \subset R^d$ and any real τ we define the local fractional energy operator

$$N_{\tau, B} = \gamma(M_w^{\tau} \cdot \chi_B), \tag{2.4}$$

where χ_B is the characteristic function of B. When $\tau = 0$, we write

$$N_{0, B} = N_B. \tag{2.5}$$

N_B is "the number of particles with momentum in B" operator. When $B = R^d$, we write (2.6)

$$N_{\tau, B} = N_{\tau}. \tag{2.6}$$

If $\tau = 0$ and $B = R^d$, we write

$$N_{\tau, B} = N. \tag{2.7}$$

Note that $K = N_1$.

If T is multiplication by the ith coordinate function, k_i, then we let P_i denote $\gamma(T)$ and write $P = (P_1, P_2, \ldots, P_d)$. P_i is "the phonon momentum in the ith direction" operator. For any positive number m, denoting the bare mass of the electron, the contribution of the electron kinetic energy to the energy of the polaron with total momentum p is the self-adjoint, closed operator, $E_m(p-P)$, where E is given by (1.2). For $f \in D(E_m(p-P)) \cap \mathfrak{F}_n$ we have

$$(E_m(p-P)f)(k_1, \cdots, k_n) = E_m(p_1 - k_{11} - \cdots - k_{n1}, \cdots, p_d - k_{1d} - \cdots - k_{nd}) \times f(k_1, \cdots, k_n), \tag{2.8}$$

where p_i and k_{ji} are the ith coordinates of p and k_j, respectively.

For g in $\mathfrak{F}_1$ we define the smooth annihilation and creation operators by

$$a(g) = \int \overline{g}(k)a(k)dk, \qquad a^*(g) = \int g(k)a^*(k)dk. \tag{2.9}$$

For such g we also define

$$R(g) = (a(g) + a^*(g))^{**}. \tag{2.10}$$

$R(g)$ is self-adjoint (Ref. 11, p. 231).

A cutoff function is a real valued, infinitely differentiable function f on R^d with compact support and satisfying $f(k) = f(-k)$.

The Hamiltonian for the polaron of total momentum p, bare mass m and cutoff function f is

$$H_{f,m}(p) = K + E_m(p-P) + R(f). \tag{2.11}$$

For an explanation of the terminology see Refs, 12, 4, or 13.

$H_{f,m}(p)$ is a self-adjoint operator which is bounded below, (Ref. 12, p. 102). Let $\lambda(f, m, p) = \inf$ (spectrum $H_{f,m}(p)$). The physical polaron mass is a positive number, m_0, which will remain fixed throughout this paper. The finite mass renormalized Hamiltonian of total momentum p, momentum cutoff f, and physical mass m_0 is $H'_f(p) = H_{f,m_0}(p) - \lambda(f, m_0, 0) + m_0$. Instead of $H'f(p)$ we will consider

$$H_f(p) = H'_f(p) - m_0 \tag{2.12}$$

so that inf (spectrum $(H_f(0))) = 0$.

$H_f(p)$ is self-adjoint and it is known (Ref. 12, p. 102) that

$$\inf(\text{spectrum } H_f(p)) \geq 0. \tag{2.13}$$

Also [according to (Ref. 12, p. 102)] zero is an eigenvalue of $H_f(0)$ with multiplicity one. Let Ψ_f be a corresponding eigenvector with norm one:

$$H_f\Psi_f = 0, \qquad \|\Psi_f\| = 1. \tag{2.14}$$

Ψ_f is an infinitesimal rest state of total momentum zero.

Let $\mathfrak{A}_0$ be the set of all finite linear combinations of operators of the form $e^{iR(g)}$, where g is an infinitely differentiable function of compact support. Then $\mathfrak{A}_0$ is an irreducible algebra and $\mathfrak{A}_0 h$ is dense for any h in $\mathfrak{F}$.

Let T be a closed operator from one Hilbert space to another. A *T-approximate sequence* for y in $D(T)$ = domain of T is a sequence $\{y_n\} \subset D(T)$ such that $y_n \to y$ and $T_{y_n} \to T_y$.

A subset $D \subset D(T)$ is a *core* for T if for each y in $D(T)$, D contains a T-approximate sequence for y (Ref. 14, p. 166).

Lemma 2.1: $\mathcal{E}$ is a core for

(i) $K(p) = K + E_{m_0}(p-p)$,

(ii) $N_\tau^{m/2}K(p)$, $m = 1, 2, \ldots, \tau \leq 1$,

(iii) $N_\tau^{n/2}E_{m_0}(p-p)$, $m = 1, 2, \ldots, \tau \leq 1$,

(iv) $N_{\tau_1}^{m_1/2}N_{\tau_2}^{m_2/2}$, $m_1, m_2 = 1, 2, \ldots, \tau_1, \tau_2 \leq 1$,

(v) $N_\tau^{m/2}(H_f(p) + r)$ $m = 1, 2, \ldots, \tau \leq 1, r$ real f a cutoff function

(vi) $N_\tau^{m/2}K$, $m = 1, 2, \ldots, \tau \leq 1$,

(vii) $N_\tau^{m/2}R(g)$, $m = 1, 2, \ldots, \tau \leq 1$, $g \in \bigcap_{n=1}^{\infty} D(N_\tau^n)$. (2.15)

Proof: The operators in (i)–(vii) are all relatively bounded (Ref. 14, p. 190) with respect to K^s for some sufficiently large integer s. Thus it suffices to prove that $\mathcal{E}$ is a case for K^s. Let $\mathcal{E}_0 = \{g = \sum_{n=0}^{\infty} g_n \in \mathcal{E}: g_n$ has compact support$\}$. Then $\mathcal{E}_0$ is a dense set of analytic vectors for K^s and so is a core for K^s by Nelson's theorem (Ref. 13, p. 583). Hence $\mathcal{E}$ is also a core.

B. N_τ Bounds for fixed momentum states

The basis for the estimates in this section is the following inequality due to Gross (Ref. 9, p. 23):

$$\|a(k_1)\cdots a(k_n)h\| \leq \prod_{j=1}^{n} |f(k_j)|w(k_j)^{-1}\|h\| + \sum_\alpha g_\alpha^n(k_1, \ldots, k_n)\|A(\alpha)H_f(0)h\|, \tag{2.16}$$

where n is a positive integer, α is a subset of $\{1, 2, \ldots, n\}$, the $\sum$ runs over all such α, h is in $\mathcal{E}$, f is a cutoff function, $A(\alpha) = \prod_{j\in\alpha} a(k_j)$, $g_{\{1\}}^1(k) = w(k)^{-1}$ and where g_α^n is defined inductively by

$$g_\alpha^{n+1}(k_1, \ldots k_{n+1}) = \left(\sum_{i=1}^{n+1} w(k_i)\right)^{-1} \sum_{i\notin\alpha} |f(k_i)|g_\alpha^n(k_1, \ldots, \hat{k_i}, \ldots, k_{n+1})$$

unless $\alpha = \{1, \ldots, n+1\}$ in which case $g_\alpha^{n+1}(k_1, \ldots, k_{n+1}) = [\sum_{i=1}^{n+1} w(k_i)]^{-1}$

(where a $\wedge$ over an element means to omit that element). The proof of (2.16) is based on the nonnegativity of $H_f(p)$ and the commutation relations of $H_f(p)$ with $a(k)$. We shall shortly return to this technique in the proof of Lemma 2.12.

For the remainder of this section f will be a fixed cutoff function and we will write $H_f(p) = H(p)$, $\Psi_f = \Psi$, and $E_{m_0} = E$.

Lemma 2.2: Ψ is in $D(K^{s/2})$ for every nonnegative integer s.

Proof: The inequality

$$E(p - k_1 - \cdots - k_n) \leq w(k_1) + \cdots + w(k_n) + |p| + m_0 \tag{2.17}$$

shows that $K^{(m+2)/2}$ and $K^{m/2}K(0)$ are relatively bounded with respect to each other. Thus

$$D(K^{(m+2)/2}) = D(K^{m/2}K(0)). \tag{2.18}$$

By definition, Ψ is in the domain of $H(0) = K + E(-P) + R(f) - \lambda(f, m_0, 0)I$. The lemma is therefore true for $s = 2$ and so for $s = 0, 1$ also. Inductively, assume the lemma is true for $s = 0, 1, 2, \ldots, t$.

That is, we have

(i) $\Psi \in D(K^{t/2})$.

Since $H(0)\Psi = 0$, we have

(ii) $\Psi \in D(K^{(t-1)/2}H(0))$.

Since $K^{(t-1)/2}R(f)$ is relatively bounded with respect to $K^{t/2}$, it is also true that

(iii) $\Psi \in D(K^{(t-1)/2}R(f))$

Subtracting (iii) from (ii) gives

(iv) $\Psi \in D(K^{(t-1)/2}K(0))$.

Now we are done with (2.18).

Lemma 2.3: For any $\tau \leq 1$, $\|N_{\tau,B}^{1/2}\Psi\| \leq \int_B |f(k)|^2 \times \omega(k)^{\tau-2}dk$.

Proof: For any h in $\mathcal{E}$ we have, by (2.16), that

$$\begin{aligned}\|N_{\tau,B}^{1/2}h\| = (N_{\tau,B}h, h) &= \left(\int_B w(k)^\tau (a(k)h, a(k)h)dk\right)^{1/2} \\ &\leq \left(\int_B w(k)^\tau |f(k)|^2 w(k)^{-2}dk\right)^{1/2}\|h\|^{1/2} \\ &\quad + \left(\int_B w(k)^{\tau-2}\|a(k)H(0)h\|^2dk\right)^{1/2}.\end{aligned}$$

Since $w(k) \geq \mu$ we have shown that for any h in $\mathcal{E}$

$$\|N_{\tau,B}^{1/2}h\| \leq \left(\int |f(k)|^2 w(h)^{\tau-2}dk\right)^{1/2}\|h\| + \mu^{\tau-2}\|N^{1/2}H(0)h\|. \tag{2.18'}$$

We saw in Lemma 2.2, that Ψ is in $D(K^{3/2})$ so by Lemma 2.1 we may choose a $K^{3/2}$-approximate sequence, $\{h_n\}$, in $\mathcal{E}$, for Ψ. $N_{\tau,B}^{1/2}$ is relatively bounded with respect to $K^{3/2}$. $N^{1/2}H(0)$ is relatively bounded with respect to $K^{1/2}H(0)$ which is relatively bounded with respect to $K^{3/2}$. Consequently $N_{\tau,B}^{1/2}h_n$ converges to $N_{\tau,B}^{1/2}\Psi$ while $N^{1/2}H(0)h_n$ converges to $N^{1/2}H(0)\Psi = 0$. Thus the lemma follows from (2.18) by taking the limit $h_n \to \Psi$ and recalling the normalization $\|\Psi\| = 1$.

We again follow Ref. 9 in making the next definition.

Definition 2.4: We have already observed that $\mathcal{E}$ is a core for $N_s^{1/2}$, where s is any open set in R^d. Thus, if h is in $D(N_s^{1/2})$, there is a sequence $\{h_n\} \subset \mathcal{E}$ such that $h_n \to h$ and $N_s^{1/2} \to N_s^{1/2}h$. The function $k \to a(h)h_n$ is in $L^2(S, \mathfrak{F})$ since h_n is in $\mathcal{E}$. since

$$\|N_s^{1/2}h_n\|^2 = (N_s h_n, h_n) = \int_S \|a(k)h_n)\|^2 dk, \tag{2.19}$$

it follows that $\{a(\cdot)h_n\}$ is a Cauchy sequence of functions in $L^2(S, \mathfrak{F})$. Thus, there is an element of $L^2(S, \mathfrak{F})$, which we denote as $k \to a(k)h$ or $a(\cdot)h$ or $a(k)h$, such that $a(\cdot)$ $a(\cdot)h_n \to a(\cdot)h$ in $L^2(S, \mathfrak{F})$. Furthermore,

$$\|N_s^{1/2}h\|^2 = \int \|a(k)h\|^2 dk. \tag{2.20}$$

Note that $a(\cdot)h$ is independent of $N_s^{1/2}$-approximate sequence, $\{h_n\} \subset \mathcal{E}$, is used. If h happens to be in $N_{\tau,s}^{1/2}$, $0 \leq \tau$, then h is also in $N_s^{1/2}$. Choosing an $N_{\tau,s}^{1/2}$-sequence for h in $\mathcal{E}$, $\{h_n\}$ and noticing that this must also be an $N_s^{1/2}$ approximate sequence gives that $w(\cdot)^{\tau/2}a(\cdot)h_n$ converges to something while $a(\cdot)h_n$ converges to $a(\cdot)h$. Since multiplication by a real continuous function is a closed operator in $L^2(S, \mathfrak{F})$ we find that, for any h in $N_{\tau,s}^{1/2}$ is relatively bounded with respect to $N_s^{1/2}$ for $\tau \leq 0$, it is also true that for any h in $N_s^{1/2}$, $\tau \leq 0$

$$\|N_{\tau,s}^{1/2}h\|^2 = \int_s w(k)^\tau \|a(k)h\|^2 dk. \tag{2.21}$$

Definition 2.5: We now define a sequence of approximate δ-functions. Put

$$\delta_n(x) = c(n)\exp[(|x|^2 - n^{-2})^{-1}] \begin{cases} \text{for } x \leq 1/n, \\ \text{for } x > 1/n, \end{cases} \tag{2.22}$$

where

$$c(n) = \int_{|x| \leq 1/n} \exp[(|x|^2 - n^{-2})^{-1}]dk. \tag{2.23}$$

for k in R^d put

$$\delta_{n,k}(x) = \delta_n(x - k). \tag{2.24}$$

These functions satisfy

$$\delta_n \in C_C^\infty(R^d), \tag{2.25}$$

$$0 \leq \delta_n(x), \quad \text{for all } x \text{ in } R^d, \tag{2.26}$$

$$\|\delta_n(x)\|_1 = \int \delta_n(x)dx = 1. \tag{2.27}$$

Given any open set U containing the origin in R^d, there is an integer $M(U)$ such that $n > M(U)$ implies support $(\delta_n) \subset U$. (2.28)

Lemma 2.6: Fix $v \in R^d$. For h in $D(N^{1/2})$ the functions h_n given by $h_n(k) = a(\delta_{n,k+v})h$ are in $L^2(S, \mathfrak{F})$ and

$$\lim h_n(\cdot) = a(\cdot + v)h \text{ in } L^2(S, \mathfrak{F}), \tag{2.29}$$

for any bounded measurable $S \subset R^d$.

Proof: First we remark that since $h \in D(N^{1/2})$, h is also in $D(N_s^{1/2})$ for any S, and consequently $a(\cdot)h$ is in $L^2(S, \mathfrak{F})$. By $a(\cdot + v)h$ we mean the translate of $a(\cdot)h$ by $-v$. Thus $a(\cdot + v)$ is in $L^2(S, \mathfrak{F})$ also. Moreover, translation is continuous in $L^2(R^d, \mathfrak{F})$ so whenever v and z are sufficiently close in R^d, then $\|a(\cdot + v)h - a(\cdot + z)h\|$ is arbitrarily small.

Now let $g \in L^2(R^d)$. Then the function $k \to \bar{g}(k)a(k)h$ is measurable from R^d to $\mathfrak{F}$. Since

$$\int \|\bar{g}(k)a(k)h\|\,dk \leq \int |g(k)|\,\|a(k)h\|\,dk \leq \|g\|_2\|N^{1/2}h\|, \tag{2.30}$$

it follows that the function $\bar{g}(\cdot)a(\cdot)h$ is strongly integrable. Next recall that $\mathcal{E}$ is a core for $N^{1/2}$ and that $A(g)$ is relatively bounded with respect to $N^{1/2}$. Consequently, we may choose $\{h_n\}$ in $\mathcal{E}$ such that

(i) $h_n \to h$,

(ii) $N^{1/2}h_n \to N^{1/2}h$,

(iii) $a(\cdot)h_n \to a(\cdot)h_n$ in $L^2(R^d, \mathfrak{F})$,

(iv) $a(g)h_n \to a(g)h$.

Furthermore, from (2.30)

(v) $\int_{R^d} \bar{g}(k)a(k)h_n dk \to \int_{R^d} \bar{g}(k)a(k)h dk$.

Thus, from (2.9)

$$a(g)h = \lim_{n\to\infty} a(g)h_n = \lim_{n\to\infty} \int \overline{g}(k)a(k)h_n dk = \int_{R^d} \overline{g}(k)a(k)dk.$$

Therefore we have verified that

$$\text{for } h \text{ in } D(N^{1/2}) \text{ and } g \text{ in } L^2(R^d), \tag{2.31}$$

the function $\overline{g}(\cdot)a(\cdot)h$ is strongly integrable and $a(g)h = \int \overline{g}(k)a(k)hdk$.

In particular

$$\begin{aligned} h_n(k) - h_n(k^1) &= \int \delta_{n,v+k}(p)a(p)hdp - \int \delta_{n,v+k}(p)a(p)hdp \\ &= \int \delta_{n,v}(p)(a(p+k)h - a(p+k^1)h)dp. \end{aligned} \tag{2.32}$$

Thus

$$\| h_n(k) - h_n(k^1)\|^2 \le \int \delta_{n,v}(p) \| a(p+k)h - a(p+h^1)h\|^2 dp$$

is arbitrarily small by the remarks at the beginning of this proof and (2.27). The functions h_n are continuous, hence weakly measurable and by the separability of $\mathfrak{F}$, strongly measurable. From the inequality

$$\|h_n(k)\|_{\mathfrak{F}} \le \| \delta_{n,v} \|_{L^2(R^d)} \|N^{1/2}h\|_{\mathfrak{F}} \tag{2.33}$$

it follows that $h_n(\cdot)$ is in $L^2(S, \mathfrak{F})$ for any bounded measurable subset S of R^d.

Finally we use (2.27) to make the estimate

$$\begin{aligned} \int_S \| a(\delta_{n,k+v})h - a(k+v)h\|dk \\ \le \int_{R^d} \int_S \delta_n(p) \| a(k+p+v)h\|^2 dkdp. \end{aligned}$$

This estimate along with the continuity of translation in $L^2(R^d, \mathfrak{F})$ gives the desired results.

Definition 2.7: Let $\mathcal{K}$ be a separable Hilbert space and S a measurable subset of R^d. Let h be in $L^2(S, \mathcal{K})$ and U a unitary operator on $\mathcal{K}$. We define a new function Uh from S to $\mathcal{K}$ by

$$(Uh)(k) = U(h(k)). \tag{2.34}$$

Uh is measurable since if g is in $\mathcal{K}$ then $k \to (Uh(k), g) = (h(k), U^{-1}g)$ is measurable. Furthermore $\| Uh\| = \| h\|$ since $\| (Uh)(k)\| = \| h(k)\|$ and $UU^{-1}h = h$ so U defines a unitary operator, which we also call U, on $L^2(S, \mathcal{K})$ by $U: h \to Uh$.

Lemma 2.8: Let h be in $D(N)$ and g in $L^2(R^d)$. Then $e^{iR(g)}h$ is in $D(N^{1/2})$ and for almost every k in R^d we have

$$a(k)e^{iR(g)}h = e^{iR(g)}a(k)h + ig(k)e^{iR(g)}h. \tag{2.35}$$

Proof: First let h be in $\mathfrak{F}^0$. Put $E_n = \sum_{s=0}^{n}(i^s/s!)R^s(g)$.

The commutation relations

$$\begin{aligned} &[a(v), a^*(z)] = (z, v), \\ &[a(v), a(z)] = 0 \end{aligned} \tag{2.36}$$

combine to give

$$[a(v), R(z)] = (z, v) \tag{2.37}$$

valid on $\mathfrak{F}^0$, for v, z in $L^2(R^d)$.

(2.37) gives

(i) $a(v)E_n = E_n a(v) + i(g, v)E_{n-1}$

also valid on $\mathfrak{F}^0$.

$\mathfrak{F}^0$ is a set of analytic vectors for $R(g)$ so

(ii) $E_n a(v)h \to e^{iR(g)}a(v)h$

and

(iii) $i(g, v)E_{n-1}h \to i(g, v)e^{iR(g)}h$.

Since $a(v)$ is a closed operator, (i)–(iii) show that for h in $\mathfrak{F}^0$, $e^{iR(g)}h$ is in $D(a(v))$ and

$$a(v)e^{iR(g)}h = e^{iR(g)}a(v)h + i(g, v)e^{iR(g)}h. \tag{2.38}$$

Since $\mathfrak{F}^0$ is a core for $N^{1/2}$ and since $N^{1/2}$ relatively bounds $a(v)$, it follows from (2.38) that if h is in $D(N^{1/2})$, then $e^{iR(g)}h$ is in $D(a(v))$ and (2.38) is valid for h in $D(N^{1/2})$.

[We have again used that $a(v)$ is closed.]

Now assume h is in $D(N)$. Then by (Ref. 9, p. 64) $e^{iR(g)}h$ is in $D(N) \subset D(N^{1/2})$. Replacing v by $\delta_{n,k}$ in (2.38), taking limits as $n \to \infty$ in (2.38) and using (2.29) shows that the square integrable function $k \to a(k)e^{iR(g)}h$ equals the function $k \to e^{iR(g)}a(k)h + ig(k)e^{iR(g)}h$. This is (2.35)

Lemma 2.9: Let $g \in \mathfrak{F}_1 = L^2(R^d)$ be in the domain of $N_{\tau,s}^{1/2}$. Then $N_{\tau,s}^{1/2}e^{iR(g)}$ is relatively bounded with respect to $N_{\tau,s}^{1/2}$. In particular, if h is in $D(N_{\tau,s}^{1/2})$, then $e^{iR(g)}h$ is in $D(N_{\tau,s}^{1/2})$ and

$$\| N_{\tau,s}^{1/2}e^{iR(g)}h\| \le \|N_{\tau,s}^{1/2}h\| + \| N_{\tau,s}^{1/2}g\| \, \| h\| . \tag{2.39}$$

Proof: First let h be in $\mathcal{E}$. Then h is in $D(N_{\tau,s})$ so by (Ref. 9, p. 64), $e^{iR(g)}h$ is in $D(N_{\tau,s}) \subset D(N_{\tau,s}^{1/2})$. According to (2.22) we may write

(i) $\| N_{\tau,s}^{1/2}e^{iR(g)}h\|^2 = \int_S w(k)^\tau \| a(k)h\|^2 dk.$

By Lemma 2.28 we have the estimate

$$\begin{aligned} \text{(ii)} \quad \| N_{\tau,s}^{1/2}e^{iR(g)}h\| &\le (\int (\| w(k)^{\tau/2}e^{iR(g)}a(k)h \\ &\quad + \| w(k)^{\tau/2}ig(k)e^{iR(g)}h\|)^2 dk)^{1/2} \\ &\le \| N_{\tau,s}^{1/2}h\| + \| N_{\tau,s}^{1/2}g\| \, \|h\|. \end{aligned}$$

Since $\mathcal{E}$ is a core for $N_{\tau,s}^{1/2}$, since $e^{iR(g)}$ is unitary and since $N_{\tau,s}^{1/2}$ is closed, we are done.

Definition 2.10: Put $R_f(p, r) = (H_f(p) + r)^{-1}$, for p in $R^2, r > 0$. In this section where we have fixed the cutoff function f we will write $R(p, r) = R_f(p, r)$.

Lemma 2.11: If $g \in \mathfrak{F}_1 = L^2(R^d)$ is also in $D(K)$, then $R(p, r)e^{iR(g)}\Psi$ is in $D(K^{3/2})$.

Proof: From the previous lemma with $\tau = 1, S = R^d$ we find that $e^{iR(g)}\Psi$ is in $D(K^{1/2})$ so that $R(p, r)e^{iR(g)}\Psi$ is in $D(K^{1/2}(H(p) + r))$. From (2.11) and (2.12) it is also true that $R(p, r)e^{iR(g)}\Psi$ is in $D(K) \subset D(K^{1/2})$. Consequently, $R(p, r)e^{iR(g)}\Psi$ is in $D(K^{1/2}(H(p)))$. Since $K^{1/2}R(g)$ is relatively bounded with respect to K, we also have that $R(p, r)e^{iR(g)}\Psi$ is in $D(K^{1/2}R(g))$ and therefore in $D(K^{1/2}K(0))$. We are done by (2.18).

Lemma 2.12: for h in $D(N_{\tau-2,s}^{1/2}(H(p) + r))$

$$\| N_{\tau,s}^{1/2}h\| \le (\int_S w(k)^{\tau-2}(f(k))^2 dk)^{1/2}\|h\| + \|N_{\tau-2,s}^{1/2}(H(p) + r)h\|.$$

Proof: On $\mathcal{E}$ the following computations are valid.

$$\begin{aligned} H(-k+p)a(k) &= a(k)H(-k+p) + [H(-k+p), a(k)] \\ &= a(k)H(-k+p) - w(k)a(k) \\ &\quad - f(k) + [E(-k+p-P), a(k)] - a(k)H(p) + a(k)H(p) \\ &= a(k)(E(-k+p-P) - E(p-P)) - w(k)a(k) - f(k) \\ &\quad + [E(-k+p-P), a(k)] + a(k)H(p). \end{aligned} \tag{2.40}$$

However,

$$\begin{aligned} a(k)(E(-k+p+P) - E(p-P)) &= a(k)E(-k+p-P) \\ &\quad - E(p-k-P)a(k) = -[E(-k)+p-P), a(k)]. \end{aligned}$$

Thus on $\mathcal{E}$

$$(H(-k+p)+r)a(k) = -w(k)a(k) - f(k) + a(k)(H(p)+r). \tag{2.41}$$

From (2.13), $H(-k+p)+r$ is nonnegative so

$$\begin{aligned} 0 &\le ((H(-k+p)+r)a(k)h, a(k)h) \\ &= -w(k)\|a(k)h\|^2 - f(k)(h, a(k)h) \\ &\quad + (a(k)(H(p)+r)h, a(k)h). \end{aligned}$$

Thus

$$\|a(k)h\| \le |f(k)|w(k)^{-1}\|h\| + \|a(k)(H(p)+r)h\|w(k)^{-1}. \tag{2.42}$$

We are done with (2.41) and (2.21) or (2.22), for h in $\mathcal{E}$. Since $\mathcal{E}$ is a core for $N_{\tau-2,s}^{1/2}(H(p)+r)$ and since $N_{\tau,s}^{1/2}$ is closed, we are done.

Lemma 2.13: suppose $g \in \mathfrak{F}_1 \cap D(N_{\tau-2,s}^{1/2})$ for some $\tau \le 1$. Then

$$\|N_{\tau,s}^{1/2}q\| \le (1+1/r)\|N_{\tau-2,s}^{1/2}f\| + \|N_{\tau-2,s}^{1/2}g\|,$$

where $q = R(p,r)e^{iR(g)}\Psi$.

Proof: By Lemma 2.11 $q \in D(K^{3/2})$. Let $\{q_n\} \subset \mathcal{E}$ be a $K^{3/2}$-approximate sequence for q. Since $N_{\tau-2,s}^{1/2}(H(p)+r)$ is relatively bounded with respect to $K^{3/2}$, $\{q_n\}$ is also a $N_{\tau-2,s}^{1/2}(H(p)+r)$ approximate sequence. Apply Lemma 2.12 to q_n and take limits. We may conclude that $\{N_{\tau-2,s}^{1/2}q_n\}$ is a Cauchy sequence. Since $N_{\tau,s}^{1/2}$ is closed we may also conclude that q $D(N_{\tau,s}^{1/2})$ and

$$\|N_{\tau,s}^{1/2}q\| \le \|N_{\tau-2,s}^{1/2}f\| \cdot \|q\| + \|N_{\tau-2,s}^{1/2}(H(p)+r)q\|. \tag{2.43}$$

The desired result now follows from Lemmas 2.9 and 2.3.

Remarks on Section B:

(a) Lemma 2.2 implies that Ψ is in $D(N^s)$ for all integers s. In fact, Ψ is an entire vector for N, (Ref. 15, p. 150).

(b) If we write $\Psi = \sum_{n=0}^{\infty}\Psi_n$, where Ψ_n is in $\mathfrak{F}_n$, then the finite particle components, Ψ_n, of Ψ are continuous functions and we can use (2.1) to define the action of $a(k)$ on each Ψ_n and so on Ψ, (Ref. 15, Sec. III. 2). Then (2.16) is valid for h replaced by Ψ or Ψ_n and the last sum in 2.16 is zero since $H_f(0) = 0$. This gives further estimates:

$$|\Psi_n(k_1, \ldots, k_n)| \le (n!)^{-1/2}\prod_{i=1}^{n} f(k_i)\cdot w(k_i)^{-1}. \tag{2.44}$$

(c) Using the just mentioned variation of (2.16) one can show that Ψ is in $D(N_{\tau,s}^{1/2})$ for all τ. Then one can show that Lemma 2.13 is valid for all τ also.

(d) Nelson's method of mass renormalization[1] was used in (2.12) rather than that suggested by Gross.[9] Frolich[16] has shown that the method of Ref. 9 leads to a trivial dynamics whereas the dynamics here have a momentum dependence as will be proved in Theorem 3.11.

C. Sobolev bounds for fixed momentum states

We again follow Ref. 9 and make the next definitions. Let $C_C^{\infty}(S, \mathfrak{F})$ be the infinitely differentiable functions from S to $\mathfrak{F}$ with compact support for any open subset S of R^d. We say that h in $L^2(S, \mathfrak{F})$ is *weakly differentiable* in $L^2(S, \mathfrak{F})$ if there are functions h_i in $L^2(S, \mathfrak{F})$, $i = 1, 2, \ldots, d$, such that for all ϕ in $C_C^{\infty}(S, \mathfrak{F})$ we have

$$-\left(\frac{\partial\phi}{\partial k_j}, h\right)_s = (\phi, h_j), \tag{2.45}$$

where k_j is the jth coordinate of $k = (k_1, \ldots, k_d)$ and where, for any v and z in $L^2(S, \mathfrak{F})$,

$$(v, z)_s = \int_S (v(k), z(k))dk. \tag{2.46}$$

We denote h_j by $\partial h/\partial k_j$. Recalling Definition 2.4, we say that h in $D(N_s^{1/2})$ is *weakly differentiable* if $k \to a(k)h$ is weakly differentiable. We define the Sobolev space

$$\mathfrak{F}^1(S) = \{h \in D(N_s^{1/2}): h \text{ is weakly differentiable}\}. \tag{2.47}$$

$\mathfrak{F}^1(S)$ is a Hilbert space in the norm $\|\cdot\|_s$ where

$$\|h\|_s^2 = \|h\|_{\mathfrak{F}} + \|N_s^{1/2}h\|_{\mathfrak{F}} + \sum_{j=1}^{d}\int\left\|\frac{\partial}{\partial k_j}a(k)h\right\|^2 dk. \tag{2.48}$$

If $h \notin \mathfrak{F}^1(S)$, then put $\|h\|_s = \infty$.

In Sec. C, f will be a fixed cutoff function. We will write $H_f(p) = H(p)$, $\Psi_f = \Psi$, $E_{m_0} = E$ and $R_f(p,r) = R(p,r)$.

Proposition 2.14: [Gross, Ref. 9, p. 34]: For any open set S in R^d, Ψ is in $\mathfrak{F}^1(S)$. If $\|\Psi\| = 1$, then

$$\|\Psi\|_s \le 1 + \int_s \frac{|f(k)|^2}{w(k)^2}dk + 8d\int_s \frac{|f(k)|^2}{w(k)^4}dk + 2\int_s \frac{|\nabla f|^2}{w(k)^2}dk, \tag{2.49}$$

where ∇f denotes the gradient of f.

Lemma 2.15: Fix a vector c in R^d and normalize $\|\Psi\| = 1$.

Let $\mathcal{L}$ be a bounded function on S and let g be in $C_C^{\infty}(R^d)$. Then

$$\begin{aligned} &\int_s \mathcal{L}(k)\|a(k+c)e^{iR(g)}\Psi - a(k)e^{iR(g)}\Psi\|^2 dk \\ &\quad \le 8|c|^2\int_s \frac{\mathcal{L}(k)}{w(k)^2}\frac{|f(k+c)|^2}{w(k+c)^2}dk \\ &\quad\quad + 2\int_s \frac{\mathcal{L}(k)}{w(k)^2}|f(k) - f(k+c)|^2 dk \\ &\quad\quad + 2\int_s \mathcal{L}(k)|g(k+c) - g(k)|^2 dk. \end{aligned}$$

Proof: Let Q be the left side of the inequality in the statement of the lemma. Then by Lemma 2.8 we may write

$$\begin{aligned} Q \le \int_s \mathcal{L}(k)\{&2\|e^{iR(g)}(a(k+c)\Psi - a(k)\Psi)\|^2 \\ &+ 2\|i(g(k+c) - g(k))e^{iR(g)}\Psi\|^2\}dk. \end{aligned}$$

Since $e^{iR(g)}$ is unitary and $\|\Psi\| = 1$, we have

$$\begin{aligned} Q \le \int_s \mathcal{L}(k)2\|a(k+c)\Psi - a(k)\Psi\|^2 dk \\ + 2\int_s \mathcal{L}(k)|g(k+c) - g(k)|^2 dk. \end{aligned}$$

The proof of Lemma 3.7 in [Gross, Ref. 9, p. 33] shows that

$$\int l(k)\| a(k+c)\Psi - a(k)\Psi\|^2 dk \le 8|c|^2 \int \frac{\mathcal{L}(k)}{w(k)^2}\frac{|f(k)+c|^2}{w(k+c)^2}dk + 2\int_s \frac{\mathcal{L}(k)}{w(k)^2}|f(k)-f(k+c)|^2 dk.$$

Combined with the previous inequality this finishes the proof.

Lemma 2.16: Fix c in R^d and g in $C_C^\infty(R^d)$. Then for any open set S in R^d we have

$$\begin{aligned}\int_s \| a(k+c)R(p,r)e^{iR(g)}\Psi - a(k)R(p,r)e^{iR(g)}\Psi\|^2 dk &\le 12|c|^2(1+1/r)\mu^{-2}\|N_{2,s-c}^{1/2} f\| + 12|c|^2\|N_{2,s-c}^{1/2} g\|\mu^{-2} \\ &+ \frac{3}{r}\int_s \frac{|f(k)-f(k+c)|^2}{w(k)^2}dk \\ &+ 24|c|^2\int_s \frac{|f(k+c)|^2}{w(k+c)^2 w(k)^4}dk \\ &+ 6\int_s \frac{|f(k)-f(k+c)|^2}{w(k)^4}dk \\ &+ 6\int_s \frac{|g(k+c)-g(k)|^2}{w(k)^2}dk.\end{aligned}$$

Proof: Let h be in $\mathcal{E}$. If k, k^1 are in R^d and if we put $\delta a = a(k^1) - a(k)$ then

$$w(k)\|(\delta a)h\| \le 2|k-k^1|\,\|a(k^1)h\| + |f(k)-f(k^1)|\,\|h\| + \|(\delta a)(H(p)+r)h\|. \quad (2.50)$$

The proof of this inequality is essentially given in [Gross, Lemma 3.6, pp. 32–33]. All the reader need do is replace $H(-k)$ by $H(-k-p)+r$, replace $E(-P+q)$ by $E(-P+q+p)$ where q may be $0, k$, or k^1, and replace $H(0)$ by $H(p)+r$ whenever these operators occur in the cited proof.

Ψ is in $D(K)$ by definition so $e^{iR(g)}\Psi$ is in $D(K^{1/2})$ by Lemma 2.9 and, putting $q = R(p,r)e^{iR(g)}\Psi$, we have

(i) $q \in D(K^{1/2}(H(p)+r)$.

But

(ii) $q \in D(H(p)+r) \subset D(K) \subset D(K^{1/2})$

so from (i) we deduce

(iii) $q \in D(K^{1/2}H(p))$.

Since $K^{1/2}R(g)$ is relatively bounded with respect to K from (ii) we see that q is in $D(K^{1/2}R(g))$ and so from (iii) it follows that

(iv) $q \in D(K^{1/2}K(p))$.

From (2.18) it now follows that $q \in D(K^{3/2})$. Since $\mathcal{E}$ is a core for $K^{3/2}$ and since both $N_s^{1/2}$ and $N_s^{1/2}(H(p)+r)$ are relatively bounded with respect to $K^{3/2}$, we may choose a $K^{3/2}$-approximate sequence, $\{h_n\}$, for q, in $\mathcal{E}$ satisfying

(a) $h_n \to q = R(p,r)e^{iR(g)}\Psi$

(b) $a(\cdot)h_n \to a(\cdot)q$ in $L^2(S,\mathfrak{F})$

(c) $a(\cdot)(H(p)+r)h_n \to a()e^{iR(g)}\Psi$, in $L^2(S,\mathfrak{F})$.

Putting $k^1 = k + c$ and $h = h_n$ in (2.50), we find

$$\| a(k+c)h_n - a(k)h_n\| \le w(k)^{-1}\{2|c|\,\|a(k+c)h_n\| + |f(k)-f(k+c)|\,\|h_n\| + \|(a(k+c)-a(k))(H(p)+r)h_n\|\}. \quad (2.51)$$

Taking the limit $n \to \infty$ in (2.51), using (a), (b), (c) above, and the fact that for any nonnegative numbers α, β, γ it is true that $(\alpha+\beta+\gamma)^2 \le 3\alpha^2 + 3\beta^2 + 3\gamma^2$, we find

$$\begin{aligned}\int_s \|(a(k+c)-a(k))q\|^2 dk &\le 12|c|^2\int_s \|a(k+c)q\|^2\frac{dk}{w(k)^2} \\ &+ 3\int_s |f(k)-f(k+c)|^2\frac{\|q\|^2}{w(k)^2}dk \\ &+ 3\int_s \|(a(k+c)-a(k))e^{iR(g)}\Psi\|^2\frac{dk}{w(k)^2}. \quad (2.52)\end{aligned}$$

If we replace the first integral with $12\mu^{-2}|c|^2\|N_{s-c}^{1/2}q\|^2$ and estimate this by Lemma 2.13, if we note that $\|q\| \le 1/r\,\|e^{iR(g)}\Psi\| = 1/r$ and if we use Lemma 2.15 to estimate the last integral, then we obtain the desired inequality.

Lemma 2.17: Let h be in $D(N_s^{1/2})$.

If

$$q_j = \lim_{\tau\to 0}\sup\frac{1}{\tau^2}\int_s \|(a(k+\tau e_j)-a(k))h\|^2 dk < \infty$$

for $j = 1, 2, \ldots, d$, where e_j is a unit vector in the jth coordinate direction, then h is weakly differentiable and

$$\left\|\frac{\partial}{\partial k_j}a(k)h\right\|^2_{L^2(S,\mathfrak{F})} \le q_j.$$

Proof: The proof of this lemma can be found in the proof of Proposition 3.8 of [Gross, Ref. 9, p. 35].

Proposition 2.18: Let S be an open set in R^d, fix p in R^d, $r > 0$ and g in $C_C^\infty(R^d)$. Normalize $\|\Psi\| = 1$. Then $R(p,r)e^{iR(g)}\Psi$ is in the Sobolev space $\mathfrak{F}^1(S)$ and

$$\begin{aligned}\|R(p,r)e^{iR(g)}\Psi\|_s^2 &\le \frac{1}{r} + \left(1+\frac{1}{r}\right)\left\|N_{2,s}^{1/2}f\right\| + \left\|N_{2,s}^{1/2}g\right\| \\ &+ 12\left(1+\frac{1}{r}\right)\mu^{-2}\left\|N_{2,s}^{1/2}f\right\| + 12\left\|N_{2,s}^{1/2}g\right\|\mu^{-2} \\ &+ \frac{3}{r}\int_s \frac{|\nabla f|^2}{w(k)^2}dk + d(24)\int_s \frac{|f(k)|^2}{w(k)^6}dk \\ &+ 6\int_s \frac{|\nabla f|^2}{w(k)^4}dk + 6\int_s \frac{|\nabla g|^2}{w(k)^2}dk\end{aligned}$$

Proof: This follows from Lemmas 2.17, 2.16 and the fact that

$$\|R(p,r)e^{iR(g)}\Psi\| \le (1/r)\|e^{iR(g)}\Psi\| \le (1/r)\|\Psi\| = (1/r).$$

Remarks on Sec. C

(a) Proposition 2.14 asserts that Ψ is weakly differentiable. In Ref. 15 it was shown that the finite particle components of Ψ, $\Psi_n \in \mathfrak{F}_n(\Psi\sum_{n=0}^\infty \Psi_n)$ are actually differentiable functions and the map $k \to a(k)\Psi$ is differentiable from R^d to $\mathfrak{F}$.

(b) As suggested by Lemma 2.15, $e^{iR(g)}\Psi$ is weakly differentiable if g is in $C_C^\infty(R^d)$ and

$$\frac{\partial}{\partial k_j}ake^{iR(g)}\Psi = e^{iR(g)}\frac{\partial}{\partial k_i}a(k)\Psi + i\frac{\partial g}{\partial k_i}(k)e^{iR(g)}\Psi.$$

3. REMOVING THE CUTOFF IN TWO SPACE DIMENSIONS

For this last section we will restrict our attention to the case of two space dimensions, $d = 2$. We assume the normalization $\|\Psi_f\| = 1$.

A. The case of fixed total momentum

In this section $S = S_r$ will denote a cube in R^2: $S_r = \{(k_1, k_2): |k_i| < r, i = 1, 2\}$ for $r > 0$. By $\mathfrak{F}_n(S)$ we mean the n-fold symmetric tensor product of $L^2(S)$. By $\mathfrak{F}(S)$ we mean Fock space over $L^2(S)$: $\mathfrak{F}(S) = \sum_{n=0}^{\infty} \mathfrak{F}_n(S)$. We will identify $\mathfrak{F}_n(S)$ with a closed subspace of $\mathfrak{F}_n = \mathfrak{F}_n(R^2)$ by extending functions initially defined on S^n to be zero on $R^{2\cdot n} - S^n$. $\mathfrak{F}(S)$ is also identified with a closed subspace of $\mathfrak{F}$ in this manner. It now makes sense to write $\|h\|_s$, where h is in $\mathfrak{F}(S)$ and $\|\cdot\|$ is given in (2.48). Let $\mathfrak{F}^1(L^2(S)) = \{h \in \mathfrak{F}(S): \|h\|_s < \infty\}$, With norm $\|\cdot\|_s$.

Lemma 3.1: Bounded subsets of $\mathfrak{F}^1(L^2(S))$ are precompact in $\mathfrak{F}(S)$.

Proof: According to [Gross, Ref. 9, p. 37] there is an operator L on $\mathfrak{F}(S)$ with domain $D(L) = \mathfrak{F}^1(L^2(S))$ satisfying

$$\|h\|_s = \|Lh\|_{\mathfrak{F}(S)} \tag{3.1}$$

for all h in $\mathfrak{F}(S)$.

Let B be a bounded subset of $\mathfrak{F}^1(L^2(S))$. Then by (3.1) LB is bounded in $\mathfrak{F}(S)$ and since L^{-1} is compact, $L^{-1}LB = B$ is precompact.

Let $P(S)$ be the projection of $\mathfrak{F}$ onto $\mathfrak{F}(S)$. On $\mathfrak{F}_0$, $P(S) = I$. On $\mathfrak{F}_n$, $P(S)$ is just multiplication by X_n, the characteristic function of S^n. If h is in $\mathcal{E}$ then

$$a(k)P(S)h = \begin{cases} P(S)a(k)h & k \in S \\ 0 & k \notin S. \end{cases}$$

If h is in $D(N_\tau^{1/2})$, $h = \sum h_n$, $h_n \in \mathfrak{F}_n$, then so is $\sum \chi_n h_n = P(S)h$, and $\|N_{\tau,S}^{1/2} P(S)h\| \le \|N_{\tau,S}^{1/2} h\|$. Thus, if h is in $D(N_S^{1/2})$ then so is $P(S)h$ and $a(k)P(S)h = P(S)a(k)h$ for almost every k in S where if $g: R^2 \to \mathfrak{F}$ then $(P(S)g)(k) = P(S)g(k)$. Since $P(S)$ is bounded $P(S)$ leaves $C_C^\infty(R^2)$ invariant and

$$P(S)\frac{\partial}{\partial k_j} g = \frac{\partial}{\partial k_j} P(S)g$$

for all g in $C_C^\infty(R^2)$. Consequently, if $h \in \mathfrak{F}$ is weakly differentiable then so is $P(S)h$ and for almost every k in S,

$$\frac{\partial}{\partial k_j} a(k)P(S)h = P(S)\frac{\partial}{\partial k_j} a(k)h.$$

Corollary 3.2: Let $\{f_n\}$ be a sequence of cutoff functions satisfying

$$\sup_n \|f_n\|_{L^2(S)} < \infty \tag{3.2}$$

and

$$\sup_n \|\nabla f_n\|_{L^2(S)} < \infty. \tag{3.3}$$

Fix g in $C_C^\infty(R^2)$, p in R^2 and $r > 0$.

Then the sequences

$$\{P(S)\Psi_{f_n}\} \tag{3.4}$$

and

$$\{P(S)(H_{f_n}(p) + r)^{-1} e^{iR(g)}\Psi_{f_n}\} \tag{3.5}$$

are precompact in $\mathfrak{F}$.

Proof: The proof consists of Lemma 3.1 along with the estimates of propositions 2.18 and 2.14, and noting that $\mathfrak{F}(S)$ is a closed subspace of $\mathfrak{F}$.

Lemma 3.3: Let $\{h_n\} \subset \mathfrak{F}$ and suppose

$$\sup \{\|N_\tau^{1/2} h_n\|\} < \infty \tag{3.6}$$

for some $\tau > 0$. Then for every $\epsilon > 0$ there is a cube S such that for all n

$$\|(I - P(S))h_n\| < \epsilon. \tag{3.7}$$

Proof: Let $M_1 = \sup_n \|N_\tau^{1/2} h_n\|$ and let $\epsilon > 0$ be given.

$w(k)$ is a positive radial function which increases monotonically to ∞ as $|k| \to \infty$, so there is a cube S such that for k outside S,

$$w(k)^\tau > M_1^2/\epsilon^2 = M. \tag{3.8}$$

Put $S^{n,c} = R^{2n} - S^n$ and

$$S_i^n = R^2 \times \ldots \times R^2 \times R^2 - S \times R^2 \times \ldots \times R^2, \tag{3.9}$$

where of the n factors in S_i^n only the ith is $R^2 - S$. Then

$$\begin{aligned} \|(I - P(S))h_n\|^2 &= \sum_{m=1}^{\infty} \int_{S^{m,c}} |h_{nm}(\bar{k})|^2 d\bar{k} \\ &\le \sum_{m=1}^{\infty} \sum_{i=1}^{m} \int_{S_i^m} w(k_i) M^{-1} |h_{nm}(\bar{k})|^2 d\bar{k} \\ &\le \|N_\tau^{1/2} h_n\|^2 M^{-1} \le M_1^2 M^{-1} = \epsilon^2, \end{aligned}$$

where $\bar{k}$ denotes a generic element of R^{n-2} with components $(k_1, \ldots, k_n)$, $k_i \in R^2$ and $h_n = \sum_{m=0}^{\infty} h_{nm}$ with $h_{nm} \in \mathfrak{F}_m$.

Corollary 3.4: Let $\{f_n\}$ be a sequence of cutoff functions satisfying

$$\sup_n \int_{R^2} w(k)^{\tau-2} |f_n(k)|^2 dk < \infty, \tag{3.10}$$

for some $1 > \tau > 0$. Then for every $\epsilon > 0$ there is a cube $S(\epsilon)$ such that

$$\|(I - P(S(\epsilon)))h_n\| < \epsilon \tag{3.11}$$

for all n where either $\{h_n\} = \{\Psi_{f_n}\}$ or $\{h_n\} = \{(H_{f_n}(p) + r)^{-1} e^{iR(g)}\Psi_{f_n}\}$, for $g \in C_C^\infty(R^2)$, $r > 0$ and $p \in R^2$.

Proof: The proof consists of applying the previous lemma with the estimates of Lemmas 2.13 and 2.3 and the hypothesis (3.10).

Proposition 3.5: Let $\{f_n\}$ be a sequence of cutoff functions satisfying (3.2), (3.3) and (3.10). Fix g in $C_C^\infty(R^2)$, p in R^2 and $r > 0$. Then, in every subsequence Λ_0 of $\{1, 2, \ldots,\}$ there is another subsequence $\Lambda \subset \Lambda_0$ such that

$$\{\Psi_n: n \in \Lambda\} \tag{3.12}$$

and

$$\{R_n(p, r)e^{iR(g)}\Psi_n: n \in \Lambda\} \tag{3.13}$$

are convergent sequences in $\mathfrak{F}(R^2)$, where we have put

$$\Psi_n = \Psi_{f_n} \text{ and } R_n(p, r) = (H_{f_n}(p) + r)^{-1}.$$

Proof: For each cube S and real $r > 0$ we may, by corollary 3.2, find a sequence $\Lambda(g, p, r, S) \subset \Lambda_0$ such that

$$\{P(S)h_n: n \in \Lambda(g, p, r, S\} \tag{3.14}$$

are convergent sequences in $\mathfrak{F}$ where either $\{h_n\} = \{\Psi_n\}$ or $\{h_n\} = \{R_n(p,r)e^{iR(g)}\Psi_n\}$. By the diagonal process and the fact that $0 < a < b$ implies $S_a \subset S_b$ which implies $\|P(S_a)X\| \leq \|P(S_b)X\|$, we may find a sequence $\Lambda(g, p, r)$ such that

$$\{P(S)h_n: n \in \Lambda(g; p, r)\} \tag{3.15}$$

are convergent in $\mathfrak{F}$ for every cube S.

Let $\epsilon > 0$ be given. By corollary 3.4, let S be a cube such that

$$\|(I - P(S))h_n\| < \epsilon/4. \tag{3.16}$$

By (3.15) choose an N such that $n, m > N$ imply $\|P(S)h_n - P(S)h_m\| < \epsilon/2$. Then, for $n, m > N$, $\|h_n - h_m\| \leq \|P(S)(h_n - h_m)\| + \|(I - P(S))h_n\| + \|(I - P(S))h_m\| < \epsilon$.

Lemma 3.6: Let Λ be a subsequence of $\{1, 2, \cdots\}$ such that $\{\Psi_n: n \in \Lambda\}$ converges to a limit Ψ_∞. If $\{R_n(p,r)e^{iR(g)}\Psi_n: n \in \Lambda_0\}$ converges for some infinite $\Lambda_0 \subset \Lambda$ then so does $\{R_n(p,r)e^{iR(g)}\Psi_\infty: n \in \Lambda_0\}$. Furthermore both limits agree.

Proof: Both assertions follow from the triangle inequality and the estimate $\|R_n(p,r)e^{iR(g)}(\Psi_n - \Psi_\infty)\| \leq 1/r\|\Psi_n - \Psi_\infty\|$.

Definition 3.7: Let g be a C^∞, real function on R^1 that is one on $(-1/2, 1/2)$, takes values between one and zero and vanishes outside $[-1, 1]$. For each positive integer m define g_m on R^1 by

$$g_m(x) = \begin{cases} 1 \text{ if } x < m \\ g(1/4 + x - m) \text{ if } x \geq m. \end{cases}$$

Then g_m is C^∞, is one on $(-\infty, m]$, takes values between one and zero and vanishes outside $(-\infty, m + 3/4)$. We now define the *realistic cutoff functions* f_n, *having real coupling constant* λ, by

$$f_m(x) = \lambda w(x)^{-1/2} g_m(|x|) \tag{3.17}$$

for x in R^2. Note that:

(a) f_n is a cutoff function;

(b) support $(f_m) \subset \{x: |x| < m + 1\}$;

(c) $f_m(x) = \lambda w(x)^{-1/2}$ if $|x| < m$;

(d) $|f_m(x)| \leq \lambda w(x)^{-1/2}$;

(e) f_m has the same sign as λ;

(f) $\sup_m \sup_i \|\partial/\partial x_i f_m\|_\infty < \infty$;

(g) $\sup_m \|\nabla f_m\|_{L^2(S)} < \infty$ for any bounded measurable subset $S \subset R^2$;

(h) $\sup_m \|f_m\|_{L^2(S)} < \infty$ for any bounded measurable subset $S \subset R^2$;

(i) $\sup_m \|N_\tau^{1/2} f_m\|^2 < \infty$ for any $\tau < 1$.

Of (a)–(i) only (i) requires a proof:

$$\|N_\tau^{1/2} f_m\|^2 = \int_{R^2} w(k)^\tau |f_m(k)|^2 dk$$
$$\leq \lambda^2 \int_0^\infty \int_0^{2\pi} (r^2 + \mu^2)^{(\tau-3)/2} r \, d\theta \, dr$$
$$\lambda^2 \cdot \frac{2\pi}{1 - \tau} \cdot \mu^{\tau - 1} \text{ if } \tau < 1.$$

Thus $\{f_n\}$ satisfies (3.2), (3.3), and (3.10). For the rest of Sec. 3, $\{f_n\}$ will denote this particular sequence of cutoff functions. We will also put $\Psi_n = \Psi_{f_n}, H_{f_n}(p) = H_n(p)$ and $R_n(p,r) = (H_n(p) + r)^{-1}$. Normalize so that $\|\Psi_n\| = 1$.

Theorem 3.8: There is a subsequence Λ in every sequence $\Lambda_0 \subset \{1, 2, \ldots\}$ and a bounded operator $R_\infty(p,r)$ for each p in R^2 and complex r not in $(-\infty, 0]$ such that

(i) $\lim_{\substack{n\to\infty \\ n\in\Lambda}} \Psi_n$ exists; we let Ψ_∞ denote this limit, $\|\Psi_\infty\| = 1$;

(ii) strong-$\lim_{\substack{n\to\infty \\ n\in\Lambda}} R_n(p,r) = R_\infty(p,r)$.

Furthermore, if $\tau > 0, R_\infty(p,r)$ is both self-adjoint and nonnegative.

Proof: By proposition 3.5, in any sequence Λ_1 we can find, for each p in $R^2, r > 0$ and g in $C_C^\infty(R^2)$, another sequence $\Lambda(p, r, g) \subset \Lambda_0$ such that the sequences (3.12) and (3.13) converge. Choose a particular p_0, r_0, and g_0 and let $\Psi_\infty = \lim_{n\to\infty} \{\Psi_n: n \in \Lambda(p_0, r_0, g_0)\}$. By Lemma 3.6 we have

$$\lim_{n\in\Lambda(p_0,r_0,g_0)} R_n(p_0,r_0)e^{iR(g_0)}\Psi_\infty = \lim_{n\in\Lambda(p_0,r_0,g_0)} R_n(p_0,r_0)e^{iR(g_0)}\Psi_n. \tag{3.18}$$

Let J be a countable dense subset of $\mathfrak{F}$ consisting of finite linear combinations of elements of the form $e^{iR(g)}\Psi_\infty, g \in C_C^\infty(R^2)$. Let $e^{iR(g_0)}\Psi_\infty \in J$. Then, by the diagonal process we can choose a sequence $\Lambda(p,r)$ of $\Lambda(p_0, r_0, g_0)$ such that

$$\lim_{\substack{n\to\infty \\ n\in\Lambda(p,r)}} R_n(p,r)v \text{ exists}, \tag{3.19}$$

for all v in J. Since $\{R_n(p,r)\}$ is a sequence of uniformly bounded linear operators converging strongly on a dense set we may conclude

$$\lim_{\substack{n\to\infty \\ n\in\Lambda(p,r)}} R_n(p,r)h \tag{3.20}$$

exists for all h in $\mathfrak{F}$. The set of all complex numbers, c, such that $(H_n(p) - c)^{-1}$ exists and are uniformly bounded is open in the complex plane [Ref. 14, 427], and contains all c not in $[0, \infty)$ because the $H_n(p)$ are nonnegative, self-adjoint operators. Consequently, this set of boundedness [Ref. 14, 427] is connected. Fix p and let $\Lambda(p) = \Lambda(p, r_0)$ for some $r_0 > 0$. The set of all c's such that $\{(H_n(p) - c)^{-1}: n \in \Lambda(p)\}$ converges strongly is known to be a union of some of the components of the set of boundedness, [Ref. 14, p. 427]. In this case, the set of boundedness is itself the only component and so the set of complex c's such that $\{(H_n(p) - c)^{-1}\}$ converges strongly is either empty or is the set of boundedness. Since $-r_0$ is in this set it can't be empty and we may conclude that

$$\lim_{\substack{n\to\infty \\ n\in\Lambda(p)}} R_n(p,r)h \tag{3.21}$$

exists for all complex r not in $(-\infty, 0]$ and h in $\mathfrak{F}$.

Let Q be a countable dense subset of R^2. By the diagonal process we may choose a sequence Λ such that

$$\lim_{\substack{n\to\infty \\ n\in\Lambda}} R_n(p,r)h \tag{3.22}$$

exists for all p in Q and complex r not in $(-\infty, 0]$.

Let $\epsilon > 0$ be given. Put $\alpha(r) = r^{-1}$ if $r > 0$ and $\alpha(r) = |\mathrm{Im}(r)|^{-1}$ if r is not real. Let $p \in R^2$ be arbitrary. Choose $g \in Q$ such that $|p-q| < \epsilon/3\alpha^2\|h\|$.

Then the resolvent equation

$$R_n(p,r) - R_n(q,r) = R_n(p,r)(E(q-P) - E(p-P))R_n(q,r)$$

and $\|(E(q)-E(p))^{**}\| < |p-q|$ give the estimate $\|(R_n(p,r)-R_n(q,r))h\| < \epsilon/3$. Choose N so large that $n, m > N$ imply $\|(R_n(q,r)-R_m(q,r)h\| < \epsilon/3$. Then along with the triangle inequality and the previous inequality we find that for $n, m > N$, $\|(R_m(p,r) - R_n(p,r)h\| < \epsilon$ so that

$$\lim_{\substack{n\to\infty \\ n\in\Lambda}} R_n(p,r)h$$

exists, for all p in R^2, complex r not in $(-\infty, 0]$ and h in $\mathfrak{F}$. We define

$$R_\infty(p,r)h = \lim_{\substack{n\to\infty \\ n\in\Lambda}} R_n(p,r)h.$$

$R_\infty(p,r)$ has norm bounded by $\alpha(r)$ since this is true of each $R_n(p,r)$. If $r > 0$, then $R(p,r)$ is symmetric and nonnegative since each $R_n(p,r)$ is.

Lemma 3.9: $\underset{\substack{r\to\infty \\ r\ real}}{\text{Strong limit}}\ r\,R_\infty(p,r) = I$, uniformly for p in compact subsets of R^2.

Proof: First we give the proof for $p = 0$. If suffices to prove convergence on the set J of Theorem 3.8. Let $h \in J$.

Then

$$h = \lim_{\substack{n\to\infty \\ n\in\Lambda}} h_n, \text{ where } h_n = (e^{iR(g_1)} + \ldots + e^{iR(g_m)})\Psi_n$$

for some $g_i \in C_C(R^2)$, $i = 1, \ldots, m$. Note that $R_n(0,r)H_n(0)h_n = h_n - rR_n(0,r)h_n$. Since $h_n \to h$, it follows that $R_n(0,r)H_n(0)h_n$ converges for $n \in \Lambda$ and

$$\|\lim_{n\to\infty}(H_n(0)+r)^{-1}H_n(0)h_n\| \le 1/r \sup_n \|H_n(0)h_n\|. \tag{3.23}$$

By (Ref. 9, p. 66), (3.23) yields a number $\infty > M > 0$ such that

$$\|\lim R_n(0,r)H_n(0)h_n\| < 1/rM. \tag{3.24}$$

Next observe that $\|R_n(0,r)h - R_n(0,r)h_n\| \le \|h - h_n\| \to 0$ so

$$\lim_{\substack{n\to\infty \\ n\in\Lambda}} R_n(0,r)h = \lim_{\substack{n\to\infty \\ n\in\Lambda}} R_n(0,r)h.$$

Consequently,

$$\begin{aligned} r\cdot\lim_{n\to\infty} R_n(0,r)h &= \lim_{n\to\infty} h_n - \lim_{n\to\infty} R_n(0,r)H_n(0)h_n \\ &= h - \lim_{n\to\infty} R_n(0,r)H_n(0)h_n \end{aligned}$$

and by (3.24)

$$\lim_{r\to\infty} r\,R_\infty(0,r)h = h.$$

Notice that from the resolvent equation, $\|R_n(p.r) - R_n(q,s)\| \le |(p-q)/(r\cdot s)|$. Taking the limit $n \to \infty, n \in \Lambda$ gives

$$\|R_\infty(p,r) - R_\infty(q,s)\| \le |(q-s)/(r\cdot s)|. \tag{3.25}$$

The observation $\|r(R_\infty(p,r) - R_\infty(0,r)\| < |p/r|$ finishes the proof.

Corollary 3.10: The null space of $R_\infty(p,r)$ is $\{0\}$ for all r not in $(-\infty, 0]$ and p in R^2.

Proof: The null space of $R(p,r)$ is independent of r and $R(p,r)$ satisfies the resolvent equation (Ref. 14, p. 428),

$$R_\infty(p,r) - R_\infty(p,s) = (s-r)R_\infty(p,r)R_\infty(p,s). \tag{3.26}$$

If $R_\infty(p,r)h = 0$ for some h then $\lim_{\substack{r\to\infty \\ r\ real}} r\,R_\infty(p,r)h = 0$. By the previous lemma $h = 0$.

Theorem 3.11: For each p in R^2, there is a self-adjoint operator $H_\infty(p)$ such that $R(p,r) = (H_\infty(p) + r)^{-1}$ for all r not in $(-\infty, 0]$. $H(p)$ is nonnegative. 0 is an eigenvalue of $H(0)$ with Ψ_∞ as an eigenvector. $D(H_\infty(p)) = D(H_\infty(g))$ and $[H_\infty(p) - H_\infty(q)]^{**} = [E(p-P) - E(q-P)]^{**}$ is bounded.

Proof: By Theorem 3.8 and Corollary 3.10, there is a closed operator $H_\infty(p)$ such that $(H_\infty(p) + r)^{-1} = R_\infty(p,r)$ for all r not in $(-\infty, 0]$, (Ref. 14, p. 428). For each $r > 0$, $H_\infty(p) + r$ is symmetric since $R_\infty(p,r)$ is self-adjoint. Thus, $H_\infty(p)$ is symmetric and since the spectrum of $H_\infty(p)$ lies in $[0, \infty)$ we may conclude that $H_\infty(p)$ is both nonnegative and self-adjoint (Ref. 14, p. 271).

$$(H_\infty(0)+r)^{-1}\Psi_\infty = \lim_{\substack{n\to\infty \\ n\in\Lambda}} R_n(0,r)\Psi_n = \lim_{\substack{n\to\infty \\ n\in\Lambda}} \Psi_n/r = \Psi_\infty/r.$$

Thus Ψ_∞ is in the range of $(H_\infty(0) + r)^{-1}$ and so in the domain of $H_\infty(0)$.

Furthermore $(H_\infty(0) + r)\Psi_\infty = r\Psi_\infty$ so $H_\infty(0)\Psi_\infty = 0$.

$$\begin{aligned} D(H_\infty(p)) &= \text{range } ([H_\infty(p)+r]^{-1}) \\ &= \{\lim_{n\to\infty}(H_n(p)+r)^{-1}h:\ h \in \mathfrak{F}\} \\ &= \{\lim_{n\to\infty}(H_n(q)+r)^{-1}h + R_n(q,r)(H_n(p) \\ &\quad - H_n(q))R_n(p,r)h : h \in \mathfrak{F}\} \\ &\subset \text{Range } (H_\infty(q)+r)^{-1} = D(H_\infty(q)). \end{aligned}$$

By symmetry, therefore, $D(H(p)) = D(H(q))$. We already know that $\|[E(p-P) - E(q-P)]^{**}\| \le |p-q|$ so all that remains to be verified is that

$$[H_\infty(p) - H_\infty(q)]^{**} = [E(p-P) - E(q-P)]^{**}. \tag{3.27}$$

To verify this we again use the resolvent equation

$$R_n(p,r) - R_n(q,r) = R_n(p,r)(E(q-P) - E(p-P))R_n(q,r).$$

Since $E(q-P) - E(p-P)$ is defined and bounded on the range of $R_n(q,r)$ for all n and since the $\{R_n(p,r)\}$ are uniformly bounded we may take limits as $n \to \infty$ in Λ and write

$$\begin{aligned} &R_\infty(p,r) - R_\infty(q,r) \\ &\quad = R_\infty(p,r)[E(q-P) - E(p-P)]^{**}R_\infty(q,r). \end{aligned} \tag{3.28}$$

Apply $H_\infty(p) + r$ to the left and $H_\infty(q) + r$ to the right giving

$$H_\infty(p) - H_\infty(q) = [E(q - P) - E(p - P)]^{**} \text{ on } D(H_\infty(q)).$$

Corollary 3.12:

(i) $\underset{\substack{n\to\infty \\ n\in\Lambda}}{\text{strong limit}} (H_n(p) + r)^{-1} = (H_\infty(p) + r)^{-1}$ for all complex r not in $(-\infty, 0]$;

(ii) $\underset{\substack{n\to\infty \\ n\in\Lambda}}{\text{strong limit}} e^{-tH_n(p)} = e^{-tH_\infty(p)}$ for all $t \geq 0$;

(iii) $\underset{\substack{n\to\infty \\ n\in\Lambda}}{\text{strong limit}} e^{itH_n(p)} = e^{itH_\infty(p)}$ for all real t.

Proof: (i) restates the previous theorem. (ii) and (iii) follow from (i) and operator theory, (Ref. 14, p. 502).

Definition: $H_\infty(p)$ is the polaron Hamiltonian without cutoffs and total momentum p.

B. The polaron with cutoffs

We have already defined the Hamiltonian for the polaron of total momentum p, $H_f(p)$. In this section we first define the Hamiltonian for the polaron with cutoffs as a direct integral of the $H_f(p)$ over p in R^2 and then state the connection of this description with the more familiar (1.6).

Lemma 3.12: Let f either be a cutoff function or $^\infty$ and put $R(p,r) = (H_f(p) + r)^{-1}$, where r is any complex number not in $(-\infty, 0]$ and p is in R^2. If h is in $L^2(R^2, \mathfrak{F})$, then the $\mathfrak{F}$-valued function on R^2, $R_f(r)h$, defined by

$$(R_f(r)h)(p) = R(p,r)h(p) \tag{3.29}$$

is also in $L^2(R^2, \mathfrak{F})$. The operator $R_f(r)$ which takes h into $R(r)h$ is bounded with norm $\leq \alpha(r)$, where $\alpha(r) = 1/r$ if $r > 0$ and $\alpha(r) = |\text{Im}r|$ if r is not real. If $r > 0$, then $R(r)$ is self-adjoint.

Proof: The lemma follows from the fact that the function $R(r)h$ is measurable, the self-adjointness of the $R(p,r)$'s for $r > 0$ and the norm estimate

$$\|R(p,r)\| \leq \alpha(r).$$

To prove that $R(r)h$ is measurable it suffices to prove that $p \to R(p,r)$ is continuous from R^2 into the bounded linear operators on $\mathfrak{F}$, with the norm topology.

This follows from:

$$\|R(p,r) - R(q,s)\| \leq |p - q| \cdot \alpha(r) \cdot \alpha(s).$$

Lemma 3.13: Let f be a cutoff function. Let $R(p,r) = (H_f(p) + r)^{-1}$ and define $R_f(r)$ as in the previous lemma. Then $R_f(r)$ is the resolvent of a nonnegative, self-adjoint operator on $L^2(R^2, \mathfrak{F})$. We denote this operator by H_f.

Proof: Since $R(p,r)$ satisfies a resolvent equation for each p, so does $R_f(r)$:

$$R_f(r) - R_f(s) = (r - s)R_f(r)R_f(s). \tag{3.30}$$

Consequently, in order to prove that $R_f(r)$ is the resolvent it will be enough to prove that the null space of $R_f(r)$ is zero for some r, (Ref. 14, p. 428). But $R_f(r)h = 0$ in $L^2(R^2, \mathfrak{F})$ iff $(R_f(r)h)(p) = R_f(p,r)h(p) = 0$ for a.e. p iff $h(p) = 0$ for a.e. p in R^2 since we know that the null space of each $R_f(p,r)$ is zero. Thus, $R_f(r)$ is the resolvent of a closed operator for which we write H_f:

$$(H_f + r)^{-1} = R_f(r) \quad \text{for all complex } r \text{ not in } (-\infty, 0]. \tag{3.31}$$

By definition, spectrum $(H_f) \subset [0, \infty)$, and so H_f is nonnegative. $R_f(s)$ is self-adjoint for $r > 0$ so H_f is symmetric. Since spectrum $(H_f) \subset [0, \infty)$, H_f is self-adjoint, (Ref. 14, p. 271).

Definition: H_f is the Hamiltonian for the polaron with cutoff f in the total momentum description.

Lemma 3.14: Let $h \in D(H_f)$. Then for almost every p in R^2, $h(p)$ is in $D(H_f(p))$ and

$$(H_f h)(p) = H_f(p)h(p).$$

Proof: h is in the range of $R_f(r)$ so that $h(p) = (R_f(r)g)(p) = R_f(p,r)g(p)$ for some g in $L^2(R^2, \mathfrak{F})$. Thus $h(p) \in D(H_f(p))$.

Furthermore $g(p) = (H_f(p) + r)h(p)$ for almost every p and $g = (H_f + r)h$.

Remarks: There is another description of the cutoff polaron which we give here. For details see Refs. 4, 12, or 15. Let H_1 denote multiplication on $L^2(R^2)$ by E_{m_0}.

Let S_k denote translation on $L^2(R^2)$ by $k \in R^2$; i.e., $(S_k h)(x) = h(x - k)$. Let V be the closed linear extension of the operator defined on the subset of $L^2(R^2) \otimes \mathfrak{F}$ consisting of finite linear combinations of elements of the form $g \otimes h$ where g is in $L^2(R^2)$ and h is in $\mathcal{E}$ by

$$V(g \otimes h) = \int_{R^2} ((S_k g) \otimes a(k)h) f(k) dk.$$

Let

$$T = H_1 \otimes I + I \otimes K + (V + V^*)^{**} - \lambda(f, m_0, 0).$$

Then T is self-adjoint on $L^2(R^2) \otimes \mathfrak{F}$ and there is a unitary operator W: $L^2(R^2) \otimes \mathfrak{F} \to L^2(R^2, \mathfrak{F})$ such that $D(T) = W^{-1}D(H_f)$ and $(WTW^{-1}h)(p) = H_f(p)h(p)$ for almost every p in R^2 and h in $W^{-1}D(H_f)$. Thus $WTW^{-1} = H_f$.

C. The physical polaron

The Hamiltonian for the physical polaron in two space dimensions can be defined in two ways: either as the direct integral of the $H_\infty(p)$ over p in R^2 or as the generalized strong limit of the H_{f_n} for $n \in \Lambda$. It will be shown that both definitions agree.

Theorem 3.15: There is a self-adjoint operator H on $L^2(R^2, \mathfrak{F})$ such that for almost every p in R^2 and h in $D(H_\infty)$, $h(p) \in D(H_\infty(p))$ and

$$(H_\infty h)(p) = H_\infty(p)h(p). \tag{3.32}$$

Proof: Let $R_\infty(r)$ be as in Lemma 3.12. The proof of Lemma 3.13, modified by replacing "f" with "∞" wherever the former occurs, demonstrates the existence of a nonnegative, self-adjoint operator, H_∞ such that $(H_\infty + r)^{-1} = R_\infty(r)$. If we modify the proof of Lemma 3.14 similarly then it is the conclusion of this proof.

Corollary 3.16: Spectrum $(H_\infty) \subset [0, \infty)$.

Proof: Given above.

Let $\{f_n\}$ be the realistic cutoff functions with coupling constant λ. Put $H_n = H_{f_n}$.

Corollary 3.17:

(i) $\underset{\substack{n\to\infty \\ n\in\Lambda}}{\text{strong limit}} (H_n + r)^{-1} = (H_\infty + r)^{-1}$ for all complex r not in $(-\infty, 0]$;

(ii) $\underset{\substack{n\to\infty \\ n\in\Lambda}}{\text{strong limit}} e^{-tH_n} = e^{-tH_\infty}$ for $t > 0$;

(iii) strong limit $e^{itH_n} = e^{itH_\infty}$ for all real t.

Proof: (ii) and (iii) will follow, once we establish (i) (Ref. 14, p. 502). We now prove (i). Since the $\{(H_n + r)^{-1}\}$ have uniformly bounded norms, in order to verify (i) it will suffice to verify convergence on a subset Z dense in $L^2(R^2, \mathfrak{F})$.

Let $x_n n = 1, 2, \ldots, s$ be characteristic functions of bounded measurable subsets E_n with finite measure and let $h_n \in \mathfrak{F}$. Define $g \in L^2(R^2, \mathfrak{F})$ by

$$g(p) = \sum_{n=1}^{s} x_n(p)h_n \tag{3.33}$$

Let Z be the set of all such g's as the E_n's, h_n's, and s's vary. Then Z is dense and we will verify convergence on Z. For g as in (3.33) we may estimate

$$\|(R_n(r) - R_\infty(r))g\|^2 \leq \sum_{j=1}^{s} \int_{E_i} \|(R_\infty(p,r) - R_n(p,r))h_j\|^2 \cdot dp. \tag{3.34}$$

By Theorem 3.8, $\|R_n(p,r)h_j - R_\infty(p,r)h_j\|^2 \to 0$ as $n \to \infty$ for each fixed p. But if $\alpha(r)$ is as in Lemma 3.12 then $\|R_n(p,r)h_j - R(p,r)h_j\|^2 \leq 4\alpha(r)\|h_j\|^2$. Since the measure of E_i is finite, we may conclude by dominated convergence that $R_n(r)g \to R_\infty(r)g$ as $n \to \infty$ in Λ.

Corollary 3.18:

$$D(H_\infty) = \{h \in L^2(R^2, \mathfrak{F}): p \to H_\infty(p)h(p) \in L^2(R^2, \mathfrak{F})\}.$$

Proof: By Theorem 3.15 we need only show $\{\} \subset D(H_\infty)$. Choose h such that $h(p) \in D(H_\infty(p)$ for a.e. p and such that the function $p \to H_\infty(p)h(p)$ is in $L^2(R^2, \mathfrak{F})$. Then, for a.e. p in R^2, $h(p) = R_\infty(p,r)g(p)$ for some $g(p) \in \mathfrak{F}$. Since $H_\infty(\cdot)h(\cdot)$ is in $L^2(R^2, \mathfrak{F})$ so is the function $g(\cdot)$ and in fact $h = (H_\infty + r)^{-1}g$ so that $h \in D(H_\infty)$.

Corollary 3.19: Zero is in the approximate point spectrum of H_∞.

Proof: Let $g \in C_C^\infty(R^2)$ have support of diameter less than ϵ and have L^2 norm one. Define a function h in $L^2(R^2, \mathfrak{F})$ by

$$h(p) = g(p)\Psi_\infty.$$

Clearly, $h(p) \in D(H_\infty)$. Furthermore, by Theorem 3.15,

$$\begin{aligned}
\|H_\infty(p)h(p) - H_\infty(q)h(q)\| &\leq \|H_\infty(p)g(p)\Psi_\infty - H_\infty(p)g(q)\Psi_\infty\| \\
&\quad + \|H_\infty(p)g(q)\Psi_\infty - H(q)g(q)\Psi_\infty\| \\
&\leq |g(p) - g(q)|\,\|H_\infty(p)\Psi_\infty\| \\
&\quad + \|g\|_\infty\|H_\infty(p)\Psi_\infty - H_\infty(q)\Psi_\infty\| \\
&\leq |g(p) - g(q)|\,\|H_\infty(p)\Psi_\infty\| + \|g_\infty\|\,|p - q|. \\
&\leq |g(p) - g(q)|\,\|H_\infty(0)\Psi_\infty\| + |g(p) - g(q)| \\
&\quad \|(H_\infty(p) - H_\infty(0))\Psi_\infty\| + \|g_\infty\|\,|p - q| \\
&\leq |g(p) - g(p)|\,|p| + \|g\|_\infty|p - q|.
\end{aligned}$$

Thus $p \to H_\infty(p)h(p)$ is in $C_C(R^2, \mathfrak{F})$ and so $p \to H_\infty(p)h(p)$ is in $L^2(R^2, \mathfrak{F})$. Thus, by the previous corollary, $h \in D(H_\infty)$.

Furthermore, since $\|g\| = 1$ we have

$$\begin{aligned}
\|H_\infty h\|^2 &\leq \int \|H_\infty(p)h(p)\|^2 dp \\
&= \int \|(H_\infty(p) - H_\infty(0))\Psi_\infty\|^2 |g(p)|^2 dp \\
&\leq \int_{\text{supp}(g)} |p|^2 |g(p)|^2 dp \leq \epsilon^2\|g\|^2 = \epsilon^2.
\end{aligned}$$

Since ϵ is arbitrary we have shown that 0 is in the approximate point spectrum of H_∞.

Corollary 3.20: For every h in $D(H_\infty)$ there is sequence $\{h_n\}$ with $h_n \in D(H_n)$ and $h_n \to h$ while $H_n h_n \to H_\infty h$ as $n \to \infty$ in Λ.

Proof: Generalized strong convergence of the self-adjoint operators H_n (i.e., strong convergence of their resolvents) implies graph convergence of the H_n, (Ref. 17, p. 404).

ACKNOWLEDGMENTS

This paper presents the main results contained in the author's doctoral dissertation written at Cornell University during 1971 under Leonard Gross. The author wishes to thank Leonard Gross not only for providing the subject of the dissertation, and for valuable help in applying many of the methods suggested, but also for further suggestions and an early copy of Ref. 9 which considerably shortened the dissertation into the form presented here. The author would also like to thank J. Frohlich for pointing out the necessity of a self-energy renormalization rather than the mass renormalization used in the author's thesis. (See the remark at the end of Sec. 2B.)

[1]E. Nelson, J. Math. Phys. 5, 1190 (1964).
[2]J. P. Eckmann, Commun. Math. Phys. 18 247 (1970).
[3]J. Frohlich (private communication).
[4]J. T. Cannon, J. Funct. Anal. 8, 101 (1971).
[5]J. Frohlich, "Mathematical Discussion of Models with Persistant Vacuum", E.T.H. preprint (1971).
[6]Albeverio "Strong Asymptotic Convergence of Fields and States for Physical Particles in Models of Q.F.T.", Princeton preprint (1971).
[7]S. Albeverio, Helv. Phys. Acta 45, 303 (1972).
[8]J. Frohlich, "On the Infrared Problem in a Model of Scalar Electrons and Massless Scalar Bosons I, II", E T.H. (1972).
[9]L. Gross, "The Relativistic Polaron without Cutoffs", preprint, Cornell University (1972).
[10]A. D Sloan, Bull. Am. Phys. Soc. 78, 407 (1972)
[11]J. M. Cook, Trans. Amer. Math. Soc., 74, 222 (1953)
[12]L. Gross, J. Funct Anal. 10, 52 (1972).
[13]E. Nelson, Ann. Math. 70, 572 (1959).
[14]T. Kato, *Perturbation of Linear Operators* (Springer, New York, 1966).
[15]A. D. Sloan, "The Relativistic Polaron without Cutoffs in Two Space Dimensions", Cornell University thesis, (Ithaca, 1971).
[16]J. Frohlich, "A Note on a Paper of L. Gross", E.T.H. Preprint (1972).
[17]J. Glimm and A. Jaffe, Commun. Pure Appl. Math. 22, 401 (1969),

Depolarization in nonuniform multilayered structures—full wave solutions*

E. Bahar

Electrical Engineering Department, University of Nebraska, Lincoln, Nebraska 68508
(Received 26 July 1973)

In view of the recent impetus to produce rigorous solutions to more realistic models of pertinent propagation problems over a wide range of frequencies, we present in this paper full wave solutions to the problem of radio wave propagation in nonuniform multilayered structures. The electromagnetic properties of the media, the geometry of the irregular structure, and the electromagnetic source distributions are assumed to be arbitrary three-dimensional functions of position. Generalized field transforms are employed to provide a basis for the expansion of the transverse electromagnetic fields and Maxwell's equations are reduced to a set of first-order coupled differential equations for the forward and backward, vertically and horizontally polarized wave amplitudes. For open structures the complete wave spectrum includes the radiation term, the lateral waves, and the surface waves or trapped waveguide modes. For structures bounded by impedance walls (or perfect electric or magnetic walls $\mu/\epsilon \to 0$ and $\epsilon/\mu \to 0$, respectively) the fields are expressed exclusively in terms of waveguide modes. Exact boundary conditions are imposed at all the interfaces of the structure and the general solutions are not limited by the (approximate) surface impedance concept. The full wave approach employed is not restricted by frequency considerations. It is applicable to very broad classes of problems in which no single constituent of the total formal solution dominates. The full wave solutions may be applied to problems such as (i) propagation of ground waves over irregular and inhomogeneous terrain, (ii) scattering by rough surfaces and objects of finite dimensions, and (iii) propagation of guided waves in nonuniform artificial waveguides as well as in irregular ducts in the earth's crust or in the ionosphere.

1. INTRODUCTION

In recent years there has been a considerable growth of civil and military interest in the development of more reliable communication and detection systems. The potential for developing radio wave methods for remote sensing (above and below the earth's crust) and the need to develop hardened communication systems have contributed much to this renewed interest. This has been paralleled by the remarkable advances that have been made in the availability of high powered, very low frequency electromagnetic sources that are capable of radiating deeper into the earth's crust as well as commercially available transmitters operating at optical frequencies. These developments along with the ready access to large, versatile, digital computers have provided considerable impetus to produce rigorous solutions to more realistic models of pertinent propagation problems over a wide range of frequencies.

To this end full wave solutions to the problem of radio wave propagation in nonuniform multilayered structures have been derived recently.[1] However in this recent work it has been assumed at the outset that the electromagnetic and geometric parameters of the layered structure are independent of one Cartesian coordinate variable (see Fig. 1). In order to apply our present analysis to a wider class of more realistic problems this restrictive assumption is not made here. Thus, in this work, the medium of the ith layer of the structure is characterized by the electromagnetic parameters $\epsilon_i(x, z)$ and $\mu_i(x, z)$ which may in general be complex to account for medium losses. For instance, if ϵ_{ir} and σ_i are the dielectric coefficient and the conductivity of the ith medium,

$$\epsilon_i(x, z) = \epsilon_{ir}(x, z) - i\sigma_i(x, z)/\omega . \tag{1.1}$$

The interface between medium i and $i + 1$ is described by the surface $y = h_{i,i+1}(x, z)$ and the thickness of the ith layer is

$$H_i(x) = h_{i-1,i}(x, z) - h_{i,i+1}(x, z), \qquad i = 1, 2, \ldots, m + 1. \tag{1.2}$$

For convenience we shall assume that the nonuniform structure is excited by arbitrary three-dimensional distributions of both electric and magnetic sources $\bar{J}$ and $\bar{M}$, respectively.

For the purposes of the analysis, generalized field transforms similar to those derived earlier[2] are employed. The generalized field transforms are used to reduce Maxwell's equations into a set of first-order coupled differential equations for the vertically and horizontally polarized, forward and backward wave amplitudes.

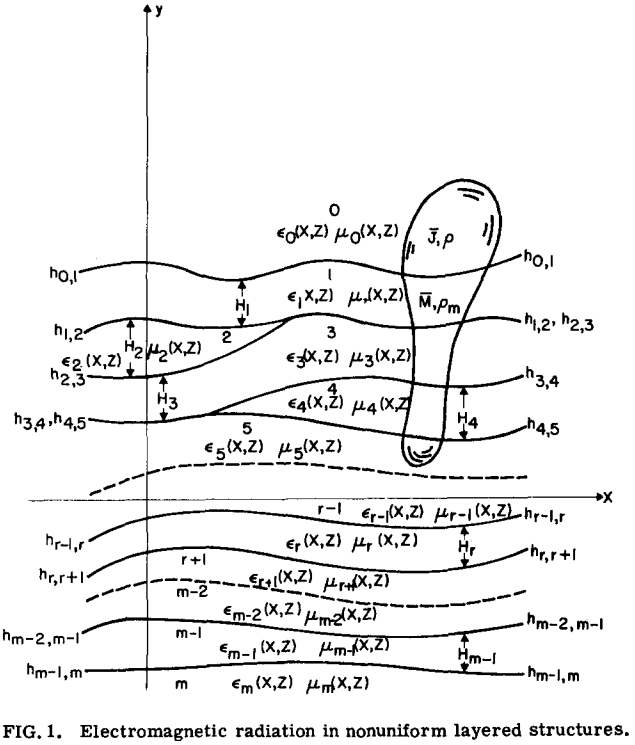


FIG. 1. Electromagnetic radiation in nonuniform layered structures.

Since no exact analytical expressions for the characteristic functions are known when ϵ, μ, and $h_{i+1,i}$ are arbitrary functions of both x and z, the basis functions used in this work, in general, do not have the desired property of being orthogonal. Nevertheless, they exhibit quasi-orthogonal properties that yield tractable solutions to the general problem.

Since exact boundary conditions are imposed at all the interfaces of the structure, the solutions are not limited by the applicability of the surface impedance concept.

2. FORMULATION OF THE PROBLEM

For the general class of problems considered in this paper there are no axes of symmetry. A right-hand Cartesian coordinate system is used, with the y axis normal to the reference plane from which the heights of the layers' interfaces are measured (see Fig. 1). For special cases in which the variations in the electromagnetic and geometrical parameters of the structure depend strongly upon direction, it is judicious to orient the one axis (for instance the z axis) in the direction of least change in these parameters. Thus when considering the problem of propagation across a coast line we orient the z axis parallel to the coast line.

Maxwell's equations for the transverse components of the electric and magnetic fields $\bar{E}_T$ and $\bar{H}_T$, respectively, are

$$-\frac{\partial \bar{E}_T}{\partial x} = i\omega\mu(\bar{H}_T \times \bar{a}_x) - \frac{1}{i\omega}\nabla_T\left(\frac{1}{\epsilon}\nabla_T \cdot (\bar{H}_T \times \bar{a}_x)\right) + \bar{M}_T \times \bar{a}_x + \frac{1}{i\omega}\nabla_T \frac{J_x}{\epsilon} \quad (2.1a)$$

and

$$-\frac{\partial \bar{H}_T}{\partial x} = i\omega\epsilon(\bar{a}_x \times \bar{E}_T) - \frac{1}{i\omega}\nabla_T\left(\frac{1}{\mu}\nabla_T \cdot (\bar{a}_x \times \bar{E}_T)\right) + \bar{a}_x \times \bar{J}_T + \frac{1}{i\omega}\nabla_T \frac{M_x}{\mu} \quad (2.1b)$$

in which the operator ∇_T is

$$\nabla_T = \bar{a}_y \frac{\partial}{\partial y} + \bar{a}_z \frac{\partial}{\partial z} \quad (2.1c)$$

and the transverse vectors are

$$\bar{A}_T = \bar{a}_y A_y + \bar{a}_z A_z. \quad (2.1d)$$

At all the m interfaces of the nonuniform layered structure (Fig. 1) the tangential components of the electric and magnetic fields are continuous. The corresponding boundary conditions for the electric and magnetic fields are expressed in terms of the unit vectors $\bar{n}_{i,i+1}, \bar{m}_{i,i+1}$, and $\bar{l}_{i,i+1}$ at the interface between medium i and medium $i+1$ (see Fig. 2). The normal to the surface $h_{i,i+1}$ is

$$\bar{n}_{i,i+1} = \left(-\frac{\partial h_{i,i+1}}{\partial x}, 1, -\frac{\partial h_{i,i+1}}{\partial z}\right) \Bigg/ \left[\left(\frac{\partial h_{i,i+1}}{\partial x}\right)^2 + 1 + \left(\frac{\partial h_{i,i+1}}{\partial z}\right)^2\right]^{1/2} \quad (2.2a)$$

The vector $\bar{m}_{i,i+1}$ is in the direction of the projection of $\bar{n}_{i,i+1}$ in the transverse y, z plane, thus

$$\bar{m}_{i,i+1} = \left(0, 1, -\frac{\partial h_{i,i+1}}{\partial z}\right) \Bigg/ \left[1 + \left(\frac{\partial h_{i,i+1}}{\partial z}\right)^2\right]^{1/2}. \quad (2.2b)$$

The unit vector $\bar{l}_{i,i+1}$ is given by

$$\bar{l}_{i,i+1} = \bar{a}_x \times \bar{m}_{i,i+1}. \quad (2.2c)$$

Thus, the tangential electric fields at the interfaces of the structure are continuous provided that

$$[\bar{n} \times \bar{E}]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \quad i = 1, \ldots m. \quad (2.3a)$$

From (2.3a) we obtain the following boundary conditions for the transverse electromagnetic fields:

$$[\bar{l} \cdot \bar{E}]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \quad \text{thus} \quad \left[E_y \frac{\partial H_{i-1,i}}{\partial z} + E_z\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0 \quad (2.3b)$$

and

$$[\bar{l} \cdot (\bar{n} \times \bar{E})]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \quad \text{thus} \quad \left[E_x + \bar{m} \cdot \bar{E}_T \frac{\partial h_{i-1,i}}{\partial x}\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \quad (2.3c)$$

in which

$$E_x = (1/i\omega\epsilon)[\nabla_T \cdot (\bar{H}_T \times \bar{a}_x) - J_x].$$

Similarly, since the tangential magnetic fields are continuous, we obtain the following boundary conditions:

$$[\bar{l} \cdot \bar{H}]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \quad \text{thus} \quad \left[H_y \frac{\partial h_{i-1,i}}{\partial z} + H_z\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0 \quad (2.4a)$$

and

$$[\bar{l} \cdot (\bar{n} \times \bar{H})]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \quad \text{thus} \quad \left[H_x + \bar{m} \cdot \bar{H}_T \frac{\partial h_{i-1,i}}{\partial x}\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0 \quad (2.4b)$$

in which

$$H_x = (1/i\omega\mu)[\nabla_T \cdot (\bar{a}_x \times \bar{E}_T) - M_x]. \quad (2.4c)$$

Arbitrary three dimensional source distributions may be regarded as superpositions of electric and magnetic point sources

$$\bar{J} = I\delta(\bar{r} - \bar{r}_0) \quad (2.5a)$$

and

$$\bar{M} = K\delta(\bar{r} - \bar{r}_0), \quad (2.5b)$$

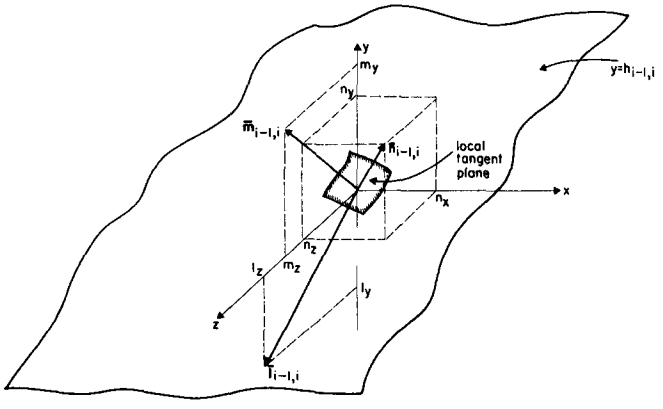


FIG. 2. Direction of the unit vectors $\bar{l}_{i,i+1}, \bar{m}_{i,i+1}, \bar{n}_{i,i+1}$ and $\bar{n}_{i,i+1}$ with respect to the local tangent plane.

in which I and K are measured in amperes and volts, respectively, and $\delta(\bar{r}-\bar{r}_0)$ is the three-dimensional Dirac delta function

$$\delta(\bar{r}-\bar{r}_0)=\delta(x-x_0)\delta(y-y_0)\delta(z-z_0). \tag{2.5c}$$

For the purpose of our analysis we employ the following completeness and orthogonal relationships derived from the familiar Fourier transform:

$$\delta(z-z_0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-iw(z-z_0)}dw \tag{2.6a}$$

and

$$\delta(w-w')=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{iz(w-w')}dz \tag{2.6b}$$

in which $\delta(\alpha-\beta)$ is the one-dimensional Dirac delta function. Using generalized Fourier transforms derived for an $m+1$ layered structure we obtain the following completeness and orthogonal relationships[3]:

$$\delta(y-y_0)=\sum_v Z^P N^P\psi^P(v,y)\psi^P(v,y_0)$$
$$\equiv\int_{-\infty}^{\infty}Z^P N_0^P\psi^P(v,y)\psi_0^P(v,y_0)dv_0$$
$$+\int_{-\infty}^{\infty}Z^P N_m^P\psi_m^P(v,y)\psi_m^P(v,y_0)dv_m$$
$$+\sum_{n=1}^{N}Z^{Pn}\psi_s^{Pn}(v,y)\psi_s^{Pn}(v,y_0) \tag{2.7a}$$

and

$$\int_{-\infty}^{\infty}Z^P N_q^P\psi_q^P(v,y)\psi_r^P(v',y)dy=\Delta(v,v')$$
$$=\delta_{q,r'}\begin{cases}\delta(v,v'), & v'\neq v_s,\\ \delta_{v,v_s}, & v'=v_s,\end{cases} \tag{2.7b}$$

where the superscript P equals V or H and the subscripts q and r are equal to o, m, or s and $\delta_{q,r}$ is the Kronecker delta. The basis functions are

$$R_{P0}^{Dh}\psi_0^P(v,y)=\begin{cases}\exp(iv_0y)+R_{P0}^{Dh}\exp(-iv_0y), & \text{for medium 0}\\ \prod_{q=1}^{r}(T_{Pq-1}^D/T_{Pq}^D)\exp\left(i\sum_{q=1}^{r}v_{q-1,q}h_{q-1,q}\right)\\ \times[\exp(iv_ry)+R_{Pr}^{Dh}\exp(-iv_ry)], & \text{for medium } r=1,2,3,\ldots,m,\end{cases} \tag{2.8a}$$

$$R_{Pm}^{Uh}\psi_m^P(v,y)=\begin{cases}\prod_{q=1}^{m-r}(T_{Pm+1-q}^U/T_{Pm-q}^{UH})\exp\left(i\sum_{q=1}^{m-r}v_{m-q,m+1-q}h_{m-q,m+1-q}\right)\\ \times[\exp(-iv_ry)+R_{Pr}^{Uh}\exp(iv_ry)], & \text{for medium } r=0,1,2,\ldots,m-1,\\ \exp(-iv_my)+R_{Pm}^{Uh}\exp(iv_my), & \text{for medium } m,\end{cases} \tag{2.8b}$$

and

$$\psi_s^{Pn}(v,y)=\psi_s^{Pn}(v,h_{0,1})\begin{cases}\exp[-iv_0^n(y-h_{0,1})], & \text{for medium 0,}\\ \frac{1}{T_{P1}^{DH}}\exp(-iv_1^nh_{0,1})[\exp(iv_1^ny)+R_{P1}^{Dh}\exp(-iv_1^ny)], & \text{for medium 1,}\\ \frac{1}{T_{P1}^{DH}}\exp(-iv_1^nh_{0,1})\prod_{q=2}^{r}(T_{Pq-1}^D/T_{Pq}^{DH}) & \\ \exp\left\{i\sum_{q=2}^{r}v_{q-1,q}^nh_{q-1,q}\right\}[\exp(iv_r^ny)+P_{Pr}^{Dh}\exp(-iv_r^ny)], & \text{for medium } r=2,3,\ldots,m,\end{cases} \tag{2.8c}$$

where

$$[\psi_s^{Pn}(v,h_{0,1})]^2=\left[\frac{u}{iZ_0^Pv_0}\frac{d}{du}\frac{1}{R_{P0}^U}\right]_{v=v_n}, \tag{2.8d}$$

R_{Pi}^D is the reflection coefficient at the $i, i+1$ interface for waves incident from above, and R_{Pi}^U is the reflection coefficient at the $i-1, i$ interface for waves incident from below (see Fig. 1). Thus for $P=V$ or H

$$R_{Pm}^D=0,\quad R_{Pi}^D=(R_{i+1,i}^P+R_{Pi+1}^{DH})/(1+R_{i+1,i}^PR_{Pi+1}^{DH}),\quad i=0,1,\ldots,m-1 \tag{2.9a}$$

and

$$R_{P0}^U=0,\quad R_{Pi}^U=(R_{i-1,i}^P+R_{Pi-1}^{UH})/(1+R_{i-1,i}^PR_{Pi-1}^{UH}),\quad i=1,2,\ldots,m,$$

in which $R_{i,i\pm1}^V$ and $R_{i,i\pm1}^H$ are the two media Fresnel reflection coefficients for vertically and horizontally polarized waves, respectively.

$$R_{i+1,i}^V=-R_{i,i+1}^V=(v_i\epsilon_{i+1}-v_{i+1}\epsilon_i)/(v_i\epsilon_{i+1}+v_{i+1}\epsilon_i),$$
$$R_{i+1,i}^H=-R_{i,i+1}^H=(v_i\mu_{i+1}-v_{i+1}\mu_i)/(v_i\mu_{i+1}+v_{i+1}\mu_i) \tag{2.9b}$$

and

$$R_{Pi}^{DH}=R_{Pi}^D\exp(-i2v_iH_i),\quad R_{Pi}^{Dh}=R_{Pi}^D\exp(i2v_ih_{i,i+1}),$$
$$R_{Pi}^{UH}=R_{Pi}^U\exp(-i2v_iH_i),\quad R_{Pi}^{Uh}=R_{Pi}^V\exp(-i2v_ih_{i-1,i}). \tag{2.9c}$$

The transmission coefficients are

$$T_{Pi}^D=1+R_{Pi}^D,\quad T_{Pi}^U=1+R_{Pi}^U,\ T_{Pi}^{DH}=1+R_{Pi}^{DH},$$
$$T_{Pi}^{UH}=1+R_{Pi}^{UH} \tag{2.9d}$$

and the normalization coefficients are

$$N_q^P=\begin{cases}R_{P0}^{Dh}/2Z_0^P, & q=0,\\ R_{Pm}^{UH}/2\,2_m^P, & q=m,\\ 1, & q=s.\end{cases} \tag{2.9e}$$

The two infinite integrals in the completeness relationships (2.7a) are associated with the radiation and the lateral wave terms (the continuous parts of the wavenumber spectrum), while the finite sum in (2.7a) is associated with the surface waves or trapped waveguide modes (discrete part of the wavenumber spectrum). The

relationship between the generalized transforms and the familiar Fourier transforms has already been established.[3] The infinite integrals are associated with branch cuts at $\mathrm{Im}(v_0) = 0$ and $\mathrm{Im}(v_m) = 0$ and the surface wave terms are associated with the residues of the poles at $1/R^D_{P0} = 0$ (or $1/R^U_{Pm} = 0$). The modal equation that determines the surface wave parameters v_s^n for vertically and horizontally polarized waves is given by

$$1 - R^U_{Pi} R^D_{Pi} \exp(-iv_i H_i) = 0, \quad \mathrm{Im}(v) \le 0 \tag{2.10a}$$

for $i = 1, 2, 3, \cdots$ *or* $m - 1$ and $P = V$ and H for the vertically and horizontally polarized waves, respectively. The parameter $v_{i-1,i}$ is defined as

$$v_{i-1,i} = v_{i-1} - v_i. \tag{2.10b}$$

From (2.1) it follows that for this problem the basis (characteristic) functions $\bar{e}_T$ and $\bar{h}_T$ satisfy the differential equations

$$\left(1 + \frac{\epsilon}{k^2}\nabla_T \frac{1}{\epsilon}\nabla_T \cdot\right)(\bar{h}_T \times \bar{a}_x) = \frac{u}{\omega\mu}\bar{e}_T \tag{2.11a}$$

and

$$\left(1 + \frac{\mu}{k^2}\nabla_T \frac{1}{\mu}\nabla_T \cdot\right)(\bar{a}_x \times \bar{e}_T) = \frac{u}{\omega\epsilon}\bar{h}_T, \tag{2.11b}$$

in which the characteristic value u and the wavenumber k_i for medium i are

$$u = (k_i^2 - v_i^2 - w^2)^{1/2}, \quad \mathrm{Im}(u) \le 0 \tag{2.11c}$$

and

$$k_i = \omega(\mu_i\epsilon_i)^{1/2}, \quad \mathrm{Im}(k_i) \le 0. \tag{2.11d}$$

The boundary conditions for $\bar{e}_T$ and $\bar{h}_T$ are

$$\bar{e}_T \cdot \bar{a}_y \left[\frac{\partial h_{i-1,i}}{\partial z} - \bar{e}_T \cdot \bar{a}_z\right]_{h^-_{i-1,i}}^{h^+_{i-1,i}} = 0, \tag{2.11e}$$

$$\left[\frac{1}{i\omega\epsilon}\nabla_t \cdot (\bar{h}_T \times \bar{a}_x) + \bar{n} \cdot \bar{e}_T \frac{\partial h_{i-1,i}}{\partial x}\right]_{h^-_{i-1,i}}^{h^+_{i-1,i}} = 0, \tag{2.11f}$$

$$\left[\bar{h}_T \cdot \bar{a}_y \frac{\partial h_{i-i,i}}{\partial z} - \bar{h}_T \cdot \bar{a}_z\right]_{h^-_{i-1,i}}^{h^+_{i-1,i}} = 0, \tag{2.11g}$$

and

$$\left[\frac{1}{i\omega\mu}\nabla_t \cdot (\bar{a}_x \times \bar{e}_T) + \bar{n} \cdot \bar{h}_T \frac{\partial h_{i-1,i}}{\partial x}\right]_{h^-_{i-1,i}}^{h^+_{i-1,i}} = 0. \tag{2.11h}$$

For the special case in which the electromagnetic and geometric parameters of the structure are independent of x and z, expressions for $\bar{e}_T$ and $\bar{h}_T$ satisfying (2.11) have been derived and biorthogonal relationships satisfied by these functions have been established.[1] There are two sets of solutions to (2.11), one corresponding to vertically polarized waves ($\bar{e}^V_T$ and $\bar{h}^V_T$), and the second corresponding to horizontally polarized waves ($\bar{e}^H_T$ and $\bar{h}^H_T$).

$$\bar{e}^V_T = Z^V\left(\bar{a}_y \psi^V(v,y) - \frac{\bar{a}_z iw}{u^2 + w^2}\frac{\partial\psi^V(v,y)}{\partial y}\phi(w,z)\right), \tag{2.12a}$$

$$\bar{h}^V_T = \bar{a}_z \psi^V(v,y)\phi(w,z), \tag{2.12b}$$

$$\bar{e}^H_T = \bar{a}_z \psi^H(v,y)\phi(w,z), \tag{2.12c}$$

and

$$\bar{h}^H_T = Y^H\left(-\bar{a}_y\psi^H(v,y) + \frac{\bar{a}_z iw}{u^2 + w^2}\frac{\partial\psi^H(v,y)}{\partial y}\right)\phi(w,z), \tag{2.12d}$$

in which $\phi(w,z) = \exp(-iwz)$, and for medium i the transverse wave impedance for vertically polarized waves is

$$Z^V(v,y) \to Z^V_i = (u^2 + w^2)/u\omega\epsilon_i \tag{2.13a}$$

and the transverse wave admittance for horizontally polarized waves is

$$Y^H = (u^2 + w^2)/u\omega\mu. \tag{2.13b}$$

These characteristic functions (2.12) provide a basis for the full wave expansion of the transverse components of the electromagnetic field. Thus we have

$$\bar{E}_T(x,y,z) = \sum_v \int_{-\infty}^{\infty} [E^V(x,v,w)\bar{e}^V_T + E^H(x,v,w)\bar{e}^H_T]dw \tag{2.14a}$$

and

$$\bar{H}_T(x,y,z) = \sum_v \int_{-\infty}^{\infty} [H^V(x,v,w)\bar{h}^V_T + E^H(x,v,w)\bar{h}^H_T]dw. \tag{2.14b}$$

For the general problem in which the electromagnetic and geometric parameters of the structure are arbitrary functions of x and z, closed form analytic expressions for $\bar{e}_T$ and $\bar{h}_T$ satisfying (2.11) are not known. Hence in this paper we employ basis functions that "locally" satisfy the differential equations (2.11a) and (2.11b) together with the boundary conditions (2.11e) to (2.11h). By this it is meant that while μ_i, ϵ_i, and $h_{i-1,i}$ are assumed to be functions of x and z, expressions involving their derivatives are ignored in order to determine the "local" basis functions. Thus in the expressions for the basis functions $\bar{e}_T$ and $\bar{h}_T$ (2.12), the scalar functions $\psi^P(v,y)(P = V \text{ or } H)$ are not only explicitly functions of y but also implicitly functions of x and z through the electromagnetic and geometric parameters of the structure μ_i, ϵ_i and $h_{i-1,i}$. When the structure's parameters are functions of x but not of z, the basis functions maintain their biorthogonal relationships. Thus in this case it can be shown that for $P = V$ or H[1]

$$E^P(x,v,w) = \int_{-\infty}^{\infty} \bar{E}_T(v,y,z)\cdot(\bar{h}^T_P \times \bar{a}_x)dydz \tag{2.15a}$$

and

$$H^P(x,v,w) = \int_{-\infty}^{\infty} \bar{H}_T(x,y,z)\cdot(\bar{a}_x \times \bar{e}^T_P)dydz, \tag{2.15b}$$

where the complementary (reciprocal) basis functions are

$$\bar{e}^T_V = Z^V N^V\left(\bar{a}_y\psi^V(v,y) + \frac{\bar{a}_z iw}{u^2 + w^2}\frac{\partial\psi^V(v,y)}{\partial y}\right)\phi^c(w,z) \equiv \bar{e}_V\phi^c \tag{2.16a}$$

$$\bar{h}^T_V = \bar{a}_z N^V \psi^V(v,y)\phi^c(w,z) \equiv \bar{h}_V\phi^c \tag{2.16b}$$

$$\bar{e}^T_H = \bar{a}_z N^H \psi^H(v,y)\phi^c(w,z) \equiv \bar{e}_H\phi^c, \tag{2.16c}$$

and

$$\bar{h}^T_H = Y^H N^H\left(-\bar{a}_y\psi^H(v,y) - \frac{\bar{a}_z iw}{u^2 + w^2}\frac{\partial\psi^H(v,y)}{\partial y}\right)\phi^c(w,z) \equiv \bar{h}_H\phi^c, \tag{2.16d}$$

in which $\phi^c(w,z) = (1/2\pi)\exp(iwz)$ and use has been made of the biorthogonal relationships

$$\left.\begin{aligned}&\int_{-\infty}^{\infty} e_T^P \cdot (h_Q^T \times \bar{a}_x)'dy\,dz\\ &\int_{-\infty}^{\infty} h_T^Q \cdot (\bar{a}_x \times \bar{e}_P^T)'dy\,dz\end{aligned}\right\} = \delta_{P,Q}\Delta(v-v')\delta(w-w'), \tag{2.17}$$

for P and Q equal to V or H. The primes associated with with some of the terms in (2.17) indicate that for these terms the variables are u', v', w'. The expression $\Delta(v, v')$ is defined in (2.7b). For the general case considered in this paper the biorthogonal relationships (2.17) are not satisfied. Consequently, appropriate expressions for the field transforms E^P and H^P ($P = V$ or H), that replace (2.15), need to be found.

3. THE FIELD TRANSFORMS

In order to obtain the appropriate expression for the electric and magnetic transverse field transforms E^P and H^P ($P = V$ or H), respectively, it is necessary first to determine the values of the surface integrals (2.17) for the general three-dimensional case considered in this paper. Using the orthogonal properties of the scalar functions ψ^P (2.7b) and the completeness relationship (2.6b) it can be readily shown that for $P = Q$ (2.17) is still satisfied, thus for $P = V$ or H

$$\left.\begin{aligned}&\int_{-\infty}^{\infty} e_T^P \cdot (h_P^T \times a_x)'dy\,dz\\ &\int_{-\infty}^{\infty} h_T^P \cdot (\bar{a}_x \times e_P^T)'dy\,dz\end{aligned}\right\} = \Delta(v-v')\delta(w-w'). \tag{3.1a}$$

Furthermore, since $\bar{e}_T^H$ and $(\bar{a}_x \times \bar{e}_H^T)$ are orthogonal to $(\bar{h}_V^T \times \bar{a}_x)$ and $\bar{h}_T^V$, respectively, we also have

$$\int_{-\infty}^{\infty} \bar{e}_T^H \cdot (\bar{h}_V^T \times \bar{a}_x)'dy\,dz = \int_{-\infty}^{\infty} \bar{h}_T^V \cdot (\bar{a}_x \times \bar{e}_H^T)'dy\,dz = 0. \tag{3.1b}$$

On integrating with respect to y it can be shown that

$$\begin{aligned}A_H^V(v', w'; v, w) \equiv{}& \int_{-\infty}^{\infty} \bar{e}_T^V \cdot (\bar{h}_H^T \times \bar{a}_x)'dydz\\ ={}& \frac{i(w-w')}{2\pi}\int_{-\infty}^{\infty}\Bigg(\frac{(u^2+w^2)(u^2+w^2)'N^{H'}}{[(u^2+w^2-(u^2+w^2)']uu'}\\ &\times \sum_{i=1}^{m}\psi^V(v, h_{i-1,i})\psi^H(v', h_{i-1,i})(k_i^{-2}-k_{i-1}^{-2})\Bigg)\\ &\times \exp[-iz(w-w')]dy.\end{aligned} \tag{3.1c}$$

Similarly it can be shown that

$$\int_{-\infty}^{\infty} h_T^H \cdot (\bar{a}_x \times \bar{e}_V^T)'dydz \equiv A_V^H(v', w'; v, w), \tag{3.1d}$$

where A_V^H can be obtained from A_H^V by employing the duality relationships in electromagnetic theory. Thus, to get A_V^H from (3.1c), we set

$$\psi^V(v, y) \to -\psi^H(v, y), \quad \psi^H(v, y) \to \psi^V(v, y)$$
and
$$N^V \leftrightarrow N^H, \quad Z^V \leftrightarrow 1/Y^H. \tag{3.1e}$$

In the integrand of (3.1c) the term in the square bracket is implicitly a function of z through the parameters μ_1, ϵ_i, and $h_{i-1,k}$. For the special case when the parameters of the layered structure are independent of z it can readily be shown on employing (2.6a) that A_H^V and A_V^H vanish by virtue of the coefficient $(w - w')$ that multiplies the integral. On the basis of the relationship (3.1) it can be shown that the field transforms E^P and H^P ($P = V$ or H) in (2.14) are

$$E^V(x, v', w') = \iint_{-\infty}^{\infty} \bar{E}_T(x, y, z) \cdot (\bar{h}_V^T \times \bar{a}_x)'dy\,dz, \tag{3.2a}$$

$$H^H(x, v', w') = \iint_{-\infty}^{\infty} \bar{H}_T(x, y, z) \cdot (\bar{a}_x \times \bar{e}_V^T)'dy\,dz, \tag{3.2b}$$

$$\begin{aligned}&E^H(x, v', w') + \sum_v \int_{-\infty}^{\infty} A_H^V(v', w'; v, w)E^V(x, v, w)dw\\ &\qquad = \iint_{-\infty}^{\infty} \bar{E}_T(x, y, z) \cdot (\bar{h}_H^T \times \bar{a}_x)'dy\,dz,\end{aligned} \tag{3.2c}$$

and

$$\begin{aligned}&H^V(x, v', w') + \sum_v \int_{-\infty}^{\infty} A_V^H(v', w'; v, w)H^H(x, v, w)dw\\ &\qquad = \iint_{-\infty}^{\infty} \bar{H}_T(x, y, z) \cdot (\bar{a}_x \times \bar{e}_V^T)'dy\,dz.\end{aligned} \tag{3.2d}$$

On substituting (3.2a) and (3.2b) into (3.2c) and (3.2d) we can also express the transforms E^H and H^V exclusively in terms of the transverse electromagnetic fields $\bar{E}_T$ and $\bar{H}_T$. Thus even though the basis functions $\bar{e}_T^P$ and $\bar{h}_T^P$ are not orthogonal to the complementary basis functions $\bar{e}_P^T$ and $\bar{h}_P^T$, nevertheless due to the quasi-orthogonal properties exhibited by these functions, (3.1), it is possible to invert the full wave expansions for the transverse electromagnetic fields $\bar{E}_T$ and $\bar{H}_T$ (2.14) and obtain explicit formulas for the field transforms E^P and H^P in terms of $\bar{E}_T$ and $\bar{H}_T$.

In our work it is convenient to express the field transforms E^P and H^P in terms of forward and backward wave amplitudes a^P and b^P, respectively, as follows:

$$H^P = a^P + b^P \text{ and } E^P = a^P - b^P \text{ for } P = V \text{ or } H. \tag{3.3}$$

4. VERTICALLY AND HORIZONTALLY POLARIZED WAVE AMPLITUDES

In this section we convert Maxwell's equations for the transverse electromagnetic fields $\bar{E}_T$ and $\bar{H}_T$ (2.1) into a coupled set of equations for the vertically and horizontally polarized, forward and backward wave amplitudes a^P and b^P (3.3). To this end we substitute the complete full wave expansions for $\bar{E}_T$ and $\bar{H}_T$ (2.14) into Eq. (2.1a) and scalar multiply the equation by $(\bar{h}_V^T \times \bar{a}_x)'$ and integrate with respect to y and z over the entire y, z plane. The expressions for the scalar functions ψ^P are in general piecewise continuous functions, thus the integrations with respect to y must be performed separately in each layer of the structure. Furthermore the transverse field expansions (2.14) do not converge uniformlly at all points of the y, z plane when the parameters of the layered structure are functions of x and z. Hence in general it is not permissible to interchange orders of integration (summation) and differentiation. Thus, we have

$$\begin{aligned}-\int_{-\infty}^{\infty} \frac{\partial E_T}{\partial x} \cdot (\bar{h}_V^T \times \bar{a}_x)'dydz ={}& -\frac{d}{dx}\int_{-\infty}^{\infty} \bar{E}_T \cdot (\bar{h}_V^T \times \bar{a}_x)'dydz\\ &+ \int_{-\infty}^{\infty} \bar{E}_T \cdot \frac{\partial}{\partial x}(\bar{h}_V^T \times \bar{a}_x)'dydz\\ &- \sum_{i=1}^{m}\int_{-\infty}^{\infty}[\bar{E}_T \cdot (\bar{h}_V^T \times \bar{a}_x)']_{h_{i-1,i}^-}^{h_{i-1,i}^+} \tan\theta_{i-1,i}\,dl_{i-1,i},\end{aligned} \tag{4.1a}$$

in which $dl_{i-1,i}$ is the line element along the intersection of the y, z plane and the surface $y = h_{i-1,i}$ and

$$\bar{m}_{i-1,i} \cdot \bar{n}_{i-1,i} = \cos\theta_{i-1,i}. \tag{4.1b}$$

Thus

$$\tan\theta_{i-1,i}\,dl_{i-1,i} = \frac{\partial h_{i-1,i}}{\partial x}\,dz. \tag{4.1c}$$

Furthermore, using Green's theorem in two dimensions, we get

$$\int_{-\infty}^{\infty}\nabla_T\frac{J_x}{\epsilon}\cdot(\bar{h}_V^T\times\bar{a}_x)'\,dydz = -\int_{-\infty}^{\infty}\frac{J_x}{\epsilon}\nabla_T\cdot(\bar{h}_V^T\times\bar{a}_x)'dydz - \sum_{i=1}^{m}\int_{-\infty}^{\infty}\left[\frac{J_x}{\epsilon}(\bar{h}_V^T\times\bar{a}_x)'\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}\cdot\bar{m}_{i-1,i}\,dl_{i-1,i}, \tag{4.2a}$$

and

$$\int_{-\infty}^{\infty}\nabla_T\frac{1}{\epsilon}\nabla_T\cdot(\bar{H}_V^T\times\bar{a}_x)\cdot(\bar{h}_V^T\times\bar{a}_x)'dydz = \int_{\infty}^{\infty}(\bar{H}_T\times\bar{a}_x)\cdot\nabla_T\frac{1}{\epsilon}\nabla_T\cdot(\bar{h}_V^T\times\bar{a}_x)'dydz - \sum_{i=1}^{m}\left[\frac{1}{\epsilon}\nabla_T\cdot(\bar{H}_T\times\bar{a}_x)(\bar{h}_V^T\times\bar{a}_x)' - \frac{1}{\epsilon}\nabla_T\cdot(\bar{h}_V^T\times\bar{a}_x)'(\bar{H}_T\times\bar{a}_x)\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}\cdot\bar{m}_{i-1,i}\,dl_{i-1,i}. \tag{4.2b}$$

Employing (3.1) and the exact boundary conditions for the transverse electromagnetic fields (2.3c, d) and (2.4a) and noting that

$$\left[\frac{1}{\epsilon}\nabla_T\cdot(\bar{h}_V^T\times\bar{a}_x)\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}=0, \tag{4.3a}$$

$$\left(1+\frac{\epsilon}{k^2}\nabla_T\frac{1}{\epsilon}\nabla_T\cdot\right)(\bar{h}_T^V\times\bar{a}_x) = \frac{u}{\omega u}\bar{e}_V^T+\frac{\epsilon}{k^2}\phi^c\bar{a}_z\frac{\partial}{\partial z}\left[\frac{1}{\epsilon}\nabla_T\cdot(\bar{h}^V\times\bar{a}_x)\right], \tag{4.3b}$$

and at any interface $y=h_{i-1,i}$

$$\bar{E}_T\cdot(\bar{h}_V^T\times\bar{a}_x)' = (\bar{E}_T\cdot\bar{m})(\bar{h}_V^T\times\bar{a}_x\cdot\bar{m}) - (\bar{m}\times\bar{E}_T)\cdot[(\bar{h}_V^T\times\bar{a}_x)\times\bar{m}], = (\bar{E}_T\cdot\bar{m})(\bar{h}_V^T\cdot\bar{l}) - (\bar{E}_T\cdot\bar{l})(\bar{h}_V^T\cdot\bar{m}), \tag{4.3c}$$

and

$$\bar{H}_T\times\bar{a}_x\cdot\bar{m}=\bar{H}_T\cdot\bar{l}, \tag{4.3d}$$

we can show that (2.1a) reduces to

$$-\frac{d}{dx}E^V(x,v',w') - iu'H^V(x,v',w') - \sum_v\int_{-\infty}^{\infty}iu'A_V^HH^H\,dw + \sum_v\int_{-\infty}^{\infty}(C_V^VE^V+C_V^HE^H)dw + \sum_v\int_{-\infty}^{\infty}(F_V^VH^V+F_V^HH^H)\,dw = f^V(x), \tag{4.4a}$$

in which A_V^H is given by (3.1d),

$$C_V^P(v',w';v,w) = \int_{-\infty}^{\infty}\bar{e}_T^P\cdot\frac{\partial}{\partial x}(\bar{h}_V^T\times\bar{a}_x)'dydz + \sum_{i=1}^{m}\int_{-\infty}^{\infty}(\bar{e}_T^P\cdot\bar{l}_{i-1,i})(\bar{h}_V^T\cdot\bar{m}_{i-1,i})']_{h_{i-1,i}^-}^{h_{i-1,i}^+}\frac{\partial h_{i-1,i}}{\partial x}dz, \tag{4.4b}$$

$$F_V^P(v',w',v,w) = -\int_{-\infty}^{\infty}\frac{1}{iw}(\bar{h}_T^P\cdot\bar{a}_y)\phi^c\frac{\partial}{\partial z}\frac{1}{\epsilon}\nabla_T\cdot(\bar{h}_V\times\bar{a}_x)'dydz, \tag{4.4c}$$

and

$$f^V(x) = \int_{-\infty}^{\infty}[M_z\bar{h}_V^{T\prime}\cdot\bar{a}_z - \frac{1}{i\omega\epsilon}J_x\nabla_T\cdot(\bar{h}_V^T\times\bar{a}_x)']dydz. \tag{4.4d}$$

Scalar multiply (2.1b) by $(\bar{a}_x\times\bar{e}_V^T)'$ and integrate with respect to y and z over the entire y,z plane. Following the same procedure used to derive (4.4a) and noting that

$$\left[\frac{1}{\mu}\nabla_T\cdot(\bar{a}_x\times\bar{e}_V^T)=\frac{1}{\mu}\phi^c\frac{\partial}{\partial z}(\bar{e}_V\cdot\bar{a}_y)\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}, \tag{4.5a}$$

$$\left(1+\frac{\mu}{k^2}\nabla_T\frac{1}{\mu}\nabla_T\cdot\right)(\bar{a}_x\times\bar{e}_V^T) = \frac{u}{\omega\epsilon}\bar{h}_V^T+\frac{\phi^c}{k^2}\left[\mu\bar{a}_z\frac{\partial}{\partial z}\left(\frac{1}{\mu}\nabla_T\cdot(\bar{a}_x\times\bar{e}_V) + \left(\bar{a}_y\frac{\partial}{\partial y}+\bar{a}_z iw\right)\frac{\partial}{\partial z}(\bar{e}_V\cdot\bar{a}_y)\right)\right], \tag{4.5b}$$

and at any interface $y=h_{i-1,i}$,

$$(\bar{a}_x\times\bar{E}_T)\cdot\bar{m} = -\bar{E}_T\cdot\bar{l}, \tag{4.5c}$$

and

$$(\bar{H}_T\cdot\bar{a}_x\times\bar{e}_V^T) = (\bar{H}_T\cdot\bar{m})(\bar{a}_x\times\bar{e}_V^T\cdot\bar{m}) - (\bar{m}\times\bar{H}_T)\cdot[(\bar{a}_x\times\bar{e}_V^T)\times\bar{m}] = -(\bar{H}_T\cdot\bar{m})(\bar{e}_V^T\cdot\bar{l}) + (\bar{H}_T\cdot\bar{l})(\bar{e}_V^T\cdot\bar{m}), \tag{4.5d}$$

it can be shown that (2.1b) reduces to

$$-\frac{d}{dx}H^V(x,v',w') - iu'E^V(x,v',w') - \frac{d}{dx}\sum_v\int_{-\infty}^{\infty}A_V^HH^Hdw + \sum_v\int_{-\infty}^{\infty}(D_V^VH^V+D_V^HH^H)dw + \sum_v\int_{-\infty}^{\infty}(G_V^VE^V+G_V^HE^H)dw = g^V(x), \tag{4.6a}$$

in which

$$D_V^P(v',w';v,w) = \int_{-\infty}^{\infty}\bar{h}_T^P\cdot\frac{\partial}{\partial x}(\bar{a}_x\times\bar{e}_V^T)'dydz - \sum_{i=1}^{m}\int_{-\infty}^{\infty}[(\bar{h}_T^P\cdot\bar{l}_{i-1,i})(\bar{e}_V^T\cdot\bar{m}_{i-1,i})']_{h_{i-1,i}^-}^{h_{i-1,i}^+}\frac{\partial h_{i-1,i}}{\partial x}dz, \tag{4.6b}$$

$$G_V^P(v',w';v,w) = \int_{-\infty}^{\infty}\frac{1}{i\omega}(\bar{a}_x\times\bar{e}_T^P)\cdot\phi^c\left[\bar{a}_z\frac{\partial}{\partial z}\left(\frac{1}{\mu}\nabla_T\cdot(\bar{a}_x\times\bar{e}_V)'\right) + \frac{1}{\mu}(\bar{a}_y\frac{\partial}{\partial y}+\bar{a}_z iw)\frac{\partial}{\partial z}(\bar{e}_V\cdot\bar{a}_y)'\right]dydz - \sum_{i=1}^{m}\int_{-\infty}^{\infty}\left[\frac{1}{i\omega\mu}(\bar{e}_T^P\cdot\bar{l}_{i-1,i})\phi^c\frac{\partial}{\partial z}(\bar{e}_V\cdot\bar{a}_y)'\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}dl_{i-1,i} \tag{4.6c}$$

$$g^V(x) = \int_{-\infty}^{\infty}\bar{J}_T\cdot\bar{e}_V^T\,dydz - \int_{-\infty}^{\infty}\frac{1}{i\omega\mu}M_x\nabla_T\cdot(\bar{a}_x\times\bar{e}_V^T)'dydz \tag{4.6d}$$

On scalar multiplying (2.1a) and (2.1b) by $(\bar{h}_H^T\times\bar{a}_x)'$ and $(\bar{a}_x\times\bar{e}_H^T)'$, respectively, and integrating with respect to y and z over the y,z plane, we obtain in a similar manner

$$-\frac{d}{dx}E^H(x,v',w') - iu'H^H(x,v',w') - \frac{d}{dx}\sum_v\int_{-\infty}^{\infty}A_H^VE^Vdw + \sum_v\int_{-\infty}^{\infty}[C_H^VE^V+C_H^HE^H]dw + \sum_v\int_{-\infty}^{\infty}[F_H^VH^V+F_H^HH^H]dw = f^H(x), \tag{4.7a}$$

$$-\frac{d}{dx}H^H(x,v',w') - iu'E^H(x,v',w') - \sum_v\int_{-\infty}^{\infty}iu'A_H^VE^Vdw + \sum_v\int_{-\infty}^{\infty}[D_H^VH^V+D_H^H]dw + \sum_v\int_{-\infty}^{\infty}[G_H^VE^V+G_H^HE^H]dw = g^H(x), \tag{4.7b}$$

in which A_V^H is defined in (3.1c) and

$$C_H^P(v',w';v,w) = \int_{-\infty}^{\infty} \bar{e}_T^P \cdot \frac{\partial}{\partial x}(\bar{h}_H^T \times \bar{a}_x)' dydz + \sum_{i=1}^{m} \int_{-\infty}^{\infty} [\bar{e}_T^P \cdot \bar{l}_{i-1,i})(\bar{h}_H^T \cdot \bar{m}_{i-1,i})']_{h_{i-1,i}^-}^{h_{i-1,i}^+} \frac{\partial h_{i-1,i}}{\partial x} dz, \tag{4.8a}$$

$$D_H^P(v',w';v,w) = \int_{-\infty}^{\infty} \bar{h}_T^P \cdot \frac{\partial}{\partial x}(\bar{a}_x \times \bar{e}_H^T)' dydz - \sum_{i=1}^{m} \int_{-\infty}^{\infty} [(\bar{h}_T^P \cdot \bar{l}_{i-1,i})(\bar{e}_H^T \cdot \bar{m}_{i-1,i})]_{h_{i-1,i}^-}^{h_{i-1,i}^+} \frac{\partial h_{i-1,i}}{\partial x} dz, \tag{4.8b}$$

$$F_H^P(v',w';v,w) = \int_{-\infty}^{\infty} \frac{1}{i\omega}(\bar{h}_T^P \times \bar{a}_x) \cdot \phi^c \bar{a}_z \frac{\partial}{\partial z}\left[\frac{1}{\epsilon}\nabla_T \cdot (\bar{h}_V \times \bar{a}_x)' - \frac{1}{\epsilon}\left(\bar{a}_y \frac{\partial}{\partial y} + \bar{a}_z i\omega\right)\frac{\partial}{\partial z}(\bar{h}_V \cdot \bar{a}_y)\right] dydz - \sum_{i=1}^{\infty} \int_{-\infty}^{\infty} \left[\frac{1}{i\omega\epsilon}(\bar{h}_T^P \cdot \bar{l}_{i-1,i})\phi^c \frac{\partial}{\partial z}(\bar{h}_H \cdot \bar{a}_y)\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+} dl_{i-1,i}, \tag{4.8c}$$

$$G_H^P(v',w',v,w) = \int_{-\infty}^{\infty} \frac{1}{i\omega}(\bar{e}_T^P \cdot \bar{a}_y)\phi^c \frac{\partial}{\partial z}\frac{1}{\mu}\nabla_T \cdot (\bar{a}_x \times \bar{e}_v)' dydz, \tag{4.8d}$$

$$f^H(x) = \int_{-\infty}^{\infty} \bar{M}_T \cdot \bar{h}_H^T dydz - \int_{-\infty}^{\infty} \frac{1}{i\omega\epsilon} J_x \nabla_T \cdot (\bar{h}_H^T \times \bar{a}_x) dydz, \tag{4.8e}$$

$$g^H(x) = \int_{-\infty}^{\infty} J_z(\bar{e}_H^T \cdot \bar{a}_z) dydz - \int_{-\infty}^{\infty} \frac{1}{i\omega\mu} M_x \nabla_T \cdot (\bar{a}_x \times \bar{e}_H^T) dydz. \tag{4.8f}$$

Equations (4.7) and (4.8) can be shown to be related to (4.4) and (4.6) through the duality relationships in electromagnetic theory. $[(\bar{a}_x \times \bar{e}_T^P) \leftrightarrow (\bar{h}_T^P \times \bar{a}_x), \bar{J} \leftrightarrow \bar{M}, \mu \leftrightarrow \epsilon]$. When the structures parameters are independent of z, A_Q^P, F_Q^P, and G_Q^P vanish and the expressions for C_Q^P and D_Q^P reduce to those derived earlier.[1] While C_Q^P and D_Q^P are $\bar{E}$ field to $\bar{E}$ field and $\bar{H}$ field to $\bar{H}$ field coupling coefficients respectively F_Q^P and G_Q^P are mixed type coupling coefficients. To derive the coupled equations for the forward and backward vertically and horizontally polarized wave amplitudes we substitute (3.3) into (4.4a), (4.6a), (4.7a), and (4.7b). Thus it can be shown that for $P = V$ or H,

$$-\frac{da^P}{dx} - iua^P - \tfrac{1}{2}\frac{d}{dx}\sum_Q \sum_{v'} \int_{-\infty}^{\infty} [A_P^Q(a^Q \pm b^Q)]dw' - \tfrac{1}{2}\sum_Q \sum_{v'} \int_{-\infty}^{\infty} iu'[A_P^Q(a^Q \pm b^Q)]dw' = \sum_Q \sum_{v'} \int_{-\infty}^{\infty} [S_{PQ}^{BA} a^Q + S_{PQ}^{BB} b^Q]dw' - A^P, \tag{4.9a}$$

$$-\frac{db^P}{dx} + iua^P - \tfrac{1}{2}\frac{d}{dx}\sum_Q \sum_{v'} [A_P^Q(a^Q \pm b^Q)]dw' + \tfrac{1}{2}\sum_Q \sum_{v'} \int_{-\infty}^{\infty} iu'[A_P^Q(a^Q \pm b^Q)]dw' = \sum_Q \sum_{v'} \int_{-\infty}^{\infty} (S_{PQ}^{AA} a^Q + S_{PQ}^{AB} b^Q)dw' + B^P, \tag{4.9b}$$

in which the upper sign is used for $P = V$ and the lower sign for $P = H$. The transmission scattering coefficients are

$$S_{PQ}^{\beta\alpha}(v,w;v',w') = -\tfrac{1}{2}[C_P^Q(v,w/v',w') + D_P^Q(v,w;v',w') \pm F_P^Q(v,w;v',w') \pm G_P^Q(v,w;v',w')] \tag{4.10a}$$

and the reflection scattering coefficients are

$$S_{PQ}^{\beta\beta}(v,w;v',w') = \tfrac{1}{2}[C_P^Q(v,w/v',w') - D_P^Q(v,w;v',w') \mp F_P^Q(v,w;v',w') \pm G_P^Q(v,w;v',w')], \tag{4.10b}$$

where the upper sign is used for $\alpha = A$ and $\beta = B$ and the lower sign for $\alpha = B$ and $\beta = A$. The source terms are

$$A^P = -(f^P + g^P)/2 \quad \text{and} \quad B^P = -(f^P - g^P)/2 \tag{4.10c}$$

for $P = V$ or H and A_Q^P is defined by (3.1) for $P \neq Q$ and

$$A_P^P(v,w;v',w') = 0, \quad P = V \text{ or } H. \tag{4.10d}$$

5. CONCLUDING REMARKS

Employing generalized field transforms, Maxwell's equations for inhomogeneous media, with arbitrary distributions of electromagnetic sources, are converted into coupled sets of first-order ordinary differential equations for the vertically and horizontally polarized, forward and backward wave amplitudes. The basis functions used for the full wave expansions of the fields are shown to exhibit a quasiorthogonal relationship with a complementary set of basis functions when the propagation medium is an inhomogeneous layered structure of nonuniform thickness. This quasiorthogonal property renders a tractable solution to the general problem even when the electromagnetic and geometric parameters of the structure are arbitrary three-dimensional functions of position.

The full wave solutions are not limited by the convenient surface impedance concept nor are they restricted by frequency considerations. They are applicable to a large class of propagation problems in nonuniform structures for which no single constituent of the total full wave solution dominates. These solutions permit the use of more realistic models of pertinent propagation problems. They include scattering by rough surfaces and irregularly shaped objects and propagation over irregular terrain and in nonuniform ducts. Since the full wave expansions for the fields account for the radiation, lateral wave, and trapped waveguide modes of the structure, the solutions developed in this paper may be applied to open as well as closed structures.

* The research reported in this paper has been supported by the National Science Foundation and the Engineering Research Center of the University of Nebraska. The author wishes to thank A. Edison for his comments. The manuscript was typed by Mrs. E. Everett.

[1] E. Bahar, J. Math. Phys. 14, 1502 (1973).
[2] E. Bahar, J. Math. Phys. 14, 1510 (1973).
[3] E. Bahar, J. Math. Phys. 14, 1024 (1973).

On inverse scattering

V. H. Weston

Division of Mathematical Sciences, Purdue University, West Lafayette, Indiana 47907
(Received 22 August 1973)

In a previous paper the inverse problem associated with a hyperbolic dispersive partial differential equation with smooth coefficients was considered. The inverse problem (the determination of the coefficients) was formulated in terms of a dual set of integral equations involving measurable quantities, the kernels of the transmission, and reflection operators. These equations contained an unknown parameter which occurs in a linear manner. A better approach to determine this parameter is presented here. It involves an auxiliary equation, which is used to eliminate the unknown parameter from the integral equations. It is shown that the resulting system has a unique solution for a certain class of scattering problems. These uniqueness results are then strengthened when an additional equation is employed to reduce the dual set of integral equations to a single integral equation.

1. INTRODUCTION

In a previous paper,[1] which we shall refer to as Paper 1, we considered the inverse problem associated with the partial differential equation

$$u_{xx} - u_{tt} + A(x)u_x + B(x)u_t + C(x)u = 0, \tag{1}$$

with smooth coefficients that vanished outside the interval $0 < x < l$. Scattering operators involving the measurable quantities, the reflection and transmission kernel, $R_{\pm}$, $T_{\pm}$ were introduced, where the signs correspond to the direction of the incident wave with relation to the positive x axis. The inverse problem (the determination of the coefficients A, B, and C) was formulated in terms of a dual set of integral equations in the variable t with x fixed. These equations contained an unknown parameter which occurs linearly. A technique for computing the unknown parameter was developed involving the solution of a second-order ordinary differential equation. The question of uniqueness was partially answered in that it was pointed out that Neumann series would converge for x sufficiently close to 0 or l. The question of uniqueness was further pursued[2] where it was pointed out that one needed additional equations or restraints.

Here we develop a better method of obtaining the unknown parameter. An auxiliary equation is derived which is used to eliminate the unknown parameter from the system of integral equations. Once their solution is found, the unknown parameter is found by integration, from the auxiliary equation. From a stability standpoint, this is a better approach than that given in Paper 1. In addition, when this auxiliary equation is employed, the treatment of uniqueness is better than that given in Ref. 2. Existence of a solution is shown, and sufficient conditions on the coefficients, for uniqueness, are obtained.

It will also be shown that when an additional equation is employed, uniqueness can be guaranteed under very general conditions, although in this case the integral equation to be solved may become unwieldy for computational purposes.

2. FUNDAMENTAL INTEGRAL EQUATIONS AND AUXILIARY CONDITION

To simplify notation we shall set

$$K_-(x, t; 0) = f(t),$$
$$K_+(x, t; l) = g(t)G(x).$$

The equations given by Eqs. (24a) and (24b) of Paper 1 become

$$G(x)^{-1}L_+(x - t) + S_+(x, t) - \int_{-x}^{x} f(y)S_+(y, t)dy = \begin{cases} g(t), & \text{(2a)} \\ 0, & \text{(2b)} \end{cases}$$

holding in the respective ranges $x \le t \le 2l - x$ and $2l - x < t$, and where

$$S_+(y, t) = R_+(y + t) + \int_x^t R_+(y + s)L_+(s - t)ds.$$

The equations given by Eqs. (25a) and (25b) of Paper 1 become

$$L_-(x - t) + G(l)G(x)^{-1}S_-(x, t) + G(l)\int_x^{2l-x} g(y)S_-(y, t)dy = \begin{cases} -f(t) & \text{(3a)} \\ 0 & \text{(3b)} \end{cases}$$

for the respective ranges $-x \le t \le x$ and $t < -x$, and where

$$S_-(y, t) = R_-(y + t) + \int_t^x R_-(s + y)L_-(s - t)ds$$

and the unknown parameter is given by $G(x)^{-1}$.

In order to obtain the proper auxiliary equation, we shall employ either Eq. (2b) for the range $t > 2l + x$, or Eq. (3b) for the range $t < -2l + x$. As was pointed out in Paper 1, these equations become independent of the variable t for the ranges quoted above. Their explicit form will be required here. To achieve this first note that the quantity S_+ can be written in the form

$$S_+(y, t) = M_+(y + t) - \int_{-y}^{x} R_+(y + s)L_+(s - t)ds$$

Since

$$-M_+(y + t) = L_+(s - t) = \text{const} = L_+(-2l)$$

for the ranges $-x \le y \le x$, $-y \le s \le x$, $t \ge 2l + x$, it follows that

$$S_+(y, t) = -L_+(-2l)\left(1 + \int_0^{x+y} R_+(u)du\right)$$

and Eq. (2b) reduces to the following for $t \ge 2l + x$:

$$L_+(-2l)\left[G(x)^{-1} - \left(1 + \int_0^{2x} R_+(u)du\right) + \int_{-x}^{x} f(y)[1 + \int_0^{x+y} R_+(u)du]dy\right] = 0. \tag{4a}$$

In a similar manner Eq. (3b) reduces to the following for $t \le -2l + x$,

$$L_-(2l)\left[G(l)^{-1} - G(x)^{-1}\left(1 + \int_{2x}^{2l} R_-(u)du\right) - \int_x^{2l-x} g(y)[1 + \int_{x+y}^{2l} R_-(u)du]dy\right] = 0. \quad (4b)$$

For present purposes we shall assume that $L_+(-2l) \neq 0$, in which case we will employ Eq. (4a) as the auxiliary equation. If $L_+(-2l) = 0$, but $L_-(2l) \neq 0$, then Eq. (4b) may be used; but to employ it, a modification of the unknown quantities must be made. One should solve for the quantities $K_+(x, y; l)$ and $G(x)K_-(x, y; 0)$ in this case. The vanishing of both constants occurs only in extreme pathological cases.

Thus with $L_+(-2l) \neq 0$, the auxiliary equation (4a) will be given in the form

$$G(x)^{-1} = 1 + \int_0^{2x} R_+(u)du - \int_{-x}^{x} f(y)\left(1 + \int_0^{x+y} R_+(u)du\right)dy. \quad (5)$$

We will require the following alternative form for Eqs. (2a) and (3a), which are obtained by employing the relations Eqs. (6) and (12) of Paper 1,

$$g(t) + \int_x^t T_+(s-t)g(s)ds = -G(x)^{-1}T_+(x-t) + R_+(x+t) - \int_{-x}^{x} R_+(y+t)f(y)dy, \quad (6a)$$

$$f(t) + \int_t^x T_-(s-t)f(s)ds = T_-(x-t) - G(l)G(x)^{-1}R_-(x+t) - G(l)\int_x^{2l-x} R_-(y+t)g(y)dy. \quad (6b)$$

The system of equations to be solved is (6a), (6b), and (5). Equation (5) is used to eliminate $G(x)$ from the other two equations. Once these resulting equations are solved, Eq. (5) is then used to obtain $G(x)$. The coefficient $B(x)$ is determined from $G(x)$ by the relation

$$G(x) = \exp\left(-\int_0^x B(\tau)d\tau\right)$$

and the remaining coefficients are obtained from the resulting expression for $f(x) = K_-(x, x; 0)$ given in paper 1.

3. EXISTENCE AND UNIQUENESS OF SYSTEM (5), (6a), AND (6b)

If we use Eq. (5) to eliminate $G(x)^{-1}$ from Eqs. (6a) and (6b), the resulting system of integral equations expressed in general operator form, is given by

$$(\mathbf{I} + \mathbf{A})\Psi = \chi, \quad (7)$$

where Ψ and χ are vector-valued functions, and in particular

$$\Psi = \begin{pmatrix} f \\ g \end{pmatrix}.$$

The operator $\mathbf{A}$ is expressed in matrix form in terms of the compact operators A_{ij}, as follows

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Employing the real Hilbert space of square integrable functions with inner-product given by

$$(\Psi, \tilde{\Psi}) = \int_x^{2l-x} g(l)\tilde{g}(l)dl + \int_{-x}^{x} f(l)\hat{f}(l)dl$$

the homogeneous adjoint system corresponding to Eq. (7) is given by

$$(\mathbf{I} + \mathbf{A}^*)\Psi^* = 0, \quad (8)$$

where

$$\mathbf{A}^* = \begin{pmatrix} A_{11}^* & A_{12}^* \\ A_{21}^* & A_{22}^* \end{pmatrix}, \qquad \Psi^* = \begin{pmatrix} f^* \\ g^* \end{pmatrix}.$$

The explicit form of Eq. (8) is given by

$$g^*(t) + \int_t^{2l-x} T_+(t-s)g^*(s)ds = -G(l)\int_{-x}^{x} R_-(s+t)f^*(s)ds, \quad (9a)$$

$$f^*(t) + \int_{-x}^{t} T_-(t-s)f^*(s)ds + \int^{2l-x} g^*(s)R_+(s+t)ds = -g^*(x)\left(1 + \int_0^{x+t} R_+(y)dy\right), \quad (9b)$$

for the respective ranges $x \leq t \leq 2l - x$ and $-x \leq t \leq x$.

The necessary and sufficient conditions for existence of a solution of Eq. (7) is given by

$$(\Psi^*, \chi) = 0,$$

where Ψ^* is a solution to Eq. (8). It can be shown that this is reducible to the explicit form

$$f^*(x) = 0. \quad (10)$$

Equation (7) will have a unique solution if and only if the homogeneous adjoint system (8) has only the trivial solution.

To investigate the solutions of the adjoint system, we will employ the definitions of the operators with kernels $R_\pm$ and $T_\pm$ as given in Paper 1, to express Eqs. (9a) and (9b) in terms of an initial-boundary value problem associated with Eq. (1). First we replace the fixed constant x by the constant μ through the relation

$$\mu = 2l - 2x,$$

in order not to confuse it with the variable x in Eq. (1).

Set

$$f^*(l) = u_-^i(l + l - \tfrac{1}{2}\mu),$$

$$g^*(l) = u_+^i(l - l - \tfrac{1}{2}\mu)\exp(-\gamma_+),$$

where

$$\gamma_\pm = \tfrac{1}{2}\int_0^l [A(s) \pm B(s)]ds.$$

Let $u_+^i(x - t - \mu)$ for $x \leq 0$, and $u_-^i(x + t)$ for $x \geq l$ represent incident waves propagating in the direction of the positive and negative x axis, respectively, such that

$$\begin{aligned} &u_+^i(s) = 0, && \text{for } s > 0, \\ &u_+^i(s) = u_+^i(-\mu + 0), && \text{for } s \leq -\mu, \\ &u_-^i(s) = 0, && \text{for } s < 0 \text{ and } s > 2l - \mu, \end{aligned} \quad (11)$$

and where u_+^i and u_-^i are arbitrary C^2 (C^1 piecewise) functions except for the possibility of jump discontinuities at $s = 0$, and a jump discontinuity in $u_\pm^i(s)$ only, at $s = 2l - \mu$.

These incident waves will give rise to reflected waves $u_+^r(x+t+\mu), u_-^r(x-t)$, and transmitted waves $u_+^t(x-t-\mu), u_-^t(x+t)$ in the appropriate half-spaces $x \le 0$, and $x \ge l$.

Employing the operators given in Paper 1, Eqs. (9a) and (9b) reduce to

$$u_+^t(l-t-\mu) + u_-^r(l+t) = 0 \qquad l-\mu \le t \le l$$

$$u_-^t(t) + u_+^r(t+\mu) + u_+^i(-t-\mu) = 0 \qquad 0 \le t \le 2l-\mu,$$

respectively.

Upon superposition, it is seen that the system of adjoint integral equations correspond to the initial-boundary value problem associated with Eq. (1), where the initial conditions are given by

$$u(x,t) = u_+^i(x-t-\mu) + u_-^i(x+t) \quad \text{for } t < -\mu$$

and boundary conditions

$$u(0,t) = u_x(0,t) = 0, \qquad 0 \le t \le 2l-\mu,$$

$$u(l,t) = u_x(l,t) = 0, \qquad l-\mu \le t \le l$$

and where the solution $u(x,t)$ must be C^2 except for the possibility of a jump discontinuity along the characteristics $x-t=\mu, x+t=0, x+t=2l-\mu$, and u_+^i and u_-^i must satisfy conditions (11).

The result that $u_x(0,t)=0$ follows from the fact that $u(x,t)$ has the form $h(x+t)$ for the domain $x \le 0$, $0 \le x+t \le 2l-\mu$, since the incident wave u_+^i is constant here.

One may proceed to show that the boundary conditions yield $u(x,t) \equiv 0$ in the triangles bounded by $x=0, x=l$, $x-t=0$, and $x+t=2l-\mu$. In fact, $u(x,t)=0$ along the sides of the two triangles given by $x-t=0$. Hence there is no jump discontinuity in $u(x,t)$ across the characteristic $x+t=2l-\mu$ at the point $x=l-\mu/2$, $t=l-\mu/2$. Since jump discontinuities $[u]$ associated with Eq. (1) propagate along the characteristic $x+t=2l-\mu$ according to the exponential law

$$[u] = \text{const} \exp\left(-\tfrac{1}{2}\int_0^x [A(s)+B(s)]ds\right),$$

if follows that there is no jump discontinuity along the characteristic $x+t=2l-\mu$. This implies that $u_-^i(2l-\mu-0=0$, hence $f^*(l-\frac{1}{2}\mu)=0$ which is equivalent to Eq. (10). Thus the solution to system (7) exists.

One can show further that $u(x,t) \equiv 0$ in the region given by $x+t>0, x-t<\mu$. Because of the possibility of a jump discontinuity along the characteristics $x-t=\mu, x+t=0$, one cannot directly proceed to show that u_+^i and u_-^i are identically zero. However from the initial conditions and Eq. (11), it is seen that $u(x,t)$ must vanish in the domain $x-t>\mu, x+t<0$ as well as the domain $x+t \ge 2l-\mu, x-t \ge 2l$. This implies that the solution $u(x,t)$ of Eq. (1) in the quarter space $x-t>\mu$, $x+t>0$, satisfying the boundary conditions

$$u = u_0 \exp\left(-\tfrac{1}{2}\int_{\mu/2}^x [A(s)+B(s)]ds\right)$$

$$u = u_0 \exp\left(-\tfrac{1}{2}\int_{\mu/2}^x [A(s)-B(s)]ds\right)$$

along the respective characteristics $x+t=0$ and $x-t=\mu$, where $u_0 = u(\mu/2, -\mu/2)$, must vanish in the domain $x+t \ge 2l-\mu, x-t \ge 2l$. Since $u(x,t) = u_0 R(0,\mu;\xi,\eta)$ where $R(0,\mu;\xi,\eta)$ is the Riemann function associated with Eq. (1), where $\xi = x+t, \eta = x-t$ are characteristic coordinates, it is seen that if $R(0,\mu;\xi,\eta)$ does not vanish in the domain $\xi \ge 2l-\mu$, $\eta \ge 2l$, then we must have $u_0 = 0$. This implies that $u(x,t) \equiv 0, u_-^i = 0$, hence both f^* and g^* must vanish. On combining results, we have

Theorem: When $L_+(-2l) \ne 0$, and the reflection and transmission operators correspond to Eq. (1), the solution of the system of equations (5), (6a), and (6b) exists. Furthermore, it is unique, if the Riemann function $R(0,\mu;2l-\mu,2l)$ associated with (1), does not vanish where $\mu = 2l-2x$.

Corollary: Sufficient (but not necessary) conditions for the solution to be unique are given by either

(i) $B \le 0, \quad C + \frac{1}{2}(B'-A') + \frac{1}{4}(B^2-A^2) \le 0,$

or

(ii) $B \ge 0, \quad C - \frac{1}{2}(B'+A') + \frac{1}{4}(B^2-A^2) \le 0,$

where A, B, C are the coefficients of Eq. (1).

Proof: We shall briefly outline the proof for case (i). Setting

$$R(0,\mu;\xi,\eta) = v(\xi,\eta) \exp\left(\tfrac{1}{2}\int_{\mu/2}^{(\xi+\eta)/2} [B(s)-A(s)]ds\right)$$

the resulting differential equation and boundary conditions for v expressed in characteristic coordinates can be represented in terms of the following integral equation for $\xi \ge 0, \eta \ge \mu$:

$$v = \exp\left(-\tfrac{1}{2}\int_\mu^\eta B(\tau/2)d\tau\right) - \int_0^\xi \int_\mu^\eta [\tfrac{1}{2}Bv_\xi + \tfrac{1}{4}Dv]d\xi' d\eta',$$

where

$$D = C + \tfrac{1}{2}(B'-A') + \tfrac{1}{4}(B^2-A^2).$$

This equation can be solved by successive approximations[3] and it follows that for case (i), v and v_ξ will be positive for $\xi \ge 0, \eta \ge \mu$. It immediately follows that $R(0,\mu;2l-\mu,2l)$ does not vanish.

The system of equations (5), (6a), and (6b) may be a natural choice for numerical treatment since they involve explicitly the measured quantities $R_\pm$ and $T_\pm$. However since the system is not always unique, we need to employ additional information. In the next section we will show how, by including Eq. (2b) we can obtain stronger uniqueness results.

4. EXTENSION OF RESULTS

We shall employ Eq. (2b) in conjunction with Eqs. (2a) and (3a) to eliminate the unknown quantity $g(t)$ and so obtain a single integral equation for $f(t)$. In the term

$$G(l)\int_x^{2l-x} R_-(y+t)g(y)dy$$

on the right-hand side of Eq. (6b), we replace $g(y)$ by the expression on the right-hand side of Eq. (2a), and on account of Eq. (2b), we extend the range of integration of y from x to $2l-t$. In this manner we obtain the resulting equation

$$\begin{aligned} f(t) &+ \int_t^x T_-(s-t)f(s)ds \\ &- G(l)\int_{-x}^x f(s)\int_{x+s}^{2l+s-t} R_+(y)P_-(t-s+y)dyds \\ &= T_-(x-t) - G(l)G(x)^{-1}P_-(x+t) \\ &- G(l)\int_{2x}^{2l+x-t} R_+(y)P_-(t-x+y)dy, \end{aligned} \tag{12}$$

for $-x \le t \le x$, where $P_-(s)$, defined by Eq. (21) in Paper 1, is given by

$$P_-(s) = R_-(s) + \int_s^{2l} R_-(y)L_+(s-y)dy.$$

Thus Eq. (12) and (5) constitute the second system of equations. Equation (5) is used to eliminate $G(x)^{-1}$ from (12). Once the resulting equation is solved for $f(t)$, $G(x)^{-1}$ is computed from Eq. (5).

If we employ the relation which is equivalent to Eq. (22) of Paper 1,

$$Q_-(\eta) = \int_0^{2l+\eta} R_+(y)P_-(y-\eta)dy$$

and in addition Eq. (23) from Paper 1,

$$T_-(\eta) = G(l)[Q_-(\eta) - L_+(-2l)] \quad \text{for } \eta \ge 0,$$

it can be shown that Eq. (5) and Eq. (12) yield the result that

$$f(-x) = 0.$$

We will show next that this system has a unique solution.

On eliminating $G(x)^{-1}$, from Eq. (12) we obtain the corresponding homogeneous equation,

$$f(t) + \int_t^x T_-(s-t)f(s)ds - G(l)\int_{-x}^x k(t,s)f(s)ds = 0, \qquad -x \le t \le x,$$

where

$$k(t,s) = P_-(x+t)\left(1 + \int_0^{x+s} R_+(y)dy\right) + \int_{x+s}^{2l+s-t} R_+(y)P_-(t-s+y)dy.$$

The adjoint equation is given by

$$f^*(t) + \int_{-x}^t T_-(t-s)f^*(s)ds - G(l)\int_{-x}^x k(s,t)f^*(s)ds = 0, \qquad -x \le t \le x. \quad (13)$$

To show that this has only the trivial solution, we want to interpret the equation in terms of an initial boundary value problem associated with Eq. (1). Replace the fixed parameter x by $\mu/2$, and set

$$t = t' - \tfrac{1}{2}\mu, \qquad f^*(t) = u_-^i(t').$$

Then Eq. (13) becomes, for $0 \le t' \le \mu$, on dropping the primes on t'.

$$e^{\gamma_+}\left(u_-^i(t) + \int_0^t T_-(t-s)u_-^i(s)ds\right) = e^{\gamma_-}\int_0^\mu \tilde{k}(s,t)u_-^i(s)ds, \quad (14)$$

where

$$\tilde{k}(s,t) = P_-(s)\left(1 + \int_0^t R_+(y)dy\right) + \int_t^{2l+t-s} R_+(y)P_-(s-t+y)dy.$$

It is easily seen that if $u_-^i(x+t)$ represents an incident wave propagating in the direction of negative x axis, such that $u_-^i(s) = 0$ for $s < 0$, and $s > \mu$, then the left-hand side of Eq. (14) corresponds to the transmitted wave $u_-^t(x+t)$ at $x = 0$. Note that $u_-^i(s)$ may have a jump discontinuity at $s = 0$ or $s = \mu$.

To interpret the right-hand side, we must set $u_+^i(x-t-2l)$ to be a wave propagating in the direction of the positive x axis such that $u_+^i(s) = 0$ for $s > 0$, and such that it produces a transmitted wave $u_+^t(x-t-2l)$ which nullifies the reflected wave $u_-^r(x-t)$ at $x = l$, for $0 \le t \le 2l$. From the Appendix, it follows that

$$u_+^i(\eta) = -e^{\gamma_-}\int_0^{-\eta} u_-^i(s)P_-(s+2l+\eta)ds, \qquad -\mu \le \eta \le 0,$$

$$= -e^{\gamma_-}\int_0^{\mu} u_-^i(s)P_-(s+2l+\eta)ds, \qquad -2l \le \eta \le -\mu.$$

We will require that $u_+^i(s)$ be constant for $s \le -2l$, such that it is continuous at $s = -2l$, yielding

$$u_+^i(\eta) = -e^{\gamma_-}\int_0^\mu u_-^i(s)P_-(s)ds, \qquad s \le -2l.$$

This incident wave will produce a reflected wave $u_+^r(x+t+2l)$, which at $x = 0$ has the form

$$u_+^r(t+2l) = \int_{-2l}^0 R_+(t+2l+s)u_+^i(s)ds + u_+^i(-2l)\int_0^t R_+(y)dy.$$

Combining these results we see that the right-hand side of Eq. (14) is equal to

$$-u_+^i(-t-2l) - u_+^r(t+2l),$$

for $0 \le t \le \mu$, where in this range u_+^i is constant.

The adjoint integral equation is thus equivalent to the initial-boundary value problem for $u(x,t)$ which satisfies Eq. (1), where the initial conditions are

$$u(x,t) = u_+^i(x-t-2l) + u_-^i(x+t), \qquad \text{for } t < -2l,$$

where u_+^i and u_-^i must vanish in the domains indicated above, and boundary conditions are

$$u_x(0,t) = u(0,t) = 0, \qquad \text{for } 0 \le t \le \mu,$$

$$u_x(l,t) = u(l,t) = 0, \qquad \text{for } -l \le t \le l.$$

Since u_+^i is continuous, the only possible jump discontinuities may occur along the characteristics $x+t = 0$ and $x+t = \mu$.

As for the previous case, one may start from the boundary conditions and show that there can be no jump discontinuity along $x+t = \mu$. However, unlike the previous case, one can show further that there can be no jump discontinuity along $x+t = 0$, and hence the only possible solution to the inital boundary value problem is the trivial solution $u = u_+^i = u_-^i = 0$.

Thus the adjoint equation has only the trivial solution, hence the original equation is unique.

Theorem: If $L_+(-2l) \ne 0$, system (5) and (12) has a unique solution for $f(t)$, which vanishes at $t = -x$.

For completeness we should mention that the solution of system (5) and (12) will satisfy the original set of Eqs. (2a), (2b), (3a), and (3b). Since it does not follow directly, we will outline the approach.

Because of the relationships between the scattering operators given by Eqs. (22) and (23) of Paper 1, it can be shown that the right-hand side of Eq. (12) vanishes not only for $t = -x$ as was indicated above, but for all $t \le -x$. Hence Eq. (12) may be used to define $f(t) = 0$ for $t \le -x$. We employ the left-hand side of Eq. (2a) to

define a function $\bar{g}(t)$ for all $t \geq x$. Equation (5) then indicates that $\bar{g}(t)$ vanishes for $t \geq 2l + x$. With the function $\bar{g}(t)$ so defined Eq. (12) may be split up to yield equations for $f(t)$ similar to Eq. (3a) and Eq. (3b) but with the range of integration in the terms containing $\bar{g}(y)$ from $y = x$ to $y = 2l - t$. The resulting sets of equations are then used to eliminate $f(t)$. In the resulting single integral equation for $\bar{g}(t)$, the properties of the scattering operators given by Eqs. (17) and (19) of Paper 1, are used to simplify the kernel. It can then be shown on using the result that $\bar{g}(t) = 0$ for $t \geq 2l + x$, that the integral equation reduces to a homogeneous Volterra integral equation for $2l - x \leq t \leq 2l + x$. It follows immediately that $\bar{g}(t) = 0$ in this range, and the resulting set of equations correspond to Eqs. (2a), (2b), (3a), and (3b).

APPENDIX

Let $u_-^i(x + t)$ be an arbitrary incident wave propagating in the direction of the negative x axis, such that $u_-^i(s) = 0$ for $s < 0$. This will generate a reflected wave $u_-^r(x - t)$ in the half-space $x \geq l$, such that $u_-^r(s) = 0$ for $s \geq 2l$, and

$$u_+^r(\eta) = \int_0^{2l-\eta} R_-(s + \eta)u_-^i(s)ds.$$

Let $u_+^i(x - t - 2l)$ be an incident wave propagating in the direction of the positive x axis, such that $u_+^i(s) = 0$, $s > 0$, and chosen so that it produces a transmitted wave $u_+^t(x - t - 2l)$ in the half-space $x \geq l$ which nullifies $u_-^r(x - t)$

$$u_-^r(x - t) + u_+^t(x - t - 2l) = 0 \quad x \geq l.$$

We want to express u_+^i in terms of u_-^i. This is achieved by noting that, $\eta \leq 0$,

$$\begin{aligned} u_+^i(\eta) &= e^{\gamma_-}\left(u_+^t(\eta) + \int_\eta^0 L_+(\eta - s)u_+^i(s)ds\right) \\ &= -e^{\gamma_-}\left(u_-^r(\eta + 2l) + \int_\eta^0 L_+(\eta - s)u_-^r(s + 2l)ds\right) \\ &= -e^{\gamma_-}\int_0^{-\eta} u_-^i(s)P_-(s + 2l + \eta)ds, \end{aligned}$$

where

$$P_-(\eta) = R_-(\eta) + \int_\eta^{2l} R_-(y)L_+(\eta - y)dy.$$

[1]V. H. Weston, J. Math. Phys. 13, 1952 (1972).
[2]V. H. Weston and R. J. Kreuger, J. Math. Phys. 14, 406 (1973).
[3]P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964).

Exact formulas for 2 × n arrays of dumbbells

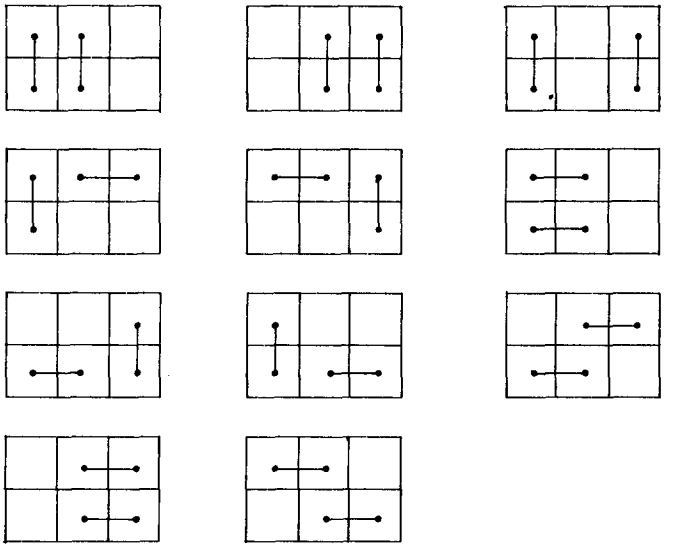
R. C. Grimson

Mathematics Department, Elon College, North Carolina
(Received 10 September 1973)

Several exact results are given for the problem of enumerating arrangements of q indistinguishable dumbbells on a $2 \times n$ array of compartments.

1. INTRODUCTION

McQuistan and Lichtman[1] have investigated the following dimer problem which has bearing on several areas of physics. Consider a $2 \times n$ rectangular array of compartments (a lattice space) and q dumbbell-shaped objects, •——•. Let $A(q, n)$ be the number of ways in which the q dumbbells may be placed in the array such that the two ends of each dumbbell are in two horizontally or vertically adjacent compartments and no two dumbbells have ends which share a compartment. For example example if $q = 2$ and $n = 3$, the possibilities are:



In this case $A(2, 3) = 11$. The following recurrence is known[1]:

$$A(q, n) = A(q, n-1) + 2A(q-1, n-1) + A(q-1, n-2) - A(q-3, n-3). \quad (1)$$

Clearly $A(q, n) = 0$ if $q > n$ so the array of numbers $A(q, n)$ is triangular, part of which is given by the following table:

n \ q	0	1	2	3	4	5
0	1					
1	1	1				
2	1	4	2			
3	1	7	11	3		
4	1	10	29	26	5	
5	1	13	56	94	56	8

McQuistan and Lichtman remark that exact solutions for problems of this sort (2 or more rows and/or 2 or more dimensions) have been obtained for only very special cases,[2,3] i.e., a 2-dimensional array completely covered with dumbbells. In many other investigations, approximation methods have been utilized.

In this paper we obtain some explicit formulas for $A(q, n)$, some generating functions, another representation of the problem, and another recurrence. Also we shall see how $A(q, n)$ is related to other well-known functions.

2. THE GENERATING FUNCTION AND ITS COEFFICIENTS

Put

$$f_n(x) = \sum_{q=0}^{n} A(q, n) x^q.$$

Then by (1) and a little manipulation,

$$f_{n+3}(x) = (2x + 1) f_{n+2}(x) + x f_{n+1}(x) - x^3 f_n(x) \quad (n \geq 0). \quad (2)$$

For example,

$$f_0(x) = 1,$$
$$f_1(x) = 1 + x,$$
$$f_2(x) = 1 + 4x + 2x^2$$
$$f_3(x) = 1 + 7x + 11x^2 + 3x^3,$$

which are verified by the above table. Next put

$$G(x, y) = \sum_{n=0}^{\infty} f_n(x) y^n. \quad (3)$$

Using (2) and the first 3 equations under (2) as initial conditions, we find

$$G(x, y) = (1 - xy)/(1 - 2xy - y - xy^2 + x^3 y^3). \quad (4)$$

Thus, (4) gives the ordinary bivariate generating function for $A(q, n)$. Incidentally, with $y = 1$, (4) becomes (15) of the McQuistan–Lichtman paper.[1] Clearly the usual methods of expanding (4) yield fairly complicated formulas. Of these, one of the more compact is, by the multinomial theorem,

$$A(q, n) = \sum_{i=0}^{1} \sum_{\substack{b+c+3d=q+i \\ a+2b+c+3d=n+i}} (-1)^{d+i} 2^c \binom{a+b+c+d}{a, b, c, d},$$

where $\binom{a+b+c+d}{a, b, c, d}$ is a multinomial coefficient.

Put

$$A = 1 - y,$$
$$B = -2y - y^2,$$
$$C = y^3,$$

so that, by (4),

$$G(x, y) = \frac{1 - xy}{1 - x} \left(1 + \frac{B}{A} x + \frac{C}{A} x^3\right)^{-1}. \quad (5)$$

Since A, B, and C are functions of y only, we may expand

the right side of (5) to obtain $G(x,y)$ as a power series in x whose coefficients involve y [compare with (3)]. This may be expressed by

$$G(x,y) = \frac{1-xy}{1-y}\sum_{k=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}\left(-\frac{B}{A}\right)^{k-j}\left(-\frac{C}{A}\right)^{j}x^{k+2j}. \quad (6)$$

If we carry through with the rather tedious details of expanding the right side of (6) in terms of y, thus getting a power series of x and y, we find that the coefficient of $x^q y^n$ is another (as to be expected) complicated formula. But this time we observe some interesting results. The coefficient is

$$A(q,n) = \sum_{k=0}^{q}\sum_{j=0}^{[q/3]}(-1)^j 2^{q-3j-k}\binom{q-2j}{j}\binom{q-3j}{k}\binom{n-2j-k}{q-2j}$$
$$-\sum_{k=0}^{q-1}\sum_{j=0}^{[(q-1)/3]}(-1)^j 2^{q-3j-k-1}\binom{q-2j-1}{j}\binom{q-3j-1}{k}\binom{n-2j-k-1}{q-2j-1},$$

where $[x]$ is the largest integer less than or equal to x. Thus, we may view $A(q,n)$ as a polynomial in n of degree q. The coefficient of n^q appears in the first double sum when $j = 0$. In this case ($j = 0$) the first double sum becomes

$$\sum_{k=0}^{q} 2^{q-k}\binom{q}{k}\frac{1}{q!}(n-k)_q.$$

But the coefficient of n^q in $(n-k)_q$ is 1 and

$$\sum_{k=0}^{q} 2^{q-k}\binom{q}{k}\frac{1}{q!} = \frac{3^q}{q!}.$$

Therefore, $A(q,n)$ is a polynomial in n as follows.

$$A(q,n) = \frac{3^q}{q!}n^q + C_1 n^{q-1} + \cdots + C_{q-1}n + C_q, \quad (7)$$

where the C's depend on q only. If we put $\Delta A(q,n) = A(q,n+1) - A(q,n)$, then (7) implies the recurrence

$$\Delta^q A(q,n) = 3^q.$$

3. ANOTHER FORMULA AND ASSOCIATED FUNCTIONS

The occupation of 2 × n arrays with dumbbells may be expressed in terms of an occupancy problem with restricted positions. Consider 3 sets of cells labeled as follows:

$$\underbrace{\overline{1,n+1}\ \overline{2,n+2}\cdots\overline{n,2n}}_{\text{I}} \mid \underbrace{\overline{1,2}\ \overline{2,3}\cdots\overline{n-1,n}}_{\text{II}}$$
$$\underbrace{\overline{n+1,n+2}\ \overline{n+2,n+3}\cdots\overline{2n-1,2n}}_{\text{III}} \quad (8)$$

Let the first n cells, I, represent the n vertical pairs of compartments of the 2 × n array, let the next set, II, of $n-1$ cells represent the $n-1$ horizontally adjacent pairs of compartments in the first row and let the final $n-1$ cells, III, be the horizontally adjacent pairs of compartments of the second row. Thus, (8) is equivalent to

1	2	3	4	. . .	n
$n+1$	$n+2$	$n+3$	$n+4$	. . .	$2n$

Each of the cells of (8) has 2 "labeling" numbers. Therefore, $A(q,n)$ is the number of ways we may distribute q like objects, one per cell, into (8) such that no 2 cells containing objects have a labeling number in common. The advent that 2 occupied cells do share a labeling number is equivalent to the ends of 2 dumbbells sharing a compartment, a situation which is forbidden.

Given that j of the q objects are in certain of the cells of I, there are u and v ($u + v = q - j$) objects that are to be distributed among the nonforbidden positions of II and III respectively. For example, if $n = 11, j = 3$, and the checks denote the cells of I occupied by the 3 objects, then the set of cells I, II, and III are

I $\quad$ — √ √ — — — √ — — — —
1 2 3 4 5 6 7 8 9 10 11
gap of 1 gap of 3 gap of 4

(deleting the second labeling number of each cell),

II $\quad$ x x x — — x x — — —
1,2 2,3 3,4 4,5 5,6 6,7 7,8 8,9 9,10 10,11,

III $\quad$ x x x — — x x — — —

(deleting the labeling numbers)

where the x's denote forbidden positions. Clearly II and III are always identical. If u objects occupy II then $q - j - u = v$ objects occupy III. We define a gap of m to be m successive unoccupied cells of I. Thus a gap of m gives rise to $m-1$ permissible cells in II and III. But of those $m-1$ permissible positions, no 2 adjacent ones may be occupied. The number of ways that we may place i objects in $m-1$ cells so that no 2 adjacent cells are occupied is

$$\binom{m-i}{i}.$$

(A well-known elementary fact asserts that the number of ways that r plus signs and s minus signs may be arranged in a row so that no 2 minus signs are adjacent is $\binom{r+1}{s}$; the above sentence is equivalent to this). If cell a_1 of I is occupied and if the next cell in I to the right of a_1 which is occupied is a_2 then the corresponding gap is $a_2 - a_1 - 1$. Allowing j to range from 0 to q and accounting for all possible distribution of the j objects in I we find

$$A(q,n) = \sum_{j=0}^{q}\sum_{\substack{a_0<a_1<\cdots<a_j<n+1 \\ u+v=q-j}} f(j,u)f(j,v), \quad (9)$$

where $a_0 = 0$ (the other a's are indices) and where

$$f(j,u) = f(u; a_0, a_1, \ldots, a_j) = \sum_{i_1+\cdots+i_{j+1}=u} \binom{a_1-a_0-1-i_1}{i_1}\binom{a_2-a_1-1-i_2}{i_2} \cdots \quad (10)$$

$$\binom{a_j-a_{j-1}-1-i_j}{i_j}\binom{n-a_j-i_{j+1}}{i_{j+1}}.$$

Thus (9) provides $A(q,n)$ with a more harmonious formula than do some of the earlier equations.

Further insight into the intricate nature of $A(q,n)$ may be made through a study of $f(j,u)$. Actually, some properties of $f(j,u)$ are well known. Putting

$$b_i = \begin{cases} a_i - a_{i-1} - 1 & (i = 1, 2, \ldots, j) \\ n - a_j & (i = j+1) \end{cases}$$

it suffices to define and examine

$$g(u; b_1, \ldots, b_r) = \sum_{i_1+\cdots+i_r=u} \binom{b_1-i_1}{i_1} \cdots \binom{b_r-i_r}{i_r}.$$

Then

$$\sum_{u=0}^{\infty} g(u; b_1, \ldots, b_r) z^u = \prod_{k=1}^{r} \sum_{i_k=0}^{b_k} \binom{b_k-i_k}{i_k} z^{i_k}. \quad (11)$$

But the function

$$u_b(z) = \sum_{i=0}^{b} \binom{b-i}{i} z^i \quad (12)$$

is familiar. The numbers $u_b(1)$ are the Fibonacci numbers. The polynomial $u_b(z)$ has been extensively studied by many investigators.[4] Two expressions for $u_b(z)$ are

$$u_b(z) = (-1)^b x^{b/2} U_b(i/2\sqrt{x}), \quad (*)$$

where $i = \sqrt{-1}$ and $U_b(z) = \sin(b+1)\theta/\sin\theta (z = \cos\theta)$; $U_b(z)$ is a Chebyshev polynomial.

$$u_b(z) = 2^{-b-1}\alpha^{-1}[(1+\alpha)^{b+1} - (1-\alpha)^{b+1}], \quad \alpha = (1+4x)^{1/2}. \quad (**)$$

[Compare (**) with the Binet form of the Fibonacci numbers.]

Using (11) and (12) it is easily seen that

$$\sum_{b_1,\ldots,b_r=0}^{\infty} \sum_{n=0}^{\infty} g(u; b_1, \ldots, b_r) z^u y_1^{b_1} \cdots y_r^{b_r} = \prod_{i=1}^{r} (1 - y_i - z y_i^2)^{-1}.$$

In a subsequent paper we shall show how the ideas presented in the latter part of this paper may be utilized and extended to enumerate arrangements of q dimers on an $m \times n$ lattice where it is not necessary to assume that the dimers are numerous enough to completely cover the lattice.

[1] R. B. McQuistan and S. J. Lichtman, J. Math. Phys. 11, 3095 (1970).
[2] M. E. Fisher, Phys. Rev. 124, 1664 (1961).
[3] P. W. Kasteleyn, Physica 27, 1209 (1961).
[4] J. Riordan, *Combinatorial Identities* (Wiley, New York, 1968), pp. 75, 76, 242.

Physical applications of multiplicative stochastic processes. II. Derivation of the Bloch equations for magnetic relaxation

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 29 August 1973)

The multiplicative stochastic process treatment of the time development of the density matrix for a subsystem in contact with a heat reservoir is applied to the specific problem of the relaxation of a nuclear magnetic moment which is interacting with a fluctuating magnetic environment. A model for the fluctuating interaction Hamiltonian, appropriate for the magnetic moment case, is presented, and the Bloch equations for nuclear magnetic relaxation are constructed as a consequence. Agreement with empirical observations is noted.

I. INTRODUCTION

Consider a subsystem in contact with a heat reservoir. The Hamiltonians for the subsystem and for the reservoir will be denoted by $\mathbf{H}_S$ and $\mathbf{H}_R$, respectively. It will be assumed that the state of the reservoir is given, on the average, by its equilibrium state throughout all time. Therefore, the interaction between the subsystem and the reservoir will be represented by a stationary, purely random, Gaussian interaction Hamiltonian, $\tilde{\mathbf{H}}(t)$.[1-4]

Latin indices will be used to denote eigenstates of the subsystem Hamiltonian:

$$\mathbf{H}_S|i\rangle = E_i|i\rangle. \tag{1}$$

Greek indices will be used to denote reservoir eigenstates:

$$\mathbf{H}_R|\alpha\rangle = E_\alpha|\alpha\rangle. \tag{2}$$

The identity matrices for the subsystem eigenstate manifold and for the reservoir eigenstate manifold are denoted by 1_S and 1_R, respectively. The total Hamiltonian which acts in the direct product manifold of the subsystem manifold and the reservoir manifold is given by

$$\mathbf{H}_{\text{total}} = \mathbf{H}_S \otimes 1_R + 1_S \otimes \mathbf{H}_R + \tilde{\mathbf{H}}(t). \tag{3}$$

Clearly, $\tilde{\mathbf{H}}(t)$ acts in the direct product manifold since it provides the interaction coupling. It has a matrix representation in the direct product manifold given by

$$\tilde{H}_{i\alpha j\beta}(t) \equiv \langle\alpha|\langle i|\tilde{\mathbf{H}}(t)|j\rangle\beta\rangle. \tag{4}$$

The Schrödinger wavefunction $\Psi(t)$ may be expanded in terms of direct product basis states giving

$$\Psi(t) = \sum_i \sum_\alpha C_{i\alpha}(t)|i\rangle|\alpha\rangle. \tag{5}$$

The density matrix for the complete system is defined by

$$\rho_{i\alpha j\beta}(t) \equiv C^*_{i\alpha}(t)C_{j\beta}(t). \tag{6}$$

Two physically motivated assumptions are imposed in order to arrive at a dynamical equation for the density matrix for the subsystem only. The assumptions are: (a) the interaction is energy conserving which means

$$\tilde{H}_{i\alpha j\beta}(t) = 0 \quad \text{unless } E_i + E_\alpha = E_j + E_\beta \tag{7}$$

and (2) the averaged total density matrix factors for all times into a product of a time-dependent averaged subsystem density matrix and a time-independent averaged reservoir density matrix[3]:

$$\langle\rho_{i\alpha j\beta}(t)\rangle \to \langle\rho_{ij}(t)\rangle\langle\rho_{\alpha\beta}\rangle, \tag{8}$$

where

$$\langle\rho_{\alpha\beta}\rangle = (1/Q_R)\exp(-E_\alpha/K_BT)\delta_{\alpha\beta}, \tag{9}$$

with

$$Q_R = \sum_\alpha \exp(-E_\alpha/K_BT), \tag{10}$$

where K_B is Boltzmann's constant and T is the temperature.

The resulting dynamical equation for the averaged subsystem density matrix is[3]

$$\frac{d}{dt}\langle\rho_{ij}(t)\rangle = -i(E_i - E_j)\langle\rho_{ij}(t)\rangle - T_{iji'j'}\langle\rho_{i'j'}(t)\rangle, \tag{11}$$

where $T_{iji'j'}$ is defined by[3]

$$T_{iji'j'} \equiv \sum_\alpha \sum_{\alpha'} R_{i\alpha j\alpha i'\alpha'j'\alpha'}(1/Q_R)\exp(-E_{\alpha'}/K_BT) \tag{12}$$

in which $R_{i\alpha j\beta i'\alpha'j'\beta'}$ is defined by[3]

$$R_{i\alpha j\beta i'\alpha'j'\beta'} \equiv \delta_{ii'}\delta_{\alpha\alpha'}\sum_\theta\sum_{\theta'} Q_{j\beta\theta\theta'\theta\theta'j'\beta'} + \delta_{jj'}\delta_{\beta\beta'}\sum_\theta\sum_{\theta'} Q_{\theta\theta'i\alpha i'\alpha'\theta\theta'} - 2Q_{j\beta j'\beta'i'\alpha'i\alpha}, \tag{13}$$

wherein $Q_{i\alpha j\beta i'\alpha'j'\beta'}$ is determined by the second moments of $\tilde{\mathbf{H}}(t)$ through the definition[3]

$$\langle\tilde{H}_{i\alpha j\beta}(t)\tilde{H}_{i'\alpha'j'\beta'}(s)\rangle = 2Q_{i\alpha j\beta i'\alpha'j'\beta'}\delta(t-s). \tag{14}$$

The assumption of energy conservation may be used to show

$$T_{iji'j'} = T^*_{i'j'ij}\exp[-(E_j + E_i - E_{j'} - E_{i'})/2K_BT] \tag{15}$$

which is a generalization of the detailed balancing condition for a thermally buffered system.

These formal considerations provide a context in which to present the stochastic description of the relaxation of a magnetic moment in an environment which produces a fluctuating magnetic field. In the next section a specific model for the interaction piece of the Hamiltonian will be introduced, and the Bloch equations will be constructed from the interaction.

II. DERIVATION OF BLOCH'S EQUATIONS

Denoting the x, y, and z components of the averaged magnetic moment by M_x, M_y, and M_z, the Bloch equations are given by

$$\frac{d}{dt}M_z(t) = -\frac{1}{T_1}(M_z(t) - M_z(\infty)), \tag{16}$$

$$\frac{d}{dt}M_x(t) = -\frac{1}{T_2}M_x(t) - (E_\uparrow - E_\downarrow)M_y(t), \tag{17}$$

$$\frac{d}{dt}M_y(t) = -\frac{1}{T_2}M_y(t) + (E_\uparrow - E_\downarrow)M_x(t), \tag{18}$$

in which T_1 is the longitudinal relaxation time and T_2 is the transverse relaxation time. $M_z(\infty)$ is the asymptotic equilibrium value of $M_z(t)$, and is nonzero if the whole system is at temperature T, and in the presence of a constant external magnetic field along the z axis. Under these conditions, $M_z(\infty)$ for a spin $\frac{1}{2}$ magnetic moment will be given by

$$M_z(\infty) = \frac{\exp(-E_\uparrow/K_BT) - \exp(-E_\downarrow/K_BT)}{\exp(-E_\uparrow/K_BT) + \exp(-E_\downarrow/K_BT)}, \tag{19}$$

where $E_\uparrow$ and $E_\downarrow$ are the energies for the magnetic moment when it is parallel and antiparallel to the external field.

Redfield has reviewed the problem of deriving the Bloch equations from a density matrix equation, which equation is often referred to as Redfield's equation.[5] The derivation of Redfield's equation utilizes 2nd-order perturbation theory for short time intervals, and makes minimal use of the stochastic properties of the interaction.[5] The derivation of the equation (11)[3] used in this paper makes no perturbation series approximations and depends critically upon the stochastic nature of the interaction. Kubo has emphasized the difficulty in obtaining the Boltzmann factors in $M_z(\infty)$ when a stochastic approach to this problem is attempted.[6] In this present paper it is seen how the Boltzmann factors appear naturally through the factorization of the total density matrix for a subsystem and a reservoir.[3]

The structure of the interaction Hamiltonian to be used here depends upon the following considerations:

(1) The magnetic moment will correspond to a spin $\frac{1}{2}$ object.
(2) The effect of the reservoir will be manifested by an effective spin $\frac{1}{2}$ reservoir magnetic moment.
(3) The interaction will have the same form as in the Heisenberg ferromagnetic interaction, but with fluctuating coupling constants.
(4) Energy will be conserved by the interaction.

Both the subsystem and the reservoir correspond with two state manifolds, and consequently Pauli spin matrices may be used to represent both the magnetic moment and the effective reservoir magnetic moment. The interaction Hamiltonian is, then

$$\tilde{H}_{i\alpha j\beta}(t) = [\sigma^x_{ij}\sigma^x_{\alpha\beta}\tilde{h}^x(t) + \sigma^y_{ij}\sigma^y_{\alpha\beta}\tilde{h}^y(t)]\delta_{i\beta}\delta_{j\alpha} + \sigma^z_{ij}\sigma^z_{\alpha\beta}\tilde{h}^z_{i\alpha}(t), \tag{20}$$

wherein the Pauli spin matrices are

$$\sigma^x_{ij}, \sigma^x_{\alpha\beta} \to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y_{ij}, \sigma^y_{\alpha\beta} \to \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z_{ij}, \sigma^z_{\alpha\beta} \to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{21}$$

The factor $\delta_{i\beta}\delta_{j\alpha}$ in (20) provides for energy conservation during the transition, and it appears with the x and y components only because the z component terms automatically conserve energy. Both $\tilde{h}^x(t)$ and $\tilde{h}^y(t)$ represent fluctuating coupling coefficients. They factor completely in the x and y terms because the interaction Hamiltonian is Hermitian. The z component, however, contains different coupling coefficients for each possible choice of indices and is, therefore, given with appropriate indices. The Hermiticity of the z component terms is guaranteed by its form. It is assumed that all six coupling coefficients are purely random, stationary, Gaussian stochastic processes with average values of zero. Furthermore, no cross correlations are assumed to exist, and the x and y fluctuations have identical second moments. These conditions are expressed by

$$\langle\tilde{h}^x(t)\rangle = \langle\tilde{h}^y(t)\rangle = \langle\tilde{h}^z_{i\alpha}(t)\rangle = 0 \quad \text{for } i = 1, 2 \text{ and } \alpha = 1, 2, \tag{22}$$

$$\langle\tilde{h}^x(t)\tilde{h}^y(S)\rangle = \langle\tilde{h}^x(t)\tilde{h}^z_{i\alpha}(S)\rangle = \langle\tilde{h}^y(t)\tilde{h}^z_{i\alpha}(S)\rangle = \langle\tilde{h}^z_{i\alpha}(t)\tilde{h}^z_{j\beta}(S)\rangle = 0, \tag{23}$$

for $i = 1, 2$ $j = 1, 2$ $\alpha = 1, 2$ and $\beta = 1, 2$ but not both $i = j$ and $\alpha = \beta$,

$$\langle\tilde{h}^x(t)\tilde{h}^x(S)\rangle = \langle\tilde{h}^y(t)\tilde{h}^y(S)\rangle = 2Q^{x,y}\delta(t - S)$$
and
$$\langle\tilde{h}^z_{i\alpha}(t)\tilde{h}^z_{i\alpha}(S)\rangle = 2Q^z\delta(t - S). \tag{24}$$

In a constant external magnetic field directed along the z axis, the magnetic moment takes on two energy values for its spin parallel or antiparallel with the external field. These energies are given by

$$E_\uparrow = +g\beta H \quad \text{and} \quad E_\downarrow = -g\beta H, \tag{25}$$

where g is the nuclear g factor, β is the nuclear magneton, and H is the external field strength, for the case of a nuclear magnetic moment such as a proton. Equations (25) and (20) lead to a special instance of (11) with the requirement that $T_{iji'j'}$ be determined from (20) using (22), (23), and (24).

In order to get an expression for $T_{iji'j'}$ Eqs. (12), (13), and (14) will be used, in reverse order. An outline of this computation follows.

Equations (14), (20), (23), and (24) lead to

$$Q_{i\alpha j\beta i'\alpha' j'\beta'} = (\sigma^x_{ij}\sigma^x_{\alpha\beta}\sigma^x_{i'j'}\sigma^x_{\alpha'\beta'} + \sigma^y_{ij}\sigma^y_{\alpha\beta}\sigma^y_{i'j'}\sigma^y_{\alpha'\beta'}) \times Q^{x,y}\delta_{i\beta}\delta_{j\alpha}\delta_{i'\beta'}\delta_{j'\alpha'} + \sigma^z_{ij}\sigma^z_{\alpha\beta}\sigma^z_{i'j'}\sigma^z_{\alpha'\beta'}Q^z\delta_{ii'}\delta_{\alpha\alpha'}. \tag{26}$$

In order to calculate $R_{i\alpha j\beta i'\alpha' j'\beta'}$ according to (13), some special cases of (26) are needed.

$$\sum_\theta\sum_{\theta'} Q_{j\beta\theta\theta'\theta\theta' j'\beta'} = \sum_\theta\sum_{\theta'}(\sigma^x_{j\theta}\sigma^x_{\beta\theta'}\sigma^x_{\theta j'}\sigma^x_{\theta'\beta'} + \sigma^y_{j\theta}\sigma^y_{\beta\theta'}\sigma^y_{\theta j'}\sigma^y_{\theta'\beta'}) \times Q^{x,y}\delta_{j\theta'}\delta_{\theta\beta}\delta_{\theta\beta}\delta_{j'\theta'} + \sigma^z_{j\theta}\sigma^z_{\beta\theta'}\sigma^x_{\theta j'}\sigma^z_{\theta'\beta'}Q^z\delta_{j\theta}\delta_{\beta\theta'}$$
$$= \delta_{jj'}\delta_{\beta\beta'}(\sigma^x_{j\beta}\sigma^x_{\beta j}\sigma^x_{\beta j}\sigma^x_{j\beta} + \sigma^y_{j\beta}\sigma^y_{\beta j}\sigma^y_{\beta j}\sigma^y_{j\beta})Q^{x,y} + \sigma^z_{jj}\sigma^z_{\beta\beta}\sigma^z_{jj}\sigma^z_{\beta\beta}Q^z, \tag{27}$$

$$\sum_\theta\sum_{\theta'} Q_{\theta\theta' i\alpha i'\alpha'\theta\theta'} = \delta_{ii'}\delta_{\alpha\alpha'}(\sigma^x_{\alpha i}\sigma^x_{i\alpha}\sigma^x_{i\alpha}\sigma^x_{\alpha i} + \sigma^y_{\alpha i}\sigma^y_{i\alpha}\sigma^y_{i\alpha}\sigma^y_{\alpha i})Q^{x,y} + \sigma^z_{i'i}\sigma^z_{\alpha'\alpha}\sigma^z_{i'i'}\sigma^z_{\alpha'\alpha'}Q^z. \tag{28}$$

Therefore, it follows that

$$R_{i\alpha j\beta i'\alpha' j'\beta'} = \delta_{ii'}\delta_{\alpha\alpha'}\delta_{jj'}\delta_{\beta\beta'}(\sigma^x_{j\beta}\sigma^x_{\beta j}\sigma^x_{\beta j}\sigma^x_{j\beta} + \sigma^y_{j\beta}\sigma^y_{\beta j}\sigma^y_{\beta j}\sigma^y_{j\beta} + \sigma^x_{\alpha i}\sigma^x_{i\alpha}\sigma^x_{i\alpha}\sigma^x_{\alpha i} + \sigma^y_{\alpha i}\sigma^y_{i\alpha}\sigma^y_{i\alpha}\sigma^y_{\alpha i})Q^{x,y}$$
$$+ (\delta_{ii'}\delta_{\alpha\alpha'}\sigma^z_{jj}\sigma^z_{\beta\beta}\sigma^z_{jj}\sigma^z_{\beta\beta'} + \delta_{jj'}\delta_{\beta\beta'}\sigma^z_{i'i}\sigma^z_{\alpha'\alpha}\sigma^z_{i'i'}\sigma^z_{\alpha'\alpha'})Q^z$$
$$- 2(\sigma^x_{jj'}\sigma^x_{\beta\beta'}\sigma^x_{i'i}\sigma^x_{\alpha'\alpha} + \sigma^y_{jj'}\sigma^y_{\beta\beta'}\sigma^y_{i'i}\sigma^y_{\alpha'\alpha}) \times Q^{x,y}\delta_{j\beta'}\delta_{\beta j}\delta_{i'\alpha}\delta_{\alpha' i} - 2\sigma^z_{jj'}\sigma^z_{\beta\beta'}\sigma^z_{i'i}\sigma^z_{\alpha'\alpha}Q^z\delta_{ii'}\delta_{\beta\alpha'}. \tag{29}$$

To get $T_{iji'j'}$ according to (12) requires

$$\sum_\alpha R_{i\alpha j\alpha i'\alpha' j'\beta'} = \delta_{ii}\delta_{jj'}\delta_{\alpha'\beta}(\sigma^x_{j\alpha}\sigma^x_{\alpha' j}\sigma^x_{\alpha' j}\sigma^x_{j\alpha'} + \sigma^y_{j\alpha'}\sigma^y_{\alpha' j}\sigma^y_{\alpha' j}\sigma^y_{j\alpha'} + \sigma^x_{\alpha' i}\sigma^x_{i\alpha'}\sigma^x_{i\alpha'}\sigma^x_{\alpha' i} + \sigma^y_{\alpha' i}\sigma^y_{i\alpha'}\sigma^y_{i\alpha'}\sigma^y_{\alpha' i})Q^{x,y}$$
$$+ (\delta_{ii'}\sigma^z_{jj}\sigma^z_{\alpha'\alpha'}\sigma^z_{jj}\sigma^z_{\alpha'\beta'} + \delta_{jj'}\sigma^z_{i'i}\sigma^z_{\alpha'\beta}\sigma^z_{i'i'}\sigma^z_{\alpha'\alpha'})Q^z$$
$$- 2(\sigma^x_{jj'}\sigma^x_{i'\beta}\sigma^x_{i'i}\sigma^x_{\alpha' i'} + \sigma^y_{jj'}\sigma^y_{i'\beta}\sigma^y_{i'i}\sigma^y_{\alpha' i})Q^{x,y}\delta_{i'j}\delta_{i\alpha'}\delta_{j\beta'}$$
$$- 2\sigma^z_{jj'}\sigma^z_{\alpha'\beta}\sigma^z_{i'i}\sigma^z_{\alpha'\alpha'}Q^z\delta_{ij}. \tag{30}$$

Therefore,

$$T_{ij i'j'} = \sum_{\alpha}\sum_{\alpha'} R_{i\alpha j\alpha i'\alpha' j'\alpha'} \frac{1}{Q_R} \exp\left(-\frac{E_{\alpha'}}{K_B T}\right)$$
$$= \delta_{ii'}\delta_{jj'}\left\{2\left[1 - \frac{1}{Q_R}\exp\left(-\frac{E_j}{K_B T}\right)\right]\right.$$
$$\left. + 2\left[1 - \frac{1}{Q_R}\exp\left(-\frac{E_i}{K_B T}\right)\right]\right\} Q^{x,y}$$
$$+ \delta_{ii'}\delta_{jj'} 2Q^z - 2\delta_{i'j'}\delta_{ij} 2(1 - \delta_{jj'})Q^{x,y} \frac{1}{Q_R}$$
$$\times \exp\left(-\frac{E_i}{K_B T}\right) - 2\delta_{ij}\delta_{i'j'}\delta_{ii'} Q^z, \tag{31}$$

wherein (21) was used as well as $(1/Q_R)\exp(-E_1/K_BT) + (1/Q_R)\exp(-E_2/K_BT) = 1$. Several special cases of (31) will be used and are given by

$$T_{ijij} = 2Q^{x,y} + 2Q^z \quad \text{for } i \neq j, \tag{32}$$

$$T_{iiii} = 4Q^{x,y}[1 - (1/Q_R)\exp(-E_i/K_BT)], \tag{33}$$

$$T_{iii'i'} = -4Q^{x,y}(1/Q_R)\exp(-E_i/K_BT) \quad \text{for } i \neq i', \tag{34}$$

$$T_{ijji} = 0. \tag{35}$$

All other choices of indices for $T_{ij i'j'}$ lead to zero values as with (35). This situation greatly simplifies Eq. (11).

The x, y, and z components of the magnetic moment are given, on the average, in terms of the averaged density matrix, $\langle \rho_{ij}(t)\rangle$, by

$$M_x(t) = g\beta \sum_{ij} \sigma^x_{ij}\langle \rho_{ij}(t)\rangle, \tag{36}$$

$$M_y(t) = g\beta \sum_{ij} \sigma^y_{ij}\langle \rho_{ij}(t)\rangle, \tag{37}$$

$$M_z(t) = g\beta \sum_{ij} \sigma^z_{ij}\langle \rho_{ij}(t)\rangle. \tag{38}$$

Using (21) gives

$$M_x(t) = g\beta(\langle \rho_{12}(t)\rangle + \langle \rho_{21}(t)\rangle), \tag{39}$$

$$M_y(t) = g\beta i(\langle \rho_{21}(t)\rangle - \langle \rho_{12}(t)\rangle), \tag{40}$$

$$M_z(t) = g\beta(\langle \rho_{11}(t)\rangle - \langle \rho_{22}(t)\rangle). \tag{41}$$

From (11) and (32)–(35) it follows that

$$\frac{d}{dt}\langle \rho_{12}(t)\rangle = -i(E_\downarrow - E_\uparrow)\langle \rho_{12}(t)\rangle - T_{1212}\langle \rho_{12}(t)\rangle, \tag{42}$$

$$\frac{d}{dt}\langle \rho_{21}(t)\rangle = -i(E_\uparrow - E_\downarrow)\langle \rho_{21}(t)\rangle - T_{2121}\langle \rho_{21}(t)\rangle, \tag{43}$$

$$\frac{d}{dt}\langle \rho_{11}(t)\rangle = -T_{1111}\langle \rho_{11}(t)\rangle - T_{1122}\langle \rho_{22}(t)\rangle, \tag{44}$$

$$\frac{d}{dt}\langle \rho_{22}(t)\rangle = -T_{2211}\langle \rho_{11}(t)\rangle - T_{2222}\langle \rho_{22}(t)\rangle. \tag{45}$$

The combination of (32)–(35), (39)–(41), and (42)–(45) gives

$$\frac{d}{dt}M_x(t) = (E_\downarrow - E_\uparrow)M_y(t) - 2(Q^{x,y} + Q^z)M_x(t), \tag{46}$$

$$\frac{d}{dt}M_y(t) = -(E_\downarrow - E_\uparrow)M_x(t) - 2(Q^{x,y} + Q^z)M_y(t), \tag{47}$$

$$\frac{d}{dt}M_z(t) = -4Q^{x,y}(M_z(t) - M_z(\infty)), \tag{48}$$

where

$$M_z(\infty) \equiv \frac{1}{Q_R}\left[\exp\left(-\frac{E_\uparrow}{K_BT}\right) - \exp\left(-\frac{E_\downarrow}{K_BT}\right)\right]. \tag{49}$$

Equations (46)–(49) are clearly equivalent with Eqs. (16)–(19) when the identifications

$$1/T_1 \equiv 4Q^{x,y} \quad \text{and} \quad 1/T_2 \equiv 2(Q^{x,y} + Q^z) \tag{50}$$

are made.

It should be noted that when there is no external, constant magnetic field present, then $E_\uparrow = E_\downarrow$, and it would be expected in isotropic media that the second moments of the fluctuating magnetic field produced by the reservoir would be isotropic, which means

$$Q^z = Q^{x,y}. \tag{51}$$

Therefore, (50) shows that $T_1 = T_2$ and the relaxation is isotropic. In the presence of an external, constant field, or in the case of anisotropic media Q^z and $Q^{x,y}$ will not necessarily remain equal, and both $Q^z > Q^{x,y}$ and $Q^{x,y} > Q^z$ lead to the special cases $T_1 > T_2$ and $T_1 \approx \frac{1}{2}T_2$, respectively. These special cases are observed experimentally.[7]

III. CONCLUSION

The theory of multiplicative stochastic processes has been used to arrive at a density matrix description of a subsystem in contact with a thermal reservoir. In this paper an application to the problem of the relaxation of a magnetic moment interacting with the fluctuating magnetic environment of a reservoir has been made. Kubo's dilemma concerning how to properly include the Boltzmann factors has been resolved. Moreover, the derivation of Bloch's equations did not require the approximation of 2nd-order perturbation theory for short time intervals because the stochastic properties of the fluctuating quantities were fully utilized.

The quantitative determination of Q^z and $Q^{x,y}$ in terms of the exact microscopic interactions involved remains as a problem. This problem may be raised in the more general setting of Sec. I with respect to quantitatively computing $Q_{i\alpha j\beta i'\alpha' j'\beta'}$ in Eq. (14) from a microscopic theory of the exact interactions. This problem will be treated elsewhere.

ACKNOWLEDGMENT

This work was partially supported by NSF Grant GP39309.

[1] R. F. Fox, J. Math. Phys. 13, 1196 (1972).
[2] R. F. Fox, J. Math. Phys. 13, 1726 (1972).
[3] R. F. Fox, J. Math. Phys. 14, 20 (1973)
[4] R. F Fox, J. Math. Phys. 14, 1187 (1973).
[5] A. Redfield, *Advances in Magnetic Resonance* (Academic, New York, 1965), Vol. 1 pp. 1–32.
[6] R. Kubo and N. Hashitsume, Prog. Theor. Phys. Suppl. 46, 210 (1970).
[7] A. Redfield (private communication, 1973).

On normalization problems of the path integral method*

Gregory S. Um

Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11790
(Received 23 August 1973)

Ambiguities of the path and of normalization in Feynman's path integral method are discussed. The investigation shows that the Feynman path integral method possesses inherent ambiguities, which can be resolved by a prescription which agrees completely with the Schrödinger equation.

INTRODUCTION

Recently, many[1] authors have attempted to derive quantum mechanical equations from the path integral method. One difficulty was that one does not get unique equations. It was claimed that the ambiguity resided in the choice of path and the *approximation* (of the short time action) used. In the next section, we point out that this ambiguity is the logical consequence of the Feynman theory and that the ambiguity of normalization apparently unnoticed till now, is equivalent to the ambiguity of path, in the sense that every ambiguity in the resulting equations as a consequence of the path ambiguity—can also be obtained by an alterations in the normalization.

Pauli and later DeWitt[2] approached the Feynman theory in a different way. They first started from the properties of the propagator function K and together with the pre-determined equation, they tried to obtain an explicit expression of K for small time intervals. It worked fine for simple cases, but when DeWitt investigated the most general Lagrangian, the asymptotic form of K did not quite satisfy the equation obtained from the canonical quantization procedure.

Cheng[3] has obtained an equation using the classical action where there is no ambiguity of path. The result was the same as that of DeWitt except a coefficient of same term has the factor $\frac{1}{6}$ instead of $\frac{1}{12}$.

This paper is aimed to resolve all the above difficulties. The organization of the paper is as follows. In Sec. 1, we will study the basic Feynman formalism and the inherent ambiguities. In Sec. 2, we will follow Pauli and find a new function K_0 which will satisfy Schrödinger equation. In the brief Sec. 3, we will show the normalization obtained from Sec. 2 will cancel Cheng's extra term, thereby obtaining Schrödinger equation.

1. PRELIMINARIES

The basic definition of the propagator function K is

$$K = \int_{x'}^{x''} \exp[(i/\hbar)\int_{t'}^{t''} L(\dot{x}, x, t)dt]\,Dx, \tag{1.1}$$

where $\int Dx$ is the sum over *all* the paths allowed by quantum mechanics. The condition is that the paths should go in one direction[4] in time only. This can be incorporated into (1.1) by the lattice calculation, i.e., by dividing up the time interval into infinitesimal ones and for each time vary the positions along the position axis. In the relativistic case, the condition is presumably different and causality may have to be incorporated as shown by Feynman.[5]

In the lattice calculation, we write (see Appendix C),

$$K(x'', t'': x', t') = \lim_{\epsilon\to 0} \int\cdots\int \frac{\exp[(i/\hbar)S_0(x'', t'': x^{\alpha-1}, t^{\alpha-1})]}{A(t'', x'': t^{\alpha-1}, x^{\alpha-1})} \times \cdots \frac{\exp[(i/\hbar)S_0(x^1, t^1: x', t')]}{A(t^1, x^1: t', x')}\, d^N x^{\alpha-1}\cdots d^N x^1, \tag{1.2}$$

where $d^N x^i$ is the volume element in the generalized coordinate space $x^i = (x_1^i, \ldots, x_N^i)$, $\epsilon = t'' - t'/\alpha$, and A is the normalization function. (The meaning of S_0 will be explained below). In order that (1.2) is valid, we must have the limiting condition

$$\lim_{\epsilon\to 0} \int_{x'}^{x''} \exp\left(\frac{i}{\hbar}\int_{t'}^{t''} L dt\right) Dx = \lim_{\epsilon\to 0} \frac{\exp[(i/\hbar)S_0(x'', t'': x', t')]}{A(t'', x'': t', x')}, \tag{1.3}$$

where S_0 means the action evaluated by a path or an approximation. Each time in the integrand of (1.2) must be the limit of an exact K for sum path and sum S. This is the center of (1.3). Therefore, if the left-hand side of (1.3) is definitely known, for different S_0, A has to be adjusted, and vice versa. Indeed as shown in the Appendix, fixing the normalization and varying S_0 is equivalent to fixing a path and varying the normalization.[6] If we consider the classical path, S_0 becomes $S_{\rm cl}$, the classical action, and we have

$$K(x'', t'': x', t') = \frac{1}{F(t'', x'': t', x')} \exp[(i/\hbar)S_{\rm cl}(x'', t'': x', t') \tag{1.4}$$

for some F. Therefore, for classical path,

$$A(t''x'': t'x') \underset{\epsilon\to 0}{\leftarrow} F(t'', x''; t'x').$$

Now let us examine

$$\psi(x'', t'') = \int K(x'', t'': x', t')\psi(x', t')d^N x'. \tag{1.5}$$

If K is a predetermined function, then (1.5) will yield a unique set of wavefunctions for given boundary conditions, i.e., it will yield unique quantum mechanical equations. However, K is only partially determined up to this point because one wishes to get A of (1.3) from (1.5). It is unjustified to expect both to get A and unique equations. If the equation is written as

$$i\hbar\frac{\partial}{\partial t}\psi = f_1(x, t)\frac{\partial^2}{\partial x^2}\psi + f_2(x, t)\frac{\partial}{\partial x}\psi + f_3\psi \tag{1.6}$$

then, Appendix A shows that the Lagrangian will determine f_1 and f_2 uniquely but not f_3.[6] Therefore, additional information is needed to determine A. The next logical step, then, is to assume an additional information that the equation obtained from the commutation relations are valid. However, we shall see that even here A is not altogether unique.

2. PAULI APPROACH

We consider the most general classical Lagrangian

$$L = \tfrac{1}{2}g_{ij}\dot{x}^i\dot{x}^j + a_i\dot{x}^i + v, \tag{2.1}$$

where g_{ij}, a_i, v are functions of x and t. The corresponding Hamiltonian is

$$H = \tfrac{1}{2}g^{-1/4}(p_i - a_i)g^{1/2}g^{ij}(p_j - a_j)g^{-1/4} + v \qquad (2.2)$$

where $g = \det(g_{ij})$ and $g^{ik}g_{kj} = \delta^i{}_j$. Equation (2.2) can be directly translated into quantum mechanical equation if we use

$$g^{1/4}p_i g^{-1/4} = \frac{\hbar}{i}\frac{\partial}{\partial x^i}, \qquad (2.3)$$

which is obtained from commutation relations of p_i and x_i.

The result of the canonical quantization is

$$\left(i\hbar\frac{\partial}{\partial t} - H\right)\psi \equiv 0_{x,t}\psi \equiv i\hbar\frac{\partial}{\partial t}\psi + \tfrac{1}{2}\hbar^2 g^{-1/2}\frac{\partial}{\partial x^i}\left(g^{1/2}g^{ij}\frac{\partial}{\partial x^j}\psi\right) - i\hbar a^i\frac{\partial}{\partial x^i}\psi - \tfrac{1}{2}i\hbar g^{-1/2}\left(\frac{\partial}{\partial x^i}g^{1/2}a^i\right)\psi - \tfrac{1}{2}a^i a_i\psi - v\psi = 0, \qquad (2.4)$$

where $a^i = g^{ij}a_j$. Following Pauli's approach starting from the Hamilton-Jacobi equation, we can construct an explicit expression of the classical action of (2.1). The result is

$$S_{\mathrm{cl}}(x'', t'': x', t') = \frac{1}{\epsilon}\left[\tfrac{1}{2}g'_{ij}\Delta x^i\Delta x^j + \tfrac{1}{12}B'_{ijk}\Delta x^i\Delta x^j\Delta x^k + \tfrac{1}{72}C'_{ijkl}\Delta x^i\Delta x^j\Delta x^l\Delta x^k + O(\Delta^5 x)\right] + a'_i\Delta x^i + \tfrac{1}{4}\left(a'_{i,j} + a'_{j,i} + \frac{\partial g'_{ij}}{\partial t'}\right)\Delta x^i\Delta x^j - (a'^i a'_i + v')\epsilon + O(\Delta^3 x, \epsilon^2) + O(\epsilon\Delta x)\cdots \qquad (2.5)$$

where

$$\Delta x^i = x''^i - x'^i,$$
$$B'_{ijk} = g'_{ij,k} + g'_{jk,i} + g'_{ki,j},$$
$$C'_{ijkl} = g'_{ij,kl} + g'_{kl,ij} + g'_{ik,lj} + g'_{il,jk} + g'_{jk,il} - g'^{mn}([ij,m]'[kl,n]' + [ik,m]'[lj,n]' + [il,m]'[jk,n]'),$$

and

$$[ij,k] = \tfrac{1}{2}(g_{ij,k} + g_{ik,j} - g_{jk,i}). \qquad (2.6)$$

Pauli and DeWitt consider

$$K_c(x'', t'': x', t') = [(2\pi i\hbar)^{N/2}]^{-1}g''^{-1/2}D^{1/2}(x'', t'': x', t')g'^{-1/4} \times \exp[(i/\hbar)S_{\mathrm{cl}}(x'', t'': x', t')], \qquad (2.7)$$

where $D = \det(-\partial^2 S_{\mathrm{cl}}/\partial x''^i\partial x'^j)$. From (2.5) we can get the explicit expression for K_C,

$$K_c(x'', t'': x', t') = [(2\pi i\hbar\epsilon)^{N/2}]^{-1}(1 + \tfrac{1}{12}R'_{ij}\Delta x^i\Delta x^j + \text{higher orders}) \times \exp[(i/\hbar)S_{\mathrm{cl}}], \qquad (2.8)$$

where

$$R'_{ij} = -g'^{kl}R'_{ikjl},$$

and

$$R'_{ikjl} = \tfrac{1}{2}(g'_{ij,kl} - g'_{il,kj} - g'_{kj,il} + g'_{kl,ij}) + g'^{mn}([ij,m]'[kl,n]' - [kj,m]'[il,n]').$$

If we operate K_c by $O_{x'',t''}$, we get

$$O_{x'',t''}K_c(x'', t'': x', t') = [\tfrac{1}{12}\hbar^2 R' + O(\Delta x) + O(\epsilon)]K_c, \qquad (2.9)$$

where $R' = g'^{ij}R'_{ij}$.

The rest of the properties of K_c are (also of K):

1. $K_c^*(x'', t'': x', t') = K_c(x', t': x'', t'')$ (2.10)

which is demanded by the Hermiticity of H and property 2,

2. $K_c(x'', t'': x', t') \to \delta(x'' - x', t')$ as $\epsilon \to 0$ (2.11)

where $\delta(x'' - x', t')$ is the delta function in the curvilinear space.

The $\frac{1}{12}R'\hbar^2$ term of (2.9) is incompatible with the requirements (1.6) and (2.11). To remedy this, DeWitt has formed a new Hamiltonian which has the additional $\frac{1}{12}R'\hbar^2$ to the original H of (2.4), thus removing the unsatisfactory features. This is a rather arbitrary step, and in the spirit of the foregoing section, it is more natural to modify K_c. To give an example, we consider

$$K_0(x'', t'': x', t') = (1 + \tfrac{1}{12}i\hbar R'\epsilon + O(\epsilon^2))K_c \qquad (2.12)$$

Then, since the new factor only depends on x' and ϵ,

$$O_{x'',t''}K_0 = \left(i\hbar\frac{\partial}{\partial t''}\left(\frac{i\hbar R'\epsilon}{12}\right) + O(\epsilon) + O_{x'',t''}\right)K_c \times (1 + \tfrac{1}{12}i\hbar R'\epsilon) = (-\tfrac{1}{12}\hbar^2 R' + \tfrac{1}{12}\hbar^2 R' + O(\Delta x) + O(\epsilon))K_0, \qquad (2.13)$$

which satisfies (2.4). However, (2.12) is not the only possible one. We write

$$K(x'', t'': x', t') = [(2\pi i\hbar\epsilon)^{N/2-1}(1 + i\hbar\epsilon A' + E'_{ij}\Delta x^i\Delta x^j + \text{higher order}) \times \exp[(i/\hbar)S_{\mathrm{cl}}(x'', t'': x', t')], \qquad (2.14)$$

where A and E_{ij} involve derivatives of g_{ij} so that K will reduce to the well known form when g_{ij}'s are constant. The Appendix shows that if $-A' - g'^{ij}E'_{ij} + R'/6 = 0$, then K of (2.14) satisfies (2.4).

Note that, (2.14) also satisfies (2.10) and (2.11)[7] up to order ϵ and $\Delta^2 x$. The higher-order terms are not necessary if we use the lattice calculation. However, only one of (2.16) satisfies the group property up to *order* ϵ, and therefore, only one is the asymptotic K up to order ϵ.

3. SHORT-TIME CALCULATION

In his calculation, Cheng did not recognize the ambiguity of normalization and he considers

$$K = [(2\pi i\hbar\epsilon)^{N/2}]^{-1}\exp[(i/\hbar)S_{\mathrm{cl}}(x'', t'': x', t')] \qquad (3.1)$$

for small ϵ. From the arguments so far given, it is clear why we will not get Eq. (2.4)

Suppose we do the same calculation using K of (2.14). To make it easier, we first write K as

$$K(x'', t'': x', t') = [(2\pi i\hbar\epsilon)^{N/2}]^{-1}(1 + i\hbar A''\epsilon)(1 + E''_{ij}\Delta x^i\Delta x^j)\exp[(i/\hbar)S_{\mathrm{cl}}]. \qquad (3.2)$$

The difference between (3.2) and (2.14) is the order of $\Delta^3 x$ and therefore it will not contribute here.

Then we have

$$\psi(x'', t') + \epsilon\frac{\partial\psi}{\partial t'} = \frac{(1 + i\hbar A''\epsilon)}{(2\pi i\hbar\epsilon)^{N/2}}\int\exp\left(\frac{i}{2\hbar\epsilon}\Delta x^i\Delta x^j g''_{ij}\right) \times (1 + E''_{ij}\Delta x^i\Delta x^j)\left(\psi(x'', t') - \Delta x^i\frac{\partial\psi}{\partial x''^i} + O(\Delta^2 x)\right) \times \left(\sqrt{g''} - \Delta x^i\frac{\partial\sqrt{g''}}{\partial x''^i} + O(\Delta^2 x)\right)$$

$$\times(\text{the expansion of the rest of the action})\, d\Delta x' \ldots d\Delta x^N$$
$$= (1 + i\hbar A''\epsilon)(\psi(x'', t') + (\epsilon/i\hbar)H_{x''}\psi(x'', t')$$
$$- (i\hbar\epsilon/6)R''\psi + i\hbar\epsilon E''\psi + O(\epsilon^2)), \tag{3.3}$$

where $E'' = E''_{ij}g''^{ij}$.

Again, if $A'' + E'' = R''/6$, then (3.3) reduces to (2.4).

CONCLUSION

If the integral of (1.1) can be evaluated directly, then, Feynman's formalism gives a unique K and a unique equation. However, at the present moment unambiguous calculational procedure which is really what would define the integral does not exist. The "lattice method" employed by others and also in this paper possesses certain inherent ambiguities for example the Feynman formalism leaves the scalar operator term of the equation ambiguous. Even if we fix the scalar term, the normalization is not unique. Note that the same phenomenon occurs even in a flat space, but all the different normalizations add up to the same K after the lattice calculation.

ACKNOWLEDGMENT

The author is grateful to Professor M. Dresden for his warm guidance. Moreover, his deep human understanding has been much appreciated.

APPENDIX A

We will use the Lagrangian of (2.1), except that we will consider only one dimension to simplify the matters. The Lagrangian is, then,

$$L(\dot{x}, x, t) = g(x, t)\dot{x}^2 + a(x, t)\dot{x} + v(x, t). \tag{A1}$$

We take an arbitrary path[6] $x(t)$ such that $x(t'') = x''$ and $x(t') = x'$. To simplify the notation, we define

$$\dot{x}'' = \frac{dx}{dt}\Big|_{t''}, \qquad \ddot{x}'' = \frac{d^2x}{dt^2}\Big|_{t''}, \qquad \text{etc.}$$

From $x' = x'' - \dot{x}''\epsilon + \frac{1}{2}\ddot{x}''\epsilon^2 + \cdots$, we see that

$$\dot{x}'' = \frac{\Delta x}{\epsilon} + \alpha''\frac{\Delta^2 x}{\epsilon} + \beta''\frac{\Delta^3 x}{\epsilon} + O\left(\frac{\Delta^4 x}{\epsilon}\right) + O(\epsilon) + O(\Delta x), \tag{A2}$$

where α'', β'' are the functions of x'' and t'' and they are determined from the path chosen.

Differentiating (A2) with respect to time,

$$\ddot{x}'' = \frac{\dot{x}''}{\epsilon} - \frac{\Delta x}{\epsilon^2} + \left(\dot{\alpha}'' + \frac{\partial\alpha''}{\partial x''}\dot{x}''\right)\frac{\Delta^2 x}{\epsilon} + \frac{2\alpha''\Delta x\dot{x}''}{\epsilon} - \frac{\alpha''\Delta^2 x}{\epsilon^2}$$
$$+ \frac{3\beta''\Delta^2 x}{\epsilon}\dot{x}'' + \left(\dot{\beta}'' + \frac{\partial\beta''}{\partial x''}\dot{x}''\right)\frac{\Delta^3 x}{\epsilon} - \beta''\frac{\Delta^3 x}{\epsilon^2} + \gamma$$
$$+ \theta\frac{\Delta x}{\epsilon} + \xi\frac{\Delta^2 x}{\epsilon} + \text{higher orders}, \tag{A3}$$

where

$$\frac{\partial\alpha''}{\partial x''} \equiv \frac{\partial\alpha(x, t)}{\partial x}\Big|_{x'', t''}$$

and γ, θ, ξ come from $\theta(\epsilon, \Delta x)$ term of (A2). Therefore we finally get,

$$\ddot{x}'' = 2\alpha''(\Delta^2 x/\epsilon^2) + \text{higher orders}. \tag{A4}$$

On the other hand, (A1) can be expanded as

$$L = [g'' + \frac{\partial g''}{\partial x''}\dot{x}''(t - t'') + \frac{\partial g''}{\partial t''}(t - t'')$$
$$\times\tfrac{1}{2}(t - t'')^2\left(\frac{\partial g''}{\partial x''}\ddot{x}'' + \frac{\partial^2 g''}{\partial x''^2}\right) + O(t - t'')^3]$$
$$\times(\dot{x}'' + \ddot{x}''(t - t'') + O(t - t'')^3)^2$$
$$+ \left(a'' + \frac{\partial a''}{\partial x''}\dot{x}''(t - t'') + \frac{\partial a''}{\partial t''}(t - t'') + O(t - t'')^2\right)$$
$$\times\left(\dot{x}'' + \ddot{x}''(t - t'') + O(t - t'')^2\right) + v'' + O(t - t''). \tag{A5}$$

Substituting (A2) and (A4),

$$L = g''\frac{\Delta^2 x}{\epsilon^2} + 2g''\alpha''\frac{\Delta^3 x}{\epsilon^2} + \left(\frac{\partial g''}{\partial x''}\frac{\Delta^3 x}{\epsilon^3} + 4g''\alpha''\frac{\Delta^3 x}{\epsilon^3}\right)$$
$$\times(t - t'') + O\left((t - t'')^2\left(\frac{\Delta^2 x}{\epsilon^2}, \frac{\Delta^3 x}{\epsilon^3}\right)\right) + a''\frac{\Delta x}{\epsilon}$$
$$+ a''\alpha''\frac{\Delta^2 x}{\epsilon} + v'' + O\left((t - t'')^2\left(\frac{\Delta x}{\epsilon}, \frac{\Delta^2 x}{\epsilon^2}\right)\right). \tag{A6}$$

Therefore, $S_0 = \int_{t'}^{t''} L dt$ becomes

$$S_0 = g''\frac{\Delta^2 x}{\epsilon} + 2g''\alpha''\frac{\Delta^3 x}{\epsilon} - \frac{1}{2}\left(\frac{\partial g''}{\partial x''}\frac{\Delta^3 x}{\epsilon} + 4g''\alpha''\frac{\Delta^3 x}{\epsilon}\right)$$
$$+ a''\Delta x + v''\epsilon + O\left(\frac{\Delta^3 x}{\epsilon}\right) + O(\epsilon) + O(\Delta^2 x)$$
$$= g''\frac{\Delta^2 x}{\epsilon} - \frac{1}{2}\frac{\partial g''}{\partial x''}\frac{\Delta^3 x}{\epsilon} + a''\Delta x + v''\epsilon$$
$$+ O\left(\frac{\Delta^4 x}{\epsilon}\right) + O(\epsilon) + O(\Delta^2 x). \tag{A7}$$

$g'', \partial g''/\partial x'', a''$ are fixed by the Lagrangian and the end points, and α'', β'' occur only in higher terms. Since ψ expansion is

$$\psi(x'', t') - \frac{\partial\psi}{\partial x''}\Delta x + \frac{1}{2}\frac{\partial^2\psi}{\partial x''^2}\Delta^2 x,$$

the coefficients of $\partial\psi/\partial x''$ and $\partial^2\psi/\partial x''^2$ are independent of the path variation, because they pick up only the lowes order terms. The arbitrariness comes only in the coefficient of ψ and this can be regulated by a normalization of the form $(2\pi\epsilon\hbar i)^{-N/2}(1 + i\hbar\epsilon f(x'', t'))$ which will alter just the coefficient of ψ. The arguments for the averaging approximation of the action is similar. All the ordering rules, i.e., Weyl-McCoy, symmetrization, and Born-Jordan, etc., differ only by a scalar operator when the Hamiltonian has no more than second-order derivatives. Therefore, we see that Feynman formalism accomodates all these with the ambiguity of normalization.

APPENDIX B

We consider

$$K_\epsilon = [(2\pi i\hbar\epsilon)^{N/2}]^{-1}(1 + i\hbar\epsilon A' + E'_{ij}\Delta x^i\Delta x^j)\exp(i/\hbar)S_{\text{cl}})$$

when operated by $O_{x'', t''}$.

The result is

$$i\hbar\left(-\frac{N}{2\epsilon} + i\hbar A' + \frac{i}{\hbar}\frac{\partial S}{\partial t''}\right)K + O(\epsilon)K + \left[\tfrac{1}{2}\hbar^2 g''^{ij}2E'_{ij}\right.$$

$$+\hbar^2\left(\frac{i}{\hbar}\right) g''^{ij}2E'_{jk}\Delta x^k \frac{\partial S}{\partial x''^i}$$
$$+\tfrac{1}{2}\hbar^2\left(\frac{i}{\hbar}\right) g''^{-1/2}\frac{\partial}{\partial x''^i}(g''^{1/2}g''^{ij})\frac{\partial S}{\partial x''^j}$$
$$+\tfrac{1}{2}\hbar^2 g''^{ij}\left(\frac{i}{\hbar}\right)\frac{\partial^2 S}{\partial x''^i\partial x''^j}+\tfrac{1}{2}\hbar^2 g''^{ij}\left(\frac{i}{\hbar}\right)^2\frac{\partial S}{\partial x''^i}\frac{\partial S}{\partial x''^j}$$
$$-i\hbar a''^i\left(\frac{i}{\hbar}\right)\frac{\partial S}{\partial x''^i}-\frac{i\hbar}{2}(g''^{1/2}{}_{,i}a''^i+g''^{1/2}a''^i{}_{,i})$$
$$-\tfrac{1}{2}a''^i a''_i - v''+O(\epsilon,\Delta x)\Big]K. \qquad \text{(B1)}$$

From Hamilton-Jacobi equation,

$$\frac{\partial S}{\partial t''}=-v''-\tfrac{1}{2}g''^{ij}\frac{\partial S}{\partial x''^i}\frac{\partial S}{\partial x''^j}+g''^{ij}\frac{\partial S}{\partial x''^i}a''_j-\tfrac{1}{2}g''^{ij}a''_i a''_j. \qquad \text{(B2)}$$

Using (B2) and (2.5), (B1) becomes (K will not be written down explicitly from now on):

$$-\frac{Ni\hbar}{2\epsilon}-\hbar^2A'+\hbar^2E'+2i\hbar g'^{ij}E'_{jk}\Delta x^k\left(\frac{g'_{il}\Delta x^l}{\epsilon}\right)$$
$$+\tfrac{1}{2}i\hbar(g''^{-1/2}g''^{1/2}{}_{,i}g''^{ij}+g^{ij}{}''_{,i})$$
$$\times\left(g'_{jk}\frac{\Delta x^k}{\epsilon}+\frac{1}{4\epsilon}B'_{jlk}\Delta x^l\Delta x^k+a'_i\right)$$
$$+\tfrac{1}{2}i\hbar\left(g'_{ij}+g'^{ij}{}_{,k}\Delta x^k+\frac{\partial g'^{ij}}{\partial t'}\epsilon+\tfrac{1}{2}g'^{ij}{}_{,kl}\Delta x^k\Delta x^l\right)$$
$$+\left(\frac{g'_{ij}}{\epsilon}\frac{1}{2\epsilon}B'_{ijk}\Delta x^k+\frac{1}{6\epsilon}C'_{ijkl}\Delta x^k\Delta x^l\right)+\tfrac{1}{2}i\hbar g'^{ij}$$
$$\times\tfrac{1}{2}\left(a'_{i,j}+a'_{j,i}+\frac{\partial g'_{ij}}{\partial t'}\right)-\frac{i\hbar}{2}g'^{-1/2}(g'^{-1/2}{}_{,i}a'^i$$
$$+g'^{1/2}g'^i{}_{,i})+O\left(\frac{\Delta^3 x}{\epsilon}\text{ and higher}\right). \qquad \text{(B3)}$$

We use $g''^{1/2}{}_{,i}g''^{-1/2}g''^{ij}=\frac{1}{2}g''^{lk}g''_{lk,i}g''^{ij}$ and evaluate everything around time t'.

Then, (B3) becomes

$$-\hbar^2A'+\hbar^2E'+2\frac{i\hbar}{\epsilon}E'_{ij}\Delta x^i\Delta x^j$$
$$+\tfrac{1}{2}i\hbar\left(\frac{g'_{jk}\Delta x^k}{\epsilon}+\frac{1}{4\epsilon}B'_{jlk}\Delta x^l\Delta x^k\right)(\tfrac{1}{2}g'^{nm}g'_{nm,i}g'^{ij}$$
$$+g'^{ij}{}_{,i}+g'^{ij}{}_{,in}\Delta x^n+\tfrac{1}{2}\Delta x^n g'^{mp}{}_{,n}g'_{mp,i}g'^{ij}$$
$$+\tfrac{1}{2}\Delta x^n g'^{mp}g'_{mp,in}g'^{ij}+\tfrac{1}{2}\Delta x^n g'^{mp}g'_{mp,i}g'^{ij}{}_{,n})$$
$$+\frac{i\hbar}{2}\left(g'^{ij}{}_{,k}g'_{ij}\frac{\Delta x^k}{\epsilon}+\tfrac{1}{2}g'^{ij}{}_{,kl}\Delta x^k\Delta x^l\frac{g'_{ij}}{\epsilon}\right.$$
$$\left.+\frac{1}{2\epsilon}g'^{ij}B'_{ijk}\Delta x^k+\frac{1}{6\epsilon}g'^{ij}C'_{ijkl}\Delta x^k\Delta x^l\right)+O\left(\frac{\Delta^3 x}{\epsilon}\right)$$
$$=\hbar^2(E'-A')$$
$$+\frac{i\hbar}{\epsilon}\Delta x^k[g'_{jk}(\tfrac{1}{2}g'^{lm}g'_{lm,i}g'^{ij}+g'^{ij}{}_{,i})$$
$$+g'^{ij}{}_{,k}g'_{ij}+\tfrac{1}{2}g'^{ij}(g'_{ij,k}+g'_{ik,j}+g'_{jk,i})]^{(1)}$$
$$+\frac{i\hbar}{\epsilon}\Delta x^i\Delta x^j[2E'_{ij}+\tfrac{1}{2}g'_{li}g'^{kl}{}_{,kj}+\tfrac{1}{4}g'^{lm}{}_{,j}g'_{lm,i}$$
$$+\tfrac{1}{4}g'^{lk}g'_{lk,ij}+\tfrac{1}{4}g'_{ki}g'^{ln}g'_{ln,m}g'^{mk}{}_{,j}$$
$$+\tfrac{1}{8}(g'_{ij,k}+g'_{jk,i}+g'_{ki,j})(\tfrac{1}{2}g'^{ln}g'_{ln,m}g'^{mk}+g'^{lk},l)$$
$$+\tfrac{1}{4}g'_{lk}g'^{lk}{}_{,ij}+\tfrac{1}{4}g'^{lk}{}_{,i}(g'_{lk,j}+g'_{lj,k}+g'_{kj,l})$$
$$+\tfrac{1}{12}g'^{lk}(g'_{ij,lk}+g'_{lk,ij}+4g'_{il,jk}-g'^{mn}[ij,m]'[kl,n]'$$
$$-2g'^{mn}[ik,m]'[lj,n]')]^{(2)}+O\left(\frac{\Delta^3 x}{\epsilon}\right).$$

$g^{ij}{}_{,k}=-g^{il}g^{jm}g_{lm,k}$ and therefore $[\ ]^1$ is easily seen to be zero.

By using

$$\frac{\partial^2}{\partial x^i\partial x^k}(g'^{lk}g'_{il})=0,\text{ etc.},$$

$[\ ]^2$ becomes

$$2E'_{ij}-\tfrac{1}{2}g'^{kl}g'_{li,kj}+\tfrac{1}{2}g'^{kn}g'^{lm}g'_{nm,k}g'_{il,j}$$
$$+\tfrac{1}{2}g'^{kn}g'^{lm}g'_{mn,j}g'_{li,k}-\tfrac{1}{4}g'_{lm,i}g'^{mk}g'^{ln}g'_{nk,j}$$
$$+\tfrac{1}{4}g'^{lk}g'_{lk,ij}-\tfrac{1}{4}g'^{ln}g'_{ln,m}g'^{mq}g'_{qi,j}$$
$$+\tfrac{1}{8}(g'_{ij,k}+2g'_{jk,i})(\tfrac{1}{2}g'^{ln}g'_{ln,m}g'^{mk}-g'^{km}g'^{ln}g'_{mn,l})$$
$$+\tfrac{1}{4}(-g'_{lk,ij}g'^{lk}+2g'_{lk,i}g'^{kn}g'^{lm}g'_{mn,j})$$
$$-\tfrac{1}{4}g'^{ln}g'^{km}g'_{nm,i}(g'_{lk,j}+2g'_{ij,k})$$
$$+\tfrac{1}{12}g'^{lk}(g'_{ij,lk}+g'_{lk,ij})+\tfrac{1}{3}g'^{lk}g'_{il,jk}$$
$$-\tfrac{1}{48}g'^{lk}g'^{mn}(2g'_{im,j}-g'_{ij,m})(2g'_{kn,l}-g'_{kl,n})$$
$$-\tfrac{1}{24}g'^{lk}g'^{mn}(2g'_{im,k}-g'_{ik,m})(2g'_{ln,j}-g'_{lj,n}).$$

If we group the above terms, we precisely get $2E'_{ij}-(R''_{ij}/6)$.

Therefore,

$$O_{x'',t''}K=\Big[-\hbar^2A'+\hbar^2E'+2E'_{ij}\Delta x^i\Delta x^j\frac{i\hbar}{\epsilon}$$
$$-\frac{i\hbar R'_{ij}}{6\epsilon}\Delta x^i\Delta x^j+O\left(\frac{\Delta^3 x}{\epsilon}\right)+O(\epsilon)+O(\Delta x)\Big]K \qquad \text{(B4)}$$

For

$$\lim_{\epsilon\to 0}\int O_{x'',t''}K(x'',t'':x',t')\psi(x',t')d^Nx',$$

we expand all the relevent quantities in terms of x'', t', ϵ, and Δx. The procedure is the same as that of the short-time calculation. The result is

$$\lim_{\epsilon\to 0}\int O_{x'',t''}K\psi(x',t')d^Nx'$$
$$=(-\hbar^2A''+\hbar^2E''-2E''_{ij}g''^{ij}\hbar^2+\tfrac{1}{6}R''_{ij}g''^{ij})\psi(x'',t').$$

Therefore, $A''+E''=\frac{1}{6}R''$.

APPENDIX C

We show that if g_{ij} is not constant, then, F of (1.4) depends also on x'', x'. The argument is similar for A of (1.3). We consider the one-dimensional case and $a_i=v_i=0$. Expanding the Lagrangian around the classical path $\bar{x}(t)$ [$\bar{x}(t'')=x'', \bar{x}(t')=x'$], we have

$$K=\int_{x'}^{x''}\exp\left[\frac{i}{\hbar}\int_{t'}^{t''}\left(g(\bar{x},t)+\frac{\partial g(\bar{x},t)}{\partial\bar{x}}y(t)\right.\right.$$
$$\left.+\tfrac{1}{2}\frac{\partial^2 g(\bar{x},t)}{\partial x\partial x}y(t)y(t)+\text{higher order}\right)(\dot{\bar{x}}^2+2\dot{\bar{x}}\dot{y}$$
$$\left.+\dot{y}^2)dt\right]Dx, \qquad \text{(C1)}$$

where y is a small variational path with $y(t') = y(t'') = 0$.

Therefore,

$$K = \int_{x'}^{x''} Dx \exp\left(\frac{i}{\hbar} S_{\mathrm{cl}}\right) \exp\left[\frac{i}{\hbar} \int_{t'}^{t''} \left(\frac{\partial g(\bar{x}, t)}{\partial \bar{x}} y(t)\dot{\bar{x}}^2 + \cdots\right)\right].$$

Equation (C1) can be written as

$$K = \exp\left(\frac{i}{\hbar} S_{\mathrm{cl}}\right) \int_0^0 Dy \exp\left[\frac{i}{\hbar} \int_{t'}^{t''} \left(2g(\bar{x}, t)\dot{\bar{x}}\dot{y} + g(\bar{x}, t)\dot{y}^2 + \frac{\partial g}{\partial \bar{x}} \dot{y}^2 + \cdots\right) dt\right].$$

If g only depends on t, $\int_0^0 Dy$ is only a function of t'' and t'. However, if g depends on x, since $\int_0^0 Dy$ is a functional of $g(\bar{x}, t)$ evaluated at t'' and t', it will depend on x'' and x'. Therefore, the normalization is allowed to have x'', x' dependence when g_{ij} is position-dependent.

* Supported in part by the National Science Foundation Grant No. GP-32998X.

[1]L. Cohen, J. Math. Phys. 11, 3296 (1970); E. Kerner and W. C. Sutcliffe, 11, 391 (1970); I. W. Mayes and J. S. Dowker, J. Math. Phys. 14, 434 (1973).
[2]B. S. DeWitt, Rev. Mod. Phys. 29, 377 (1957). W. Pauli, "Feldquantisierung", lecture notes (1950–1951).
[3]K. S. Cheng, J. Math. Phys. 13, 1723 (1972).
[4]Preferably forward in time. This can be done by making $K(x'', t'' : x', t') = 0$ for $\epsilon < 0$. See R. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[5]Feynman and Hibbs, p. 35.
[6]We consider only the arbitrary paths dealt by the authors in Ref. 1. All these paths have the velocity at the end points in the form of (A2).
[7]See DeWitt or Appendix B.
[8]This does not necessarily mean that we should take K^* for $\epsilon < 0$.

On the conditions that a vector field vanish outside a given radius*

B. Bosco†

Istituto di Fisica Teorica, Università di Firenze, Istituto Nazionale di Fisica Nucleare, Sezione di Firenze, Firenze, Italy

Kelvin Laboratory, Department of Natural Philosophy, Glasgow University, Glasgow, Scotland

M. T. Sacchi

Istituto di Fisica dell'Università, Cagliari, Italy
(Received 27 March 1973)

The constraints that the curl and divergence of a vector field must satisfy in order that the field be identically zero outside a given radius are investigated. These result in two equations which connect the multipoles of the components of the curl and the multipoles of the divergence of the field for each given l. In an alternative proof of this constraint it is shown that the same relations still hold for the full space provided that the radial functions, which are the coefficients of the expansion of the field in vector spherical harmonics, satisfy certain conditions at infinity.

1. INTRODUCTION

A physical way of introducing a vector field is by prescribing its curl and divergence. The uniqueness of this decomposition is well known. Furthermore, the physical meaning of the irrotational and solenoidal parts of the field need not be stressed here. In order to quote a straightforward example, all the relevant physical quantities in electromagnetic theory, like the radiation emitted by a current, are connected to the divergence and to the curl of the current itself.

The origin of this paper was an investigation with the purpose of seeing what kind of general constraints are imposed on the current responsible for a given type of radiation.

In many problems these currents are restricted to a finite region of the space. This condition which is quite general imposes a constraint on the current itself since the curl and the divergence of the current cannot be assigned arbitrarily otherwise the current will only go to zero as r^{-2} for large r. While this is quite obvious[1] and well known, it seems to us that in the past it has been overlooked,[2] and the detailed relations between the curl and the divergence of a vector field that must be satisfied for the field to vanish outside a given region, to the best of our knowledge, have never been derived.[3]

We shall in the present paper derive a set of equations which represent the necessary and sufficient conditions for a vector field to vanish outside a given radius.

This will be done in Sec. 2 making use of the Helmholtz representation.

In Sec. 3 an alternative derivation of the theorem is given using the expansion of the vector field in vector spherical harmonics.

In Sec. 4 some explicit examples are worked out.

2. HELMHOLTZ'S THEOREM AND THE VANISHING OF THE VECTOR FIELD

Let $\mathbf{j}(\mathbf{r})$ be the vector field which vanishes outside a given radius r_0. The actual shape of the region in which $\mathbf{j}(\mathbf{r})$ is different from zero may be completely arbitrary, since we can always enclose it in a sphere of radius r_0. Our starting point is the well-known Helmholtz's theorem[4] which allows us to write $\mathbf{j}(\mathbf{r})$ as a sum of an irrotational part $\nabla\varphi(\mathbf{r})$ and of a solenoidal part $\nabla \times \mathbf{A}$:

$$\mathbf{j}(\mathbf{r}) = \nabla\varphi(\mathbf{r}) + \nabla \times \mathbf{A}(\mathbf{r}) \tag{2.1}$$

with the supplementary condition

$$\nabla \cdot \mathbf{A} = 0. \tag{2.2}$$

Obviously Eqs. (2.1) and (2.2) imply

$$\nabla \cdot \mathbf{j}(\mathbf{r}) = \Delta_2 \varphi(\mathbf{r}), \tag{2.3}$$

$$\nabla \times \mathbf{j}(\mathbf{r}) = -\Delta_2 \mathbf{A}(\mathbf{r}). \tag{2.4}$$

The solutions of Eqs. (2.3) and (2.4) are given by the Poisson formulas

$$\varphi(\mathbf{r}) = -\frac{1}{4\pi} \int \frac{\nabla \cdot \mathbf{j}(\xi)}{|\mathbf{r} - \xi|}\, d^3\xi, \tag{2.5}$$

$$\mathbf{A}(\mathbf{r}) = \frac{1}{4\pi} \int \frac{\nabla \times \mathbf{j}(\xi)}{|\mathbf{r} - \xi|}\, d^3\xi \tag{2.6}$$

It is easy to verify that condition (2.2) is automatically satisfied by expression (2.6). We are now mainly concerned in imposing the condition that outside the radius r_0, the vector field $\mathbf{j}(\mathbf{r})$ resulting from the insertion of expressions (2.5) and (2.6) into (2.1) should be zero.

The simplest way to proceed is first to use the classical Legendre expansion:

$$\frac{1}{|\mathbf{r} - \xi|} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{(2l+1)} \frac{\xi^l}{r^{l+1}} Y^*_{lm}(\vartheta', \varphi') Y_{lm}(\vartheta, \varphi) \tag{2.7}$$

valid for $r > \xi$, where ϑ', φ' and ϑ, φ are the angles of the vectors $\boldsymbol{\xi}$ and $\mathbf{r}$, respectively.

By inserting Eq. (2.7) into expressions (2.5) and (2.6) we get

$$\varphi(\mathbf{r}) = -\sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{2l+1} \frac{1}{r^{l+1}} D^{lm} Y_{lm}(\vartheta, \varphi), \tag{2.8}$$

$$\mathbf{A}(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{2l+1} \frac{1}{r^{l+1}} \mathbf{R}^{lm} Y_{lm}(\vartheta, \varphi), \tag{2.9}$$

where we have introduced the notations

$$D^{lm} = \int \xi^l Y^*_{lm}(\vartheta', \varphi') \nabla \cdot \mathbf{j}(\xi)\, d^3\xi, \tag{2.10}$$

$$\mathbf{R}^{lm} = \int \xi^l Y^*_{lm}(\vartheta', \varphi') \nabla \times \mathbf{j}(\mathbf{r})\, d^3\xi. \tag{2.11}$$

Of course, Eqs. (2. 8) and (2. 9) are valid only for $r > r_0$, but that will be enough for our purposes.

In order to proceed further we will need the following formula:[5]

$$\nabla\left(\frac{Y_{lm}(\vartheta,\varphi)}{r^{l+1}}\right) = -\left(\frac{l+1}{2l+1}\right)^{1/2}\left(\frac{d}{dr}-\frac{l}{r}\right)\frac{1}{r^{l+1}}\,\mathbf{Y}_{l,\,l+1,\,m}(\vartheta,\varphi) + \left(\frac{l}{2l+1}\right)^{1/2}\left(\frac{d}{dr}+\frac{l+1}{r}\right)\frac{1}{r^{l+1}}\,\mathbf{Y}_{l,\,l-1,\,m}(\vartheta,\varphi) = [(2l+1)(l+1)]^{1/2}\,\frac{1}{r^{l+2}}\,\mathbf{Y}_{l,\,l+1,\,m}(\vartheta,\varphi). \tag{2. 12}$$

This formula is a particular case of Olsons' formulas and the $\mathbf{Y}_{l,\lambda,m}$, where $\lambda = l-1, l, l+1$, are the vector spherical harmonics as defined in A83.[5] With the help of Eq. (2. 12) we can write at once

$$\nabla\varphi(\mathbf{r}) = -\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{1}{r^{l+2}}\,D^{lm}\,\mathbf{Y}_{l,\,l+1,m}(\vartheta,\varphi). \tag{2. 13}$$

Since $\mathbf{R}^{lm}$ are constant vectors, we get also

$$\nabla\times\mathbf{A}(\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{1}{r^{l+2}}\,\mathbf{Y}_{l,\,l+1,m}(\vartheta,\varphi)\times\mathbf{R}^{lm}. \tag{2. 14}$$

In order to make explicit the cross product, we must write down the definition of the vector spherical harmonics which appears in (2. 14):

$$\mathbf{Y}_{l,l+1,m}(\vartheta,\varphi) = \sum_{\mu q} Y_{l+1,\mu}(\vartheta,\varphi)\,\mathbf{e}_q(l+1\;\mu 1 q\,|\,l+1\;1lm). \tag{2. 15}$$

Here the $\mathbf{e}_q$ are the unit spherical vectors and $(l+1\;\mu 1 q\,|\,l+1\;1lm)$ are the Clebsch–Gordan coefficients both being defined as in A. Equations (2. 13) and (2. 14) are valid for $r > r_0$, and therefore they must add up to zero if the vector field $\mathbf{j}(\mathbf{r})$ has to be zero in that region.

We shall define the spherical components of the vectors $\mathbf{R}^{lm}$ as:

$$R_{-1}^{lm} = (1/\sqrt{2})(R_x^{lm} - iR_y^{lm}),$$
$$R_0^{lm} = R_z^{lm} \quad R_{+1} = -(1/\sqrt{2})(R_x^{lm} + R_y^{lm}) \tag{2. 16}$$

In terms of these quantities the three spherical components of $\nabla\times\mathbf{A}$ can be written down as

$$(\nabla\times\mathbf{A})_{-1} = -\sum_{l}\sum_{m}\sum_{\mu}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{1}{r^{l+2}}\,Y_{l+1\;\mu} \times\{(l+1\;\mu 10\,|\,l+1\;1lm)R_{-1}^{lm} + (l+1\;\mu 11\,|\,l+1\;1lm)R_0^{lm}\}, \tag{2. 17}$$

$$(\nabla\times\mathbf{A})_0 = \sum_{l}\sum_{m}\sum_{\mu}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{1}{r^{l+2}}\,Y_{l+1\;\mu} \times i\{(l+1\;\mu\;1\;-1\,|\,l+1\;1lm)R_{-1}^{lm} - (l+1\;\mu 11\,|\,l+1\;1lm)R_{+1}^{lm}\}, \tag{2. 18}$$

$$(\nabla\times\mathbf{A})_1 = \sum_{l}\sum_{m}\sum_{\mu}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{1}{r^{l+2}}\,Y_{l+1,\mu} \times i\{(l+1\;\mu\;1\;-1\,|\,l+1\;1lm)R_0^{lm} + (l+1\;\mu 10\,|\,l+1\;1lm)R_1^{lm}\}. \tag{2. 19}$$

Similarly one can obtain the three spherical components of the vector $\nabla\varphi$ [Eq. (2. 13):

$$(\nabla\varphi)_{-1} = \sum_{lm\mu}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{D^{lm}}{r^{l+2}} \times(l+1\;\mu 11\,|\,l+1\;1lm)Y_{l+1,\mu}(\vartheta,\varphi), \tag{2. 20}$$

$$(\nabla\varphi)_0 = -\sum_{lm\mu}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{D^{lm}}{r^{l+2}} \times(l+1\;\mu 10\,|\,l+1\;1lm)Y_{l+1,\mu}(\vartheta,\varphi), \tag{2. 21}$$

$$(\nabla\varphi)_{+1} = \sum_{lm\mu}\left(\frac{l+1}{2l+1}\right)^{1/2}\frac{D^{lm}}{r^{l+2}} \times(l+1\;\mu 1-1\,|\,l+1\;1lm)Y_{l+1,\mu}(\vartheta,\varphi). \tag{2. 22}$$

The change in sign in (2. 20) and (2. 22) is due to the property of the unit spherical vectors:

$$\mathbf{e}_q^*\cdot\mathbf{e}_{q'} = (-1)^q\,\mathbf{e}_{-q}\cdot\mathbf{e}_{q'} = \delta_{qq'}. \tag{2. 23}$$

With the explicit expressions for $(\nabla\times\mathbf{A})_q$ given by Eqs. (2. 17)–(2. 19) and $(\nabla\varphi)_q$ given by (2. 20)–(2. 22) we can write the three equations compelling the field $\mathbf{j}(\mathbf{r})$ to be zero for $r > r_0$:

$$(\nabla\varphi(\mathbf{r}))_q = -(\nabla\times\mathbf{A}(\mathbf{r}))_q \quad \text{with } q = -1, 0, 1. \tag{2. 24}$$

From the resulting equations, by performing the sum over μ, and using the orthogonality of the spherical harmonics, one gets

$$D^{lm+1}(l+1\;m11\,|\,l+1\;1l\,m+1) = -i\{-(l+1\;m11\,|\,l+1\;1l\,m+1)R_0^{l\,m+1} - (l+1\;m10\,|\,l+1\;1lm)R_{-1}^{lm}\}, \tag{2. 25}$$

$$-D^{lm}(l+1\;m10\,|\,l+1\;1lm) = i\{-(l+1\;m\;1-1\,|\,l+1\;1l\,m-1)R_{-1}^{l\,m-1} + (l+1\;m11\,|\,l+1\;1l\,m+1)R_1^{l\,m+1}\}, \tag{2. 26}$$

$$D^{lm-1}(l+1\;m1-1\,|\,l+1\;1l\,m-1) = -i\{(l+1\;m1-1\,|\,l+1\;1l\,m-1)R_0^{l\,m-1} + (l+1\;m10\,|\,l+1\;1lm)R_1^{lm}\}. \tag{2. 27}$$

One notices that all the Clebsch–Gordan coefficients entering into these three equations have simple expressions. Upon substitutions of the explicit expressions and simplification one gets the final equations:

$$R_{-1}^{lm} = [(l-m)/2(l+m+1)]^{1/2}(R_0^{lm+1} + iD^{lm+1}), \tag{2. 28}$$

$$R_1^{lm} = [(l+m)/2(l-m+1)]^{1/2}(R_0^{lm-1} - iD^{lm-1}), \tag{2. 29}$$

$$iD^{lm} = [(l+m)/2(l-m+1)]^{1/2}\,R_{-1}^{lm-1} - [(l-m)/2(l+m+1)]^{1/2}R_1^{lm+1}. \tag{2. 30}$$

Equations (2. 28)–(2. 30) are the main result of this work. One sees at once that only two of them are independent. For reasons which will become clear in the next section we shall take as independent equations the first two, i.e., Eqs. (2. 28) and (2. 29). Let us now observe that Eqs. (2. 28) and (2. 29) are both necessary and sufficient conditions for the vector field $\mathbf{j}(\mathbf{r})$ to vanish outside the given radius.

That these relations are necessary conditions follows from the way they have been derived, i.e., by demanding that the field be identically zero outside the given region. But they are also sufficient, for, once they are satisfied, all the coefficients of the Laurent series in r vanish identically.

A last remark is in order here.

The constraints imposed on the field by the fact that it is vanishing outside a given region are formally independent of the radius of the sphere enclosing the domain in which the field is different from zero. This is obvious physically since these constraints are relations among integrals which are different from zero only in the domain in which the field is different from zero.

This argument can also be seen from the derivation of our relations (2. 28) and (2. 29). The whole point was, as far as the radius was concerned, to choose a radius r_0 such that the entire domain in which the field is nonvanishing is enclosed in the sphere of radius r_0, and then for $r > r_0$ to impose the condition that the field be identically zero.

It is quite clear that whatever radius $R > r_0$ we had chosen, our proof would have proceeded in the same way and would have ended with the same results.

This consideration suggests that Eqs. (2. 28) and (2. 29) are probably still valid for an infinite radius provided that certain conditions are satisfied. This turns out to be the case as the alternative proof that we give in next section will make clear.

3. ALTERNATIVE PROOF

Although the guess made at the end of Sec. 2 about the validity of Eqs. (2. 28) and (2. 29) for infinite radius could be put on more solid ground directly even with the Helmoltz's representation, we shall outline in this section an alternative proof of the same relations. Let us assume that the vector field $\mathbf{j}(\mathbf{r})$ is sufficiently well-behaved that it can be expanded in vector spherical harmonics:

$$\mathbf{j}(\mathbf{r}) = \sum_{JM} \{ f_{JM}(r)\, \mathbf{Y}_{J,J+1\,M}(\vartheta,\varphi) + g_{JM}(r)\, \mathbf{Y}_{J,J+1\,M}(\vartheta,\varphi) + h_{JM}(r)\, \mathbf{Y}_{J,J-1\,M}(\vartheta,\varphi) \}. \quad (3.1)$$

The curl and the divergence of $\mathbf{j}(\mathbf{r})$ can then be expressed in the following compact forms (A84):

$$\begin{aligned}\nabla \times \mathbf{j}(\mathbf{r}) = \sum_{JM} i\{ & L[J+2]\, f_{JM}(r)\, [J/(2J+1)]^{1/2}\, \mathbf{Y}_{J,JM}(\vartheta,\varphi) \\ & + L[-J]\, g^{(r)}_{JM}\, [J/(2J+1)]^{1/2}\, \mathbf{Y}_{J,J+1\,M}(\vartheta,\varphi) \\ & + L[J+1] g^{(r)}_{JM}\, [(J+1)/(2J+1)]^{1/2} \\ & \times \mathbf{Y}_{J,J-1\,M}(\vartheta,\varphi) + L[-(J-1)]\, h_{JM}(r) \\ & \times [(J+1)/(2J+1)]^{1/2}\, \mathbf{Y}_{JJM}(\vartheta,\varphi)\}, \end{aligned} \quad (3.2)$$

$$\begin{aligned}\nabla \cdot \mathbf{j}(\mathbf{r}) = \sum_{JM} \{ & [J/(2J+1)]^{1/2}\, L[-(J-1)]\, h_{JM}(r) \\ & - [(J+1)/(2J+1)]^{1/2}\, L[J+2]\, f_{JM}(r)\}\, Y_{JM}(\vartheta,\varphi). \end{aligned} \quad (3.3)$$

In Eqs. (3. 2) and (3. 3), the operator

$$L[x] = \frac{d}{dr} + \frac{x}{r} \quad (3.4)$$

has been introduced.

From Eqs. (3. 2) and (3. 3) the multipole components

$$R^{lm}_q = \int r^l\, Y^*_{lm}(\vartheta,\varphi)\, (\nabla \times \mathbf{j}(\mathbf{r}))_q\, d^3r, \quad (3.5)$$

$$D^{lm} = \int r^l\, Y^*_{lm}(\vartheta,\varphi) \nabla \cdot \mathbf{j}(\mathbf{r})\, d^3r \quad (3.6)$$

can be obtained.

The result is

$$\begin{aligned} R^{lm}_1 = i\{ & [(l+m)(l+m-1)(l-1)(2l+1)/2l(2l-1)]^{1/2} \\ & \times G^{l+1}_{l-1,\,m-1} + [(l+m)(l-m+1)(2l+1)/2l]^{1/2}\, H^{l+1}_{l,\,m-1}\}, \end{aligned} \quad (3.7)$$

$$\begin{aligned} R^{lm}_0 = i\{ & [(l-1)(2l+1)(l+m)(l-m)/l(2l-1)]^{1/2}\, G^{l+1}_{l-1,m} \\ & - m[(2l+1)/l]^{1/2}\, H^{l+1}_{lm}\}, \end{aligned} \quad (3.8)$$

$$\begin{aligned} R^{lm}_{-1} = i\{ & [(2l+1)(l-1)(l-m-1)(l-m)/2l(2l-1)]^{1/2} \\ & \times G^{l+1}_{l-1,\,m+1} - [(2l+1)(l+m+1)(l-m)/2l]^{1/2}\, H_{l,\,m+1}, \end{aligned} \quad (3.9)$$

$$D^{lm} = -[l(2l+1)]^{1/2}\, H^{l+1}_{l,m}. \quad (3.10)$$

In Eqs. (3. 7)–(3. 10) the following definitions have been introduced:

$$H^{l+1}_{l,m} = \int_0^{r_0} r^{l+1}\, h_{l,m}(r)\, dr, \quad (3.11)$$

$$G^{l+1}_{l-1,\,m} = \int_0^{r_0} r^{l+1}\, g_{l-1,m}(r)\, dr, \quad (3.12)$$

and use has been made of the fact that since the field is vanishing outside the domain contained in a sphere of radius r_0 the following quantities must also vanish for all l and m and $r \geq r_0$:

$$f_{l,m}(r) r^{l+2}, \quad g_{lm}(r) r^{l+2}, \quad h_{l,m}(r) r^{l+2}. \quad (3.13)$$

Equations (3. 7)–(3. 10) express the four quantities R^{lm}_q and D^{lm} in parametric form, by means of the parameters $G^{l+1}_{l-1,\,m}$ and $H^{l+1}_{l,m}$ for a given l and m.

A simple algebric manipulation allows the elimination of these parameters and one ends up again with Eqs. (2. 28) and (2. 29). Although it would have been difficult to foresee our basic equations from this direct calculation, this proof has the advantage that it established in clear way the conditions under which the stated theorem holds.

They are:

(a) The field $j(r)$ can be expanded in the series (3. 1) of vector spherical harmonics.

(b) The radial coefficients are such that the quantities (3. 13) are zero for $r \geq r_0$.

The condition (b) allows an immediate extension of the conditions under which our theorem is valid. In fact it is obvious that if the radial coefficients are such that

$$\left.\begin{aligned} &\lim_{r\to\infty} r^{l+2} f_{lm}(r) = 0 \\ &\lim_{r\to\infty} r^{l+2} g_{lm}(r) = 0 \\ &\lim_{r\to\infty} r^{l+2} h_{lm}(r) = 0 \end{aligned}\right\} \quad \text{for all } l \text{ and } m, \quad (3.14)$$

nothing is changed in our formulas, and we are still left with Eqs. (2. 28) and (2. 29). Therefore, provided the conditions (3. 14) are satisfied, our theorem is valid if the region is the whole space.

Before closing this section it is worthwhile to analyze the physical meaning of the two parameters $G^{l+2}_{l,m}$ and $H^{l+1}_{l,m}$.

Equation (3. 10) already dictates the significance of $H^{l+1}_{l,m}$: Apart from a known function of l, this parameter is the multipole (l, m) of the divergence of the field.

In order to find the meaning of $G^{l+2}_{l,m}$, let us compute the following quantity:

$$S^{l,m} = \int r^l Y^*_{lm}(\vartheta, \varphi) \nabla \cdot (\mathbf{r} \times \mathbf{j})\, d^3r \tag{3. 15}$$

This integral, when $\mathbf{j}$ is a current, is the static magnetic multipole (l, m) of the current.

Its evaluation in terms of the radial coefficients can be accomplished if one observes that by a vector identity $S^{l,m}$ can be written as

$$S^{l,m} = -\int r^l Y^*_{lm}(\vartheta, \varphi)\, \mathbf{r} \cdot \nabla \times \mathbf{j}\, d^3r. \tag{3. 16}$$

The following formula is valid (A84):

$$(\mathbf{r}/r) Y_{lm}(\vartheta, \varphi) = -[(l+1)/(2l+1)]^{1/2} \mathbf{Y}_{l,l+1,m}(\vartheta, \varphi) + [l/(2l+1)]^{1/2} \mathbf{Y}_{l,l-1,m}(\vartheta, \varphi). \tag{3. 17}$$

It is easily seen that by use of Eqs. (3. 2) and (3. 17), and due to the orthogonality of the vector spherical harmonics one has

$$S^{lm} = i \int r^{l+3} \left[\left(\frac{d}{dr} - \frac{l}{r}\right) g_{l,m}(r) \frac{[l(l+1)]^{1/2}}{2l+1} - \left(\frac{d}{dr} + \frac{l+1}{r}\right) g_{l,m}(r) \frac{[l(l+1)]^{1/2}}{2l+1}\right] dr, \tag{3. 18}$$

or

$$S^{l,m} = -i[l(l+1)]^{1/2} \int r^{l+2} g_{l,m}(r)\, dr = -i[l(l+1)]^{1/2}\, G^{l+2}_{l,m}. \tag{3. 19}$$

Equations (3. 10) and (3. 18) therefore complete our picture by telling us the physical meaning of the parameters $H^{l+1}_{l,m}$ and $G^{l+2}_{l,m}$. They are, respectively, apart from some given function of l, the static electric and magnetic multipole (l, m) of our vector field $\mathbf{j}(\mathbf{r})$.

4. SOME EXPLICIT EXAMPLES

In this section we shall verify on some elementary examples the validity of our relations among the multipoles of the curl and of the divergence. We shall consider two examples, one solenoidal and one irrotational.

First we consider the closed loop field

$$\mathbf{j}(\mathbf{r}) = I\, \delta(\Theta - \pi/2)\, \delta(r - a)\, \eta_\varphi, \tag{4. 1}$$

where η_φ is the unit vector in the φ direction and I is a constant. Physically this may represent a current of intensity I flowing in a circle of radius a centered at the Z axis.[6]

Obviously one has

$$\nabla \cdot \mathbf{j} = 0 \tag{4. 2}$$

so that all $D^{lm} = 0$.

Furthermore, an explicit calculation shows that the only nonzero multipoles of the components of the curl are

$$R^{l,-1}_{-1} = I[(2l+1)(l+1)\pi/2l]^{1/2} a^{l+1} P^1_{l-1}(0), \tag{4. 3}$$

$$R^{l,1}_{1} = I[(2l+1)\, l(l+1)\, \pi/2]^{1/2}\, a^{l+1}(l-1) P_{l-1}(0), \tag{4. 4}$$

$$R^{l,0}_{0} = -I\, a^{l+1} [\pi(2l+1)]^{1/2}\, l P_l(0). \tag{4. 5}$$

By remembering the following relations,

$$P^0_l(0) = (l-1) P^{-1}_{l-1}(0),$$

$$l\, P^0_l(0) = -P^1_{l-1}(0),$$

it is easily seen that Eqs. (2. 28) and (2. 29) are satisfied. As a second example let us consider the field[7]

$$\mathbf{j}(\mathbf{r}) = \eta_r \frac{I(r)}{2\pi r^2} [\delta(\cos\Theta - 1) - \delta(\cos\Theta + 1)], \tag{4. 6}$$

where η_r is the unit vector in the r direction, the δ functions force the field to be zero everywhere but along the polar axis, and $I(r)$ is zero for $r \geq r_0$.

For a particular choice of the function $I(r)$, (4. 6) may describe an antenna. However, for our purpose we shall not need any further specification of $I(r)$.

The only multipoles of $\nabla \cdot \mathbf{j}$ different from zero are the ones with $m = 0$, which are

$$D^{l0} = \left(\frac{2l+1}{4\pi}\right)^{1/2} [P_l(1) - P_l(-1)] \int_0^{r_0} r^l \frac{dI(r)}{dr}\, dr \tag{4. 7}$$

while the only R^{lm}_q different from zero are

$$R^{l,1}_{1} = +\frac{i}{\sqrt{2}} \left[\frac{(2l+1)(l-1)!}{4\pi(l+1)!}\right]^{1/2} [P_l(1) - P_l(-1)] \times \int_0^{r_0} I(r) r^{l-1}\, dr, \tag{4. 8}$$

$$R^{l,-1}_{-1} = -\frac{i}{\sqrt{2}} \left[\frac{(2l+1)(l+1)!}{4\pi(l-1)!}\right]^{1/2} l(l+1)[P_l(1) - P_l(-1)] \times \int_0^{r_0} I(r) r^{l-1}\, dr. \tag{4. 9}$$

Now due to the fact that $I(r_0) = 0$ we have identically

$$\int_0^{r_0} r^l \frac{dI(r)}{dr}\, dr = -l \int_0^{r_0} r^{l-1} I(r)\, dr$$

so that one sees again that (2. 28), (2. 29) are satisfied.

5. CONCLUSIONS

A full and detailed investigation has been made of the relationship which the curl and the divergence of a vector field must satisfy in order the field vanish outside a given region. In particular the alternative proof we have given in Sec. 3 shows that this region may be the full space and provides the conditions which the radial functions, which are the coefficients of the field's expansion in spherical harmonics, must then satisfy. The alternative proof also provides a quite natural parametric representation of the four quantities R^{lm}_q and D^{lm} in terms of two parameters G^{l+2}_{lm} and $H^{l+1}_{l,m}$ which are directly related through Eqs. (3. 14) and (3. 22) to S^{lm} and D^{lm}.

These two parameters are the ones which by virtue of our Eqs. (2. 28)–(2. 30) determine the multipoles of the components of the curl of the field. Furthermore, they can be identified respectively with the magnetic and electric multipoles (l, m) of the field $\mathbf{j}(\mathbf{r})$. While this observation may be taken as the starting point for a general formulation of a theory of the field's multipoles, we believe that the theorem derived in this paper may be useful in other respects.

For example the electric and magnetic multipoles of soft radiation generated by a given current are determined by the quantities D^{lm} and S^{lm} respectively. If,

therefore, the angular distribution and the polarization of such radiation is measured, one can obtain D^{lm} and S^{lm} and by the theorem derived in this paper one can determine the multipole components of the curl of the current itself. The theorem, therefore, will provide a test for any proposed model for such a current.[8]

Conversely, the theorem may prove useful in constructing models of currents which radiate in prescribed way, for example with a fixed ratio of magnetic to electric multipoles.

ACKNOWLEDGMENTS

One of the authors (B. B.) wishes to acknowledge many useful discussions with many physicists. Beside those quoted in Ref. 3 he wishes to thank Professor R. Blankenbecler for an enlightening conversation and encouragement. Finally he would like to thank Professor G. Bishop for the warm hospitality extended to him at the Department of Natural Philosophy of Glasgow University where this work has been completed, and Dr. R. R. Whitehead for reading the manuscript.

* Supported in part by NATO Research Grant No. 553.

† On leave of absence from University of Florence. Present and permanent address: Istituto di Fisica dell'Università, Florence (Italy).

[1]See for example J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (Wiley, New York, 1952), Appendix B, p. 806, Footnote 1.

[2]The only paper we were able to find which makes a quantitative use of a current being zero outside a given radius is the one of French and Shimamoto, Phys. Rev. **91**, 898 (1953).

[3]We are indebted for discussions on this point to many physicists. Particularly we wish to thank Professor C. L. Biedenharn, Professor G. Feldman, Professor H. Feshbach, and Professor B. Ferretti for interesting discussions on this point.

[4]For a mathematical discussion of this theorem see for example P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 52.

[5]See A R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1957), p. 84. From here on this book shall be referred as A followed by the page number.

[6]See for example Ref. 4, p. 1881.

[7]See for example J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962), p. 563.

[8]B. Bosco, Phys. Lett. B **40**,171 (1972). This paper contains some misprints which are corrected in the present paper.

Analyticity in the coupling strength

P. R. Graves-Morris and J. F. Rennison

Mathematical Institute, University of Kent, Canterbury, England
(Received 6 September 1972)

We briefly review the methods which have been used to establish the domains of analyticity of the partial wave scattering amplitude in the presence of a Yukawa potential as a function of the coupling strength, and the methods available for proving the convergence of Padé approximants to the Neumann series of the partial wave Lippmann–Schwinger equation. We then give a complete proof, using the Banach space technique first used by Lovelace, that the scattering amplitude is meromorphic in the coupling strength, and use Pommerenke's or Beardon's theorem to deduce the domains of convergence of the Padé approximants.

1. INTRODUCTION

In this paper, we briefly review the progress which has been made in proving the convergence of Padé approximants to the Neumann series of the partial wave Lippman-Schwinger[1] equations in the presence of Yukawa type potentials. We present a self-contained and explicit derivation of what appears to be the strongest result pertaining to the physical situation. We prove using Beardon's theorem that the $[M,N]$ approximants to the Neumann series converge uniformly as $M \to \infty$ with $M \leqslant \alpha N$ for some $\alpha > 0$ to the solution of the Lippmann-Schwinger equation in any compact set containing no poles of the solution or limit points of the poles of the approximants. We prove, using Pommerenke's theorem, that the $[M,N]$ approximants to the Neumann series converge in capacity with $\lambda^{-1} \leqslant N/M \leqslant \lambda$ for some $\lambda \geqslant 1$. The method was first used by Caser, Piquet, and Vermeulen for the T matrix[2]; it was simplified and applied to the K matrix by one of us[3] and is being used to solve the Faddeev equations of K decays.

The theorem we prove can certainly be derived by taking appropriate sections from the published literature, but some of the literature is incorrect and has been superseded. There are several different approaches to the problem, and we wish to use methods which will generalize. Generalization to nonsingular short-range potentials is straightforward in all methods, but not all methods generalize to treat the Blankenbecler-Sugar equation,[4] the Bethe-Salpeter equation or equations with nonlocal or energy-dependent potentials.

(i) Banach space methods seem the most natural way of approaching the problem. Lovelace used a spectral representation of the kernel to prove the kernel compact on the Banach space C_1 of bounded differentiable functions with bounded continuous derivatives on $[0,\infty)$.[5] It is now known that not all compact operators on an arbitrary Banach space can be approximated in norm by operators of finite rank. With this possibility in mind Lovelace used a framework in which it is implicit that such an approximation exists. To be precise, he considers

$$\hat{K}_l(s) = (-1/\pi)\int_0^\infty ds'\,[\hat{\Delta}_l(s')/(s'-s)],$$

where

$$\langle p\,|\hat{\Delta}_l|q\rangle = 2\pi^2 V_l(p, s^{1/2})s^{1/2}\delta(q - s^{1/2})$$

and s does not lie on the positive real axis. $\hat{\Delta}_l(s)$ is a rank 1 operator for any value of s. It is then true that the integral converges and can be uniformly approximated by a Riemann sum, which is a finite rank operator. Muskhelishvili's[6] theorem shows that when the derivative with respect to s' of the integrand without the factor $(s'-s)^{-1}$ is uniformly bounded in the neighborhood of any given point on the positive real axis, then the limit in norm of $\hat{K}_l(s)$ as s tends to the real axis from above can be taken. Then the integral becomes the kernel of the Lippmann-Schwinger T-matrix equation.

Since we are using the uniform topology, the kernel of the Lippmann-Schwinger equation can also be approximated by a kernel of finite rank when the $i\epsilon$ limit is taken.

(ii) Hilbert space methods using the space of square integrable functions can be used to prove that the kernel of the scattering equations when suitably rewritten, is compact. Scadron, Weinberg, and Wright[7] consider local potentials, $V(r)$ and take a (possibly complex) square root $V^{1/2}(r)$. They define $\tilde{\Psi}(\mathbf{r}) = V^{1/2}(\mathbf{r})\,|\Psi(\mathbf{r})\rangle$ where $|\Psi(\mathbf{r})\rangle$ is a scattering state for which $|\Psi(\mathbf{r})|^2$ is the measurable and finite probability density, and so, for finite range potentials, $\int|\Psi(\mathbf{r})|^2 d\mathbf{r} < \infty$. Thus they need only consider operators defined on the Hilbert space of L^2 functions.

To calculate the T matrix, use

$$T = V^{1/2}(1 - \lambda V^{1/2} G V^{1/2})^{-1} V^{1/2}.$$

Provided $V^{1/2}GV^{1/2}$ exists and is an L^2 kernel, then the resolvent exists and is meromophic in λ, and its momentum space representation gives the full T matrix. The methods given by the authors are rigorous in the coordinate space representation, for potentials of short range. The kernel is not self-adjoint unless $V(r)$ is of one sign. The results carry over to the momentum space representation by using the unitary Fourier-Plancherel[8] transformation, which leaves the spectrum unchanged.

(iii) Symmetrization and subtraction methods can also be used to prove meromorphy of the solution by reducing the Lippmann-Schwinger equation and its variants to nonsingular equations to which Fredholm theory applies. Taylor's subtraction method[9] has the advantage that it generalizes easily to the Bethe-Salpeter equation and all three methods[9-11] lead to integral equations with continuous kernels. These can be solved by matrix inversion methods, provided the p-space representation of the potentials have enough partial derivatives. The solutions of the nonsingular equations given by the subtraction methods[9,11] are easily related to the solution of the Lippmann-Schwinger equation by quadrature formulas which manifestly preserve the meromorphic character of the solution as a function of g.

There is a very general theorem due to Poincaré[12] which establishes the analyticity in g of the Jost functions of the Schrödinger equation, which leads directly to meromorphy of the S matrix; this method may also apply to other scattering equations like the Bethe-Salpeter

equation in the elastic region, but it does not generalize to the Blankenbecler-Sugar equation.

There are two special cases which allow strong results to be deduced. When the potential is positive, i.e., repulsive, Masson[13] has used the Schrödinger equation to show that for positive energies the K matrix is an extended series of Stieltjes in the g plane. The analysis of Scadron, Weinberg and Wright[7] also leads to real equations with real solutions whenever $V(r)$ has one sign. If one works at negative energies, the Lippmann-Schwinger equation can be symmetrized, and the Hilbert space theorems for L^2 kernels apply. Thus T is meromorphic in g and the poles lie on the real g axis. An eigenfunction expansion shows that T can be related to an extended series of Stieltjes for any potential, by treating the first term of the series separately. A somewhat roundabout numerical method of solving the Lippmann-Schwinger equation is to solve it for negative energies and then extrapolate back to the physical region[14]; this is claimed to work accurately over moderate energy ranges.

(iv) There are two approaches to the method of establishing convergence of the Padé approximants. If the series is a series of Stieltjes[15] or an extended series of Stieltjes,[13] then there is no difficulty, because the poles of the $[N, N-1]$ approximants are restricted to a particular region of the g plane and all their residues are positive. If the kernel is a kernel of finite rank, Chisholm[16] has proved convergence. It is likely that similar results apply to compact (completely continuous) kernels, but there are some difficulties about uniformity of convergence in the proof, and we are grateful to Professor Chisholm for a discussion of the point. The most powerful method in the general case appears to be that of using Pommerenke's or Beardon's theorem[7] on the convergence of Padé approximants to meromorphic functions. Pommerenke's theorem grew from the theorem of Nuttall about convergence in measure.[18] Once it is established that the kernel of an integral equation is a compact operator, then the resolvent is meromorphic in the g plane.[19]

Only when convergence of the $[N, N-1]$ approximants to the Neumann series is proved can one establish the convergence of Lanczos' minimal iteration method to the solution.[20]

We present an analysis of the Lippmann-Schwinger equation which is similar to Lovelace's, but borrows a technique of the subtraction method. Our method is explicit, to the extent that we construct a finite rank approximant to the kernel of the Lippmann-Schwinger equation, and it shares with Lovelace's the advantage of being easily generalized to other scattering equations. Our method is conceptually simpler than Lovelace's because we do not use operator valued integrals, and we prove explicitly that the Yukawa potential satisfies the conditions of our hypotheses. We apply our methods to the K-matrix equation, because it is easier to do numerical calculations with real numbers; the methods apply equally well to the t-matrix equation, and this is briefly discussed in Sec. 3.

2. CONSTRAINTS ON THE POTENTIALS

We consider first the Lippmann-Schwinger equation for the partial wave K matrix written in momentum space:

$$K(p,k') = gV(p,k') - (2g/\pi)\int_0^\infty V(p,q)[q^2dq/(k^2-q^2)]K(q,k').$$

Suppose that k' is fixed at an arbitrary positive value. Set $V(p,k') = \phi(p)$ and seek a solution $f(p) = K(p,k')$. Then in this notation, we must solve

$$f(p) = g\phi(p) + (2g/\pi)\int_0^\infty V(p,q)[q^2dq/(q^2-k^2)]f(q).$$

In Sec. 3, we shall prove the convergence of certain Padé approximants of the Neumann series of this equation to the solution of the space C_1 in a region D of the coupling constant plane.

Our method works for potentials satisfying

(i) $|V(p,q)| \leq \Phi(q)$ and $\left|\frac{\partial V}{\partial p}(p,q)\right| \leq \Phi(q)$,

for all $p, q \geq 0$, where $\int_0^\infty \Phi(q)\,dq < \infty$.

(ii) V has continuous second partial derivatives for $p \geq 0$, $q \geq 0$ and, for any $\Lambda > 0$,

$$\frac{\partial V(p,q)}{\partial q} \to 0, \qquad \frac{\partial^2 V(p,q)}{\partial p\,\partial q} \to 0$$

as $p \to \infty$ uniformly for $0 \leq q \leq \Lambda$. We now show that the Yukawa potential satisfies these conditions. For $p, q \geq 0$ write $s = 2pq/(p^2+q^2+\mu^2)$ so that $0 \leq s < 1$. The momentum space representation of the unit Yukawa potential in the lth partial wave is

$$V(p,q) = (2pq)^{-1}Q_l(1/s), \qquad l = 0, 1, 2, \ldots .$$

Since $Q_l(1/s)$ has a zero of order $l+1$ at $s = 0$,[21] we can write

$$V(p,q) = (p^2+q^2+\mu^2)^{-1}F(s),$$

where $F(s) = s^{-1}Q_l(1/s)$ is analytic for $|s| < 1$.

Further[22], we can express F in the form

$$F(s) = A(s) + B(s)\log[(1+s)/(1-s)],$$

for $\frac{1}{2} \leq s < 1$, where A and B are bounded functions with bounded derivatives in this range.

In showing that V satisfies conditions (i), we make separate estimates for the regions $0 \leq s \leq \frac{1}{2}$ and $s > \frac{1}{2}$. Note that for any number r for which $0 < r < 1$, the contour $s = r$ is a branch of a hyperbola with asymptotes $rp^2 - 2pq + rq^2 = 0$, as shown in Fig. 1. In particular,

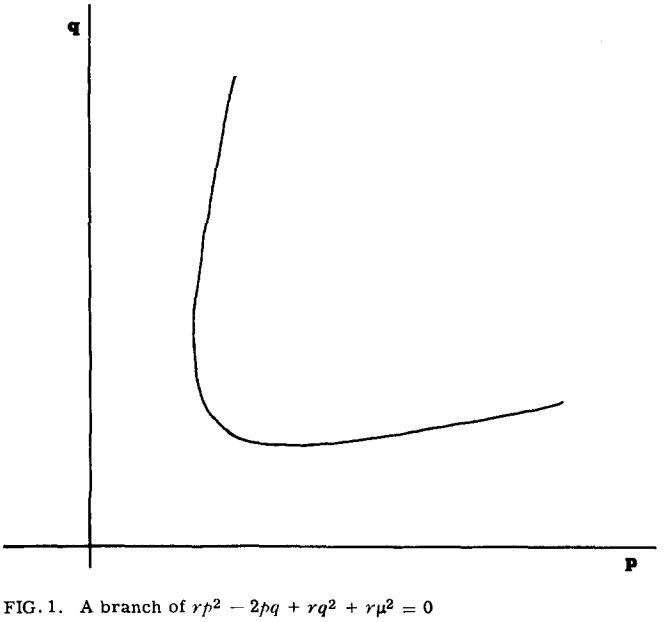


FIG. 1. A branch of $rp^2 - 2pq + rq^2 + r\mu^2 = 0$

the region $s > \frac{1}{2}$ lies between the lines $q = (2 \pm \sqrt{3})p$ so that $\frac{1}{4}p < q < 4p$ in this region. For all $p, q \geq 0$,

$$\frac{\partial V}{\partial p} = \frac{1}{p^2 + q^2 + \mu^2} \frac{\partial F}{\partial p} - \frac{2p}{(p^2 + q^2 + \mu^2)^2} F,$$

$$\frac{\partial F}{\partial p} = F' \frac{\partial s}{\partial p},$$

$$\frac{\partial s}{\partial p} = \frac{2q}{p^2 + q^2 + \mu^2} - \frac{4p^2 q}{(p^2 + q^2 + \mu^2)^2}.$$

Note that $|\partial s/\partial p| \leq 6/\sqrt{q^2 + \mu^2}$, for all $p, q \geq 0$.

The region $0 \leq s \leq \frac{1}{2}$: In this region F and F' are bounded, therefore $V(p, q) = O((q^2 + \mu^2)^{-1})$ and $\partial V(p, q)/\partial p = O((q^2 + \mu^2)^{-3/2})$ uniformly in p.

The region $\frac{1}{2} < s < 1$: In this region

$$\frac{1 + s}{1 - s} = \frac{(p + q)^2 + \mu^2}{(p - q)^2 + \mu^2} \leq \frac{25q^2 + \mu^2}{\mu^2}.$$

Hence $F(s) = O(1 + \log^+ q)$ and

$$V(p, q) = O((1 + \log^+ q)/(q^2 + \mu^2)) \quad \text{uniformly in } p.$$

Also, since

$$\left|\frac{\partial}{\partial p} \log\left(\frac{1 + s}{1 - s}\right)\right| = \left|\frac{2(p + q)}{(p + q)^2 + \mu^2} - \frac{2(p - q)}{(p - q)^2 + \mu^2}\right| \leq \frac{2}{\mu},$$

$$\frac{\partial F}{\partial p} = A'(s)\frac{\partial s}{\partial p} + B'(s)\frac{\partial s}{\partial p} \log\left(\frac{1 + s}{1 - s}\right) + B(s)\frac{\partial}{\partial p} \log\left(\frac{1 + s}{1 - s}\right) = O(1),$$

and $\partial V/\partial p = O((1 + \log^+ q)/(q^2 + \mu^2))$ uniformly in p.

Hence V satisfies (i). To verify (ii), observe that the region $0 \leq q \leq \Lambda$, s remains bounded and so do F, F', and F''. Simple calculations show

$$\frac{\partial V}{\partial q} = O\left(\frac{1}{p^3}\right) \quad \text{and} \quad \frac{\partial^2 V}{\partial p \partial q} = O\left(\frac{1}{p^4}\right)$$

uniformly in $0 \leq q \leq \Lambda$.

3. OPERATORS ON THE BANACH SPACE

Let C_1 be the space of all complex valued bounded, differentiable functions f on $[0, \infty)$ with bounded continuous derivative. Then C_1 is a Banach space under the norm

$$\|f\| = \sup |f(p)| + \sup |f'(p)|.$$

We shall show that the formula

$$(Tf)(p) = \int_0^\infty V(p, q)[q^2/(q^2 - k^2)]f(q)\,dq$$

defines a bounded linear operator T of C_1 into itself which can be approximated arbitrarily closely in norm by an operator of finite rank.

For any $\Lambda > 2k$, decompose Tf as

$$\begin{aligned}(Tf)(p) &= \int_0^\Lambda \frac{V(p, q) - V(p, k)}{q - k} \frac{q^2}{q + k} f(q)\,dq \\ &\quad + V(p, k) \int_0^\Lambda \frac{q^2}{q^2 - k^2} f(q)\,dq + \int_\Lambda^\infty V(p, q) \frac{q^2}{q^2 - k^2} f(q)\,dq \\ &= (T_1 f)(p) + (T_2 f)(p) + (T_3 f)(p),\end{aligned}$$

The existence for each f in C_1 of the principal value integral defining $T_2 f$ follows from the fact that f is differentiable at k. Further, since

$$\begin{aligned}\left|\int_0^\Lambda \frac{q^2}{q^2 - k^2} f(q)\,dq\right| &\leq \left|\int_0^\Lambda \frac{f(q) - f(k)}{q - k} \frac{q^2 dq}{q + k}\right| \\ &\quad + |f(k)| \left|\int_0^\Lambda \frac{q^2 dq}{q^2 - k^2} dq\right| \\ &\leq \|f\| \left(\int_0^\Lambda \frac{q^2 dq}{q + k} + \left|\int_0^\Lambda \frac{q^2 dq}{q^2 - k^2}\right|\right),\end{aligned}$$

and $V(q, k)$ is a C_1 function of q, T_2 is a bounded operator of rank 1 on C_1.

Next, given $\epsilon > 0$, choose $\Lambda > 2k$ so that $\int_\Lambda^\infty \Phi(q)\,dq < \epsilon$. Then

$$\left|(T_3 f)(p)\right| \leq \int_\Lambda^\infty |V(p, q)|\,2|f(q)|\,dq \leq 2\|f\|\epsilon$$

and

$$\left|(T_3 f)'(p)\right| \leq \int_\Lambda^\infty \left|\frac{\partial V(p, q)}{\partial p}\right| 2|f(q)|\,dq \leq 2\|f\|\epsilon,$$

so

$$\|T_3 f\| \leq 4\|f\|\epsilon,$$

and T_3 is a bounded linear operator on C_1 with $\|T_3\| \leq 4\epsilon$.

To complete the proof of the assertion about T, it is thus sufficient to prove that T_1 is a bounded operator on C_1 and that it can be approximated in norm by operators of finite rank.

Observe that the kernel

$$\begin{aligned}U(p, q) &= \frac{V(p, q) - V(p, k)}{q - k} \frac{q^2}{q + k} \quad (q \neq k) \\ &= \frac{\partial V(p, k)}{\partial q} \cdot \frac{k}{2} \quad (q = k)\end{aligned}$$

which defines T_1 has continuous first partial derivatives in $p \geq 0$, $0 \leq q \leq \Lambda$. For each such p and q, we can find q' and q'' in $[0, \Lambda]$ such that

$$U(p, q) = \frac{\partial V(p, q')}{\partial q} \frac{q^2}{q + k},$$

$$\frac{\partial U(p, q)}{\partial q} = \frac{\partial^2 V(p, q'')}{\partial p \partial q} \frac{q^2}{q + k},$$

so, by condition (ii), $U(p, q) \to 0$ and $\partial U(p, q)/\partial p \to 0$ as $p \to \infty$ uniformly for $0 \leq q \leq \Lambda$. In particular, U and $\partial U/\partial p$ are bounded in $0 \leq q \leq \Lambda$ and it follows easily that T_1 is a bounded operator on C_1.

We next show that, given $\epsilon > 0$, one can choose functions $r_1 \cdots r_N$ in C_1 and polynomials $q_1 \cdots q_N$ such that if

$$A(p, q) = \sum_{i=1}^N r_i(p)\, q_i(q)$$

then

$$\left.\begin{aligned}&|U(p, q) - A(p, q)| < \epsilon \\ &\left|\frac{\partial U(p, q)}{\partial p} - \frac{\partial A(p, q)}{\partial p}\right| < \epsilon\end{aligned}\right\} \quad \text{for all } p \geq 0 \text{ and } 0 \leq q \leq \Lambda. \tag{1}$$

First choose $X > 1$ so that $|U(p, q)| < \epsilon/9$ and $|\partial U(p, q)/\partial p| < \epsilon/9$ for $p \geq X$ and $0 \leq q \leq \Lambda$.

By the Weierstrass theorem, we can find polynomials $p_1 \cdots p_N$; $q_1 \cdots q_N$ such that if

$$B(p, q) = \sum_{i=1}^N p_i(p)\, q_i(q),$$

then

$$\left| \frac{\partial U(p,q)}{\partial p} - B(p,q) \right| < \frac{\epsilon}{9X}$$

for $0 \leq p \leq X+1$ and $0 \leq q \leq \Lambda$. Put $P_i(p) = \int_X^p p_i(t)\,dt$ and

$$C(p,q) = \sum_{i=1}^{N} P_i(p)\, q_i(q)$$

so that $\partial C/\partial p = B$. Also,

$$U(p,q) - C(p,q) = \int_X^p \left(\frac{\partial U(t,q)}{\partial p} - B(t,q) \right) dt + U(X,q)$$

so $|U(p,q) - C(p,q)| < 2\epsilon/9$ for $0 \leq p \leq X+1$ and $0 \leq q \leq \Lambda$.

Now choose any element s of C_1 such that $s(p) = 1$ for $0 \leq p \leq X$, $s(p) = 0$ for $p \geq X+1$ and $0 \leq s(p) \leq 1$ and $|s'(p)| \leq 2$ for all $p \geq 0$. Define $r_i = sP_i$ so that $A = sC$.

For $0 \leq p \leq X$, $A = C$ and $\partial A/\partial p = B$ so that the inequalities (1) are satisfied for $0 \leq q \leq \Lambda$.

For $X < p < X+1$ and $0 \leq q \leq \Lambda$, we have both $|U| < \epsilon/9$ and $|U\text{-}C| < 2(\epsilon/9)$, so $|C| < \epsilon/3$. Similarly, $|B| < 2(\epsilon/9)$. Hence $|U\text{-}A| \leq |U| + |A| \leq |U| + |C| \leq 4(\epsilon/9)$. Also, $\partial A/\partial p = sB + s'C$, so

$$\left| \frac{\partial A}{\partial p} \right| < |B| + 2\,|C| < \frac{8\epsilon}{9}$$

and

$$\left| \frac{\partial U}{\partial p} - \frac{\partial A}{\partial p} \right| \leq \left| \frac{\partial U}{\partial p} \right| + \left| \frac{\partial A}{\partial p} \right| < \epsilon.$$

For $p \geq X+1$, $A = \partial A/\partial p = 0$ so again (1) holds for $0 \leq q \leq \Lambda$ by choice of X.

Now consider the operator Q on C_1 defined by

$$(Qf)(p) = \int_0^\Lambda A(p,q) f(q)\,dq = \sum_{i=1}^{N} \left(\int_0^\Lambda q_i(q) f(q)\,dq \right) r_i(p)$$

Clearly, Q is a bounded operator of finite rank on C_1. Also, $\| T_1\text{-}Q \| \leq 2\epsilon\Lambda$ which completes the proof.

We now turn to solving the equation

$$f = g\phi + (2/\pi)\, gTf.$$

Given any $\epsilon > 0$, we may express $T = Q + N$, where Q is an operator of finite rank and $\|N\| < \epsilon$.

The equation can be written

$$[1 - (2/\pi)\, gN]f = g\phi + (2/\pi)\, gQf.$$

Now for $|g| < \pi/2\epsilon$, $[1 - (2/\pi)\, gN]^{-1}$ exists as a bounded operator on C_1, so we solve

$$f = [1 - (2/\pi)\, gN]^{-1} g\phi + (2/\pi)\, g[1 - (2/\pi)\, gN]^{-1} Qf.$$

This can be done by solving a finite system of linear equations and yields f as a meromorphic function of g in $|g| < \pi/(2\epsilon)$. Since ϵ is arbitrary, f is meromorphic in the whole g plane. Also, since $(1 - 2gT/\pi)^{-1}$ exists for $|g| < \pi/2\|T\|$, f is analytic at $g = 0$. Thus Pommerenke's theorem and an inspection of the proof of Beardon's theorem imply, respectively:

(1) Let $\lambda \geq 1$. If $[M,N]f$ is any sequence of Padé approximants to the Neumann series for f with $\lambda^{-1} \leq N/M \leq \lambda$ and D is any compact set, then $[M,N]f$ converges to f in capacity on D as $M \to \infty$. (By convention, M is the degree of the denominator and N is the degree of the numerator of the approximants.) More explicitly, for each $\epsilon > 0$, the capacity of the set $\{g \epsilon D: |[M,N]f(g) - f(g)| \geq \epsilon\}$ tends to zero as $M \to \infty$. The definition of capacity is given by Pommerenke and Hille.[17]

(2) If $[M,N]f$ is any sequence of Padé approximants to the Neumann series for f with $\alpha N \geq M$ for some $\alpha > 0$ and D is any compact set which contains no poles of f or limit points of the poles of the approximants, then $[M,N]f$ converges to f uniformly on D as $M \to \infty$.

We also observe that this method applies to the equation for the T matrix

$$t(p,k') = V(p,k') - \frac{2}{\pi} \int_0^\infty V(p,q) \frac{q^2\,dq}{k^2 - q^2 + i\epsilon}\, t(q,k')$$
$$= V(p,k') + ikV(p,k')\, t(k,k')$$
$$- \frac{2P}{\pi} \int_0^\infty V(p,q)\, \frac{q^2\,dq}{k^2 - q^2}\, t(q,k').$$

We should redefine T so as to include the extra rank 1 operator, and the proof is otherwise unchanged.

[1] B. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950).
[2] S. Caser, C. Piquet, and J. L. Vermeulen, Nucl. Phys. **14**, 119 (1969).
[3] P. R. Graves-Morris, Nuovo Cimento A **4**, 91 (1971).
[4] R. Blankenbecler and R. Sugar, Phys. Rev. **142**, 1051 (1966).
[5] C. Lovelace, Phys. Rev. **135**, B1225 (1964).
[6] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953), p. 38.
[7] M. Scadron, S. Weinberg, and J. Wright, Phys. Rev. **135**, B202 (1964).
[8] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955).
[9] J. G. Taylor, Nuovo Cimento Suppl. 1, 1002 (1963); P. R. Graves-Morris, Phys. Rev. Lett. **16**, 201 (1966); Erratum p. 494; M. M. Broido and J. G. Taylor, J. Math. Phys. **10**, 184 (1969).
[10] H. P. Noyes, Phys. Rev. Lett. **15**, 538 (1965). K. L. Kowalski, Phys. Rev. Lett. **15** 798 (1965).
[11] T. W. Osborne, Phys. Rev. D **3**, 395 (1971); P. R. Graves-Morris, Nuovo Cimento A **4**, 535 (1971).
[12] H. Poincaré, Acta Math. **4**, 201 (1884).
[13] D. Masson, J. Math. Phys. **8**, 2308 (1967).
[14] R. W. Haymaker and L. Schlessinger, in *The Padé Approximant in Theoretical Physics*, edited by G. A. Baker and J. L. Gammel (Academic, New York, 1970), p. 257 and references therein.
[15] G. A. Baker, *Advances in Theoretical Physics*, edited by K. A. Bruckner (Academic, New York, 1965), Vol. 1, p. 1.
[16] J. S. R. Chisholm, J. Math. Phys. **4**, 1506 (1963).
[17] A. F. Beardon, J. Math. Anal. Appl. **21**, 344 (1968); Ch. Pommerenke, J. Math. Anal. Appl. (to be published); E. Hille, *Analytic Function Theory* (Ginn, London, 1959), Vol. II.
[18] J. Nuttall, J. Math. Anal. Appl. **31**, 147 (1970).
[19] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience, New York, 1967), p. 579
[20] C. R. Garibotti and M. Villani, Nuovo Cimento A **59**, 107 (1969); C. Lanczos, J. Res. Natl. Bur. Stand. **45**, 255 (1950).
[21] *Handbook of Mathematical Functions*, edited by M. Abromowitz and I. Stegun (Dover, New York, 1965), p. 332, Eq. 8-1-3.
[22] See Ref. 21, p. 334, Eq. 8-6-19.

Multipole expansions and plane wave representations of the electromagnetic field*

A. J. Devaney and E. Wolf

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 31 May 1973; revised manuscript received 2 August 1973)

A new and conceptually simple derivation is presented of the multipole expansion of an electromagnetic field that is generated by a localized, monochromatic charge-current distribution. The derivation is obtained with the help of a generalized plane wave representation (known also as the angular spectrum representation) of the field. This representation contains both ordinary plane waves, and plane waves that decay exponentially in amplitude as the wave is propagated. The analysis reveals an intimate relationship between the generalized plane wave representation and the multipole expansion of the field and leads to a number of new results. In particular, new expressions are obtained for the electric and magnetic multipole moments in terms of certain components of the spatial Fourier transform of the transverse part of the current distribution. It is shown further that the electromagnetic field at all points outside a sphere that contains the charge-current distribution is completely specified by the radiation pattern (i.e., by the field in the far zone). Explicit formulas are obtained for all the multipole moments in terms of the radiation pattern.

1. INTRODUCTION

Multipole expansions of the electromagnetic field are employed extensively both in classical electrodynamics and in the quantum theory of radiation. Such expansions originated in a restrictive form in the classical theory of diffraction by a sphere, especially in the work of Clebsch,[1] Mie,[2] Debye,[3] and Bromwich,[4] as a "partial wave" expansion.[5] The first fairly general formulation is implicit in the work of Hansen.[6] Important contributions were later made by Heitler,[7] Kramers,[8] Franz,[9] Wallace,[10] Blatt and Weisskopf,[11] Bouwkamp and Casimir,[12] Nisbet,[13] and Wilcox.[14] The methods employed in some of the treatments include tensor analysis, operator calculus, and group theory, though more elementary derivations have been given.[10,12]

In a well-known paper dealing with the differential equations of mathematical physics, Whittaker[15] introduced a multipole expansion of a *source free* scalar wave field. He first expressed the solution of the homogeneous wave equation as a superposition of homogeneous plane waves. On expanding the amplitude function of the plane waves in a series of spherical harmonics he was then lead to a multipole expansion of the free field. Whittaker's analysis is conceptually very simple, and gives a clear understanding of the relationship between the representation of a free field as a superposition of plane waves and its representation in terms of multipole fields.

In the present paper, we show that Whittaker's method may be extended to electromagnetic fields generated by localized charge-current distributions. To illustrate the essential feature of the technique, we first show how a generalized plane wave expansion for a scalar field generated by a localized source distribution may be obtained. The generalized plane wave expansion contains not only ordinary (homogeneous) waves, but also waves that decay in amplitude as the wave is propagated (the so-called evanescent waves, well known from the theory of total internal reflection). Such generalized plane wave expansions have been playing an increasingly important role in recent years in optics and classical electrodynamics, and are generally known as angular spectrum representations.[16] We show that the amplitudes of all the plane waves in this representation are expressible in terms of the spatial Fourier transform, and its analytic continuation, of the source distribution. By expanding the amplitude function of the plane waves in a series of spherical harmonics we are immediately lead to the required multipole expansion. In Sec. 4 we consider the electromagnetic field generated by a localized charge-current distribution. The only essential difference arises from the fact that the amplitudes in the generalized plane wave expansion of the electromagnetic field are now vectors, and the appropriate basic set of functions in terms of which they are expanded are the vector spherical harmonics instead of the ordinary spherical harmonics. In the concluding section (Sec. 5), we show that the angular spectrum representation provides also a new insight into the well-known relationship between the multipole expansion and the so-called Debye potential representation of the electromagnetic field generated by a localized charge-current distribution.

Our analysis not only provides a new derivation of the multipole expansion of the electromagnetic field, but it also reveals an intimate relationship between the multipole expansion and the generalized plane wave representation of the field. From this fact some new results readily follow. In particular, we obtain new expressions for the electric and magnetic multipole moments in terms of certain components of the spatial Fourier transform of the transverse part of the current distribution. We also show that the electromagnetic field at all points outside a sphere that encloses the source distribution is completely specified by the radiation pattern (the far field), and we derive formulas for all the multipole moments in terms of the radiation pattern. Moreover, as we show elsewhere, our results lead readily to some interesting new theorems on properties of fields generated by localized charge-current distributions and on localized charge-current distributions that do not give rise to any radiation.[17]

2. WHITTAKER'S REPRESENTATION OF A SOURCE-FREE MONOCHROMATIC SCALAR WAVE FIELD

We begin with a brief review of Whittaker's derivation of the multipole expansion of a source-free, monochromatic scalar wave field.

Whittaker[15] showed that a wide class of solutions of the reduced wave equation (the Helmholtz equation),

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0 \tag{2.1}$$

(k = real constant), may be expressed in the form

$$\psi(\mathbf{r}) = \frac{k}{4\pi} \int_{-\pi}^{\pi} d\beta \int_{0}^{\pi} d\alpha \, \sin\alpha \hat{\psi}(\mathbf{s}) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{2.2}$$

where $\mathbf{s}$ is a real unit vector with Cartesian components

$$s_x = \sin\alpha \cos\beta, \quad s_y = \sin\alpha \sin\beta, \quad s_z = \cos\alpha. \tag{2.3}$$

The right-hand side of (2.2) expresses $\psi(\mathbf{r})$ as a superposition of homogeneous plane waves propagating in all possible directions and, hence, we will refer to it as the homogeneous plane wave expansion of $\psi(\mathbf{r})$.

Whittaker showed further, that the homogeneous plane wave expansion of the multipole field[18]

$$\Lambda_l^m(\mathbf{r}) = j_l(kr) Y_l^m(\theta, \phi), \tag{2.4}$$

where $j_l(kr)$ is the spherical Bessel function of order l and $Y_l^m(\theta, \phi)$ is the spherical harmonic of degree l and order m, is[19]

$$\Lambda_l^m(\mathbf{r}) = (-i)^l \frac{1}{4\pi} \int_{-\pi}^{\pi} d\beta \int_{0}^{\pi} d\alpha \, \sin\alpha \, Y_l^m(\alpha, \beta) e^{ik\mathbf{s}\cdot\mathbf{r}}. \tag{2.5}$$

Here r, θ, ϕ are the spherical polar coordinates of the field point $\mathbf{r}$, referred to the same system of axes as the unit vector $\mathbf{s}$, so that

$$x = r \sin\theta \cos\phi, \quad y = r \sin\theta \sin\phi, \quad z = r \cos\theta. \tag{2.6}$$

Returning to the general case, Whittaker expanded the plane wave amplitude function $\hat{\psi}(\mathbf{s})$ in a series of spherical harmonics,

$$\hat{\psi}(\mathbf{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-i)^l a_l^m Y_l^m(\alpha, \beta), \tag{2.7}$$

and on substituting from (2.7) into (2.2) and using (2.5) he obtained an alternative representation of $\psi(\mathbf{r})$, namely the *multipole expansion*

$$\psi(\mathbf{r}) = k \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_l^m \Lambda_l^m(\mathbf{r}). \tag{2.8}$$

The *multipole moments* a_l^m may, of course, be expressed from (2.7) in terms of the plane wave amplitudes $\hat{\psi}(\mathbf{s})$ by using the fact that the spherical harmonics form a orthonormal set over the unit sphere. The result is

$$a_l^m = i^l \int_{-\pi}^{\pi} d\beta \int_{0}^{\pi} d\alpha \, \sin\alpha \, \hat{\psi}(\mathbf{s}) Y_l^{m*}(\alpha, \beta), \tag{2.9}$$

and shows that the multipole moments are, apart from the trivial factor i^l, simply the projections of the plane wave amplitudes $\hat{\psi}(\mathbf{s})$ onto the set of the spherical harmonics $Y_l^m(\alpha, \beta)$.

It should be noted that since each plane wave in (2.2), and each multipole field in (2.8), obey the Helmholtz equation (2.1) throughout the whole space, each of the two representations is a *mode expansion* of the general solution of that equation. The relations (2.7) and (2.9) establish an intimate connection between the two representations. Unfortunately, Whittaker's elegant results are of limited applicability since they are essentially existence theorems. They leave unanswered the question of how the amplitudes $\hat{\psi}(\mathbf{s})$, or the multipole moments a_l^m, are to be determined in any particular case. Moreover, the representation (2.2) [and, consequently, (2.8)] is valid only for source-free fields. Consequently, Whittaker's derivation of the multipole expansion is of limited use in electromagnetic theory where one must frequently deal with fields generated by sources that are not situated at infinity.

In this paper we will show that the insight gained from Whittaker's analysis leads readily to a new, and basically simple, derivation of the multipole expansion of wavefields generated by a localized source distribution, and also to explicit expressions for the multipole moments. The derivation is based on a generalized plane wave expansion which includes not only homogeneous plane waves, but also evanescent plane waves (i.e., plane waves that decay exponentially in amplitude in a particular direction). In Sec. 3 we consider the scalar field and in Sec. 4 we will treat the electromagnetic field.

3. MULTIPOLE EXPANSION OF A MONOCHROMATIC SCALAR WAVE FIELD GENERATED BY A LOCALIZED SOURCE DISTRIBUTION

Let us consider a real, monochromatic scalar wave field

$$\psi(\mathbf{r}, t) = \Re(\psi(\mathbf{r}) e^{-i\omega t}),$$

generated by a source distribution

$$\rho(\mathbf{r}, t) = \Re(\rho(\mathbf{r}) e^{-i\omega t}),$$

in the infinite free space. Here, ω is a real positive constant and $\Re$ denotes the real part. Then $\psi(\mathbf{r})$ and $\rho(\mathbf{r})$ are related by the differential equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \tag{3.1}$$

where

$$k = \omega/c, \tag{3.2}$$

c being the velocity of propagation. We will assume that the source distribution $\rho(\mathbf{r})$ is a continuous function of position and is confined to a finite region around the origin. Hence, $\rho(\mathbf{r}) \equiv 0$ when $r \equiv |\mathbf{r}| > R$, where R is some real constant.

The field $\psi(\mathbf{r})$ generated by the source distribution $\rho(\mathbf{r})$ is identified with that particular solution of Eq. (3.1) which behaves at infinity as an outgoing spherical wave, and is well known to be given by

$$\psi(\mathbf{r}) = \int_{r' \le R} \rho(\mathbf{r}') \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} d^3r'. \tag{3.3}$$

Now, the spherical wave which enters as the kernel of the integral transform (3.3) may be expressed in the following form due to Weyl:[20]

$$\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} = \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \, \sin\alpha \, e^{ik\mathbf{s}\cdot(\mathbf{r}-\mathbf{r}')}, \tag{3.4}$$

where $\mathbf{s} = \mathbf{s}(\alpha, \beta)$ is again a unit vector with Cartesian components given by (2.3). However, the polar angle α is no longer necessarily real, but takes on all values on the contours C^+ and C^- in the complex α plane shown in Fig. 1. It is understood that in (3.4) the contour C^+ applies when $z - z' > 0$, and the contour C^- applies when $z - z' < 0$, z and z', being, of course, the Cartesian z coordinates of the field point $\mathbf{r}$ and the integration point $\mathbf{r}'$, respectively.

On substituting from (3.4) into (3.3), and on interchanging the order of integration, we obtain the following representation of $\psi(\mathbf{r})$:[21]

$$\psi(\mathbf{r}) = \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \, \sin\alpha \, \hat{\psi}(\mathbf{s}) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{3.5}$$

where the spectral amplitudes $\hat{\psi}(\mathbf{s})$ are given in terms of the source distribution by the formula

$$\hat{\psi}(\mathbf{s}) = \int_{r' \le R} \rho(\mathbf{r}') e^{-ik\mathbf{s}\cdot\mathbf{r}'} d^3r'. \tag{3.6}$$

In (3.5), the contour C^+ is used when $z > R$ and C^- when $z < -R$. It is not difficult to show that the interchange of the order of integration in deriving (3.5) is justified provided $|z| > R$, i.e., provided the field point is outside a strip bounded by planes parallel to the z plane and containing the source. However, since the orientation of our coordinate system is arbitrary, such a representation may be used to represent the field outside *any* strip bounded by two parallel planes tangential to the sphere of radius R, centered at the origin.

The representation (3.5), just like Whittaker's representation (2.2) of the source-free field, expresses $\psi(\mathbf{r})$ as a superposition of plane waves. However, since in (3.5) the unit vector $\mathbf{s}$ is real on a portion of the contour of integration and complex on the rest of the contour, (3.5) includes both *homogeneous* plane waves (corresponding to real $\mathbf{s}$) and *inhomogeneous* (called also evanescent) plane waves (corresponding to complex[22] $\mathbf{s}$). Integrals of this type are said to represent the wave field in the form of an *angular spectrum of plane waves*.

The spectral amplitudes $\hat{\psi}(\mathbf{s})$ as defined in (3.6) are intimately related to the threefold Fourier transform $\tilde{\rho}(\mathbf{K})$ of the source distribution $\rho(\mathbf{r})$:

$$\tilde{\rho}(\mathbf{K}) = \int_{r' \leq R} \rho(\mathbf{r}') e^{-i\mathbf{K}\cdot\mathbf{r}'} d^3r'. \tag{3.7}$$

On comparing (3.6) with (3.7), we conclude that $\hat{\psi}(\mathbf{s})$ is simply $\tilde{\rho}(\mathbf{K})$ with $\mathbf{K} = k\mathbf{s}$; i.e.,

$$\hat{\psi}(\mathbf{s}) = \tilde{\rho}(k\mathbf{s}). \tag{3.8}$$

In the definition of the Fourier transform $\tilde{\rho}(\mathbf{K})$, the variable $\mathbf{K}$ conjugate to $\mathbf{r}'$ is, of course, real. However, since the integral on the right-hand side of (3.7) extends over a finite domain (because, by hypothesis, $\rho \equiv 0$ when $r' > R$) and, moreover, since ρ was assumed to be continuous, it follows that $\tilde{\rho}(\mathbf{K})$ is the boundary value, on the real K_x, K_y, K_z axes, of an entire analytic function of three complex variables[23] K_x, K_y, K_z. Consequently, the relation (3.8) is valid for all (real and complex) unit vectors $\mathbf{s}$ and, moreover, shows that $\hat{\psi}$ is the boundary value (on the contours C^+ and C^-) of an entire analytic function of three complex variables.

Following Whittaker's treatment of the free field, we next expand the spectral amplitudes $\hat{\psi}(\mathbf{s})$ into a series of spherical harmonics, viz.,

$$\hat{\psi}(\mathbf{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-i)^l a_l^m Y_l^m(\alpha, \beta), \tag{3.9}$$

where the expansion coefficients (multipole moments) a_l^m are simply the projections of $\hat{\psi}(\mathbf{s})$ onto the spherical harmonics:

$$a_l^m = i^l \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \, \sin\alpha \, \hat{\psi}(\mathbf{s}) Y_l^{m*}(\alpha, \beta). \tag{3.10}$$

It is important to note that although the multipole moments, defined by (3.10), depend explicitly only on those spectral amplitudes $\hat{\psi}(\mathbf{s})$ which are associated with real $\mathbf{s}$ (i.e., those corresponding to homogeneous plane waves in the angular spectrum representation), the expansion (3.9) is valid for all unit vectors associated with the complex contours $C^{\pm}$; this result is a consequence of the fact that $\hat{\psi}(\mathbf{s})$ is the boundary value of an entire analytic function.

We can also readily express the multipole moments a_l^m in terms of the source distribution $\rho(\mathbf{r})$. To do this, we simply substitute (3.6) into (3.10) and interchange the order of integration. We then obtain the following expressions for the multipole moments:

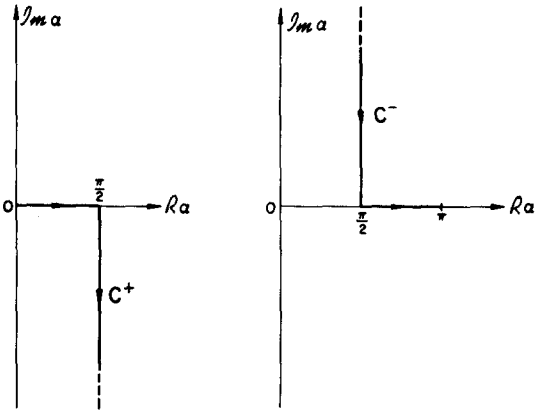


FIG. 1. The α-contours of integration C^+ and C^-.

$$a_l^m = \int_{r' \leq R} d^3r' \rho(\mathbf{r}') i^l \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \, \sin\alpha \, Y_l^{m*}(\alpha, \beta) e^{-ik\mathbf{s}\cdot\mathbf{r}'},$$
$$= 4\pi \int_{r' \leq R} \rho(\mathbf{r}') \Lambda_l^{m*}(\mathbf{r}') d^3r'. \tag{3.11}$$

In deriving (3.11) we have made use of (2.5).

Next, we substitute the expansion (3.9) into the representation (3.5) of $\psi(\mathbf{r})$ and interchange the order of integration and summation. We than obtain, in analogy with (2.8), the following series expansion for $\psi(\mathbf{r})$:

$$\psi(\mathbf{r}) = k \sum_{l=0}^{\infty} \sum_{l=-m}^{m} a_l^m \Pi_l^m(\mathbf{r}), \tag{3.12}$$

where

$$\Pi_l^m(\mathbf{r}) = (-i)^l \frac{i}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \, \sin\alpha \, Y_l^m(\alpha, \beta) e^{ik\mathbf{s}\cdot\mathbf{r}}. \tag{3.13}$$

Now, it is not difficult to show that the expression on the right-hand side of (3.13) is precisely the angular spectrum representation of the scalar multipole field of degree l and order m, i.e.,

$$\Pi_l^m(\mathbf{r}) = h_l^{(+)}(kr) Y_l^m(\theta, \phi), \tag{3.14}$$

where (r, θ, ϕ) are the spherical polar coordinates of the field point $\mathbf{r}$ and $h_l^{(+)}$ is the spherical Hankel function[19] of the first kind of order l. This result, which is the counterpart of Whittaker's result expressed by (2.4), appears to have been first stated by Erdélyi,[24] and is proved in Appendix A of the present paper. Thus, we see that (3.12) is the *multipole expansion* of $\psi(\mathbf{r})$ and a_l^m are the corresponding *multipole moments*.

Since the angular spectrum representation (3.5), from which the multipole expansion (3.12) was derived, converges only throughout the two half-spaces $z > R$ and $z < -R$, it would appear that this expansion is valid only at field points $\mathbf{r}$ situated in these two regions. However, as is well known, the multipole expansion (3.12) represents the field correctly at every point outside the source region, i.e., for all values of $r > R$. For the sake of completeness, this result is verified in Appendix C.

The angular spectrum representation also yields readily the far zone approximation for $\psi(\mathbf{r})$. If we let $kr \to \infty$ in a fixed direction specified by a unit vector

$$\mathbf{u}_r = \mathbf{r}/r, \tag{3.15}$$

we have

$$|\mathbf{r} - \mathbf{r}'| \sim r - \mathbf{r}' \cdot \mathbf{u}_r, \tag{3.16}$$

and Eq. (3.3) then gives

$$\psi(r\mathbf{u}_r) \sim \tilde{\rho}(k\mathbf{u}_r) e^{ikr}/r \quad (kr \to \infty), \tag{3.17}$$

where, as before, $\tilde{\rho}(\mathbf{K})$ is the Fourier transform of $\rho(\mathbf{r})$ [Eq. (3.7)]. Now according to (3.8), $\tilde{\rho}(k\mathbf{u}_r)$ is precisely the spectral amplitude $\hat{\psi}(\mathbf{u}_r)$ of ψ, so that (3.17) may be expressed in the form

$$\psi(r\mathbf{u}_r) \sim \hat{\psi}(\mathbf{u}_r)(e^{ikr}/r) \quad (kr \to \infty). \tag{3.18}$$

This formula shows that the *radiation pattern of the field* is precisely given by the spectral amplitude function $\hat{\psi}(\mathbf{u}_r)$ for real unit vectors $\mathbf{u}_r$ (corresponding to points on the contours C^+ and C^- that coincide with portions of the *real* α axis). In other words, the radiation pattern of the field in any particular direction $\mathbf{u}_r$ is equal to the complex amplitude of a certain plane wave in the angular spectrum representation of the field, namely the plane wave propagated in the direction $\mathbf{u}_r$. This result may be understood in physical terms as follows: The angular spectrum representation (3.5) expresses $\psi(\mathbf{r})$ at every point $\mathbf{r}$ outside the source region $r \leq R$ as a superposition of plane waves, both homogeneous and evanescent ones. As the field point $\mathbf{r} = r\mathbf{u}_r$ gradually moves further away from the source region in any fixed direction $\mathbf{u}_r$, the evanescent waves gradually die out because of their exponential amplitude decay. The contribution of the homogeneous waves gradually decreases also, but for a different reason, namely because they progressively cancel each other out by destructive phase interference. In the asymptotic limit as $kr \to \infty$, only the single homogeneous plane wave in the angular spectrum that is propagated in the direction $\mathbf{u}_r$ survives, and it is this wave that determines the behavior of $\psi(\mathbf{r})$ in the far zone, as Eq. (3.18) shows. This argument may be made more rigorous with the help of the principle of stationary phase.[25]

If in Eq. (3.18) we express $\hat{\psi}(\mathbf{u}_r)$ in terms of the multipole moments [Eq. (3.9)] we obtain, at once, the asymptotic approximation for ψ, valid in the far zone, (i.e., as $kr \to \infty$):

$$\psi(r\mathbf{u}_r) \sim \frac{e^{ikr}}{r} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-i)^l a_l^m Y_l^m(\theta, \phi) \quad (kr \to \infty), \tag{3.19}$$

where (θ, ϕ) are the spherical polar coordinates of the unit direction vector $\mathbf{u}_r$.

Before concluding this section it is worthwhile to stress the following points: The angular spectrum representation (3.5), and the multipole expansion (3.12), are *mode expansions* in the sense that they express the field $\psi(\mathbf{r})$ in terms of certain elementary fields (plane wave fields and multipole fields, respectively), each of which satisfies the same equation as does $\psi(\mathbf{r})$ outside the source region, namely the Helmholtz equation $(\nabla^2 + k^2)\psi = 0$. The range of validity of each of the two expansions is different. The angular spectrum expansion represents $\psi(\mathbf{r})$ outside the strip $|z| \leq R$, the multipole expansion represents it outside the sphere $r \leq R$. The expansion coefficients in the two representations are related by Eqs. (3.9) and (3.10).

4. MULTIPOLE EXPANSION OF A MONOCHROMATIC ELECTROMAGNETIC FIELD GENERATED BY A LOCALIZED CHARGE-CURRENT DISTRIBUTION

Let us now turn our attention to the electromagnetic field. Just as in the scalar case, two situations are to be distinguished, namely when the field is source free and when it is generated by a localized source distribution. We will discuss here only the second case.[26]

We consider then a real, monochromatic, electromagnetic field

$$\mathbf{E}(\mathbf{r}, t) = \Re(\mathbf{E}(\mathbf{r})e^{-i\omega t}), \quad \mathbf{H}(\mathbf{r}, t) = \Re(\mathbf{H}(\mathbf{r})e^{-i\omega t}),$$

generated by a charge-current distribution[27]

$$\rho(\mathbf{r}, t) = \Re(\rho(\mathbf{r})e^{-i\omega t}), \quad \mathbf{j}(\mathbf{r}, t) = \Re(\mathbf{j}(\mathbf{r})e^{-i\omega t}),$$

in the infinite free space. As before, ω is a real positive constant and $\Re$ denotes the real part. We assume that $\rho(\mathbf{r})$ and $\mathbf{j}(\mathbf{r})$ are continuous and continuously differentiable functions of position and vanish identically when $r > R$, where R is some real constant. From the Maxwell equations it readily follows that $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ satisfy (in Gaussian system of units) the equations[28]

$$(\nabla^2 + k^2)\mathbf{E}(\mathbf{r}) = -4\pi[i(k/c)\mathbf{j}(\mathbf{r}) - \nabla\rho(\mathbf{r})], \tag{4.1a}$$

$$(\nabla^2 + k^2)\mathbf{H}(\mathbf{r}) = -4\pi[(1/c)\nabla \times \mathbf{j}(\mathbf{r})]. \tag{4.1b}$$

The fields $\mathbf{E}, \mathbf{H}$ generated by the charge-current distribution $\rho, \mathbf{j}$ are identified with those particular solutions of Eqs. (4.1a) and (4.1b) that behave at infinity as outgoing spherical waves.

It is clear from Eqs. (4.1) that each Cartesian component of the electromagnetic field vectors $\mathbf{E}, \mathbf{H}$ satisfy inhomogeneous Helmholtz equations of the form (3.1). Consequently, we may apply the results established in the preceding section to each of the Cartesian components. In particular, it follows by making use of (3.5) and (3.6) that the field vectors have the following angular spectrum representations,[29] valid throughout the two half-spaces $z < -R$ and $z > R$:

$$\mathbf{E}(\mathbf{r}) = \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha \, \sin\alpha \, \hat{\mathbf{E}}(\mathbf{s}) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{4.2a}$$

$$\mathbf{H}(\mathbf{r}) = \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha \, \sin\alpha \, \hat{\mathbf{H}}(\mathbf{s}) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{4.2b}$$

where the spectral amplitude vectors $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ are given by

$$\hat{\mathbf{E}}(\mathbf{s}) = \int_{r' \leq R} [(ik/c)\mathbf{j}(\mathbf{r}') - \nabla\rho(\mathbf{r}')] e^{-ik\mathbf{s}\cdot\mathbf{r}'} d^3r', \tag{4.3a}$$

$$\hat{\mathbf{H}}(\mathbf{s}) = \int_{r' \leq R} [(1/c)\nabla \times \mathbf{j}(\mathbf{r}')] e^{-ik\mathbf{s}\cdot\mathbf{r}'} d^3r'. \tag{4.3b}$$

We may express $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ in simpler forms by introducing the three-dimensional Fourier transforms of $\mathbf{j}$ and ρ:

$$\tilde{\mathbf{j}}(\mathbf{K}) = \int_{r' \leq R} \mathbf{j}(\mathbf{r}') \, e^{-i\mathbf{K}\cdot\mathbf{r}'} \, d^3r', \tag{4.4a}$$

$$\tilde{\rho}(\mathbf{K}) = \int_{r' \leq R} \rho(\mathbf{r}') \, e^{-i\mathbf{K}\cdot\mathbf{r}'} d^3r'. \tag{4.4b}$$

Then (4.3a) and (4.3b) readily give the following expressions for the spectral amplitudes:

$$\hat{\mathbf{E}}(\mathbf{s}) = -(i/c)k\mathbf{s} \times [\mathbf{s} \times \tilde{\mathbf{j}}(k\mathbf{s})], \tag{4.5a}$$

$$\hat{\mathbf{H}}(\mathbf{s}) = (i/c)k\mathbf{s} \times \tilde{\mathbf{j}}(k\mathbf{s}). \tag{4.5b}$$

In deriving Eq. (4.5a) from Eq. (4.3a) we made use of the equation $\nabla \cdot \mathbf{j}(\mathbf{r}) - ick\rho(\mathbf{r}) = 0$ which expresses the conservation of the charge. From Eqs. (4.5) it follows that

$$\hat{\mathbf{E}}(\mathbf{s}) = -\mathbf{s} \times \hat{\mathbf{H}}(\mathbf{s}), \tag{4.6a}$$

$$\mathbf{s} \cdot \hat{\mathbf{E}}(\mathbf{s}) = \mathbf{s} \cdot \hat{\mathbf{H}}(\mathbf{s}) = 0. \tag{4.6b}$$

Consequently, for each $\mathbf{s}$, the terms in the integrands of (4.2a) and (4.2b) are plane waves that satisfy the homogeneous Maxwell equations for a monochromatic field

with the frequency $\omega = kc$ throughout the whole space. Thus, the angular spectrum representation (4.2) is, throughout its domain of validity ($|z| > R$), a mode expansion of the electromagnetic field. We note in passing that, in analogy with the scalar case, the spectral amplitudes $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ are, according to Eqs. (4.5), simply related to the spatial Fourier transform, and its analytic continuation, of the current distribution $\mathbf{j}(\mathbf{r})$.

We could now follow the procedure employed in Sec. 3 to expand $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ into a series of spherical harmonics with constant (vector) coefficients. If these expansions were then substituted into Eqs. (4.2), and use was made of Eq. (3.13), we would obtain expansions of the two field vectors as series of the *scalar* multipole fields $\Pi_l^m(\mathbf{r})$. Unfortunately, although such expansions have been discussed in the literature[30] they are not very useful. One reason for this is that the individual terms in the expansion need not have a vanishing divergence and, consequently, such a representation will not provide a mode expansion of the electromagnetic field outside the source region. It is well known that the proper generalization, which is a mode expansion, is an expansion in terms of so-called *electromagnetic multipole fields* [defined by Eqs. (4.15) below], these being the appropriate electromagnetic vector analogs of the scalar multipole fields.

To obtain the expansions for $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ in terms of the electromagnetic multipole fields, we first expand the spectral amplitudes $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ in terms of the *vector* spherical harmonics. A vector spherical harmonic $\mathbf{Y}_l^m(\alpha, \beta)$ of degree l and order m may be defined in terms of the ordinary spherical harmonics $Y_l^m(\alpha, \beta)$ by means of the formula

$$\mathbf{Y}_l^m(\alpha, \beta) = \mathcal{L}_s Y_l^m(\alpha, \beta), \tag{4.7}$$

where $\mathcal{L}_s$ is the "orbital angular momentum operator:

$$\mathcal{L}_s = -ik\mathbf{s} \times \nabla_{ks} = -i\left(\mathbf{u}_\beta \frac{\partial}{\partial \alpha} - \frac{1}{\sin\alpha}\mathbf{u}_\alpha \frac{\partial}{\partial \beta}\right). \tag{4.8}$$

Here ∇_{ks} denotes the gradient operator in $k\mathbf{s}$ space, $\mathbf{u}_\alpha$, $\mathbf{u}_\beta$ being unit vectors in the positive α and β directions, respectively. A discussion of the main properties of these functions can be found, in Refs. 11 and 31. Here, we only remark that the vector spherical harmonics $\mathbf{Y}_l^m$ are everywhere tangent to the unit sphere [i.e., $\mathbf{s} \cdot \mathbf{Y}_l^m(\alpha, \beta) = 0$], that they form an orthogonal set in the sense that

$$\int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin\alpha\, \mathbf{Y}_l^{m*}(\alpha, \beta) \cdot \mathbf{Y}_{l'}^{m'}(\alpha, \beta) = l(l+1)\delta_{l,l'}\ \delta_{m,m'}, \tag{4.9}$$

and that they, *together with the associated functions* $\mathbf{s} \times \mathbf{Y}_l^m(\alpha, \beta)$, form a complete orthogonal basis[32] for all well-behaved vector functions $\mathbf{F}(\mathbf{s})$ defined on the unit sphere $s^2 = 1$ and tangential to it (i.e., such that $\mathbf{s} \cdot \mathbf{F}(\mathbf{s}) = 0$). Thus, in particular, $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ defined by Eqs. (4.5) may each be expanded in terms of these two sets of vector functions and we have, in analogy with (3.9):

$$\hat{\mathbf{E}}(\mathbf{s}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} (-i)^l [a_l^m\, \mathbf{s} \times \mathbf{Y}_l^m(\alpha, \beta) + b_l^m \mathbf{Y}_l^m(\alpha, \beta)], \tag{4.10a}$$

$$\hat{\mathbf{H}}(\mathbf{s}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} (-i)^l [-a_l^m \mathbf{Y}_l^m(\alpha, \beta) + b_l^m\, \mathbf{s} \times \mathbf{Y}_l^m(\alpha, \beta)], \tag{4.10b}$$

where we have made use of the relation (4.6a). The factor $(-i)^l$ is included in these expansions to lead to the conventional definition of the expansion coefficients (the multipole moments). The l summations begin now with $l = 1$ rather than with $l = 0$, since there are no vector spherical harmonics of zero degree.

By making use of the orthogonality relation (4.9) we may express the expansion coefficients a_l^m and b_l^m in terms of the spectral amplitudes $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ of the angular spectrum representation (4.2). The result is

$$a_l^m = -\frac{i^l}{l(l+1)} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin\alpha\, \hat{\mathbf{H}}(\mathbf{s}) \cdot \mathbf{Y}_l^{m*}(\alpha, \beta), \tag{4.11a}$$

$$b_l^m = \frac{i^l}{l(l+1)} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin\alpha\, \hat{\mathbf{E}}(\mathbf{s}) \cdot \mathbf{Y}_l^{m*}(\alpha, \beta). \tag{4.11b}$$

We may also express a_l^m and b_l^m in terms of the Fourier transform of the current density itself by substituting into (4.11) from (4.5). We then obtain the following expressions[33] for a_l^m and b_l^m:

$$a_l^m = -\frac{i^{l+1}}{l(l+1)}\left(\frac{k}{c}\right) \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin\alpha [\mathbf{s} \times \tilde{\mathbf{j}}(k\mathbf{s})] \cdot \mathbf{Y}_l^{m*}(\alpha, \beta), \tag{4.12a}$$

$$b_l^m = -\frac{i^{l+1}}{l(l+1)}\left(\frac{k}{c}\right) \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin\alpha \{\mathbf{s} \times [\mathbf{s} \times \tilde{\mathbf{j}}(k\mathbf{s})]\} \cdot \mathbf{Y}_l^{m*}(\alpha, \beta). \tag{4.12b}$$

Again, by analogy with the procedure that we employed in our treatment of the scalar case in Sec. 3, we substitute the expansions (4.10) of the spectral amplitude vectors $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ into the angular spectrum representations (4.2) of the fields, interchange the order of integration and summation and we then obtain the following series expansions for $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$:

$$\mathbf{E}(\mathbf{r}) = \sum_{l=1}^{\infty} \sum_{m=-1}^{l} [a_l^m \mathbf{E}_{lm}^e(\mathbf{r}) + b_l^m \mathbf{E}_{lm}^h(\mathbf{r})], \tag{4.13a}$$

$$\mathbf{H}(\mathbf{r}) = \sum_{l=1}^{\infty} \sum_{m=-1}^{l} [a_l^m \mathbf{H}_{lm}^e(\mathbf{r}) + b_l^m \mathbf{H}_{lm}^h(\mathbf{r})], \tag{4.13b}$$

where

$$\mathbf{E}_{lm}^e(\mathbf{r}) = \mathbf{H}_{lm}^h(\mathbf{r}) = (-i)^l \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha \sin\alpha [\mathbf{s} \times \mathbf{Y}_l^m(\alpha, \beta)] e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{4.14a}$$

$$\mathbf{E}_{lm}^h(\mathbf{r}) = -\mathbf{H}_{lm}^e(\mathbf{r}) = (-i)^l \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha \sin\alpha\, \mathbf{Y}_l^m(\alpha, \beta)] e^{ik\mathbf{s}\cdot\mathbf{r}} \tag{4.14b}$$

We show in Appendix B that the integrals on the right-hand side of Eqs. (4.14) are the usual *electromagnetic multipole fields*, i.e.,

$$\mathbf{E}_{lm}^e(\mathbf{r}) = \mathbf{H}_{lm}^h(\mathbf{r}) = \nabla \times \{\nabla \times [\mathbf{r}\, \Pi_l^m(\mathbf{r})]\}, \tag{4.15a}$$

$$\mathbf{E}_{lm}^h(\mathbf{r}) = -\mathbf{H}_{lm}^e(\mathbf{r}) = ik\nabla \times [\mathbf{r}\, \Pi_l^m(\mathbf{r})], \tag{4.15b}$$

where $\Pi_l^m(\mathbf{r})$ is, as before, the scalar multipole field defined by Eq. (3.14). $\mathbf{E}_{lm}^e(\mathbf{r})$, $\mathbf{H}_{lm}^e(\mathbf{r})$ are the fields generated by an electric multipole and the fields $\mathbf{E}_{lm}^h(\mathbf{r})$, $\mathbf{H}_{lm}^h(\mathbf{r})$ are those generated by a magnetic multipole, each of degree l and order m. Thus, we see that Eqs. (4.13) are the usual *multipole expansions* of the electric and magnetic fields generated by a localized charge-current distribution, the coefficients a_l^m and b_l^m being the electric and magnetic *multipole moments*, respectively. As is well known, each electromagnetic multipole

field satisfies the homogeneous Maxwell equations (for a monochromatic field of frequency $\omega = kc$) throughout the whole space, with the exception of the origin. Thus, Eqs. (4.13) are true *mode expansions* of the field $\mathbf{E}, \mathbf{H}$ outside the source region. Although we have derived the multipole expansion (4.13) from the angular spectrum representation, which is valid only when $|z| > R$, it is, in fact, valid everywhere outside the source region, i.e., for all $r > R$. The reason for this is essentially the same as in the scalar case. (See Appendix C.)

We see now that the multipole expansion of the electromagnetic field generated by a localized charge-current distribution arises naturally when the spectral amplitudes of the angular spectrum representation of the field are expressed as a series of vector spherical harmonics [Eqs. (4.10)]. The multipole moments are simply the projections of these spectral amplitudes onto the set of vector spherical harmonics [Eqs. (4.11)]. The multipole moments are also seen to be projections of certain components, (namely, those for which $|\mathbf{K}| = k = \omega/c$), of the Fourier transform of the transverse part of the current distribution onto the vector spherical harmonics $\mathbf{Y}_l^m(\alpha, \beta)$ and $\mathbf{s} \times \mathbf{Y}_l^m(\alpha, \beta)$ [Eqs. (4.12a) and (4.12b)].

The formulas (4.12), which express the multipole moments in terms of the transverse components of the Fourier transform $\tilde{\mathbf{j}}$ of the current density appear to be new. It is not difficult to express the multipole moments in a more conventional form involving the current density $\mathbf{j}$ directly. For this purpose, we first substitute into Eqs. (4.12) from (4.4) and obtain the formulas

$$a_l^m = \frac{i^{l+1}}{l(l+1)}\left(\frac{k}{c}\right)\int_{r' \le R} d^3r'\, \mathbf{j}(\mathbf{r}') \cdot \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha\, \sin\alpha\, \mathbf{s} \times \mathbf{Y}_l^{m*}(\alpha, \beta) e^{-ik\mathbf{s}\cdot\mathbf{r}'}, \quad (4.16a)$$

$$b_l^m = \frac{i^{l+1}}{l(l+1)}\left(\frac{k}{c}\right)\int_{r' \le R} d^3r'\, \mathbf{j}(\mathbf{r}') \cdot \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha\, \sin\alpha\, \mathbf{Y}_l^{m*}(\alpha, \beta) e^{-ik\mathbf{s}\cdot\mathbf{r}'}. \quad (4.16b)$$

In writing down the expression (4.16b) we made use of the relation $\mathbf{s} \times (\mathbf{s} \times \mathbf{Y}_l^m) = -\mathbf{Y}_l^m$, which follows from the fact that the vector spherical harmonics are everywhere tangential to the unit sphere $s^2 = 1$. Except for the nature of the α-contour of integration, the (α, β) integrals in Eqs. (4.16) are very similar to those appearing in the angular spectrum representation of the electromagnetic multipole fields [See Appendix B, Eqs. (B12) and (B13)]. In fact, if one carries out a strictly similar calculation as given in Appendix B, except that instead of the multipole field $\Pi_l^m(\mathbf{r}) = h_l^{(+)}(kr) Y_l^m(\theta, \phi)$ with source at the origin $r = 0$ one considers the source free multipole field $\Lambda_l^m(\mathbf{r}) = j_l(kr) Y_l^m(\theta, \phi)$, then comparison of Eqs. (3.13) with (2.5) shows that, except for a trivial proportionality factor $-i/2$ (which arises from the difference between coefficients on the right-hand sides of the two equations), the only change in the calculation will be the replacement of the α-contours $C^{\pm}$ by the real contour $0 \le \alpha \le \pi$. One then obtains, in place of Eqs. (B12) and (B13), the identities

$$ik\nabla \times [\mathbf{r}\Lambda_l^m(\mathbf{r})] = (-i)^l \frac{k}{4\pi}\int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha\, \sin\alpha\, \mathbf{Y}_l^m(\alpha, \beta) e^{ik\mathbf{s}\cdot\mathbf{r}}, \quad (4.17a)$$

$$\nabla \times \{\nabla \times [\mathbf{r}\Lambda_l^m(\mathbf{r})]\} = (-i)^l \frac{k}{4\pi}\int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha\, \sin\alpha\, \mathbf{s} \times \mathbf{Y}_l^m(\alpha, \beta) e^{ik\mathbf{s}\cdot\mathbf{r}}. \quad (4.17b)$$

With the help of these two identities, one readily obtains from Eqs. (4.16) the following, well-known expressions for the multipole moments:

$$a_l^m = \frac{4\pi i}{l(l+1)}\left(\frac{1}{c}\right)\int_{r' \le R} \mathbf{j}(\mathbf{r}') \cdot \{\nabla \times [\nabla \times (\mathbf{r}'\Lambda_l^{m*}(\mathbf{r}'))]\}\, d^3r', \quad (4.18a)$$

$$b_l^m = \frac{4\pi}{l(l+1)}\left(\frac{k}{c}\right)\int_{r' \le R} \mathbf{j}(\mathbf{r}') \cdot \{\nabla \times [\mathbf{r}'\Lambda_l^{m*}(\mathbf{r}')]\}\, d^3r'. \quad (4.18b)$$

Just as in the scalar case, we may readily obtain from the angular spectrum representation the far zone approximation for the electromagnetic fields $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$. In fact, if we apply the result expressed by Eq. (3.18) to each of the Cartesian components of $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ separately, we see at once that

$$\mathbf{E}(r\mathbf{u}_r) \sim \hat{\mathbf{E}}(\mathbf{u}_r)(e^{ikr}/r) \quad (kr \to \infty), \quad (4.19a)$$

$$\mathbf{H}(r\mathbf{u}_r) \sim \hat{\mathbf{H}}(\mathbf{u}_r)(e^{ikr}/r) \quad (kr \to \infty), \quad (4.19b)$$

for any real direction specified by the unit vector $\mathbf{u}_r$. The physical significance of these two relations may be readily understood from similar considerations as given in connection with the corresponding scalar equation (3.18).

If in Eqs. (4.19) we express the spectral amplitudes $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$ in terms of the multipole moments by means of Eq. (4.10), we obtain at once the usual asymptotic approximations for the electromagnetic field, valid in the far zone (i.e., as $kr \to \infty$):

$$\mathbf{E}(r\mathbf{u}_r) \sim \frac{e^{ikr}}{r}\sum_{l=1}^{\infty}\sum_{m=-l}^{l} (-i)^l [a_l^m \mathbf{u}_r \times \mathbf{Y}_l^m(\theta, \phi) + b_l^m \mathbf{Y}_l^m(\theta, \phi)], \quad (4.20a)$$

$$\mathbf{H}(r\mathbf{u}_r) \sim \frac{e^{ikr}}{r}\sum_{l=1}^{\infty}\sum_{m=-l}^{l} (-i)^l [-a_l^m \mathbf{Y}_l^m(\theta, \phi) + b_l^m \mathbf{u}_r \times \mathbf{Y}_l^m(\theta, \phi)]. \quad (4.20b)$$

Here, (θ, ϕ) are again the spherical polar coordinates of the unit direction vector $\mathbf{u}_r$.

We note, in passing, that since, according to Eq. (4.19a), the radiation pattern of the field [i.e., the vector function of $\mathbf{u}_r$ that multiplies the scalar field $\exp(ikr)/r$ in the asymptotic expansion of $\mathbf{E}(r\mathbf{u}_r)$ as $kr \to \infty$], is given precisely by the spectral amplitude vector $\hat{\mathbf{E}}(\mathbf{u}_r)$ for all real unit direction vectors $\mathbf{u}_r$, Eqs. (4.11), together with Eqs. (4.6), may also be interpreted as giving all the multipole moments in terms of the radiation pattern. Thus, we see, incidentally, that *all the multipole moments, and hence by Eqs. (4.13) the electromagnetic field at all points outside the sphere $r > R$, are completely specified by the radiation pattern.*

Finally, we may readily deduce from our results expressions for the time averaged power radiated by the source. It is given by the integral of the radial component of the time averaged Poynting vector across a limitingly large sphere Σ of radius r (with $kr \to \infty$):

$$\langle P \rangle = \frac{c}{8\pi}\,\Re \int_{\Sigma} [\mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r})] \cdot \mathbf{u}_r d\Sigma. \quad (4.21)$$

On substituting from Eqs. (4.19) into (4.21) we obtain the following expression for $\langle P \rangle$:

$$\langle P \rangle = \frac{c}{8\pi}\,\Re \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta\, \sin\theta\, [\hat{\mathbf{E}}(\mathbf{u}_r) \times \hat{\mathbf{H}}^*(\mathbf{u}_r)] \cdot \mathbf{u}_r. \quad (4.22)$$

But from the orthogonality relations between the three

vectors $\hat{\mathbf{E}}(\mathbf{u}_\gamma)$, $\hat{\mathbf{H}}(\mathbf{u}_\gamma)$, and $\mathbf{u}_\gamma$, indicated by Eqs. (4.6), we see that

$$[\hat{\mathbf{E}}(\mathbf{u}_\gamma) \times \hat{\mathbf{H}}^*(\mathbf{u}_\gamma)] \cdot \mathbf{u}_\gamma = \hat{\mathbf{E}}^*(\mathbf{u}_\gamma) \cdot \hat{\mathbf{E}}(\mathbf{u}_\gamma), \tag{4.23a}$$

$$= \hat{\mathbf{H}}^*(\mathbf{u}_\gamma) \cdot \hat{\mathbf{H}}(\mathbf{u}_\gamma). \tag{4.23b}$$

If we now substitute from (4.23a) [or (4.23b)] into (4.22), and express $\hat{\mathbf{E}}(\mathbf{u}_\gamma)$ [or $\hat{\mathbf{H}}(\mathbf{u}_\gamma)$] in the series form (4.10a) [or (4.10b)], and if we also make use of the orthogonality relations (4.9) between the vector spherical harmonics we find that

$$\langle P \rangle = \frac{c}{8\pi} \sum_{l=1}^{\infty} \sum_{m=-l}^{l} l(l+1)[|a_l^m|^2 + |b_l^m|^2]. \tag{4.24}$$

Equation (4.24) is the well-known expression for the radiated power in terms of the multipole moments.

5. THE MULTIPOLE EXPANSION AND THE DEBYE POTENTIALS

Many of the existing treatments of multipole expansions and of other problems arising in electromagnetic theory employ a representation of the electromagnetic field in terms of two scalar potentials that was introduced by Debye[3] in a well-known investigation relating to the pressure exerted by light on a homogeneous sphere composed of arbitrary material. In this concluding section we briefly show that the angular spectrum representation gives a new insight into the relation between the multipole expansion, and the Debye representation.[34]

The Debye potentials $\Pi_e(\mathbf{r})$ and $\Pi_h(\mathbf{r})$ are solutions of the scalar Helmholtz equation which yield an electromagnetic field in free space by means of the formulae (the Debye representation):

$$\mathbf{E}(\mathbf{r}) = \nabla \times [\nabla \times (\mathbf{r}\Pi_e(\mathbf{r}))] + ik\nabla \times [\mathbf{r}\Pi_h(\mathbf{r})], \tag{5.1a}$$

$$\mathbf{H}(\mathbf{r}) = -ik\nabla \times [\mathbf{r}\Pi_e(\mathbf{r})] + \nabla \times [\nabla \times (\mathbf{r}\Pi_h(\mathbf{r}))]. \tag{5.1b}$$

It is clear that the multipole expansion (4.13) may be expressed in the form (5.1) if we substitute into (4.13) the definitions (4.15) of the electromagnetic multipole fields and interchange the order of differentiation and summations. We then obtain the following expressions for the electric and magnetic Debye potentials:

$$\Pi_e(\mathbf{r}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_l^m \Pi_l^m(\mathbf{r}), \tag{5.2a}$$

$$\Pi_m(\mathbf{r}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} b_l^m \Pi_l^m(\mathbf{r}). \tag{5.2b}$$

The Debye representation is intimately connected with our decomposition (4.10) of the spectral amplitudes $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$ into series of the vector spherical harmonics $\mathbf{Y}_l^m$ and $\mathbf{s} \times \mathbf{Y}_l^m$. To see this, let us introduce in (4.10) the definition (4.7) of the vector spherical harmonic $\mathbf{Y}_l^m$ in terms of the ordinary spherical harmonic Y_l^m. If then we interchange the orders of differentiation and summation we obtain the formulas

$$\hat{\mathbf{E}}(\mathbf{s}) = \mathbf{s} \times \mathcal{L}_s \hat{A}(\mathbf{s}) + \mathcal{L}_s \hat{B}(\mathbf{s}), \tag{5.3a}$$

$$\hat{\mathbf{H}}(\mathbf{s}) = -\mathcal{L}_s \hat{B}(\mathbf{s}) + \mathbf{s} \times \mathcal{L}_s \hat{A}(\mathbf{s}), \tag{5.3b}$$

where

$$\hat{A}(\mathbf{s}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} (-i)^l a_l^m Y_l^m(\alpha, \beta), \tag{5.4a}$$

$$\hat{B}(\mathbf{s}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} (-i)^l b_l^m Y_l^m(\alpha, \beta). \tag{5.4b}$$

Consider now the scalar fields $A(\mathbf{r}), B(\mathbf{r})$, whose spectral amplitudes are $\hat{A}(\mathbf{s})$ and $\hat{B}(\mathbf{s})$, respectively:

$$A(\mathbf{r}) = \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha \sin\alpha \, \hat{A}(\mathbf{s}) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{5.5a}$$

$$B(\mathbf{r}) = \frac{ik}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha \sin\alpha \, \hat{B}(\mathbf{s}) e^{ik\mathbf{s}\cdot\mathbf{r}}. \tag{5.5b}$$

If we substitute from (5.4a) into (5.5a), interchange the order of integration and summation, and recall, also, the angular spectrum representation (3.13) of the multipole field $\Pi_l^m(\mathbf{r})$, we obtain the following expression for $A(\mathbf{r})$:

$$A(\mathbf{r}) = k \sum_{l=1}^{\infty} \sum_{m=-l}^{l} (-i)^l a_l^m \Pi_l^m(\mathbf{r}). \tag{5.6a}$$

In a strictly similar way, we obtain from (5.4b) and (5.5b), if again we use (3.13), the following expression for $B(\mathbf{r})$:

$$B(\mathbf{r}) = k \sum_{l=1}^{\infty} \sum_{m=-l}^{l} (-i)^l b_l^m \Pi_l^m(\mathbf{r}). \tag{5.6b}$$

Comparison of Eqs. (5.2) and (5.6) shows that

$$A(\mathbf{r}) = k\Pi_e(\mathbf{r}), \quad B(\mathbf{r}) = k\Pi_h(\mathbf{r}), \tag{5.7}$$

i.e., apart from the proportionality factor k, $A(\mathbf{r})$ and $B(\mathbf{r})$ are precisely the Debye potentials and, hence, the *amplitude functions* $\hat{A}(\mathbf{s})$ *and* $\hat{B}(\mathbf{s})$ *introduced in (5.3) are, apart from the proportionality factor* k, *the spectral amplitudes in the angular spectrum representation of the Debye potentials.*[35]

ACKNOWLEDGMENT

We wish to acknowledge our appreciation to Dr. George Sherman for stimulating discussions relating to the analysis presented in this paper.

APPENDIX A: ANGULAR SPECTRUM REPRESENTATION OF A SCALAR MULTIPOLE FIELD

In this Appendix, we show that the angular spectrum representation of the scalar multipole field $\Pi_l^m(\mathbf{r})$, defined by Eq. (3.14), is given by Eq. (3.13). For this purpose we will use the well known result that a multipole field $\Pi_l^m(\mathbf{r})$ of any order $m \geq 0$ may be generated from the spherical wave $\exp(ikr)/kr$, [which apart from a normalization constant is the lowest order scalar multipole field $\Pi_0^0(\mathbf{r})$], by means of the following relation[24]:

$$\Pi_l^m(\mathbf{r}) = C_l^m \left\{ \left[\frac{1}{ik} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) \right]^m P_l^{(m)} \left(\frac{1}{ik} \frac{\partial}{\partial z} \right) \right\} \frac{e^{ikr}}{kr}. \tag{A1}$$

Here, the operator

$$P_l^{(m)} \left(\frac{1}{ik} \frac{\partial}{\partial z} \right)$$

is defined by the formula

$$P_l^{(m)} \left(\frac{1}{ik} \frac{\partial}{\partial z} \right) = \frac{d^m}{du^m} P_z(u) \Big|_{u=(1/ik)\partial/\partial l}, \tag{A2}$$

where $P_l(u)$ is the Legendre polynomial of degree of l and C_l^m are normalization constants, defined as

$$C_l^m = (-1)^m (-i)^l \left(\frac{(2l+1)(l-m)!}{4\pi(l+m)!} \right)^{1/2}. \tag{A3}$$

When $m < 0$ we use the identity

$$\Pi_l^{-|m|}(r, \theta, \phi) = (-1)^{|m|} \Pi_l^{|m|}(r, \theta, -\phi), \tag{A4}$$

where, of course, (r, θ, ϕ) are the spherical polar coordinates of $\mathbf{r}$.

We now express the spherical wave $\exp(ikr)/kr$ in (A1) in the form of an angular spectrum representation, given by Weyl's formula (3.4):

$$\frac{e^{ikr}}{kr} = \frac{i}{2\pi}\int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha\, e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{A5}$$

and interchange the order of integration and differentiation [which may be shown to be justified when $|z|>0$ because the double integral in (A5) is then uniformly convergent]. We then obtain the following expression for $\Pi_l^m(\mathbf{r})$, valid when $|z|>0$ and $m \geq 0$:

$$\Pi_l^m(\mathbf{r}) = \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha\, F(\alpha,\beta) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{A6}$$

where

$$F(\alpha,\beta)e^{ik\mathbf{s}\cdot\mathbf{r}} = \frac{i}{2\pi} C_l^m \left\{\left[\frac{1}{ik}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\right]^m P_l^{(m)}\left(\frac{1}{ik}\frac{\partial}{\partial z}\right)\right\} e^{ik\mathbf{s}\cdot\mathbf{r}}. \tag{A7}$$

Now, with the Cartesian coordinates of $\mathbf{s}$ given by (2.3), we have

$$\mathbf{s}\cdot\mathbf{r} = x\sin\alpha\cos\beta + y\sin\alpha\sin\beta + z\cos\alpha, \tag{A8}$$

and we then readily obtain from (A7) the following expression for $F(\alpha,\beta)$:

$$F(\alpha,\beta) = \frac{i}{2\pi} C_l^m \sin^m\alpha\, e^{im\beta} P_l^{(m)}(\cos\alpha). \tag{A9}$$

But

$$\sin^m\alpha\, P_l^{(m)}(\cos\alpha) = P_l^m(\cos\alpha), \tag{A10}$$

where P_l^m is the associated Legendre polynomial of degree l and order m. Hence,

$$\begin{aligned} F(\alpha,\beta) &= (i/2\pi) C_l^m P_l^m(\cos\alpha) e^{im\beta} \\ &= (i/2\pi)(-i)^l Y_l^m(\alpha,\beta), \end{aligned} \tag{A11}$$

where

$$Y_l^m(\alpha,\beta) = i^l C_l^m P_l^m(\cos\alpha) e^{im\beta} \tag{A12}$$

is the spherical harmonic of degree l and order m. Finally, on substituting from (A11) into (A6) we find that

$$\Pi_l^m(\mathbf{r}) = \frac{i}{2\pi}(-i)^l \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha\, Y_l^m(\alpha,\beta) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{A13}$$

which is the desired result valid when $|z|>0$ and $m \geq 0$. Although, as previously mentioned, the interchange the order of integration and differentiation which lead to (A13) is justified only when $|z|>0$, (A13) can be shown to be valid also when $z=0$, except at the origin, in the sense of the following limit:

$$\Pi_l^m(\mathbf{r})\,|_{z=0} = \lim_{|z|\to 0} \frac{i}{2\pi}(-i)^l \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha\, Y_l^m(\alpha,\beta) e^{ik\mathbf{s}\cdot\mathbf{r}}. \tag{A13a}$$

To determine the angular spectrum representation for $\Pi_l^m(\mathbf{r})$ when $m<0$, we first express the scalar product $\mathbf{s}\cdot\mathbf{r}$ in (A13) in a more explicit form, using Eqs. (A8) and (2.6):

$$\mathbf{s}\cdot\mathbf{r} = r(\sin\theta\sin\alpha\cos(\beta+\phi) + \cos\theta\cos\alpha). \tag{A14}$$

Equation (A13) then becomes, on substitution from (A14):

$$\Pi_l^{|m|}(\mathbf{r}) = \frac{i}{2\pi}(-i)^l \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha\, Y_l^{|m|}(\alpha,\beta) \times e^{ikr[\sin\theta\sin\alpha\cos(\beta+\phi)+\cos\theta\cos\alpha]}. \tag{A15}$$

From (A14) and (A4) it follows that

$$\Pi_l^{-|m|}(\mathbf{r}) = (-1)^{|m|}\frac{i}{2\pi}(-i)^l \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha\, Y_l^{|m|}(\alpha,\beta) \times e^{ikr[\sin\theta\sin\alpha\cos(\beta-\phi)+\cos\theta\cos\alpha]}. \tag{A16}$$

If in (A16) we change the variable of integration from β to $-\beta$ and use the relation

$$Y_l^{-|m|}(\alpha,\beta) = (-1)^{|m|} Y_l^{|m|}(\alpha,-\beta),$$

and also the formula (A14), we find that

$$\Pi_l^{-|m|}(\mathbf{r}) = \frac{i}{2\pi}(-i)^l \int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha\, Y_l^{-|m|}(\alpha,\beta) e^{ik\mathbf{s}\cdot\mathbf{r}}. \tag{A17}$$

If now we set $-|m| = m$ where $m<0$, Eq. (A17) becomes formally identical with (A13). Thus, (A13) is valid for both positive and negative integers m, $(-l \leq m \leq l)$, as we wished to show.

APPENDIX B: ANGULAR SPECTRUM REPRESENTATION OF AN ELECTROMAGNETIC MULTIPOLE FIELD

We will show in this Appendix that the angular spectrum representation of an electromagnetic multipole field, defined by Eqs. (4.15), is given by Eqs. (4.14), i.e., that

$$\nabla\times\{\nabla\times[\mathbf{r}\Pi_l^m(\mathbf{r})]\} = (-i)^l \frac{ik}{2\pi}\int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha \times[\mathbf{s}\times\mathbf{Y}_l^m(\alpha,\beta)]e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{B1a}$$

$$ik\nabla\times[\mathbf{r}\Pi_l^m(\mathbf{r})] = (-i)^l \frac{ik}{2\pi}\int_{-\pi}^{\pi} d\beta \int_{C^{\pm}} d\alpha \sin\alpha \times\mathbf{Y}_l^m(\alpha,\beta)e^{ik\mathbf{s}\cdot}. \tag{B1b}$$

We derive first the formula (B1b). We have, by using a standard vector identity,

$$\begin{aligned} ik\{\nabla\times[\mathbf{r}\Pi_l^m(\mathbf{r})]\} &= ik\{\Pi_l^m(\mathbf{r})\nabla\times\mathbf{r} - \mathbf{r}\times\nabla\Pi_l^m(\mathbf{r})\} \\ &= k\mathcal{L}_r\Pi_l^m(\mathbf{r}), \end{aligned} \tag{B2}$$

where we have used the fact that $\nabla\times\mathbf{r}=0$. In (B2), $\mathcal{L}_r$ is the "orbital angular momentum operator" in $\mathbf{r}$ space, viz.

$$\mathcal{L}_r = -i\mathbf{r}\times\nabla = -i\left(\mathbf{u}_\phi\frac{\partial}{\partial\theta} - \frac{1}{\sin\theta}\mathbf{u}_\theta\frac{\partial}{\partial\phi}\right), \tag{B3}$$

where $\mathbf{u}_\phi, \mathbf{u}_\theta$ are unit vectors in the positive ϕ and θ directions, respectively. Now the operator $\mathcal{L}_r$ does not act on the radial coordinate r and, consequently, if we recall the definition (3.14) of the scalar multipole field $\Pi_l^m(\mathbf{r})$, we may express (B2) in the form

$$ik\nabla\times[\mathbf{r}\Pi_l^m(\mathbf{r})] = kh_l^+(kr)\mathbf{Y}_l^m(\theta,\phi), \tag{B4}$$

where

$$\mathbf{Y}_l^m(\theta,\phi) = \mathcal{L}_r Y_l^m(\theta,\phi) \tag{B5}$$

is the vector spherical harmonic of degree l and order m.

Next, we will make use of the following identity, which expresses a vector spherical harmonic as a linear combination of ordinary spherical harmonics:[36]

$$\mathbf{Y}_l^m(\theta,\phi) = a_- Y_l^{m+1}(\theta,\phi)\boldsymbol{\epsilon}_- + a_+ Y_l^{m-1}(\theta,\phi)\boldsymbol{\epsilon}_+ + mY_l^m(\theta,\phi)\mathbf{u}_z, \tag{B6}$$

where

$$a_- = [(l-m)(l+m+1)]^{1/2},$$
$$a_+ = [(l+m)(l-m+1)]^{1/2}, \tag{B7}$$

$$\epsilon_- = \tfrac{1}{2}(\mathbf{u}_x - i\mathbf{u}_y), \quad \epsilon_+ = \tfrac{1}{2}(\mathbf{u}_x + i\mathbf{u}_y), \tag{B8}$$

and $\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}_z$ are unit vectors in the positive x, y, and z directions, respectively. On substituting from (B6) into (B4), and recalling once again the definition (3.14) of $\Pi_l^m(\mathbf{r})$, we obtain at once the identity

$$ik\nabla \times [\mathbf{r}\,\Pi_l^m(\mathbf{r})] = k[a_-\Pi_l^{m+1}(\mathbf{r})\epsilon_- + a_+\Pi_l^{m-1}(\mathbf{r})\epsilon_+ + m\Pi_l^m(\mathbf{r})\mathbf{u}_z]. \tag{B9}$$

We now express each of the three scalar multipole fields appearing on the righthand side of (B9) in terms of the angular spectrum representation (3.13) to obtain the identity

$$ik\nabla \times [\mathbf{r}\,\Pi_l^m(\mathbf{r})] = k(-i)^l \frac{i}{2\pi}\int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha\, \sin\alpha\, \mathbf{G}(\alpha,\beta) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{B10}$$

where

$$\mathbf{G}(\alpha,\beta) = a_- Y_l^{m+1}(\alpha,\beta)\epsilon_- + a_+ Y_l^{m-1}(\alpha,\beta)\epsilon_+ + mY_l^m(\alpha,\beta)\mathbf{u}_z. \tag{B11}$$

But according to (B6), the right-hand side of (B11) is precisely the vector spherical harmonic $\mathbf{Y}_l^m(\alpha,\beta)$, i.e., $\mathbf{G}(\alpha,\beta) = \mathbf{Y}_l^m(\alpha,\beta)$. Hence, (B10) gives

$$ik\nabla \times [\mathbf{r}\,\Pi_l^m(\mathbf{r})] = (-i)^l \frac{ik}{2\pi}\int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha\, \sin\alpha\, \mathbf{Y}_l^m(\alpha,\beta) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{B12}$$

which establishes the representation (B1b).

Next, we apply the curl operator to (B12). The curl operator may be taken under the integral signs on the rhs of (B12) since the double integral may be shown to converge uniformly when $|z| > 0$. We then obtain the formula

$$\nabla \times \{\nabla \times [\mathbf{r}\,\Pi_l^m(\mathbf{r})]\} = (-i)^l \frac{ik}{2\pi}\int_{-\pi}^{\pi} d\beta \int_{C_\pm} d\alpha\, \sin\alpha\, \mathbf{s}\times\mathbf{Y}_l^m(\alpha,\beta) e^{ik\mathbf{s}\cdot\mathbf{r}}, \tag{B13}$$

which establishes the representation (B1a).

We should note that the two expansions (B1a) and (B1b) are valid for all $z \neq 0$ and, moreover, can be shown to have limiting values as $|z| \to 0$ which correctly represent the electromagnetic multipole fields on the plane $z = 0$, except at the origin [cf. Eq. (A13a) of Appendix A].

APPENDIX: C: DOMAIN OF VALIDITY OF THE MULTIPOLE EXPANSION (3.12)

In this appendix, we verify that the multipole expansion (3.12) is valid not only throughout the two half-spaces $z > R$ and $z > -R$ [where the angular spectrum expansion (3.5) converges], but is, in fact, valid throughout the exterior, $r > R$, of the source region.

As mentioned in Sec. 3, the field $\psi(\mathbf{r})$ can be represented in the form of an angular spectrum of plane waves outside *any* strip bounded by two parallel planes tangential to the sphere of radius R, which surrounds the source. For example, if we choose a new Cartesian coordinate system of axes $O\bar{X}, O\bar{Y}, O\bar{Z}$, obtained from the original system OX, OY, OZ, by a rotation about the origin O, we can represent the field at all points outside the strip $|\bar{z}| < R$ in the form (3.5), where the polar angles α, β of $\mathbf{s}$ and θ, ϕ of the field point $\mathbf{r}$ are referred to the rotated (barred) rather than the original (unbarred) system of axes. Consequently, all the analysis leading to the multipole expansion (3.12) remains valid in the rotated system.

Consider now a field point $\mathbf{r}_0$ that lies outside the source region (i.e., for which $r_0 > R$), but which is situated *within* the strip $|\bar{z}| < R$ and let us choose the rotated axes $O\bar{X}, O\bar{Y}, O\bar{Z}$ in such a way that $\mathbf{r}_0$ lies outside the strip $|\bar{z}| < R$ (see Fig. 2). Then if $r_0, \bar{\theta}_0, \bar{\phi}_0$ and $r', \bar{\theta}', \bar{\phi}'$ are the spherical polar coordinates of the field point $\mathbf{r}_0$ and of the integration point $\mathbf{r}'$, respectively, referred to the *rotated* system of axes, we have by (3.12)

$$\psi(\mathbf{r}_0) = k\sum_{l=0}^{\infty}\sum_{m=-l}^{l} \bar{a}_l^m h_l^{(+)}(kr_0) Y_l^m(\bar{\theta}_0, \bar{\phi}_0), \tag{C1}$$

where [if we also use (2.4)]

$$\bar{a}_l^m = 4\pi \int_{r'\leq R} \rho(\mathbf{r}')\, j_l(kr')\, Y_l^{m*}(\bar{\theta}', \bar{\phi}')\, d^3r'. \tag{C2}$$

We can rewrite (C1) in the form

$$\psi(\mathbf{r}_0) = k\sum_{l=0}^{\infty} h_l^{(+)}(kr_0)\left(4\pi\int_{r'\leq R} d^3r'\rho(\mathbf{r}')\, j_l(kr')\sum_{m=-l}^{l} \times Y_l^{m*}(\bar{\theta}', \bar{\phi}')\, Y_l^m(\bar{\theta}_0, \bar{\phi}_0)\right). \tag{C3}$$

Now, if r_0, θ_0, ϕ_0 and r', θ', ϕ' are the spherical polar coordinates of the field point $\mathbf{r}_0$ and the integration point $\mathbf{r}'$, respectively, referred to the *original* system of axes we have, from the addition theorem on spherical harmonics (Ref. 37, pp. 290–291)

$$\sum_{m=-l}^{l} Y_l^{m*}(\bar{\theta}', \bar{\phi}') Y_l^m(\bar{\theta}_0, \bar{\phi}_0) = \sum_{m=-l}^{l} Y_l^{m*}(\theta', \phi') Y_l^m(\theta_0, \phi_0), \tag{C4}$$

each of these two sums being equal to $(2l+1)P_l(\cos\chi)/4\pi$. Here P_l is the Legendre polynomial of degree l, and χ is the angle between the position vectors $\mathbf{r}_0$ and $\mathbf{r}'$. Making use of (C4), Eq. (C3) can be rewritten in the form

$$\psi(\mathbf{r}_0) = k\sum_{l=0}^{\infty} h_l^{(+)}(kr_0)\left(4\pi\int_{r'\leq R} d^3r'\rho(\mathbf{r}')\, j_l(kr') \times \sum_{m=-l}^{l} Y_l^{m*}(\theta', \phi') Y_l^m(\theta_0, \phi_0)\right), \tag{C5}$$

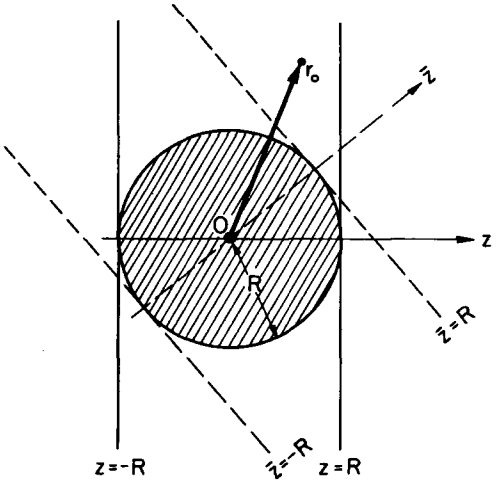


FIG. 2. Notation relating to the proof that the multipole expansion (3.12) is valid throughout the exterior $r > R$ of the source region. The point $\mathbf{r}_0$ is situated in the strip $|z| < R$ (referred to the original coordinate system) but outside the strip $|\bar{z}| < R$ (referred to the rotated system).

or

$$\psi(\mathbf{r}_0) = k \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_l^m h_l^{(+)}(kr_0) Y_l^m(\theta_0, \phi_0), \qquad \text{(C6)}$$

with [if use is also made of (2.4)]

$$a_l^m = 4\pi \int_{r' \leq R} \rho(\mathbf{r}') \Lambda_l^{m*}(\mathbf{r}') d^3\mathbf{r}'. \qquad \text{(C7)}$$

Equations (C6) is seen to be precisely Eq. (3.12) evaluated at the point $\mathbf{r}_0$, and referred to the original system of axes and Eq. (C7) is identical with Eq. (3.11). Since $\mathbf{r}_0$ can be taken to be *any* point outside the source region, we conclude that (3.12) is valid at all points $\mathbf{r}$ such that $r > R$.

* Research supported by the Air Force Office of Scientific Research and the Army Research Office (Durham). Some of the results presented here were obtained in the course of doctoral thesis research by one of the authors [A. J. Devaney, "A New Theory of the Debye Representation of Classical and Quantized Electromagnetic Fields", Ph. D. thesis, University of Rochester, Rochester, N.Y. (1971)]. A preliminary account was also presented at the Fall meeting of the Optical Society of America, held in San Francisco, California in October 1972. [Abstract WT17, J. Opt. Soc. Amer., 62, 1360 (1972).]

[1] A. Clebsch, J. Math. 61, 195 (1863).
[2] G. Mie, Ann. Phys. (Leipz.) 25, 377 (1908).
[3] P. Debye, Ann. Phys. (Leipz.) 30, 57 (1909).
[4] T. J. I'A. Bromwich, Phil. Trans. R. Soc. Lond. 220, 175 (1920).
[5] For a very readable account of the early studies of scattering of plane waves by spheres see, N. A. Logan, Proc. IEEE 63, 48 (1965). See also, a note by the same author published in the J. Opt. Soc. Am. 52, 342 (1962). For a detailed list of references to papers on multipole expansions up to 1953 see M. E. Rose, *Multipole Fields* (Wiley, New York, 1955).
[6] W. W. Hansen, Phys. Rev. 47, 139 (1935).
[7] W. Heitler, Proc. Camb. Phil. Soc. 32, 112 (1936).
[8] H. A. Kramers, Physica 10, 261 (1943).
[9] W. Franz, Z. Phys. 127, 363 (1950).
[10] P. R. Wallace, Can. J. Phys. 29, 393 (1951).
[11] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (Wiley, New York, 1952), App. B.
[12] C. J. Bouwkamp and H. B. G. Casimir, Physica 20, 539 (1954).
[13] A. Nisbet, Physica 21, 799 (1955).
[14] C. H. Wilcox, J. Math. Mech. 6, 167 (1957).
[15] E. T. Whittaker, Math. Ann. 57, 333 (1902); see also, E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1963), 4th ed., Sec. 18.6.
[16] For a discussion of this representation see, for example, C. J. Bouwkamp, Rep. Prog. Phys. 17, 41 (1954) or J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968), Sec. 3.7. The relationship between homogeneous plane wave expansions and angular spectrum representations is discussed, for certain classes of scalar wave fields, in A. J. Devaney and G. C. Sherman, SIAM Rev. 15, 765 (1973).
[17] A. J. Devaney and E. Wolf, Phys. Rev. 8, 1044 (1973).
[18] The customary term "multipole field" for the field $\Lambda_l^m(\mathbf{r})$ defined by Eq. (2.4) is rather inappropriate since this field is everywhere well-behaved. This terminology is undoubtedly responsible for incorrect statements in the literature about the behavior of such a field at the origin. [See, for example, W. Heitler, *The Quantum Theory of Radiation* (Clarendon, Oxford, 1954), 3rd ed., statement following Eq. (9) on p. 402.] It is the scalar field $\Pi_l^m(\mathbf{r})$, defined by Eq. (3.14), and the electromagnetic fields $\mathbf{E}_{lm}^{(e)}$, $\mathbf{H}_{lm}^{(e)}$ and $\mathbf{E}_{lm}^{(h)}$, $\mathbf{H}_{lm}^{(h)}$, defined by Eqs. (4.15), that behave as fields generated by a true multipole at the origin $r = 0$, and hence have an appropriate singularity at that point.
[19] We use the same definitions of the spherical harmonics and of the spherical Bessel and Hankel functions as those given in A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, third printing 1965), Vol. I, App. BII and BIV.
[20] H. Weyl, Ann. Phys. (Paris) 60, 481 (1919). Because the integrand of (3.4) is an entire analytic function of α, the α integration contours $C^{\pm}$ may be deformed in accordance with the rules of integration in the complex plane. When these contours are chosen as shown in Fig. 1, (3.4) corresponds to the following, frequently used, alternative form of the Weyl expansion:

$$\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|}{|\mathbf{r}-\mathbf{r}'|} = \frac{ik}{2\pi} \iint_{-\infty}^{\infty} \frac{1}{s_z} \exp\{ik[s_x(x-x') + s_y(y-y') + s_z|z-z'|]\}\, ds_x\, ds_y. \qquad \text{(3.4a)}$$

Here,

$$s_z = +\sqrt{1 - s_x^2 - s_y^2}, \quad \text{if } s_x^2 + s_y^2 \leq 1,$$
$$s_z = +i\sqrt{s_x^2 + s_y^2 - 1}, \quad \text{if } s_x^2 + s_y^2 > 1. \qquad \text{(3.4b)}$$

The two forms of the Weyl expansion, and the transformations required to pass from one to the other, are discussed in A. Baños, *Dipole Radiation in the Presence of a Conducting Half-Space* (Pergamon, New York, 1966), sec. 2.13.
[21] If the representation (3.4a), rather than (3.4), is employed one obtains, instead of (3.5), the following form of the angular spectrum representation, which is commonly employed in the literature:

$$\psi(\mathbf{r}) = \frac{ik}{2\pi} \iint_{-\infty}^{\infty} \hat{\phi}^{(\pm)}(s) \exp[ik(s_x x + s_y y \pm s_z z)]\, ds_x\, ds_y, \qquad \text{(3.5a)}$$

where s_z is defined by (3.4b), and the positive signs are used when $z > R$ and the negative signs when $z < -R$. The spectral amplitudes $\phi^{(\pm)}(s)$ are given by

$$\phi^{(\pm)}(\mathbf{s}) = \frac{1}{s_z} \tilde{\rho}(ks_x, ks_y, \pm ks_z), \qquad \text{(3.5c)}$$

with $\tilde{\rho}(\mathbf{K})$ being the threefold Fourier transform of the source distribution, as defined by Eq. (3.7).
[22] The effect of discarding all evanescent plane waves in the angular spectrum representation of the scalar dipole field $\pi_1^0(\mathbf{r})$ [see Eq. (3.14)] has been very clearly analyzed by W. H. Carter, Opt. Commun. 2, 142 (1970).
[23] This result is, essentially, a three-dimensional analogue of a well-known theorem that the Fourier transform of a continuous function which vanishes outside a finite interval is a boundary value of an entire analytic function. This theorem follows at once from a well known result on analytic functions defined by definite integrals [cf., E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford U.P., London, 1962), Sec. 5.5]. The multidimensional form of the theorem is the well-known Plancherel-Pólya theorem [cf., B. A. Fuks, *Introduction to the Theory of Analytic Functions of Several Complex Variables* (Amer. Math. Soc., Providence, R.I., 1963), p. 352].
[24] A. Erdélyi, Physica 4, 107 (1937).
[25] See, for example, the Appendix in K. Miyamoto and E. Wolf, J. Opt. Soc. Am. 52, 615 (1962).
[26] It is planned to present the corresponding analysis for the source-free, classical electromagnetic field in another publication. An analogous treatment of the quantized, free electromagnetic field has been given by A. J. Devaney, *Proceedings of the Third Rochester Conference on Coherence and Quantum Optics*, edited by L. Mandel and E. Wolf (Plenum, New York, 1973), p. 241.
[27] If, in addition to oscillating charges and current densities, there is a contribution from density of magnetization $\mathfrak{M}(\mathbf{r}, t) = \Re(\mathfrak{M}(\mathbf{r})\, e^{-i\omega t})$, one must replace $\mathbf{j}(\mathbf{r})$ by $\mathbf{j}(\mathbf{r}) + c\nabla \times \mathfrak{M}(\mathbf{r})$. The introduction of magnetization makes it possible to take into account the effect of spin in the corresponding quantum mechanical formulation.
[28] Conversely, it is not difficult to show that the E and H fields that are the solutions to Eqs. (4.1) and which behave at infinity as outgoing spherical waves satisfy the full set of Maxwell equations everywhere.
[29] The angular spectrum representation (4.2) may, of course, be expressed in the vectorial generalization of the alternative form discussed in Footnote 21.
[30] See, for example, F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965), Sec. 4.3.
[31] E. L. Hill, Am. J. Phys. 22, 211 (1954).
[32] This fact may readily be deduced from the customary form of the completeness theorem (discussed, for example, in Ref. 11, pp. 798–799), which may be stated as follows: An arbitrary, well-behaved vector field $\mathbf{A}(\mathbf{r})$ may be expanded in terms of three types

of vector spherical harmonics $\mathbf{X}_{lm}(\theta,\phi)$, $\mathbf{V}_{lm}(\theta,\phi)$, and $\mathbf{W}_{lm}(\theta,\phi)$ in the form

$$\mathbf{A}(\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}[f_l^m(r)\,\mathbf{X}_{lm}(\theta,\phi) + g_l^m(r)\,\mathbf{V}_{lm}(\theta,\phi) + h_l^m(r)\,\mathbf{W}_{lm}(\theta,\phi)],$$

where f_l^m, g_l^m, and h_l^m are functions of the radial coordinate $r = |\mathbf{r}|$ only. (The vector spherical harmonics $\mathbf{X}_{lm}$, $\mathbf{V}_{lm}$, and $\mathbf{W}_{lm}$ correspond to $\mathbf{Y}^m{}_{l,l,1}$, $\mathbf{Y}^m{}_{l,l+1,1}$, and $\mathbf{Y}^m{}_{l,l-1,1}$ respectively, of Ref. 11.) The vector spherical harmonic $\mathbf{X}_{lm}$ is, except for a normalization factor, the same function that we denoted by $\mathbf{Y}_l^m$; more precisely

$$\mathbf{Y}_l^m = [l(l+1)]^{1/2}\,\mathbf{X}_{lm}.$$

Moreover, one has the relations [see Eqs. (B3) and (B9) of Ref. 31):

$$\mathbf{s}\times\mathbf{Y}_l^m = i\,[l(l+1)]^{1/2}\left[\left(\frac{l}{l+1}\right)^{1/2}\mathbf{V}_{lm} + \left(\frac{l+1}{2l+1}\right)^{1/2}\mathbf{W}_{lm}\right],$$

and

$$\mathbf{s}\,Y_l^m = \left[-\left(\frac{l+1}{2l+1}\right)^{1/2}\mathbf{V}_{lm} + \left(\frac{l}{2l+1}\right)^{1/2}\mathbf{W}_{lm}\right],$$

where $\mathbf{s} = \mathbf{r}/r$ is the unit vector in the radial direction. Since, according to these relations, the vector spherical harmonics $\mathbf{Y}_l^m$, $\mathbf{s}\times\mathbf{Y}_l^m$, and $\mathbf{s}\,Y_l^m$ are linearly independent combinations of $\mathbf{X}_{lm}$ $\mathbf{V}_{lm}$, and $\mathbf{W}_{lm}$, it is clear that they too form a complete basis for the expansion of $\mathbf{A}(\mathbf{r})$. Moreover, $\mathbf{Y}_l^m$ and $\mathbf{s}\times\mathbf{Y}_l^m$ are tangential to the unit sphere ($s^2 = 1$) and $\mathbf{s}\,Y_l^m$ is perpendicular to it. Hence, the two types of vector spherical harmonics $\mathbf{Y}_l^m$ and $\mathbf{s}\times\mathbf{Y}_l^m$ form a complete set for arbitrary "tangential vector fields" $\mathbf{A}(\mathbf{r})$, i.e., an arbitrary, well-behaved, vector field $\mathbf{A}(\mathbf{r})$ such that $\mathbf{s}\cdot\mathbf{A}(\mathbf{r}) = 0$, may be expanded in terms of them.

[33]By making use of elementary vector identities and the fact that $\mathbf{s}\cdot\mathbf{Y}_l^m = 0$, one can rewrite (4.12a) in the form

$$a_l^m = -\frac{il+1}{l(l+1)}\left(\frac{k}{c}\right)\int_{-\pi}^{\pi}d\beta\int_{0}^{\pi}d\alpha\,\sin\alpha\,\{\mathbf{s}\times[\mathbf{s}\times\tilde{\mathbf{j}}(k\mathbf{s})]\}\cdot[\mathbf{s}\times\mathbf{Y}_l^{m*}(\alpha,\beta)]. \quad (4.12a')$$

Now, $-\mathbf{k}\times[\mathbf{k}\times\tilde{\mathbf{j}}(\mathbf{k})]/|\mathbf{k}|^2$ is the Fourier transform $\mathbf{j}_T(\mathbf{k})$ of the transverse part $\mathbf{j}_T(\mathbf{r})$ of the current distribution [cf., E. A. Power, *Introductory Quantum Electrodynamics* (American Elsevier, New York, 1964), Sec. 6.3]. Thus, (4.12a') and (4.12b) show that all multipole moments, (and, consequently, the field outside the source region), depend only on those Fourier components $\tilde{\mathbf{j}}_T(\mathbf{k})$ of the transverse part of the current distribution for which $|\mathbf{k}| = k = \omega/c$.

[34]For a discussion of the various aspects of the Debye representation see, for example, the papers by Bouwkamp and Casimir[12], Nisbet[13], and Wilcox[14].

[35]The decomposition (5.3) of the spectral amplitudes $\hat{\mathbf{E}}(\mathbf{s})$ and $\hat{\mathbf{H}}(\mathbf{s})$, which we obtained as a consequence of the completeness of the vector spherical harmonics $\mathbf{Y}_l^m(\alpha,\beta)$ and $\mathbf{s}\times\mathbf{Y}_l^m(\alpha,\beta)$ with respect to all well behaved fields $\mathbf{F}(\mathbf{s})$ that are orthogonal to $\mathbf{s}$ [i.e., such that $\mathbf{s}\cdot\mathbf{F}(\mathbf{s}) \equiv 0$], may also be obtained as a direct consequence of the so-called *Hodge's decomposition theorem* [See, for example, P. Bidal and G. de Rham, Commun. Math. Hel. **19**, 1 (1946).] Wilcox[14] employed this theorem in his treatment of the Debye representation, which, however, is quite different from ours.

[36]This identity follows from a well known general expression for vector spherical harmonics in terms of ordinary spherical harmonics. [See, for example, Eq. (1.5), p. 797 in Ref. 11.]

[37]See, for example, B. W. Shore and D. H. Menzel, *Principles of Atomic Spectra* (Wiley, New York, 1968), pp. 290-91.

Analytic vectors for the Poincaré group in quantum field theory

Bengt Nagel and Håkan Snellman

Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm 70, Sweden
(Received 4 September 1973)

We give a new proof of the theorem stating that in a quantum field theory with tempered field operators the dense domain of the polynomial algebra of these field operators applied to the vacuum state contains a dense invariant set of analytic vectors for the representation of the Poincaré group.

INTRODUCTION

Analytic vectors for the unitary representation $U(\mathcal{P})$ of the Poincaré group $\mathcal{P}$ in axiomatic quantum field theory have been discussed earlier by one of us[1]. In this note we use a different type of argument to give a new demonstration that the domain D_0, constructed out of the polynomials in the fields smeared with test functions from $\oplus_{n=0}^{\infty} \mathcal{S}(\mathbb{R}^{4n})$ applied to the vacuum state, contains a dense invariant set of analytic vectors for $U(\mathcal{P})$. We end with some remarks on the integrability of symmetry Lie algebras in quantum field theory.

ANALYTIC VECTORS FOR THE REPRESENTATION OF THE POINCARÉ GROUP

The axioms for a Wightman field theory of a spinless field are reproduced in.[2] In our considerations we do not make use of the locality property of the field or of the spectral condition for P^μ, the generators of translations. It will also be sufficient to have one, not necessarily unique, cyclic vacuum Ω.

Theorem 1: Given a Wightman field theory with a finite number of tempered fields transforming covariantly among themselves under the unitary representation $U(\mathcal{P}_+^\uparrow)$ of the Poincaré group $\mathcal{P}_+^\uparrow = T^4 \otimes SL(2,\mathbb{C})$. Then the dense domain D_0 generated by the polynomial algebra over the fields smeared with test functions from $\oplus_{n=0}^{\infty} \mathcal{S}(\mathbb{R}^{4n})$ applied to the vacuum state Ω, contains a dense invariant domain of analytic vectors for $U(\mathcal{P}_+^\uparrow)$.

We shall give the proof of this theorem only for a single scalar field. The generalization to fields with spin is straightforward since the matrix elements of the $SL(2,\mathbb{C})$ representations transforming the field components are polynomials in the real and imaginary parts of the matrix elements of $SL(2,\mathbb{C})$.

Let $\varphi(x_1,\ldots,x_n) \in \mathcal{S}(\mathbb{R}^{4n})$, where x_i is a four-vector. A scalar field ϕ gives a map $\mathcal{S}(\mathbb{R}^{4n}) \to \mathcal{H}$, which is linear and continuous,

$$\mathcal{S}(\mathbb{R}^{4n}) \ni \varphi \to \phi^n(\varphi)\Omega \in \mathcal{H},$$

where $\mathcal{H}$ is the Hilbert space.

For a given $\varphi \in \mathcal{S}(\mathbb{R}^{4n})$ we define a map $\mathcal{P}_+^\uparrow \to \mathcal{S}(\mathbb{R}^{4n})$ by

$$\mathcal{P}_+^\uparrow \ni (\Lambda, a) \to \varphi_{(\Lambda,a)} \in \mathcal{S}(\mathbb{R}^{4n}),$$

where

$$\varphi_{(\Lambda,a)}(x_1,\ldots,x_n) \equiv \varphi(\Lambda x_1 + a,\ldots,\Lambda x_n + a).$$

Composition gives

$$(\Lambda, a) \to \varphi_{(\Lambda,a)} \to \phi^n(\varphi_{(\Lambda,a)})\Omega = U((\Lambda,a)^{-1})\phi^n(\varphi)\Omega.$$

We shall show that if

$$\varphi(x_1,\ldots,x_n) = P(x_1,\ldots,x_n)\exp(-\tfrac{1}{2}\sum_1^n x_i^2)$$

with P a polynomial in the components of x_i, and $x_i^2 = x_i^{02} + \mathbf{x}_i^2$, this composite map is real-analytic from $\mathcal{P}_+^\uparrow$ to $\mathcal{H}$. Then, since $(\Lambda, a) \to (\Lambda, a)^{-1}$ is an analytic homeomorphism, $\phi^n(\varphi)\Omega$ is an analytic vector for $U(\mathcal{P}_+^\uparrow)$. (The set of functions described above is denoted $\{\varphi\}$.)

We start from arbitrary complex transformations (Λ, a): $x^\mu \to \Lambda^\mu_\nu x^\nu + a^\mu$ where the 20 complex numbers $(\Lambda^\mu_\nu, a^\lambda)$ are denoted by $z = (z_1,\ldots,z_{20}) \in \mathbb{C}^{20}$. We then define functions f on $\mathbb{R}^{4n} \times \mathbb{C}^{20}$:

$$f(x,z) = P(\Lambda x_1 + a,\ldots,\Lambda x_n + a)e^{-N(x,z)},$$

$$N(x,z) = \tfrac{1}{2}\sum_1^n (\Lambda x_i + a)^2.$$

In fact, we shall study a more general f of the form $Q(x,z)e^{-N(x,z)}$ where $Q(x,z)$ is a polynomial. For every fixed $x \in \mathbb{R}^{4n}$, such an f is evidently holomorphic in $\mathbb{C}^{20}$.

Lemma 1: There is an open set $V \subset \mathbb{C}^{20}$, such that $V \supset \mathcal{P}_+^\uparrow$ and $f(.,z)$ is holomorphic from V to $\mathcal{S}(\mathbb{R}^{4n})$.

Proof: Define a subset of $\mathbb{C}^{20}$: $V = \{(\Lambda, a);\ a \in \mathbb{C}^4,$ Λ such that the quadratic form $\mathrm{Re}(\Lambda x)^2 > 0\}$. Since the condition for positive definiteness is given by inequalities related to the matrix $\mathrm{Re}\Lambda^T\Lambda - T$ denotes transposition — V is obviously an open subset of $\mathbb{C}^{20}$. As every real nonsingular linear transformation takes a pos. def. form into a pos. def. form, we have in particular that $V \supset \mathcal{P}_+^\uparrow$.

Now, for every compact subset V_c of V we can find positive numbers K and η such that

$$|e^{-N(x,z)}| < Ke^{-\eta\sum_1^n x_i^2}, \quad x \in \mathbb{R}^{4n}, \quad z \in V_c.$$

Evidently $f(.,z) \in \mathcal{S}(\mathbb{R}^{4n})$ for every $z \in V$ and it remains to show that the map $V \ni z \to f(.,z) \in \mathcal{S}(\mathbb{R}^{4n})$ is holomorphic. This amounts to showing that the map is differentiable [in the topology of $\mathcal{S}(\mathbb{R}^{4n})$], i.e., that for every seminorm p in a set defining the topology of $\mathcal{S}(\mathbb{R}^{4n})$ we can, given $\epsilon > 0$, find a $\delta > 0$ such that

$$p\left(f(.,z+\Delta z) - f(.,z) - \sum_1^{20}\frac{\partial f}{\partial z_i}\Delta z_i\right) < \epsilon|\Delta z|$$

provided $|\Delta z| = (\Sigma|\Delta z_i|^2)^{1/2} < \delta$.

The set of (semi-) norms on $\mathcal{S}(\mathbb{R}^{4n})$ is taken as $\{p_{\alpha,\beta}\}$ with

$$p_{\alpha,\beta}(f) = \sup_x |x^\alpha D^\beta f(x)|$$

with an obvious multi-index notation.

Now we can write

$$f(.,z+\Delta z) - f(.,z) - \sum_1^{20}\frac{\partial f}{\partial z_i}\Delta z_i = \int_0^1 \sum_1^{20}\left(\frac{\partial f}{\partial z_i}(.,z+t\Delta z) - \frac{\partial f}{\partial z_i}(.,z)\right)\Delta z_i dt.$$

Provided $z + t\Delta z$, $t \in [0, 1]$, stays inside a compact V_c as above, the effect of the suppressing factor $e^{-N(x,z)}$ is that the term inside ( ) under the integral sign can be made small, uniformly in x, by choosing $|\Delta z|$ small enough, for any given fixed $z \in V$. Hence we get

$$p_{0,0}\left(f(.,z+\Delta z) - f(.,z) - \sum_1^{20} \frac{\partial f}{\partial z_i}\,\Delta z_i\right) < \epsilon\,|\Delta z|$$

for $|\Delta z|$ small.

For a general norm $p_{\alpha,\beta}$ we can write $p_{\alpha,\beta}(f) = p_{0,0}(x^\alpha D^\beta f)$.

Since x and z are independent arguments, we get

$$p_{\alpha,\beta}\left(f(.,z+\Delta z) - f(.,z) - \sum_1^{20} \frac{\partial f}{\partial z_i}\,\Delta z_i\right)$$
$$= p_{0,0}\left(g(.,z+\Delta z) - g(.,z) - \sum_1^{20} \frac{\partial g}{\partial z_i}\,\Delta z_i\right),$$

where $g(x,z) = x^\alpha D_x^\beta f(x,z)$. g is of the same type as f so the same argument as above applies. Thus $f(.,z)$ is holomorphic from V to $\mathcal{S}(\mathbb{R}^{4n})$.

Proof of Theorem 1: Since the map ϕ^n from $\mathcal{S}(\mathbb{R}^{4n})$ to $\mathcal{H}$ is linear and continuous it follows from Lemma 1 that the composite map

$$V \ni (\Lambda, a) \to \phi^n(\varphi_{(\Lambda,a)})\Omega \in \mathcal{H}$$

is holomorphic. Since $\mathcal{P}_{c+}$ (the complex Poincaré group with determinant $|\Lambda| = +1$) is an analytic submanifold of $\mathbb{C}^{20}$, restriction of the above map to $V \cap \mathcal{P}_{c+}$ gives again a holomorphic map. Now, $V \cap \mathcal{P}_{c+}$ is evidently an open subset of $\mathcal{P}_{c+}$ containing $\mathcal{P}_+^\uparrow$, so that restriction to $\mathcal{P}_+^\uparrow$ gives a real-analytic map. Since the set of functions $\{\varphi\}$ contains the Hermite functions in $4n$ dimensions it is obviously dense in $\mathcal{S}(\mathbf{R}^{4n})$. By taking the polynomial algebra over the field, smeared with test functions in $\{\varphi\}$ for all n, we get a dense domain of Poincaré analytic vectors contained in D_0. By further taking all $U(\mathcal{P}_+^\uparrow)$-translates of the vectors in the domain so constructed, we get the dense invariant domain of analytic vectors of Theorem 1.

REMARKS ON INTEGRABILITY OF SYMMETRY LIE ALGEBRAS

Regarding the local aspect of analytic vectors for the Poincaré algebra in connection with the integrability discussed in Ref. 1 we want to emphasize that in a field theory the integrability is often most simply shown by means of the extension[3] of a theorem by Segal[4] on the unitary implementability of groups of *-automorphisms of C^*-algebras to the case of unbounded operator algebras in field theory, usually called the "reconstruction theorem." From this theorem one can deduce, that any Lie algebra representation in $\mathcal{H}$ acting as a derivation in the set of field operators, is integrable if

(a) the generators annihilate the vacuum state Ω, and
(b) the action of the Lie algebra on the algebra of test functions (which induces the given Lie algebra representation in $\mathcal{H}$) can be integrated to a differentiable group of *-automorphisms of the algebra of test functions.

For representations of the Poincaré algebra by symmetric operators in $\mathcal{H}$ satisfying the condition (a) and transforming a finite number of fields among themselves, the condition (b) is always fulfilled.

In the case of internal symmetry algebras, the algebra of test functions can be taken as the tensor algebra generated by $\mathcal{S}(\mathbb{R}^4) \hat{\otimes} L$, where L is the index space and $\mathcal{S}(\mathbb{R}^4)$ is the covariant test function space. In this case the condition (b) is equivalent to requiring the representation in L to be integrable.

ACKNOWLEDGMENT

It is a pleasure to acknowledge fruitful discussions with our friends M. Flato and D. Sternheimer.

[1] H. Snellman, "Poincaré analytic vectors in axiomatic quantum field theory", J. Math. Phys. (to be published).
[2] R. Jost, *The general theory of quantized fields* (American Mathematical Society, Providence 1965). R. F. Streater and A. S. Wightman, *PCT, Spin & Statistics and all that* (Benjamin, New York, 1964).
[3] H. Borchers, Nuovo Cimento 24, 214 (1962). K Maurin. Bull. Acad. Polon. Sci. 11, 115, 121 (1963).
[4] I. Segal, Duke Math. J. 18, 221 (1951).

Regge trajectory structure of the Amati–Bertocchi–Fubini–Stanghellini–Tonin multiperipheral model

V. Balakrishnan

Tata Institute of Fundamental Research, Bombay, India
(Received 31 March 1972; revised manuscript received 13 February 1973)

The Regge pole spectrum of the ABFST multiperipheral model with a resonance kernel is studied, using both the forward and nonforward partial-wave integral equations. Insight into the complicated pattern of the trajectory spectrum is provided by an analysis of the singularity structure of the basic kernel using appropriate representations for it, and by an elucidation of the process of generation of trajectories in the weak-coupling limit. This also establishes a framework for understanding the characteristics of the approximate solutions to the problem that are taken up subsequently. The trace approximations for both the nonforward and forward equations are worked out in detail. The traces involved are evaluated in convenient closed forms from which all the necessary information can be extracted easily. It is found that the approximation preserves, to a fair degree of accuracy, the trajectory-generating singularity structures of the relevant kernels, and that good 'effective' trajectory positions for the leading and secondary poles are obtained in both the forward and nonforward cases. It is also shown that, in this approximation, other phenomena such as complex Regge poles, the threshold behavior of the trajectories, and the intercept and slope of the leading trajectory can be investigated in a close simulation of the actual situation. Recent factorizable approximations are then examined from the point of view of the pole spectra they lead to, and it is concluded that, by and large, they oversimplify the problem and that their shortcomings are thus more extensive than those of the trace approximation.

1. INTRODUCTION

There are very few relativistic models for which complex angular momentum analysis can be carried out, because of the enormous technical complications involved in a program of this sort. Yet, because of the importance of such studies in high-energy hadron physics, the few tractable models that exist have been the center of a great deal of attention, even if they do neglect complications arising due to spin and other quantum numbers. (Apart from the usual claim that these factors can, in principle, be included without much trouble, it is of course essential to be able to handle at least the simplified cases in the first instance.) In particular, the Amati–Bertocchi–Fubini–Stanghellini–Tonin[1] (ABFST) and later, more generalized[2] multiperipheral models[3] have been investigated in many papers in connection with high-energy behavior and continuation to complex angular momentum. Interest in a multiperipheral model like that of ABFST derives from its relevance in describing such features of high-energy scattering as Regge behavior for elastic amplitudes and total cross sections, scaling, and logarithmic growth of the multiplicity in multiparticle production reactions, to mention just a few. From a theoretical viewpoint, too, the ABFST model has as one of its important uses the "underpinning" of the assumptions involved in multi-Regge bootstraps on which modern dynamical calculations on the origin of the Pomeron and other trajectories are based.[4,5] The model is thus of basic importance.[6,7] It is further capable of accommodating refinements and modifications[6–8] that make it more complete from theoretical considerations and improve its description of the experimental situation. The question of its Regge trajectory content is therefore a pertinent one and is worthy of analytical study in its own right, given the fundamental nature of this type of model for strong interactions at very high energies.

A formalism that has proved most useful when used in conjunction with multiperipheral models is that of the Bethe–Salpeter (BS) equation. For instance, in the multi-Regge bootstraps referred to above, much use has been made of the BS framework with Regge exchange incorporated into it.[5] The formalism is also of considerable aid in the "diagonalization" of the multiperipheral model-based integral equation for absorptive parts of scattering amplitudes[9]; this is an essential step in the general program of carrying out the crossed-channel partial-wave analysis necessary to express the high-energy properties of physical scattering amplitudes in terms of singularities in the partial-wave parameter plane, taking into account the appropriate little group of the momentum-transfer vector in various kinematic regions. Now it has been found possible to derive "partial-wave" integral equations[10] for the absorptive part of the scattering amplitude in the ABFST model in both the forward ($t=0$) and nonforward ($t\neq 0$) cases, that are, with the usual assumption of a "resonance" kernel, rather similar in form to the Bethe–Salpeter equations satisfied by the partial-wave projections at $t=0$ and $t\neq 0$ of the scattering amplitude in the scalar $g\phi_1^2\phi_2$ field theory in the ladder approximation. The particular problem of deducing the *trajectory spectrum* in each of these instances is, formally, a common one. Discussions of some of the Regge trajectories of the ABFST model have been given in various approximations such as "factorizable" approximations[11,12] (for general t) and the "trace" or "first Fredholm" approximation (at $t=0$),[7,13] for example, in connection with the occurrence of complex Regge poles.[14] It will be recalled that more detailed work on the trajectory-spectrum aspect of the problem (beyond

the level of the leading Regge pole) has been done, but in a rather different context—essentially in the weak-coupling limit of the field-theoretic example referred to earlier. These investigations have been made both in perturbation theory,[15] usually using Mellin-transformed Feynman amplitudes,[16] and directly from the BS equation for the partial-wave amplitude.[17] However, even here, explicit analytic continuation into the left-half partial-wave plane[18] (or Mellin parameter plane) to isolate Regge poles has been performed[16,19–21] only for a few of the higher-lying trajectories, because of the rapidly multiplying technical complications at each stage of the (stripwise) continuation procedure. It is rather unclear from such considerations, as also from those involving the trace or factorizable approximations, exactly what the trajectory spectrum of the original problem is. We take the attitude that this is an important question, and that it is worth trying to clarify the manner of generation of Regge poles in the ABFST model. A second aspect worth examining is the extent to which the trace and factorizable approximations preserve the actual trajectory structure (insofar as the latter can be reliably deduced). These two broad objectives form the main content of the present paper. We shall find, first, that a careful consideration of the singularity structure of the kernel of the relevant integral equation in the angular momentum plane yields a good deal of information on the process of generation of the Regge trajectories, and second, that the trace approximation can be analyzed in depth (both at $t=0$ and $t\neq 0$) and is in fact a rather good approximation as compared to the various factorizable approximations. Although we shall be concerned with the (physically interesting) ABFST model, we shall often use (for convenience) the language of the field-theoretic model in the BS formalism.

The basic partial-wave equations in both the $t=0$ and $t\neq 0$ cases, from which the trajectory spectrum is to be deduced, are linear integral equations with nondegenerate kernels. Essentially because of the nondegeneracy of the kernels, the problem of determining the *full* trajectory spectrum remains unsolved. This is so even in the simplified case of the weak-coupling limit mentioned above. (Indeed, only in the special case of the Wick–Cutkosky model,[22] where the mass of the exchanged scalar meson is zero, has the full singularity structure in the partial-wave plane been elucidated.[23]) We shall use the known, exact results in the weak-coupling limit to examine the accuracy of the trajectory structures found in various approximation to the ABFST equations. We shall also occasionally use the Wick–Cutkosky example for the same purpose, although of course this case certainly does not apply in a literal sense to the ABFST model, where the exchanged (resonance) mass is generally taken to be large compared to that of the scattering particles. As we shall see, there are basic qualitative differences between the trajectory sequences in the massless-scalar and massive-scalar exchange cases; however, the former will serve as a check on some results when the appropriate limit is taken, since several exact results are known here.

The organization of this paper is as follows: Sec. 2 is devoted to writing down the integral equations concerned, introducing notation, etc. In Sec. 3, we discuss the trajectory spectrum in the weak-coupling limit both for $t=0$ and $t\neq 0$. (A method of extracting the positions of the Regge poles at $t=0$ is also incidentally indicated, that is much simpler than any of the earlier methods.) Part of this discussion will be a quick summary of known results, as its purpose is merely to provide a means for understanding the origin of the singularities in the partial-wave plane. Attention is then paid to the form of the kernel in order to point out precisely how the trajectories are generated, for the reasons given earlier. In Sec. 4, we give a full discussion of the eigenvalue problem in the trace approximation. First we consider the weak-coupling limit in this case (for the sake of the comparison referred to above), and then the physically interesting questions of threshold behavior, complex poles, and Pomeron intercept and slope. Finally we re-examine the trace approximation for the forward equation. Although this has been considered in earlier papers,[7,14] the integral representing the trace of the kernel has been evaluated only approximately. We show that this integral can be evaluated exactly and that the eigenvalue condition for the pole positions can be cast into a simple closed form, from which results such as the occurrence of complex Regge poles can be easily read off. In Sec. 5, we briefly look at various factorizable approximations from the point of view indicated above in order to get an idea of the reliability of the corresponding trajectory spectra. We close with some general concluding remarks. (Other such remarks on specific points are made at the appropriate places in the text itself.) The emphasis throughout will be on analytic aspects, adequate numerical analyses[24,25] having been carried out already.

2. THE EIGENVALUE EQUATIONS

The direct (s-) channel absorptive part of the off-shell (π-π) scattering amplitude in the ABFST model satisfies a linear inhomogeneous equation rather similar to an integral equation in each of the cases $t=0$ (forward scattering) and $t\neq 0$. A lot of work has been done in recent years on the "partial-wave" analysis of the equations in the two cases and in diagonalizing[9] the equations in the partial-wave parameters (i.e., in deriving the integral equations satisfied by each decoupled partial-wave projection). All these equations are neatly summarized in Ref. 11. Here we shall merely quote the relevant equations.

Let us first consider the general case, $t\neq 0$. With kinematics as in Ref. 11, we have ingoing momenta $P+Q/2$, $k-Q/2$, and outgoing momenta $P-Q/2$, $k+Q/2$ in the s channel, so that $s=(P+k)^2$, $t=Q^2/4$. The t-channel relative momenta squared are denoted by $u=P^2$, $v=k^2$, and held negative. The angles ψ, φ are defined by $\sin\psi=(P\cdot Q)/(tu)^{1/2}$, $\sin\varphi=(k\cdot Q)/(tv)^{1/2}$. The two-pion contribution to the s-channel absorptive part is approximated by a single resonance of mass M, with "strength" g^2. Then a partial-wave projection A_l of this absorptive part may be defined, which satisfies the linear integral equation

$$A_l(t;u,\psi;v,\varphi)$$
$$= \frac{\pi g^2}{B(\frac{1}{2}, l+1)} [(uv)^{1/2}\cos\psi\cos\varphi]^{-l-1}$$
$$\times Q_l\left(\frac{M^2 - u - v - 2(uv)^{1/2}\sin\psi\sin\varphi}{2(uv)^{1/2}\cos\psi\cos\varphi}\right)$$
$$+ \frac{g^2}{16\pi^3}\int_{-\infty}^{0} du' \int_{-\pi/2}^{\pi/2} d\psi' \left[\left(\frac{u'}{u}\right)^{1/2}\left(\frac{\cos\psi'}{\cos\psi}\right)\right]^{l+1}$$
$$\times Q_l\left(\frac{M^2 - u - u' - 2(uu')^{1/2}\sin\psi\sin\psi'}{2(uu')^{1/2}\cos\psi\cos\psi'}\right)$$
$$\times[(m_\pi^2 - t/4 - u')^2 - u't\sin^2\psi']^{-1} A_l(t;u',\psi';v,\varphi), \tag{2.1}$$

where u', ψ' are defined in terms of intermediate state momenta $(P' \pm Q/2)$ analogous to u, ψ. l is the partial-wave parameter. The Regge trajectories of the model are to be found from the zeros of the Fredholm denominator $D(l,t)$ of the kernel in Eq. (2.1). Aside from some changes of variables, the equation satisfied by A_l is similar to the Bethe–Salpeter equation for the partial-wave amplitude in the $g\phi_1^2\phi_2$ theory in the ladder approximation. For a consideration of the trajectory spectrum, we are concerned with the homogeneous integral equation. Let us write this down in the BS formalism. Let m, M denote, respectively, the masses of the scattering and exchanged particles, and g the coupling constant at the m-M-m vertex. Then, after the usual Wick rotation has been carried out and we go over to Euclidean momenta, the off-shell, homogeneous, t-channel partial-wave BS equation in the c.m. system reads[17]

$$\phi_l(p,\omega) = g^2(2\pi)^{-3}\int_0^\infty dp' \int_{-\infty}^{\infty} d\omega'$$
$$\times Q_l\left(\frac{M^2 + p^2 + p'^2 + (\omega - \omega')^2}{2pp'}\right) f(p',\omega',t)\phi_l(p',\omega'). \tag{2.2}$$

$p(=|\mathbf{p}|)$ and ω refer, respectively, to the relative momentum and energy variables. $f(p,\omega,t)$ is the product of the propagators for the two internal lines on the sides of the ladder, and is given by

$$f(p,\omega,t) = [(p^2 + \omega^2 + m^2 - t/4)^2 + t\omega^2]^{-1}. \tag{2.3}$$

We shall work with the convenient notation of Eq. (2.2) in our consideration of the Regge trajectories of the ABFST model.

We now turn to the case when $t=0$. At this point, $f(p,\omega,t)$ becomes a function of the combination $p^2+\omega^2$, and the well-known $O(4)$ symmetry of the (Wick-rotated) BS equation can be used to further simplify the eigenvalue equation (2.2).[26] What is needed is an expansion of the Q_l function in this equation in terms of hyperspherical functions.[27] As we shall need this expansion when we consider the process of generation of trajectories, we write it down here. In terms of the variables $u = p^2+\omega^2$, $\zeta = \omega/u^{1/2}$, and $\rho = (u+u'+M^2)/2(uu')^{1/2}$, we have

$$Q_l\left(\frac{M^2 + p^2 + p'^2 + (\omega-\omega')^2}{2pp'}\right)$$
$$\equiv Q_l\left(\frac{M^2 + u + u' - 2(uu')^{1/2}\zeta\zeta'}{2[uu'(1-\zeta^2)(1-\zeta'^2)]^{1/2}}\right)$$
$$= B(\tfrac{1}{2}, l+1)[(1-\zeta^2)(1-\zeta'^2)]^{(l+1)/2}$$
$$\times \sum_{n=0}^{\infty} nB(n, 2l+2)\mathcal{J}_{l+n}(\rho)C_n^{l+1}(\zeta)C_n^{l+1}(\zeta'), \tag{2.4}$$

where $C_n^{l+1}(\zeta)$ is the standard Gegenbauer polynomial and $\mathcal{J}_{l+n}(\rho)$ is the corresponding function of the second kind, given by

$$\mathcal{J}_{l+n}(\rho) = [\rho + (\rho^2-1)]^{-l-n-1}. \tag{2.5}$$

If we also expand the eigenfunction $\phi_l(u,\zeta)$ in Eq. (2.2) in terms of four-dimensional harmonics, then, exactly at $t=0$, the equation decouples to give a one-dimensional integral equation; this reads

$$\phi(u) = \frac{G^2}{(\sigma+1)}\int_0^\infty du'(u'+m^2)^{-2}\mathcal{J}_0(\rho)\phi(u'), \tag{2.6}$$

where $G^2 = g^2/16\pi^2$ and we have put $l+n=\sigma$ since the kernel depends on this combination alone. (Any statement on the spectrum at $t=0$ as a function of the "four-dimensional angular momentum" σ may now be translated into one in the l plane simply by using $l=\sigma-n$.[28])

Turning to the ABFST model in the case of forward scattering, let u, v denote the four-momenta squared of the two off-shell pions in the s-channel (u, v are held negative). Then a partial-wave projection A_λ of the s-adsorptive part of the amplitude can be *defined* in such a manner that it satisfies the equation

$$A_\lambda(u,v) = I_\lambda(u,v) + \int_{-\infty}^{0} du' A_\lambda(u,u')K_\lambda(u',v). \tag{2.7}$$

Here I_λ is the inhomogeneous term, and the kernel K_λ is given by

$$K_\lambda(u',v) = (\lambda+1)^{-1}\int_{4m_\pi^2}^{\infty} ds(u'-m_\pi^2)^{-2}C_I(s)\mathcal{J}_\lambda(\rho), \tag{2.8}$$

with $\rho = (s-u'-v)/2(u'v)^{1/2}$. $C_I(s)$ is related to the two-pion unitarity contribution (it is proportional to the $\pi\pi$ elastic cross section). We have ignored isospin here for simplicity. If, as in the $t\neq 0$ equation, we take into account only a resonance contribution by replacing $C_I(s)$ with $G^2\delta(s-M^2)$, we arrive at an eigenvalue equation[11] for the trajectories at $t=0$ that is essentially Eq. (2.6), with λ taking the place of σ. As in the nonforward case, we shall work with the (BS) notation of Eq. (2.6) in our discussion of the trajectory spectrum.

3. TRAJECTORY STRUCTURE IN THE WEAK-COUPLING LIMIT

The Regge trajectories of the model are found by solving the eigenvalue condition obtained from the homogeneous integral equations (2.2) and (2.6) for the appropriate partial-wave parameter in terms of the other quantities, namely, t, G^2, m^2, and M^2. As stated in the Introduction, we shall now briefly consider the weak-coupling limit, i.e., we shall work to $O(G^2)$ in the trajectory functions, in order to facilitate comparison with the results of subsequent approximate solutions to the problem.

A. $t \neq 0$ case

The relevant eigenvalue equation is Eq. (2.2). We expect Regge poles to be generated from the fixed poles of the kernel at $l=-1,-2,\cdots$ as the coupling is turned on, but initially we are restricted to the region $\mathrm{Re}\, l > -3/2$, which is the region of square integrability of the kernel (except for the point $l=-1$).[18] It is well known that in this region there is just one Regge pole for sufficiently small G^2. Since the residue of Q_l at $l=-1$ is unity, the position of this leading pole to $O(G^2)$ is easily seen to be given by

$$l+1=g^2(2\pi)^{-3}\int_0^\infty dp\int_{-\infty}^{\infty} d\omega f(p,\omega,t)$$
$$=4G^2(4m^2-t)^{-1}F[\tfrac{1}{2},1;\tfrac{3}{2};t/(t-4m^2)]. \qquad (3.1)$$

M^2 enters only in the higher order terms in G^2.

Now let us consider the secondary trajectories. In the expansion[29]

$$Q_l(z)=\sqrt{\pi}\sum_{r=0}^{\infty}(2z)^{-l-2r-1}\frac{\Gamma(l+2r+1)}{\Gamma(l+r+\frac{3}{2})r!} \quad (|z|>1), \qquad (3.2)$$

it is the term with $r=0$ that has a pole at $l=-2$. Before we can isolate the Regge poles expected near $l=-2$ in the weak-coupling limit, however, analytic continuation (of the original inhomogeneous equation) to that neighborhood must be performed, because the term $\sim z^{-l-1}$ in $Q_l(z)$ leads to divergences of the original representation before we can reach $l=-2$ both for $p'\sim 0$ and $p',\omega' \to \infty$. After this problem is taken care of,[19,20] one finds three Regge poles near $l=-2$ in the weak-coupling limit, one of which is the daughter of the leading Regge pole. Further stripwise analytic continuation to the neighborhood of $l=-3$ (now the $r=0$, 1 terms have to be handled properly) gives five Regge poles in that neighborhood, three of these being daughters of those near $l=-2$.

While we can, in principle, proceed to isolate all the secondary trajectories of the model by such stripwise continuation using the expansion of Eq. (3.2), in practice the procedure rapidly becomes too complicated to be tractable beyond $l=-3$. Up to this point the multiplicity of Regge poles is indeed the same as that in the $M=0$ case. In the latter model, the eigenvalue equation may be reduced to a second order ordinary differential equation,[30] and it turns out[23] that there are $(2N-1)$ Regge poles near $l=-N$ in the weak-coupling limit. Beyond $l=-3$, the $M\neq 0$ problem departs considerably from the pattern of the $M=0$ case.[31] No "one-step" continuation procedure has been found so far when M is nonzero. Therefore, although we may look upon the Regge poles of the model as being generated from the poles of Q_l at $l=-N$, the problem of explicit isolation of the Regge poles is nontrivial, and indeed the exact multiplicity of Regge poles near $l=-N$ in the weak-coupling limit is unknown for general N.

Still, we *can* make some useful statements on the trajectory spectrum in the weak-coupling limit for the case of interest to us. It is already known that for the kernel we are concerned with in this paper, the only singularities in the l plane are (Regge) poles,[18] and the somewhat simpler $t=0$ case makes it possible to put rough upper[21] and lower[32] bounds on the number of Regge poles present in the model. We proceed to make some more definite statements: while we cannot directly substitute for $Q_l(z)$ the singular but degenerate (finite-ranked) kernel $P_{N-1}(z)/(l+N)$ to solve the problem in the weak-coupling limit because of the divergence problem, the exact number of Regge poles generated by each (fixed) pole of Q_l must be related to the number of linearly independent vectors in the (p,ω) function space that can be constructed from a suitable "continuable" form of the residue P_{N-1} at that pole. An example is needed to clarify the point: For $N=2$, the residue of $Q_l(z)$ is simply $z=[M^2+p^2+p'^2+(\omega-\omega')^2]/2pp'$, which, however, is not square-integrable; but an inspection of the expansion for $Q_l(z)$, keeping in mind the divergence problem, shows that we must use $z^{-l-1}/(l+2)$, rather than just $z/(l+2)$, as the effective kernel in this instance. We can then proceed to show[20] that a suitable "continuable" form for $Q_l(z)$ near $l=-2$ in the weak-coupling approximation is the rank-three kernel $[\varphi_1(p,\omega)\varphi_2(p',\omega')+\varphi_2(p,\omega)\varphi_1(p',\omega')-\varphi_3(p,\omega)\times\varphi_3(p',\omega')]/(l+2)$, where the φ's are (l-dependent) functions given by

$$\varphi_1=(\sqrt{2}p)^{l+1}, \quad \varphi_2=[\sqrt{2}p/(M^2/2+p^2+\omega^2)]^{l+1},$$
$$\varphi_3=(p/\omega)^{l+1}.$$

The resolvent of this kernel can be easily evaluated explicitly and *then* analytically continued to the neighborhood of $l=-2$. Putting $l=-2$ everywhere except in the explicit pole factors $(l+2)^{-1}$ will then lead to the three Regge poles referred to above. For a general value of N, an examination of Eq. (3.2) shows that the pole of Q_l at $l=-N$ comes from such a pole in each of the first $[(N-1)/2]$ terms of the expansion. If we extract this pole part *without altering the l-dependent powers of z*, we find that the "residue" can be written in the surprisingly simple form $z^{-l-N}P_{N-1}(z)$. (The factor z^{-l-N} is of course what makes the kernel nonseparable.) It is this last quantity that must be used as the effective residue from which the appropriate separable kernel is constructed to extract Regge poles in the weak-coupling limit, after analytic continuation in l of the resolvent of this kernel. Further discussion of the process of generation of trajectories is given in Sec. 3C, using the expansion of Eq. (2.4) for the Q_l function.

B. $t = 0$ case

The relevant eigenvalue equation is Eq. (2.6). It is trivial to see that in the weak-coupling limit the leading Regge pole is at

$$\sigma=-1+G^2\int_0^\infty du/(u+m^2)^2=-1+G^2/m^2,$$

which is consistent with Eq. (3.1). Before we can isolate secondary trajectories, we have again to handle a problem of analytic continuation, the kernel in Eq. (2.6) being square-integrable only in the region $\mathrm{Re}\,\sigma > -2$. While no poles at $\sigma=-2,-3,\cdots$ appear explicitly in the kernel of Eq. (2.6), unlike the case of the Q_l function in the kernel of the nonforward equation, such poles could (and do) appear once we carry out the necessary analytic continuation in σ.

The eigenvalue equation (2.6) has been the subject of a number of investigations,[33] but, as in the $t\neq 0$ case, one is forced to adopt a stripwise continuation procedure in

the σ plane, and the exact multiplicity of trajectories is not known even in the weak-coupling limit. As stated earlier, the only exception is in the special case $M=0$. The function $\mathcal{J}_\sigma(\rho)$ in the kernel then reduces to the quasiseparable form

$$\theta(u-u')(u'/u)^{(\sigma+1)/2}+\theta(u'-u)(u/u')^{(\sigma+1)/2}, \tag{3.3}$$

and the equation can be exactly solved. In the weak-coupling limit, there is one pole near $\sigma=-1$, given by $\sigma+1=G^2/m^2$; and two poles near $\sigma=-N$ ($N\geq 2$), given by $\sigma+N=\pm G^2/m^2$. The corresponding statements in the l plane follow trivially. We state these exact results here because we shall subsequently check the corresponding results of the trace and various factorizable approximations against them.

In view of the importance of being able to make some exact statements at least in the case of forward scattering, it is worth studying the eigenvalue equation (2.6) in some detail. Here we shall mention very briefly some results in this direction, and indicate a new method of extracting weak coupling results with very little effort.[34] From the known singularity structure of the kernel $\mathcal{J}_\sigma$ in u and u', we can analytically continue the eigenfunction $\phi(u)$ off the real positive axis in u, and find its singularities. We find that the function $\Phi(u)=\phi(u)u^{-(\sigma+1)/2}$ has branch points at $-(nM)^2$, where $n=1,2,\cdots$.[35] A study[21] of the corresponding configuration space problem, showing how the "wavefunction" behaves near the origin, indicates that, at least up to $\sigma=-3$, the singularity of $\Phi(u)$ at $u=\infty$ is not relevant in the weak-coupling limit. This immediately leads us to compactify the region of integration by changing variables to $v=M^2/(M^2+u)$, $v'=M^2/(M^2+u')$. Denoting the new eigenfunction divided by $v^{\sigma+1}$ by Φ once again, we get from Eq. (2.6)

$$\Phi(v)=(G^2/M^2)2^{\sigma+1}(\sigma+1)^{-1}\int_0^1 dv'[1+(a-1)v']^{-2}\times H_\sigma(v,v')\Phi(v'), \tag{3.4a}$$

where $a=m^2/M^2$ and

$$H_\sigma(v,v')=[v'(1-v')]^{\sigma+1}[v+v'-vv'+\{(v+v'-vv')^2-4vv'(1-v)(1-v')\}^{1/2}]^{-\sigma-1}. \tag{3.4b}$$

Neglecting the singularities of $\Phi(v')$ at $v'=0$ and 1, the right-hand side of Eq. (3.4a) can be analytically continued to the left of $\mathrm{Re}\sigma=-2$ by converting the v'-integration to one over the contour $(1-,0+)$ encircling the branch points of H_σ at $v'=0$ and $v'=1$, and multiplying the contour integral by the factor $\exp(i\pi\sigma)/(2i\sin\pi\sigma)$. The weak-coupling expressions for the trajectories near $\sigma=-2$ or $\sigma=-3$ are then obtained by substituting these values of σ everywhere except in the pole factor coming from $1/(\sin\pi\sigma)$, the resulting integrals reducing to trivial evaluation of residues. In this simple manner we recover the known expressions for the two trajectories near each of the points $\sigma=-2$ and $\sigma=-3$, namely,

$$(\sigma+2)^2+G^2(\sigma+2)M^2/m^4-G^4/m^4=0,$$

$$(\sigma+3)^2-G^2(\sigma+3)M^2(M^2-2m^2)/m^6+G^4(M^4+2m^2M^2-m^4)/m^8=0. \tag{3.5}$$

Equations (3.5), especially the second one, are arrived at only after a great deal of algebra by any other method.[16,21]

Beyond $\sigma=-3$, the above method has first to be modified because the singularities of $\Phi(v)$ at $v=0$ and 1 also begin to play a role even in the weak-coupling limit, and the procedure becomes more complicated.[36] However, such a contour integral method, taking into account the relevant singularities of the eigenfunction, seems to be the best approach to an algebraically very complicated problem for a general value of N. In any case, the procedure makes it clear that *the behavior of the kernel at the boundary points* $v=0$ ($p^2+\omega^2\to\infty$) *and* $v=1$ ($p,\omega=0$) *is crucial* in deciding the trajectory spectrum of the model. Any approximation that alters this behavior (even if it does so in only *one* of the sets of variables p,ω and p',ω') would cause modifications not only in the positions of the Regge poles but also in their multiplicity.

C. Structure of the kernel

We now consider the kernel of the nonforward equation, in order to make some further general remarks regarding the process of generation of trajectories in the model. Here we shall not be concerned with the problem or analytic continuation discussed earlier.

To understand the spectrum, we have to go at least as far as the larger symmetry of the forward equation. We therefore consider the decomposition of Eq. (2.4). The connection between the weak-coupling limit and the poles of Q_l at negative integral values of l has already been explained. Now the pole of Q_l at $l=-N$ ($N=1,2,\ldots$) is displayed by the expansion of Eq. (2.4) as follows[20]: writing the Gegenbauer polynomials in terms of hypergeometric functions that are regular at $l=-N$, one finds that the pole arises from a factor $\Gamma(2l+n+2)$ in the numerator of the summand. Thus the first $(2N-1)$ terms of the expansion become singular at $l=-N$, while the rest of the terms remain regular and their sum converges as before. *The origin of the number* $(2N-1)$ *of Regge poles near* $l=-N$ *in the* $M=0$ *case rests in the above fact.* To see how the corresponding residue P_{N-1} is constructed, we note that the residue of the infinite series at that point is a finite sum of terms proportional to $\mathcal{J}_{n-N}$, with n running from 0 to $2N-2$. This sum may be split down the middle at the Nth term, and terms symmetrically spaced about this point may be combined by using the identity $\mathcal{J}_{-\alpha-2}=(\mathcal{J}_\alpha)^{-1}$, for any α. The residue at $l=-N$ is then a sum from $r=0$ to $r=N-1$ of terms proportional to

$$\mathcal{J}_{r-1}+(\mathcal{J}_{-r-1})^{-1}=(\rho+(\rho^2-1)^{1/2})^r+(\rho-(\rho^2-1))^{-r},$$

so that the square-root factors cancel out, in order to produce the required polynomial in z.

The term corresponding to $n=N-1$ (or $r=0$) plays a special role: in this case we have $\mathcal{J}_{-1}=1$, and this term is responsible for the sequence of Regge poles with the leading Regge pole of the model as parent. The degeneracy involved in this instance, i.e., the fact that $\mathcal{J}_{-1}=(\mathcal{J}_{-1})^{-1}$ ($=1$, independent of ρ), is the precise reason why this term generates only a *single* Regge pole (near each $-N$ in the weak-coupling limit). We can also see directly how the various trajectories near $l=-1$, -2, -3, etc. are generated. At $l=-1$, only the $n=0$

term of the sum is singular. At $l=-2$, the $n=1$ term generates the daughter of the leading pole, while $n=0$ and $n=2$ generate two new trajectories. At $l=-3$, it is the $n=2$ term that generates the pole belonging to the leading sequence; $n=1,3$ generate the daughters of the two other poles near $l=-2$, while $n=0,4$ give two new trajectories, and so on. Beyond $l=-3$, the comments made earlier suggest the possibility, when $M\neq 0$, of more than one Regge pole being generated by a single value of the index n; the pattern of Regge poles then departs from that in the case $M=0$, as already stated.

With the general discussion of this section, we now have a perspective for understanding the nature and extent of the approximations in which the trajectory spectrum will be discussed in the following sections.

4. THE TRACE APPROXIMATION

Since the eigenvalue equations for the Regge trajectories both for $t=0$ and $t\neq 0$ involve fairly complicated kernels and are not amenable to exact, analytic solution, various approximate solutions have been suggested. Of these, the trace or first Fredholm approximation (hereafter referred to as the TA) has been investigated for the forward equation both analytically and numerically: the agreement between numerical results of the TA and the exact problem is quite a convincing argument for the validity of the approximation and the "approximate factorizability" of the kernel. This point has already been discussed in several papers.[7,12,14] Our purpose here is to work out as fully as possible the trajectory spectrum in the TA for both the $t=0$ and $t\neq 0$ equations, and to explicitly calculate quantities of physical interest such as the Pomeron slope in the model. It turns out that once the basic approximation is made, the rest of the program can be carried out more or less exactly, and we shall find that a large number of the features expected in the solution of the original problem are preserved in this approximation.

A. Trace approximation for the nonforward equation

The TA for the eigenvalues of an integral operator with kernel K consists in writing the Fredholm denominator $D(G^2)$ in the truncated form $1-G^2\,\mathrm{Tr}(K)$, leaving out all the succeeding terms. Adopting this procedure for the integral equation (2.2), the ω dependence of the Q_l function drops out when the trace is taken,[37] and the eigenvalue condition may be written, after doing the integration over ω, as

$$1-\frac{G^2}{M^2}\int_0^\infty dx\, Q_l\left(1+\frac{1}{2x^2}\right)\left(x^2+a-\frac{t}{4M^2}\right)^{-1}(x^2+a)^{-1/2}=0 \tag{4.1}$$

($a=m^2/M^2$). The Regge trajectories are the solutions of this implicit equation for l. For $t\neq 0$ (we may keep it negative), and $a\neq 0$ (physically, we are interested in small positive values of a), the integral in Eq. (4.1) converges, as expected, in the region $\mathrm{Re}\,l>-3/2$.[38] Although we cannot directly evaluate it in a closed form, a power series in t is easily generated. Thus for $|t|\ll 4M^2$, Eq. (4.1) can be written as

$$1-\frac{G^2}{M^2}\sum_{n=0}^{\infty}\left(-tM^2\frac{\partial}{\partial a}\right)^n I(l,a)\frac{n!}{(2n+1)!}=0, \tag{4.2}$$

where

$$I(l,a)=\int_0^\infty dx\, Q_l\left(1+\frac{1}{2x^2}\right)(x^2+a)^{-3/2}. \tag{4.3}$$

While we can proceed directly from Eqs. (4.2) and (4.3), it is more convenient to use a representation for the trace in Eq. (4.1) discovered by Nakanishi.[39] He showed that this could be written in a form rather closely analogous to the Feynman parametric expression for an on-shell vertex diagram: the second term in Eq. (4.1) takes the form

$$G^2\int_0^1 dx_1\,dx_2\,dx_3\,\frac{\delta(1-x_1-x_2-x_3)x_3^l}{[(x_1+x_2)(1-x_3)m^2+x_3M^2-x_1x_2t]}\,. \tag{4.4}$$

Note the extra l-dependent factor in the integrand. While (4.4) is no easier to evaluate explicitly than is the integral in Eq. (4.1), it provides a very simple representation for $I(l,a)$, which is, with the help of Eq. (4.2), sufficient for our purpose. We find

$$I(l,a)=\int_0^1 dx\,\frac{x^l(1-x)}{[x+(1-x)^2a]} \tag{4.5a}$$

$$=a^{-1}\int_0^1 dx\,\frac{x^l(1-x)}{(x+R)(x+R^{-1})}\,, \tag{4.5b}$$

where

$$R=[1-(1-4a)^{1/2}]/[1+(1-4a)^{1/2}]. \tag{4.6}$$

The right-hand side of Eq. (4.5a) will be recognized as being quite similar to the Feynman parametric form for a bubble or self-energy diagram (again with an extra l-dependent factor in the integrand). It is worth giving here a simple geometrical explanation for this: the dual of the vertex diagram referred to above can be reoriented to have M^2 appear as the label of one of the external lines of another vertex diagram, with t as the internal line opposite it. At $t=0$, the new dual diagram collapses to that of a bubble diagram with M^2 labelling the external line and m^2 the internal lines. *This is why the threshold like factor R appears naturally everywhere in the trace approximation.* The form of Eq. (4.5b) will be very useful in our subsequent comparison of the TA for the nonforward equation at $t=0$ with that for the forward equation.

The integral appearing in Eq. (4.5b) can in fact be explicitly evaluated in terms of standard hypergeometric functions. After some algebra, we find

$$I(l,a)=[(l+1)(l+2)]^{-1}[(1+R)/(1-R)]$$
$$\times[R^{-1}F(1,l+1;l+3;-R^{-1})-RF(1,l+1;l+3;-R)], \tag{4.7}$$

a closed form from which various results can be extracted.

B. Weak-coupling limit

Using a convenient power-series representation for the hypergeometric functions in Eq. (4.7), we find that, in the TA, there is just one Regge pole near each of the

points $l=-N$ ($N\geq 1$) in the weak-coupling limit. In terms of the quantity $\theta=-\ln R$, we can write a compact expression for the positions of the above Regge poles in this limit, again as a power series in t. We find finally,

$$l+N=(-1)^{N-1}(G^2/m^2)(M/m)(1+c_1t+\cdots)\times[\sinh(N-\tfrac{1}{2})\theta/\sinh\theta], \quad (4.8)$$

with $c_1=(1/4m^2)[1-(\frac{2}{3})a\partial/\partial a]$, and the higher order terms in t are generated by using Eq. (4.2). The weak-coupling limit enables us to judge how the multiplicity of Regge poles has been altered by the TA. If μ_N is the actual number of Regge poles near $l=-N$ in the weak-coupling limit, the Fredholm denominator in that limit takes the form $\sum a_r[G^2/(l+N)]^r$, with $a_0=1$ and r running from 0 to μ_N. In the TA, however, the above sum is truncated at $r=1$, so that (for $N\geq 2$) the right-hand side in Eq. (4.8) is simply the *sum* of the μ_N weak-coupling limit solutions for $(l+N)$.[40] In the case $t=0$, we would improve the situation regarding the multiplicity of Regge poles by first going over to the forward equation and then making the TA: we shall find a bit later that this gives one pole near each point $\sigma=-N$ ($N\geq 1$), and therefore N Regge poles near $l=-N$. Even this is quite different from the actual number of Regge poles. However, we can say that in each of the two cases the TA *does* give a certain "effective" trajectory position in the appropriate variable.

C. Complex Regge poles

These have been shown to occur in the model by considering the forward equation in the TA[14] or in a factorizable approximation.[11] The possibility of the occurrence of this feature may in fact be understood quite generally, *without* any approximations, by observing the different regions of the l plane in which the kernel in Eq. (2.2) is square-integrable, depending on whether m^2 is zero or not. In the TA to the nonforward equation, for instance, we find that the trace [represented by the integral in Eq. (4.1)] converges in $\mathrm{Re}\,l>-3/2$ for $a\neq 0$, but for $a=0$ it converges only in $\mathrm{Re}\,l>-1$. This will immediately lead to complex Regge poles to the left of the line $\mathrm{Re}\,l=1$ when a takes on the small values ($\ll 1$) that we are interested in, because of the predictable nonholomorphic behavior of the trace at $a=0$ for general values of l. Indeed, it is easy to see from Eq. (4.1) that there occurs a term proportional to a^{l+1} (plus higher powers of a) in the trace when $t\neq 0$, if we use the fact that the Q_l function has a leading behavior $\sim x^{2l+2}$ near $x=0$. As $t\to 0$, complex Regge poles can now occur to the left of the line $\mathrm{Re}\,l=0$, in agreement with the result found from the forward equation.[41] This is also clear if we reduce the representation of Eq. (4.7) for $I(l,a)$ to the following form:

$$I(l,a)=\frac{(1+R)}{(1-R)}\left(\frac{F(1,-l-1;-l+1;-R)}{l(l+1)}-\frac{RF(1,l+1;l+3;-R)}{(l+1)(l+2)}-\frac{\pi R^l(1+R)}{\sin\pi l}\right). \quad (4.9)$$

Since $R=a+O(a^2)$ as $a\to 0$, the last term on the right in Eq. (4.9) is nonholomorphic at $a=0$, and leads to the complex Regge poles referred to above when we solve the eigenvalue condition $1-(G^2/M^2)I(l,a)=0$ for l. We do not give further details here as they can be easily worked out.

D. Threshold behavior

The "accumulation" of Regge poles near $l=-\frac{1}{2}$ as $t\to 4m^2$ is of course well known. Recently this phenomenon has been discussed[11] in the context of a factorizable approximation to the ABFST equation. However, we have already given elsewhere[42] a full analysis of the threshold behavior of the trajectories, with the original kernel retained, in the framework of the BS equation. Once again the phenomenon can be anticipated, even in the TA, simply by noting that the trace in Eq. (4.1) converges only in $\mathrm{Re}\,l>-\frac{1}{2}$ instead of $\mathrm{Re}\,l>-\frac{3}{2}$ when t is set equal to $4m^2$; and that a term proportional to $(t-4m^2)^{l+1/2}$ emerges from the integral using the same leading behavior of Q_l near $x=0$ as before. For the sake of completeness, we give just the results of a simple computation of the threshold behavior in the TA: First, the integral in Eq. (4.1) can be analytically continued to all values of l by converting it to a contour integral around the branch cut of the Q_l function on the real positive x axis. Then, exactly at $t=4m^2$, Regge poles to the right of $\mathrm{Re}\,l=-\frac{1}{2}$ can be shown to be solutions of the equation

$$1-(G^2/4m^2)\exp(-i\pi l)(\sec\pi l)A(l)\approx 0, \quad (4.10)$$

where

$$A(l)=\int_C dx\,x^{-2}(x^2+a)^{-1/2}Q_l(1+1/2x^2), \quad (4.11)$$

C being the (clockwise) contour referred to above. The solution to Eq. (4.10) in the weak-coupling limit is

$$l+\tfrac{1}{2}\approx iG^2A(-\tfrac{1}{2})/4\pi M^2=\pi G^2/2Mm, \quad (4.12)$$

correctly reproducing the weak-coupling expression for the leading Regge pole at threshold.[16,42] Again, for t very close to $4m^2$, we can show that the eigenvalue condition of Eq. (4.1) takes the form

$$C(l)[(4m^2-t)/M^2]^{l+1/2}\approx 1, \quad (4.13$$

where

$$C(l)=iA(-\tfrac{1}{2})/[\exp(-i\pi l)A(l)+(4M^2/G^2)\cos\pi l],\qquad C(-\tfrac{1}{2})=1.$$

The "accumulation" of Regge poles near $l=-\frac{1}{2}$ as $t\to 4m^2$ now follows in the standard manner from Eq. (4.13). The TA thus preserves, by and large, the threshold behavior of trajectories [although the exact form of $C(l)$ is a bit different from that given above]. This is not surprising, since the approximation does not alter the "propagator" factor $f(p,\omega,t)$ in the kernel of Eq. (2.2), and it is this factor that leads to the threshold singularity at $t=4m^2$.

E. Pomeron intercept and slope

The intercept at $t=0$ of the leading trajectory of the model, denoted by α_0, is found in the TA by solving the equation

$$I(\alpha_0,a)=M^2/G^2. \quad (4.14)$$

For a kernel more general than the resonance kernel we are working with, we are to understand by (G^2/M^2) in this equation the quantity $R=\int ds C_I(s)/s$, where $C_I(s)$

has already been introduced in Eq. (2.8).

The slope α_0' of the leading trajectory at $t=0$ is given by [see Eq. (4.2)]

$$\alpha_0' = (1/6M^2)\left[\left(\frac{\partial I}{\partial a}\right)\Big/\left(\frac{\partial I}{\partial l}\right)\right]_{l=\alpha_0}. \tag{4.15}$$

For $a \ll 1$, we find that if $0<\alpha_0<1$, then α_0 and α_0' are given by

$$[\alpha_0(1+\alpha_0)]^{-1} - \pi a^{\alpha_0}\csc\pi\alpha_0 \approx R^{-1}, \tag{4.16}$$

and

$$\alpha_0' \approx \frac{\alpha_0^2(\alpha_0+1)^2}{M^2(2\alpha_0+1)}\left(\frac{\pi\alpha_0 a^{\alpha_0-1}}{6\sin\pi\alpha_0} - [(1-\alpha_0)\alpha_0(\alpha_0+1)(\alpha_0+2)]^{-1}\right). \tag{4.17}$$

The factor a^{α_0-1} in Eq. (4.17) shows that the slope depends strongly on the ratio m^2/M^2 (m is to be identified with the pion mass, while M is typically of the order of 1 GeV).[43]

Let us now consider the interesting case of the Pomeron, with $\alpha_0=1$. We find that the quantity $I(1,a)$ can be written completely in terms of elementary functions,[44] so that Eq. (4.14) takes the form

$$R^{-1} = (1+R)^2[(1-R)^{-1}R\ln R + R^{-2}(1+R+R^2)\ln(1+R) - R^{-1}], \tag{4.18}$$

where we have expressed $I(1,a)$ as a function of the quantity R defined earlier, in Eq. (4.6). With $\alpha_0=1$, Eq. (4.18) is an *exact* equation in the framework of the TA. As a increases from 0 to ∞, R increases monotonically from 2 to ∞. For $a=(m_\pi/m_\rho)^2 \approx 0.018$, we find $R^{-1}\approx 0.47$ and $G^2 \approx 1.2$ (GeV)^2; while for $m=m_\pi$, $M=1$ (GeV), ($a=0.031$), we find $R^{-1}\approx 0.46$ and $G^2\approx 2.2$ (GeV)^2. We can also write an exact expression for the slope α_0' when $\alpha_0=1$, but this will not be solely in terms of elementary functions because $(\partial I/\partial l)$ is not, even at $l=1$. However, for the small values of a we are interested in, we can write an expansion for α_0' that reads, retaining all the necessary terms.

$$\alpha_0' \approx (2/9M^2)[-\ln a - (17/6) + ca], \tag{4.19}$$

where

$$c = -2(\ln a)^3/3 - 17(\ln a)^2/9 - (185+6\pi^2)(\ln a)/27 - 17(3\pi^2+52)/81 + O(a). \tag{4.20}$$

We note the logarithmic dependence of the slope on m^2, a feature also found in some factorizable approximations.[45] For the case $m=m_\pi$ and $M=1$ (GeV) considered above, we find $\alpha_0' \approx 0.27$ (GeV)$^{-2}$. For $a=(m_\pi/m_\rho)^2$, the term ca in Eq. (4.19) introduces an appreciable correction to the contribution of the first two terms because of the larger value of a, and we find $\alpha_0' \approx 0.59$ (GeV)$^{-2}$. It is gratifying that the TA yields such reasonable values for the slope of the leading trajectory.

F. Trace approximation for the forward equation

This was first considered in Ref. 7, and subsequently in greater detail in Ref. 14 in connection with the occurrence of complex Regge poles. Arguments in favor of the approximation have also been presented in these references. The trace concerned has been evaluated in Ref. 14, but the resulting expression is not in a closed form, although the leading terms in $a=m^2/M^2$ can be found from it; it is in fact written, after some labor, as an infinite power series in $\exp(-\alpha)$, where $\cosh\alpha = (1/2a) - 1$.[46] Here we want to show that some simple manipulations put the trace in a form which can be (i) *easily* compared with the corresponding expression from the nonforward equation, (ii) evaluated in a neat closed form, from which all the relevant information can be read off.

The eigenvalue condition for the forward equation (2.6) in the TA is

$$1 - RJ(\sigma, a) = 0,$$

where

$$J(\sigma,a) = (\sigma+1)^{-1}\int_0^\infty dv(v+a)^{-2}\left(\frac{2v}{2v+1+(4v+1)^{1/2}}\right)^{\sigma+1}. \tag{4.21}$$

It is difficult to compare this with the condition $1 - RI(l,a)=0$ obtained from the TA for the nonforward equation on setting $t=0$, if we use the original representation of Eq. (4.3) for $I(l,a)$. However, if we change the variable of integration in Eq. (4.21) to

$$x = 2v/[2v+1+(4v+1)^{1/2}]$$

and partially integrate once, we get

$$aJ(\sigma,a) = (\sigma+1)^{-1} - \int_0^1 dx\,\frac{x^{\sigma+1}}{x+(1-x)^2 a}. \tag{4.22}$$

This is to be compared with Eq. (4.5) for $I(l,a)$. The similarities and differences between the eigenvalue conditions $1-RI(\lambda,a)=0$ and $1-RJ(\lambda,a)=0$ are now quite obvious.[47] Evaluating $J(\sigma,a)$ explicitly, we find

$$aJ(\sigma,a) = (\sigma+1)^{-1} - (\sigma+2)^{-1}[(1+R)/(1-R)][R^{-1}F(1,\sigma+2;\sigma+3;-R^{-1}) - RF(1,\sigma+2;\sigma+3;-R)], \tag{4.23}$$

the analog of Eq. (4.7). As before, various results can be easily deduced from Eq. (4.23). For example, in the weak-coupling limit, there is one pole near $\sigma=-1$, -2, ... etc., given again by a simple closed expression:

$$\sigma+1 = G^2/m^2,\ \sigma+N = (-1)^{N-1}(G^2/m^2)(M^2/m^2) \times[\sinh(N-1)\theta/\sinh\theta] \quad (N\geq 2). \tag{4.24}$$

This is the analog of Eq. (4.8) ($\theta=-\ln R$ as before). We find that for $N=2$ and 3, Eq. (4.24) gives the correct *sums* of the roots for $(\sigma+2)$ and $(\sigma+3)$ in Eq. (3.5). As discussed in Sec. 4B, we expect this to happen in the TA. Again, to deduce the presence of complex poles, we first separate the term in $J(\sigma,a)$ that is singular at $a=0$, thus:

$$aJ(\sigma,a) = (\sigma+1)^{-1} - \frac{(1+R)}{(1-R)}\left(\frac{F(1,-\sigma-1;-\sigma;-R)}{(\sigma+1)} - \frac{RF(1,\sigma+2;\sigma+3;-R)}{(\sigma+2)} - \frac{\pi R^{\sigma+1}}{\sin\pi\sigma}\right). \tag{4.25}$$

This is the counterpart of Eq. (4.9). The last term on the right gives rise to complex poles to the left of $\mathrm{Re}\,\sigma=0$ for small values of a [recall that $R=a+O(a^2)$ as $a\to 0$]. The position of the leading pole (for which $\mathrm{Re}\,\sigma>0$) is given, as m^2 (or a) $\to 0$, by the approximate equation $\sigma(\sigma+1)(\sigma+2)\approx 2R$, the expression used in Ref. 7. If this pole is at $\sigma=1$, we must have $R\approx 3$. A numerical solution[25] of the exact forward equation gives $R\approx 5.3$ in

this situation. The nonforward equation gives instead the condition [see Eq. (4.16)] $\alpha_0(\alpha_0+1)\approx R$; and for $\alpha_0=1$ we must have $R\approx 2$. Finally, let us consider the $M^2=0$ limit. The exact results in this case have been stated following Eq. (3.3). The TA of this Section yields just one pole in the σ plane, given by $\sigma+1=G^2/m^2$. It thus gives the correct sum of roots for $(\sigma+N)$ to $O(G^2)$, a behavior expected of the TA, as already explained. We shall use all these points to evaluate, in the next section, the results of the factorizable approximations that have been suggested.

5. FACTORIZABLE APPROXIMATIONS

Various separable kernels have been proposed in several recent papers as replacements for the nondegenerate kernels in the forward and nonforward partial-wave integral equations of the ABFST model, in the attempt to obtain at least approximate solutions to the problem. Our object here is the very restricted one of examining certain features of the trajectory structures that emerge in these factorizable approximations, and making some critical comments on the nature of such approximations in the light of the general discussion we have given earlier on the role of the singularity structure of the basic kernel in the generation of trajectories. As in the other sections of this paper, we shall be concerned only with the trajectory functions and not with the corresponding residues. The simpler forward equation will be taken up first.

A. Factorizable approximations for the forward equation

The approximations that have been proposed replace the nondegenerate kernel $\mathcal{J}_\sigma(u,u')$ of Eq. (2.6) by a separable, rank-one kernel of the form $[f(u)g(u')]^{\sigma+1}$. The latter quantity is generally taken to match the value of $\mathcal{J}_\sigma(u,u')$ at some boundary of the region $0\leq u,u'<\infty$ of the variables u and u', since it is the behavior of the kernel at such points that determines the singularity structure.

(a) The first case we consider is the choice[11,12]

$$f(u)=u^{1/2},\quad g(u)=u^{1/2}(u+M^2)^{-1}, \tag{5.1}$$

so that $[f(u)g(u')]^{\sigma+1}$ is the limit of $\mathcal{J}_\sigma(u,u')$ as $u\to 0$; it also correctly reproduces the $u'\to\infty$ limit and has the correct power behavior for u' near 0. The choice $[f(u')g(u)]^{\sigma+1}$, with f and g as given above, leads (obviously) to the same *trajectory* structure. The eigenvalue condition may be written as

$$R^{-1}=[(\sigma+1)(\sigma+2)]^{-1}F(1,2;\sigma+3;1-a). \tag{5.2}$$

Writing the ${}_2F_1$ function in Eq. (5.2) in terms of functions of argument a yields the expected $a^\sigma/(\sin\pi\sigma)$ term that leads to complex poles, etc. These details have already been given in Ref. 11. Let us consider the eigenvalue condition of Eq. (5.2) in the weak-coupling limit. We find that the leading pole is given by $\sigma+1=G^2/m^2$, agreeing with both the TA and the exact expression, whereas the secondary poles (one near each negative integer) are given by $\sigma+N=(-1)^{N-1}(G^2/m^2)(M^2/m^2)(M^2/m^2-1)^{N-2}$, $N\geq 2$. This agrees with the TA only for $N=2$. The limit $M^2\to 0$ produces just one pole in the σ plane, given by $\sigma+1=G^2/m^2$, which also agrees with the TA. Again, for $a\to 0$ the leading pole is given by $R\approx\sigma(\sigma+1)$ so that $R\approx 2$ for $\sigma=1$, which is a bit removed from the value $R\approx 5.3$ yielded by the numerical solution to the exact equation. (However this last feature agrees with the $t=0$ limit of the TA for the nonforward equation.) We conclude that this simplest factorizable approximation is quite a reasonable one on most counts.

(b) Next, we may try to match also the large-u behavior of $\mathcal{J}_\sigma(u,u')$ by modifying our choice of f and g to[48]

$$f(u)=g(u)=Mu^{1/2}(u+M^2)^{-1}. \tag{5.3}$$

The eigenvalue condition is then

$$R^{-1}=[\Gamma(\sigma+1)\Gamma(\sigma+2)/\Gamma(2\sigma+4)]F(2,\sigma+2;2\sigma+4;1-a). \tag{5.4}$$

The leading pole is well-described in this approximation, for the weak-coupling limit gives $\sigma+1=G^2/m^2$, and the $a\to 0$ limit gives $R\approx\Gamma(2\sigma+2)/\Gamma(\sigma)\Gamma(\sigma+1)$, so that $R\approx 6$ for $\sigma=1$. However the $M^2=0$ limit is totally incorrect, which is not surprising because the form assumed for the kernel is also an approximation to $\mathcal{J}_\sigma(u,u')$ when M^2 is *large* compared to u,u'. Further, the secondary spectrum is also quite different from that of the TA or the exact equation.

(c) Finally, let us consider the choice[48]

$$f(u)=g(u)=u^{1/2}(M^2+2u)^{-1/2}, \tag{5.5}$$

which gives a kernel that is again a good approximation to $\mathcal{J}_\sigma(u,u')$ for $M^2\gg u,u'$, without totally altering the $M^2=0$ limit. The eigenvalue condition obtained is

$$R^{-1}=[2^\sigma(\sigma+1)(\sigma+2)]^{-1}F(1,2;\sigma+3;1-2a). \tag{5.6}$$

As expected, in the weak-coupling limit there is one pole near each negative integer in the σ plane, given by $\sigma+1=G^2/m^2$, $\sigma+N=(-1)^{N-1}(G^2/m^2)(M^2/m^2)(M^2/m^2-2)^{N-2}$, $N\geq 2$. This agrees with the results of the TA all the way up to $N=3$. There is one pole near $\sigma=-1$ in the $M^2=0$ limit, given by $\sigma+1=G^2/2^{\sigma+1}m^2$, which is reasonably close to the result in the TA. As for the leading pole as $a\to 0$, this is given by the solution of $R\approx 2^\sigma\sigma(\sigma+1)$, or $R\approx 4$ for $\sigma=1$. We therefore consider the approximation of Eq. (5.5) a good one.

We note that none of the above approximations properly handles the boundary in which u and u' both $\to\infty$, or at least as well as the TA does, although the choice of case (a) is again a reasonable one in this respect.

B. Factorizable approximations for the nonforward equation

The approximations that have been proposed here[11,12,48] are rather straightforward extensions of those for the forward equation. The Q_l function in the eigenvalue equation (2.2) is replaced by its asymptotic form $B(\frac{1}{2},l+1)(2z)^{-l-1}$ for large values of the argument z, i.e., by the first term in the expansion of Eq. (3.2). Next, the cross term $-2(uu')^{1/2}\xi\xi'$ in the expression $[M^2+u+u'-2(uu')^{1/2}\xi\xi']/2[uu'(1-\xi^2)(1-\xi'^2)]^{1/2}$ for z is dropped as being negligible compared to M^2. This of course reduces the integral equation to a one-dimensional equation, because the dependence of the kernel on ξ and ξ' is now in a factorized form. The intergal equation obtained still has a nondegenerate kernel because of the factor

$[(uu')^{1/2}/(M^2+u+u')]^{l+1}$. At this stage the same factorized approximations as were used in the case of the forward problem may be used, namely, the replacement of the above expression by $[f(u)g(u')]^{l+1}$, with f and g as in (a), (b), or (c) of Sec. 5A. Since the exponent that appears is $(l+1)$, the net effect is *also* equivalent to retaining just the first term in the expansion of Eq. (2.4) for Q_l, and then approximating $\mathcal{J}_l(u,u')$ by one of the above factorized forms. The resulting eigenvalue condition reads

$$R^{-1}=(l+1)^{-1}\int_0^\infty dx\,\frac{[f(M^2x)g(M^2x)]^{l+1}}{(x+a-t/4M^2)^2}$$

$$\times F\left[\tfrac{1}{2},1;l+2;\frac{-tx}{M^2(x+a-t/4M^2)^2}\right]. \qquad (5.7)$$

Considering precisely how the sequence of approximations has run in deriving Eq. (5.7), it is not at all surprising that at $t=0$ the eigenvalue condition collapses to just that found in Sec. 5A, with σ simply replaced by l. For f and g chosen as before, we can proceed to show that Eq. (5.7) gives a simple pole near each negative integer in the l plane, in the weak-coupling limit. Except for the leading Regge pole, the other trajectory functions bear little relation to either the results of the TA or the known exact expressions. In fact the "continuity" of the above eigenvalue conditon (for l) at $t=0$, namely, its coinciding with the eigenvalue condition (for σ) derived from the forward equation, itself suggests how drastic the approximation is for all but the leading trajectory.

An improved version of the factorizable approximation for the nonforward equation has also been proposed.[12] The expansion of Eq. (2.4) is used for the Q_l function in the kernel, and all the terms of the expansion are retained. A factorized approximation is used for $\mathcal{J}_{l+n}(u,u')$ for each value of n (just the choice of case (a) above for the functions f and g). The Fredholm denominator continues to be rather complicated, being obtained as an infinite determinant, each element of which is an (one-dimensional) integral. Even with this refinement the secondary trajectory spectrum is rather severely altered, there being one Regge pole near each negative integer value of l given (in the case $M=m$) by the too-simple expression $l+N\approx G^2/m^2$. As already stressed in Ref. 12, what is in fact established by the detailed consideration of such factorizable approximations is the validity of the TA for the problem at hand. We agree with this conclusion.

6. CONCLUDING REMARKS

We have studied the trajectory structure of the ABFST multiperipheral model with a resonance kernel, one of the simplest of a class of models of considerable importance in high-energy scattering, using both the forward and nonforward partial-wave integral equations. Our motivation and method of approach have already been spelt out at length in the Introduction. We have provided some insight into the complicated problem of the trajectory structure by analyzing the singularity structure of the relevant kernel using appropriate representations (expansions) for it. This procedure would remain an essential step when other, more complicated, kernels are considered. We have also worked the problem out in considerable detail in the trace or first Fredholm approximation. That this is a good approximation has already been hinted at in several works including some of those dealing with factorizable approximations. An examination of such factorizable approximation has shown us that they suffer from shortcomings that include, and, by and large, are more extensive than, those of the TA. We find that the TA preserves, to a fair degree of accuracy, the singularity structures of the relevant kernels. [For instance, in the case of the nonforward equation, the eigenvalue condition in the TA still involves a Q_l function (with its attendant singularities in l), instead of merely the first term in an expansion of this function, as in the factorizable approximation—such a term may be the correct asymptotic limit of the function for large values of the argument, but it certainly drastically distorts the residues of the kernel at its poles in the l plane.] We also find that the TA leads to good "effective" trajectory positions for the leading and secondary poles in both the forward and nonforward cases, and that other phenomena such as complex Regge poles, threshold behavior of trajectories, the intercept and slope of the leading trajectory, etc. can be investigated in a close simulation of the actual problem, and more or less exactly. The recognition of the fact that the traces concerned can be evaluated in convenient closed forms, as functions of the particular quantity R (the reason for the natural occurence of which we have explained), should also be given due consideration in this connection, for it means that we have not had to introduce further *ad hoc* assumptions in the middle of our analysis.

ACKNOWLEDGMENTS

I should like to thank Professor B.M. Udgaonkar and the members of the Theory Group at TIFR for the kind hospitality they have shown me. I am grateful to Professor N. Mukunda for reading parts of the manuscript and for his valuable suggestions. Most of Sec. 3 of this work was carried out at Brandeis University, with support from NSF Grant GP 7257, and I should like to thank Professor M. Grisaru for his help and encouragement.

[1]L. Bertocchi, S. Fubini, and M. Tonin, Nuovo Cimento 25, 626 (1962); D. Amati, A. Stanghellini, and S. Fubini, Nuovo Cimento 26, 896 (1962). See also D. Amati *et al.*, Nuovo Cimento 22, 569 (1961); C. Ceolin *et al.* Nuovo Cimento 26, 547 (1962); and L. Bertocchi *et al.*, Nuovo Cimento 27, 913 (1963).

[2]G. F. Chew, M.L. Goldberger, and F. Low, Phys. Rev. Lett. 22, 208 (1969); I.G. Halliday and L.M. Saunders, Nuovo Cimento A 60, 177 (1969); S. Pinsky and W.I. Weisberger, Phys. Rev. D 2, 1640 (1970).

[3]The literature is truly vast. In addition to the papers cited in Ref. 1–14, we give as general references the following: M.L. Goldberger, "*Multiperipheral Models and High Energy Processes*," Princeton University Report, July 1971 (unpublished); H.D.I. Abarbanel, G.F. Chew, M.L. Goldberger, and L.M. Saunders, Ann. Phys. (N.Y.) 73, 156 (1972).

[4]Some representative references on this important topic are: G.F. Chew and A. Pignotti, Phys. Rev. 176, 2112 (1968); G.F. Chew and C. De Tar, Phys. Rev. 180, 1577 (1969); L. Caneschi and A. Pignotti, Phys. Rev. 180, 1525 (1969) and Phys. Rev. 184, 1915 (1969); M.L. Goldberger, C.-I. Tan, and J.-M. Wang, Phys. Rev. 184, 1920 (1969); H.D.I.

Abarbanel *et al.*, Phys. Rev. D 4, 2988 (1971).
[5]J.S. Ball and G. Marchesini, Phys. Rev. 188, 2209, 2508 (1969).
[6]For relevant calculations using the model, further see D.M. Tow, Phys. Rev. D 2, 154 (1970); G.F. Chew and D.R. Snider, Phys. Rev. D 1, 3453 (1970) and Phys. Rev. D 3, 420 (1971); H.D.I. Abarbanel *et al.*, Phys. Rev. Lett. 25, 1735 (1970); M.L. Goldberger, D. Silverman, and C.-I. Tan, Phys. Rev. Lett. 26, 100 (1971). See also Ref. 7.
[7]G.F. Chew, T. Rogers, and D.R. Snider, Phys. Rev. D 2, 765 (1970).
[8]As an illustration of this point, see for instance D.R. Snider and D.M. Tow, Phys. Rev. D 3, 996 (1971).
[9]M. Ciafaloni, C. De Tar, and M.N. Misheloff, Phys. Rev. 188, 2522 (1969); H.D.I. Abarbanel and L.M. Saunders, Phys. Rev. D 2, 711 (1970); L.M. Saunders, O.H.N. Saxton, and C.-I. Tan, Phys. Rev. D 3, 1005 (1971).
[10]See for instance the last paper of Ref. 9.
[11]D. Silverman and P.D. Ting, Nucl. Phys. B 35, 445 (1971).
[12]J. Dash, Cambridge University Preprints DAMTP 71/12 and 71/19 (unpublished).
[13]See for instance the third paper cited in Ref. 6.
[14]S.-S. Shei, Phys. Rev. D 3, 1962 (1971).
[15]The standard reference is R.J. Eden *et al.*, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).
[16]J.C. Polkinghorne, J. Math. Phys. 5, 431 (1964); A.R. Swift, J. Math. Phys. 8, 2420 (1967); A.R. Swift, Phys. Rev. 171, 1466 (1968) and 176, 1848 (1968); I.G. Halliday and P.V. Landshoff, Nuovo Cimento A 56, 983 (1968); A.R. Swift and R.W. Tucker, Phys. Rev. D 1, 2894 (1970); H. Cheng and T.-T. Wu, Phys. Rev. D 2, 2285 (1970).
[17]N. Nakanishi, Prog. Theoret. Phys. Suppl. 43, 1 (1969), and references contained therein.
[18]B.W. Lee and R.F. Sawyer, Phys. Rev. 127, 2266 (1962); G. Tiktopoulos, Phys. Rev. 133, B1231 (1964); H. Abe, G. Konisi, and T. Ogimoto, Progr. Theoret. Phys. 32, 348 (1964).
[19]M. Fontannaz, Nuovo Cimento A 59, 215 (1969).
[20]V. Balakrishnan, Brandeis University, thesis, 1970 (unpublished).
[21]R.W. Haymaker and R. Blankenbecler, Phys. Rev. 186, 1648 (1969).
[22]G.C. Wick, Phys. Rev. 96, 1124 (1954); R.E. Cutkosky, Phys. Rev. 96, 1135 (1954).
[23]R. Gatto and P. Menotti, Nuovo Cimento A 68, 118 (1970).
[24]For the case of the *BS* equation, see for instance V. Chung and D.R. Snider, Phys. Rev. 162, 1639 (1967); R.E. Cutkosky and B.B. Deo, Phys. Rev. Lett. 19, 1256 (1967).
[25]For the case of the ABFST model, see M. Misheloff, Phys. Rev. D 3, 1486 (1971). Also see Ref. 5.
[26]D.Z. Freedman and J.-M. Wang, Phys. Rev. 153, 1596 (1967).
[27]See, for instance, Ref. 26 and also L.M. Saunders *et al.* cited in Ref. 9.
[28]G. Domokos, Phys. Rev. 159, 1387 (1967); A.H. Mueller, Phys. Rev. 172, 1516 (1968); A. Sciarrino and M. Toller, J. Math. Phys. 8, 1252 (1967). See also Ref. 26.
[29]As a general reference on special functions, etc., we have throughout used I.S. Gradshteyn and I.W. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[30]See Refs. 22 and 23; also see, for example, N.P. Chang and R.P. Saxena, Phys. Rev. 176, 2101 (1968). At $t=0$ this differential equation can be solved in terms of standard hypergeometric functions.
[31]A technical reason is that the region where *both* sets of variables (p,ω) and (p',ω') are large also becomes relevant in the generation of singularities.
[32]L. Caneschi, Nuovo Cimento Lett. 1, 70 (1969).
[33]See Refs. 20, 21, 30, 32.
[34]V. Balakrishnan and M. Grisaru (unpublished).
[35]These are the expected "thresholds." In addition, $\Phi(u)$ has a singularity at $u=0$ on the unphysical sheets reached by first encircling the singularity at $u=-M^2$, essentially because the square-root factor in $\mathcal{J}_\sigma$ then changes sign and exposes another branch-point of the kernel. The singularity structure of $\Phi(u)$ suggests that one write a spectral representation for it, and work out the corresponding equation for the weight function.[1] We have been able to compute[34] the kernel in this equation in a closed form, but our hope that this would yeild a simpler problem of analytic continuation in σ is not realized because of the recursive mechanism by which successive singularities of $\Phi(u)$ are generated.
[36]This is related to our comment in Ref. 31. From the technical point of view, we have to perform step-by-step analytic continuation (in α and β) of integrals of the type $\int_0^1 dx\, x^\alpha(1-x)^\beta \times (\ln x)^n f(x)$, where $f(x)$ is regular in the range of integration.
[37]This immediately blurs the distinction of a trajectory by its "time-parity." Although a better procedure would be to separately make the TA for the kernels, respectively, even and odd in ω (and therefore in ω'), we stick to the simpler procedure here.
[38]Formal analytic continuation by means of a contour integral is very simple in this instance, but it is not convenient for our purpose.
[39]N. Nakanishi, Phys. Rev. 130, 1230 (1963); and Phys. Rev. 131, 2841 (1963). We believe that a far more direct derivation of this relation (based on a certain representation for the Q_l function concerned) should be possible, but so far we have been unable to complete the derivation along these lines.
[40]We can check that at $t=0$, taking the limit $M^2\to 0$ [using a more convenient representation for $I(l,a)$] gives $l+N\simeq G^2/m^2$, the correct sum of roots for the Wick–Cutkosky model. Of course the TA does not give the full expression for the sum of roots to higher orders in G^2.
[41]And in fact the integral in Eq. (4.1) converges only in $\mathrm{Re}\, l>0$ if we first set $t=0$ in the integrand.
[42]V. Balakrishnan, Phys. Rev. D 4, 3205 (1971).
[43]This fact has also been pointed out in Ref. 11 in the context of their factorizable approximation. Their expression for α_0' is similar to that obtained here, except for an extra factor $2/(1+\alpha_0)$ in the last term.
[44]The algebraic branch point of $I(l,a)$ at $a=0$ for general l is converted into a logarithmic one for positive integer l.
[45]See, for instance, Ref. 11. It is remarkable that the first two terms on the right in Eq. (4.19) are precisely those obtained in the BFT[1] approximation to the ABFST equation.
[46]$\exp(-\alpha)$ will be easily recognized as our quantity R.
[47]The result of such a comparison is more specific than the general relation $I(l,a)=\sum_0^\infty J(l+n,a)$ that is derivable from Eq. (2.4).
[48]See for instance the Report cited in Ref. 3.

Branching rules for $GL(N) \supset \Sigma_m$ and the evaluation of inner plethysms

R. C. King*

Department of Mathematics, The University of Toronto, Toronto, Canada
(Received 24 April 1972; revised manuscript received 3 July 1972)

It is pointed out that the operation of inner plethysm may be defined in terms of the branching rules associated with the decomposition of an irreducible representation of the general linear group into a set of irreducible representations of the symmetric group. A variety of methods of evaluating inner plethysms are developed and a number of results tabulated. Special emphasis is placed on those results relevant to the embedding of crystallographic point groups in the rotation group in three dimensions. Symmetrized squares and cubes of some of these groups are deduced and an application to the symmetry properties of $3j$ symbols is given.

I. INTRODUCTION

Littlewood[1,2] defined a new multiplication of S functions which he later called the operation of plethysm and denoted by the symbol $\otimes$. Later still he introduced another operation on S functions which he referred to as the operation of inner plethysm[3] and denoted by the symbol $\odot$. Inner plethysms bear the same relationship to inner products of S functions as the original plethysms, now known as outer plethysms, bear to outer products of S functions.

Murnaghan,[4] on the other hand, defined a particular class of inner plethysms in terms of the symmetrization of Kronecker powers of irreducible representations of the symmetric group. Since it was well known that outer plethysms correspond to the symmetrized Kronecker powers of irreducible representations of the general linear group, he used the same symbol, $\otimes$, to denote both inner and outer plethysms, which were distinguished by the use of different symbols in the specification of irreducible representations of the symmetric and linear groups. This notation is adopted here.

As symmetrized powers, inner plethysms have an important role to play in theoretical physics, since, in particular, the problem of the reduction of a Kronecker square into its symmetric and antisymmetric parts occurs in many different contexts[5-11] some of which involve groups isomorphic with a symmetric group. Furthermore, the symmetrized cubes are the key to an understanding of the symmetry properties of the generalized $3j$ symbols appropriate to the coupling of physical states; these correspond to basis states of irreducible representations of a physical symmetry group.

However, inner plethysms were first used explicitly in the context of problems involving the reduction from a continuous symmetry group to a finite symmetry group.[12] The particular inner plethysms discussed by Murnaghan and whose evaluation was later reduced to a well-defined procedure by Littlewood[13] are of great use in analyzing the physical states of the nuclear shell model.[14]

Butler and Wybourne[15] first pointed out explicitly that outer plethysms may be identified with the branching rules associated with $GL(N) \supset GL(M)$. The first aim of this paper is to stress the fact that inner plethysms may be identified with the branching rules associated with $GL(N) \supset \Sigma_m$. This identification is made explicit in the next section in which a method of determining the algebraic rules[3] associated with the operation of inner plethysm is outlined.

In Sec. 3 a number of rather special inner plethysms are evaluated using a variety of techniques based primarily on dimensionality arguments and the properties of the well known Young–Yamanouchi orthogonal realization of the representations of the symmetric group.[7,16]

Some of these techniques are extended in Sec. 4 to give the evaluation of a class of plethysms associated with the embedding of the dihedral group, D_3, the tetrahedral group, T, the octahedreal group, O, and the icosahedral group, Y, in the three-dimensional rotation group $SO(3)$. The results obtained are not new,[6,7] but the method used is original.

It has recently been demonstrated[17] that a study of the subgroup chain $GL(N) \supset GL(M) \supset GL(M-1)$ yields some powerful techniques for the evaluation of outer plethysms. In the same way techniques for the evaluation of inner plethysms are derived in Sec. 5 by a consideration of the subgroup chain $GL(N) \supset \Sigma_m \supset \Sigma_{m-1}$ and, more generally, of the chain $GL(N) \supset \Sigma_m \supset \Sigma_s \times \Sigma_t$ with $m = s + t$.

Using all the techniques of Secs. 3–5, a table of inner plethysms is drawn up which represents the first systematic enumeration of such results. It is shown that it is a trivial matter to derive from the tables a number of results concerning the embedding of one group in another.

Finally an application of the results to the determination of the symmetry properties of $3j$ symbols is discussed and illustrated by reference to the octahedral group.

II. THE ALGEBRA OF INNER PLETHYSMS

Each of the irreducible representations (μ) of the symmetric group on m symbols, Σ_m, may be specified by a partition, μ, of m into p parts, so that $(\mu) = (\mu_1, \mu_2, \ldots, \mu_p)$ with $\mu_1 + \mu_2 + \ldots + \mu_p = m$ and $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_p > 0$. The dimension, or degree, of such a representation is denoted by $f^{(\mu)}$.

If

$$f^{(\mu)} = N \tag{2.1}$$

then the matrices of the representation (μ) of Σ_m form a

subset of the set of nonsingular $N\times N$ matrices associated with the general linear group $GL(N)$. Thus Σ_m is necessarily a subgroup of $GL(N)$ if there exists a faithful representation (μ) such that (2.1) is satisfied. If (μ) is not a faithful representation of Σ_m then the relevant subgroup is $\Sigma_m/K^{(\mu)}$, where $K^{(\mu)}$ is the normal subgroup of Σ_m consisting of the set of elements of Σ_m mapped onto the $f^\mu\times f^\mu$ unit matrix in the representation (μ). Of course in such a case the matrices of the representation (μ) furnish a faithful representation of the factor group $\Sigma_m/K^{(\mu)}$.

For example the representation (2^2) of Σ_4 is not faithful. The corresponding normal subgroup $K^{(2^2)}$ is the 4 group V, whilst the factor group $\Sigma_4/K^{(2^2)}$ is just Σ_3. The representation (2^2) of Σ_4 then furnishes the representation (21) of Σ_3 which is faithful.

In what follows it is assumed that (μ) is a faithful representation of Σ_m. In certain cases this will not be a valid assumption and the results should then be interpreted in terms of properties of the factor group $\Sigma_m/K^{(\mu)}$.

An irreducible representation $\{\nu\}$ of the general linear group in N dimensions, $GL(N)$, may be specified by a partition, ν, of n into p parts with $p\leq N$. The dimension of such a representation is denoted by $D_N\{\nu\}$.

The defining, N-dimensional, representation of $GL(N)$ is denoted by $\{1\}$, and the embedding of Σ_m in $GL(N)$ is defined by the mapping

$$\{1\}\to(\mu) \tag{2.2}$$

corresponding to the fact that the set of representation matrices $\{1\}$ contains the set (μ). The representation $\{1\}$ of $GL(N)$ is said to subduce in the subgroup Σ_m the representation (μ). Under this mapping (2.2) each irreducible representation $\{\nu\}$ of $GL(N)$ subduces a representation of Σ_m, denoted by $(\mu)\otimes\{\nu\}$, which is in general reducible in accordance with the branching rule:

$$\{\nu\}\to(\mu)\otimes\{\nu\}=\sum_\lambda p_{\mu\nu,\lambda}(\lambda), \tag{2.3}$$

where the summation is taken over all irreducible representations (λ) of Σ_m, so that λ denotes a partition of m. The representation $(\mu)\otimes\{\nu\}$ of Σ_m is said to be an inner plethysm and the determination of the coefficients $p_{\mu\nu,\lambda}$ corresponds to the evaulation of the inner plethysm which is defined by (2.3).

A check on the evaluation of the plethysm $(\mu)\otimes\{\nu\}$ is provided by the fact that the reduction procedure implies, using (2.1) that (Ref. 16, p. 71)

$$f^{(\mu)\otimes\{\nu\}}=D_N^{\{\nu\}}=D_{f^\mu}^{\{\nu\}}, \tag{2.4}$$

i.e.,

$$D_N^{\{\nu\}}=\sum_\lambda p_{\mu\nu,\lambda}f^{(\lambda)}.$$

With the definition (2.3) it follows immediately that[3]

$$(\mu)\otimes[\{\nu\}\pm\{\rho\}]=(\mu)\otimes\{\nu\}\pm(\mu)\otimes\{\rho\}, \tag{2.5}$$

and

$$(\mu)\otimes[\{\nu\}\cdot\{\rho\}]=[(\mu)\otimes\{\nu\}]\cdot[(\mu)\otimes\{\rho\}], \tag{2.6}$$

where in (2.6) the symbol $\cdot$ indicates Kronecker products of representations of $GL(N)$ and of Σ_m on the left-and right-hand sides, respectively.

The interpretation of an inner plethysm as a symmetrized Kronecker power is then seen by noting that the n-fold Kronecker power of the defining representation $\{1\}$ of $GL(N)$ given by

$$\{1\}\cdot\{1\}\cdot\ldots\cdot\{1\}=\sum_\nu f^{(\nu)}\{\nu\}, \tag{2.7}$$

where the summation is carried out over all partitions, ν, of n, subduces the n-fold Kronecker power of the representation (μ) of Σ_m by virtue of (2.1). Hence (Ref. 16, p. 71)

$$(\mu)\cdot(\mu)\cdot\ldots\cdot(\mu)=\sum_\nu f^{(\nu)}(\mu)\otimes\{\nu\}. \tag{2.8}$$

Just as $\{\nu\}$ is a symmetrized power of $\{1\}$, so $(\mu)\otimes\{\nu\}$ is a symmetrized power of (μ).

The remaining algebraic rules[3] associated with inner plethysms may be derived by a consideration of the subgroup chains

$$GL(M+N)\supset GL(M)\otimes GL(N)\supset\Sigma_m\otimes\Sigma_m\supset\Sigma_m,$$

$$GL(MN)\supset GL(M)\otimes GL(N)\supset\Sigma_m\otimes\Sigma_m\supset\Sigma_m,$$

and

$$GL(N)\supset GL(M)\supset\Sigma_m,$$

where the first links of these chains are associated with the outer products, inner products,[18] and outer plethysms,[15] respectively.

Thus the complete algebra of inner plethysms may be derived in a very trivial way from the definition afforded by (2.3). Clearly this definition may be generalized to give a class of plethysms associated with every group—subgroup combination. The algebra of such plethysms may be determined by a consideration of the appropriate subgroup chains.

III. THE EVALUATION OF INNER PLETHYSMS

Certain plethysms can be evaluated very simply using the definitions and results of Sec. 2: e.g., a comparison of (2.2) and (2.3) yields the trivial result

$$(\mu)\otimes\{1\}=(\mu), \tag{3.1}$$

while

$$(\mu)\otimes\{0\}=(m), \tag{3.2}$$

since under the mapping (2.2) the identity representation $\{0\}$ of $GL(N)$ subduces the identity representation (m) of Σ_m.

If ν is a partition of n into p parts, with $p>N=f^{(\mu)}$, then $D_{f^{(\mu)}}\{\nu\}=0$, and the dimensionality formula (2.4) implies that

$$(\mu)\otimes\{\nu\}=0. \tag{3.3}$$

For example

$$(21)\otimes\{1^3\}=0,$$

since $p=3>2=f^{(21)}$.

In the previous section the emphasis was placed on the group—subgroup reduction $GL(N)\supset\Sigma_m$; however, it is well known that certain irreducible representations of Σ_m are unimodular, whilst all such representations may be made orthogonal as in the explicit Young—Yamanouchi realization of the representations.

If an irreducible representation (μ) of Σ_m is unimodular then Σ_m may be embedded in the special linear group $SL(N)$ where $N=f^{(\mu)}$. The branching rules appropriate to the subgroup chain $GL(N)\supset SL(N)\supset\Sigma_m$ are then such that

$$\{1^N\}\to\{0\}\to(m). \tag{3.4}$$

On the other hand, if the representation (μ) is not unimodular then the one-dimensional representations $\{0\}$ and $\{1^N\}$ of $GL(N)$ must subduce distinct one-dimensional representations of Σ_m, so that for $GL(N)\supset\Sigma_m$

$$\{1^N\}\to(1^m). \tag{3.5}$$

To determine whether or not the matrices of a representation (μ) are unimodular it is only necessary to examine the Young–Yamanouchi (Ref. 7, p. 221; Ref. 16, p. 38) realization of these matrices. The group Σ_m is generated by the set of transpositions (12), (23), ..., $(m-1,m)$ whose matrix representations all have the same determinant since they belong to the same class. Consideration of the transposition $(m-1,m)$ is therefore sufficient to determine whether or not the matrices of a representation (μ) are unimodular. It is easy to see that the corresponding determinant is given by

$$|D^{\mu}(m-1,m)| = (+1)^{n_1}(-1)^{n_2}(-1)^{n_3} = (-1)^{n_2+n_3}, \tag{3.6}$$

where n_1 and n_2 are the number of standard tableaux of shape specified by the partition μ having $m-1$ and m in the same row and same column, respectively, while n_3 is the number of pairs of such tableaux related by interchanging the positions of $m-1$ and m.

The Littlewood–Richardson rule[19] for evaluating the coefficients $m_{\mu\nu,\lambda}$ associated with outer products is then such that

$$n_2+n_3=f^{(\mu/1^2)}, \tag{3.7}$$

where the division symbol (Ref. 20, p. 110) is defined by the relation:

$$(\mu)/(1^2)=\sum_{\lambda} m_{1^2\lambda,\mu}(\lambda), \tag{3.8}$$

in which the summation is carried out over all partitions λ of $m-2$. From (3.6) and (3.7) it follows that the matrix corresponding to every element of the group Σ_m in the representation (μ) will be unimodular if and only if $f^{(\mu/1^2)}$ is even. With this result (3.4) and (3.5) give the formula

$$(\mu)\otimes\{1^N\}=\begin{cases}(m) & \text{if } N=f^{(\mu)} \text{ and } f^{(\mu/1^2)} \text{ is even},\\ (1^m) & \text{if } N=f^{(\mu)} \text{ and } f^{(\mu/1^2)} \text{ is odd}.\end{cases} \tag{3.9}$$

For example

$$(31)\otimes\{1^3\}=(1^4) \quad \text{and} \quad (21^2)\otimes\{1^3\}=(4) \tag{3.10}$$

since

$$(31)/(1^2)=(2) \quad \text{and} \quad (21^2)/(1^2)=(2)+(1^2).$$

More generally by considering products of representations of $GL(N)$ and making use of (2.6), (3.3), and (3.9) it is easily shown that

$$(\mu)\otimes\{\nu\}=\begin{cases}(\mu)\otimes\{\rho\} & \text{if } \nu_N \text{ is even } \textit{or } f^{(\mu/1^2)} \text{ is even},\\ \widetilde{(\mu)\otimes\{\rho\}} & \text{if } \nu_N \text{ is odd } \textit{and } f^{(\mu/1^2)} \text{ is odd},\end{cases} \tag{3.11}$$

where $f^{\mu}=N$, $\{\nu\}=\{\nu_1,\nu_2,\ldots,\nu_N\}$ and $\{\rho\}=\{\nu_1-\nu_N,\nu_2-\nu_N,\ldots,0\}$, so that under $GL(N)\supset SL(N)$, $\{\nu\}\to\{\rho\}$. For example:

$$(31)\otimes\{21^2\}=\widetilde{(31)\otimes\{1\}}=\widetilde{(31)}=(21^2),$$

and

$$(21^2)\otimes\{21^2\}=(21^2)\otimes\{1\}=(21^2). \tag{3.12}$$

Particular cases of (3.11) corresponding to $(\mu)=(m)$ and $(\mu)=(1^m)$ are of special interest. In these cases since $f^{(m)}=f^{(1^m)}=1$, whilst $f^{(m/1^2)}=0$ and $f^{(1^m/1^2)}=f^{(1^{m-2})}=1$ it follows from (3.3) and (3.11) that

$$(m)\otimes\{\nu\}=\begin{cases}(m) & \text{if } p=1,\\ 0 & \text{if } p>1,\end{cases} \tag{3.13}$$

and

$$(1^m)\otimes\{\nu\}=\begin{cases}(m) & \text{if } p=1 \text{ and } n \text{ is even},\\ (1^m) & \text{if } p=1 \text{ and } n \text{ is odd},\\ 0 & \text{if } p>1,\end{cases} \tag{3.14}$$

where ν is a partition of n into p parts.

These results correspond to statements about the branching laws associated with $GL(1)\supset\Sigma_m/K^{(m)}$ and $GL(1)\supset\Sigma_m/K^{(1^m)}$, respectively, since the representations (m) and (1^m) of Σ_m are not faithful. In fact $K^{(m)}=\Sigma_m$ and $K^{(1^m)}=A_m$ so that (3.13) and (3.14) are associated with $GL(1)\supset\Sigma_1$ and $GL(1)\supset\Sigma_2$, respectively. It should be stressed, however, that interpreting the results in terms of symmetrized products gives a meaning to (3.13) and (3.14) as statements purely about the group Σ_m.

The last result (3.14) may be used in conjunction with the algebraic rule (2.6) to prove an important conjugacy theorem as follows. Since

$$(1^m)\cdot(\mu)=(\tilde{\mu}) \tag{3.15}$$

the rule (2.6) implies

$$(\tilde{\mu})\otimes\{\nu\}=[(1^m)\otimes\{n\}]\cdot[(\mu)\otimes\{\nu\}].$$

Application of (3.14) then yields the theorem[21]

$$(\tilde{\mu})\otimes\{\nu\}=\begin{cases}(\mu)\otimes\{\nu\} & \text{if } n \text{ is even},\\ \widetilde{(\mu)\otimes\{\nu\}} & \text{if } n \text{ is odd}.\end{cases} \tag{3.16}$$

This theorem is illustrated by the examples (3.10) and (3.12).

Quite apart from the unimodularity properties of the matrices of the representation (μ), the very existence of an orthogonal realization of these matrices is sufficient to prove that Σ_m is a subgroup of $O(N)$ as well as of $GL(N)$.

Hence[23]

$$(\mu)\otimes\{2\}=1(m)+\cdots \tag{3.17}$$

and

$$(\mu)\otimes\{1^2\}=0(m)+\cdots. \tag{3.18}$$

It is worth pointing out that the existence of the subgroup chain $GL(N)\supset O(N)\supset\Sigma_m$ implies that it is meaningful to define generalized inner plethysms of the form $(\mu)\otimes[\nu]$ which are associated with the second link of this chain. Such plethysms may be evaluated by making use of the full chain and the algebra developed in Sec. 2. For example:

$$(\mu)\otimes[n]=(\mu)\otimes(\{n\}-\{n-2\})$$
$$=(\mu)\otimes\{n\}-(\mu)\otimes\{n-2\}. \tag{3.19}$$

Furthermore, if there exists a unimodular representation of Σ_m then it is meaningful to talk about the subgroup chains $GL(N)\supset SL(N)\supset SO(N)\supset\Sigma_m$ and $GL(N)\supset O(N)\supset SO(N)\supset\Sigma_m$, so that plethysms associated with the embedding of Σ_m in the rotation group may be defined.

In the same way of course the notion of inner plethysm may be generalized somewhat to include plethysms associated with the alternating group, A_m. The relevant subgroup chain is then $GL(N)\supset\Sigma_m\supset A_m$ where the last link is associated with the branching rule[22]

$$(\mu)\to\begin{cases}\langle\mu\rangle & \text{if } (\mu)>(\tilde{\mu}),\\ \langle\tilde{\mu}\rangle & \text{if } (\mu)<(\tilde{\mu})\\ \langle\mu\rangle_+ + \langle\mu\rangle_- & \text{if } (\mu)=(\tilde{\mu}),\end{cases} \tag{3.20}$$

where in general $(\mu)\gtrless(\tilde{\mu})$ according as the first non-vanishing difference $\mu_1-\tilde{\mu}_1$, $\mu_2-\tilde{\mu}_2$, ... is positive or negative. If all such differences vanish then $(\mu)=(\tilde{\mu})$. Brackets $\langle\ \rangle$ are used to denote irreducible representations of A_m, and the subscripts + and − are used in the usual way to signify the irreducible constituents of the representation associated with a self-conjugate partition.

The use of the associativity of the operation of plethysm together with (3.20) and (3.9) gives

$$\langle\mu\rangle\otimes\{1^N\}=\langle m\rangle \text{ if } N=f^{(\mu)} \tag{3.21}$$

and

$$[\langle\mu\rangle_+ + \langle\mu\rangle_-]\otimes\{1^N\}=\langle m\rangle \text{ if } N=f^{(\mu)} \text{ and } \mu=\tilde{\mu},$$

so that using the algebra of plethysms

$$[\langle\mu\rangle_+\otimes\{1^{N/2}\}]\cdot[\langle\mu\rangle_-\otimes\{1^{N/2}\}]=\langle m\rangle, \tag{3.22}$$

where the fact that $f^{\langle\mu\rangle_+}=f^{\langle\mu\rangle_-}=N/2$ has been used in conjunction with (3.3).

If $N=2$, corresponding to the two cases $\langle\mu\rangle_\pm=\langle 21\rangle_\pm$ and $\langle\mu\rangle_\pm=\langle 2^2\rangle_\pm$, (3.22) merely implies that $\langle 21\rangle_+\cdot\langle 21\rangle_-=\langle 3\rangle$ and $\langle 2^2\rangle_+\cdot\langle 2^2\rangle_-=\langle 4\rangle$. However, for $N>2$ a dimensionality argument is sufficient to show that

$$\langle\mu\rangle_+\otimes\{1^{N/2}\}=\langle\mu\rangle_-\otimes\{1^{N/2}\}=\langle m\rangle. \tag{3.23}$$

These results imply that all representations of A_m other than $\langle 21\rangle_\pm$ and $\langle 2^2\rangle_\pm$ are unimodular.

The results (3.17) and (3.18) may be extended somewhat[23] and used in conjunction with (3.20) to give

$$\left.\begin{aligned}&\langle\mu\rangle\otimes\{2\}=1\langle m\rangle+\cdots\\ &\langle\mu\rangle\otimes\{1^2\}=0\langle m\rangle+\cdots\end{aligned}\right\}\text{if } \mu\neq\tilde{\mu}, \tag{3.24a}$$

and

$$\left.\begin{aligned}&\langle\mu\rangle_\pm\otimes\{2\}=1\langle m\rangle+\cdots\\ &\langle\mu\rangle_\pm\otimes\{1^2\}=0\langle m\rangle+\cdots\end{aligned}\right\}\text{if } \mu=\tilde{\mu} \text{ and } \tfrac{1}{2}(m-r) \text{ is even}, \tag{3.24b}$$

but

$$\left.\begin{aligned}&\langle\mu\rangle_\pm\otimes\{2\}=0\langle m\rangle+\cdots\\ &\langle\mu\rangle_\pm\otimes\{1^2\}=0\langle m\rangle+\cdots\end{aligned}\right\}\text{if } \mu=\tilde{\mu} \text{ and } \tfrac{1}{2}(m-r) \text{ is odd}. \tag{3.24c}$$

Thus the representations $\langle\mu\rangle$ and $\langle\mu\rangle_\pm$ with $\frac{1}{2}(m-r)$ even are all orthogonal and not symplectic, whilst the representations $\langle\mu\rangle_\pm$ with $\frac{1}{2}(m-r)$ odd are neither orthogonal nor symplectic. This corresponds to the fact that the former all possess real characters and real representation matrices and the latter some complex characters as was first deduced by Frobenius.[22]

These results may also be confirmed through an examination of the explicit representation matrices for the alternating group A_m.[24]

IV. BRANCHING RULES ASSOCIATED WITH THE CRYSTALLOGRAPHIC POINT GROUPS

The ease with which (3.13) and (3.14) were derived comes about as a direct result of the fact that these plethysms are associated with $GL(N)$, where $N=f^{(m)}=f^{(1^m)}=1$. The value $N=2$ arises in the cases $(\mu)=(21)$ and $(\mu)=(2^2)$. It is then necessary to deal with the representations $\{\nu\}$ of the general form $\{p+q,q\}$ so that $\nu_N=\nu_2=q$. Since $(21)/(1^2)=(1)$ and $(2^2)/(1^2)=(1^2)$ it follows from (3.11) that

$$(21)\otimes\{p+q,q\}=\begin{cases}(21)\otimes\{p\} & \text{if } q \text{ is even},\\ \widetilde{(21)\otimes\{p\}} & \text{if } q \text{ is odd},\end{cases} \tag{4.1}$$

with an exactly similar result in which (21) in replaced by (2^2).

To proceed further, it is merely necessary to note that for representations of $GL(2)$

$$\{p\}=\{p-1\}\cdot\{1\}-\{p-2\}\cdot\{1^2\}, \tag{4.2}$$

so that

$$(21)\otimes\{p\}=[(21)\otimes\{p-1\}]\cdot(21)-[(21)\otimes\{p-2\}]\cdot(1^3), \tag{4.3}$$

where use has been made of (3.1) and (3.9).

Clearly (4.3) is a recurrence relation which may be used to derive $(21)\otimes\{p\}$ from the knowledge that $(21)\otimes\{1\}=(21)$ and $(21)\otimes\{0\}=(3)$. The general result obtained in this way may be written in the form

$$(21)\otimes\{6s+r\}=(21)\otimes\{r\}+s[(3)+2(21)+(1^3)], \tag{4.4a}$$

for $r=0, 1, 2, \ldots, 5$ and $s=0, 1, 2, \ldots,$

with

$$(21)\otimes\{0\}=(3),\quad (21)\otimes\{1\}=(21),\quad (21)\otimes\{2\}=(3)+(21),$$

$$(21)\otimes\{3\}=(3)+(21)+(1^3),\quad (21)\otimes\{4\}=(3)+2(21),$$
$$(21)\otimes\{5\}=(3)+2(21)+(1^3). \tag{4.4b}$$

An exactly analogous result holds for the plethysms $(2^2)\otimes\{p\}$ with (3), (21), and (1^3) everywhere replaced by (4), (2^2), and (1^4), respectively.

Making use of the results (4.4) and the identity (3.19), it follows that:

$$(21)\otimes[3s+r]=(21)\otimes[r]\quad\text{with } r=1,\ 2,\ 3 \text{ and } s=0,\ 1,\ 2,\ \ldots, \tag{4.5a}$$

where

$$(21)\otimes[0]=(3),\quad (21)\otimes[1]=(21)\otimes[2]=(21),$$
$$(21)\otimes[3]=(3)+(1^3). \tag{4.5b}$$

Exactly the same techniques may be used in the cases for which $N=f^{(\mu)}=3$, i.e., for $(\mu)=(31)$ and $(\mu)=(21^2)$. By virtue of the conjugacy theorem (3.16) the corresponding plethysms are related so that it is only necessary to consider the slightly simpler of these two cases $(\mu)=(21^2)$. The use of (3.11) and (2.6) gives

$$\begin{aligned}(21^2)\otimes\{p+q+r,\ q+r,\ r\}&=(21^2)\otimes\{p+q,\ q\},\\ &=[(21^2)\otimes\{p+q\}]\cdot[(21^2)\otimes\{q\}]\\ &\quad-[(21^2)\otimes\{p+q+1\}]\cdot[(21^2)\otimes(q-1)].\end{aligned} \tag{4.6}$$

Furthermore the use of the subgroup chain $GL(3)\supset SO(3)\supset\Sigma_4$ allows $(21^2)\otimes\{k\}$ to be determined from a knowledge of $(21^2)\otimes[k]$, which may itself be evaluated using the well known $SO(3)$ identity:

$$[k]=[k-1]\cdot[1]-[k-1]-[k-2]. \tag{4.7}$$

It is found that

$$\begin{aligned}&(21^2)\otimes[12s+r]\\ &\quad=(21^2)\otimes[r]+s[(4)+3(31)+2(2^2)+3(21^2)+(1^4)]\end{aligned}$$
$$\text{for } r=0,\ 1,\ 2,\ \ldots,\ 11 \text{ and } s=0,\ 1,\ 2,\ \ldots, \tag{4.8a}$$

where now

$$(21^2)\otimes[r]+(21^2)\otimes[11-r]=(4)+3(31)+2(2^2)+3(21^2)+(1^4), \tag{4.8b}$$

with

$$\begin{aligned}&(21^2)\otimes[0]=(4),\quad (21^2)\otimes[1]=(21^2),\quad (21^2)\otimes[2]=(31)+(2^2),\\ &(21^2)\otimes[3]=(31)+(21^2)+(1^4),\\ &(21^2)\otimes[4]=(4)+(31)+(2^2)+(21^2),\\ &(21^2)\otimes[5]=(31)+(2^2)+2(21^2).\end{aligned} \tag{4.8c}$$

It is worth pointing out that further extension of the subgroup chain to give $GL(3)\supset SO(3)\supset\Sigma_m\supset A_m$ yields from (3.20) and (4.8) the result

$$\{31\}\otimes[6s+r]=\{31\}\otimes[r]+s[\{4\}+3\{31\}+\{2^2\}_+ +\{2^2\}_-]$$
$$\text{for } r=0,1,2,\ldots,5 \text{ and } s=0,1,2,\ldots, \tag{4.9a}$$

where

$$\begin{aligned}&\{31\}\otimes[0]=\{4\},\quad \{31\}\otimes[1]=\{31\},\\ &\{31\}\otimes[2]=\{31\}+\{2^2\}_+ +\{2^2\}_-,\\ &\{31\}\otimes[3]=\{4\}+2\{31\},\\ &\{31\}\otimes[4]=\{4\}+2\{31\}+\{2^2\}_+ +\{2^2\}_-,\\ &\{31\}\otimes[5]=3\{31\}+\{2^2\}_+ +\{2^2\}_-.\end{aligned} \tag{4.9b}$$

The enumeration of these results has been spelled out in detail in view of their important application in connection with crystallographic point groups[6,7] through the well-known isomorphisms between the tetrahedral group T and A_4 and the octahedral group O and Σ_4. The embeddings of T and O in the rotation group $SO(3)$ are defined by

$$[1]\to\{31\} \tag{4.10}$$

and

$$[1]\to(21^2), \tag{4.11}$$

respectively, and the branching rules appropriate to $SO(3)\supset T$ and $SO(3)\supset O$ are given by (4.9) and (4.8), respectively. It should be noted that the mapping

$$[1]\to(31) \tag{4.12}$$

cannot be associated with $SO(3)\supset\Sigma_4$ since the representation matrices for (31) are not unimodular as can be seen from (3.10).

To complete the story it is really necessary to examine the icosahedral group Y, which is isomorphic with A_5, and the particular dihedral group D_3, which is isomorphic with Σ_3. Suffice it to say that the embeddings of these groups in $SO(3)$ are defined by the mappings

$$[1]\to\{31^2\}_+ \tag{4.13}$$

and

$$[1]\to(21)+(1^3), \tag{4.14}$$

respectively.

Plethysms associated with these embeddings may be evaluated in the case of (4.13) by using the method involving (4.7), and in the case of (4.14) by using the algebra of plethysms and the results (4.4) or (4.5). The results take the form

$$\begin{aligned}&\{31^2\}_+\otimes[0]=\{5\},\quad \{31^2\}_+\otimes[1]=\{31^2\}_+,\\ &\{31^2\}_+\otimes[2]=\{32\},\quad (31^2)_+\otimes[3]=\{41\}+\{31^2\}_-,\\ &(31^2)_+\otimes[4]=\{41\}+\{32\},\end{aligned} \tag{4.15}$$

for the first few cases associated with the icosahedral group, and quite generally for D_3

$$\begin{aligned}[(21)+(1^3)]\otimes[6s+r]&=[(21)+(1^3)]\otimes[r]+2s[(3)+2(21)\\ &\quad+(1^3)]\quad\text{for } r=0,\ldots,5 \text{ and } s=0,1,2,\ldots,\end{aligned} \tag{4.16a}$$

with

$$[(21)+(1^3)]\otimes[r]+[(21)+(1^3)]\otimes[5-r]=2[(3)+2(21)+(1^3)], \tag{4.16b}$$

where

$$\begin{aligned}&[(21)+(1^3)]\otimes[0]=(3),\quad [(21)+(1^3)]\otimes[1]=(21)+(1^3),\\ &[(21)+(1^3)]\otimes[2]=(3)+2(21).\end{aligned} \tag{4.16c}$$

The results (4.8), (4.9), (4.15), and (4.16) are not,

TABLE I. Notation adopted by Griffith,[6] Hamermesh,[7] and Lomont[25] for the irreducible representations of the groups D_3, T, O, and Y.

Σ_3		D_3		A_4		T	
Ref.	6	7	25	Ref.	6	7	25
(3)	A_1	A_1	Γ_1	$\{4\}$	A_1	A	Γ_1
(21)	E	E	Γ_3	$\{31\}$	T	F	Γ_4
(1^3)	A_2	A_2	Γ_2	$\{2^2\}_+$	E	E	Γ_2
				$\{2^2\}_-$			Γ_3
Σ_4		O		A_5		Y	
Ref.	6	7	25	Ref.	6		25
(4)	A_1	A_1	Γ_1	$\{5\}$	A		Γ_1
(31)	T_2	F_2	Γ_4	$\{41\}$	U		Γ_4
(2^2)	E	E	Γ_3	$\{32\}$	V		Γ_5
(21^2)	T_1	F_1	Γ_5	$\{31^2\}_+$	T_1		Γ_2
(1^4)	A_2	A_2	Γ_2	$\{31^2\}_-$	T_2		Γ_3

of course, new, but it is notable that they have been obtained without recourse to character tables. To make comparison with the results of Griffith (Ref. 6, p. 389), Hamermesh (Ref. 7, p. 339), and Lomont (Ref. 25, p. 145), it is convenient to note the variations in notation given in Table I for the representations of the groups D_3, T, O, and Y.

V. RECURRENCE TECHNIQUES FOR EVALUATING INNER PLETHYSMS

It has been demonstrated elsewhere[17] that the evaluation of outer plethysms may be accomplished by a consideration of the subgroup chain $GL(N)\supset GL(M)\supset GL(M-1)$. In the same way, for inner plethysms, consideration of the branching rules associated with the subgroup chain $GL(N)\supset\Sigma_m\supset\Sigma_{m-1}$ indicates that

$$\{1\}\to(\mu)\to(\mu)/(1), \tag{5.1}$$

and more generally

$$\{\nu\}\to(\mu)\otimes\{\nu\}\to[(\mu)\otimes\{\nu\}]/(1), \tag{5.2}$$

where use has been made of the branching rule (Ref. 26, p. 390 associated with $\Sigma_m\supset\Sigma_{m-1}$. Hence

$$[(\mu)\otimes\{\nu\}]/(1)=[(\mu)/(1)]\otimes\{\nu\}. \tag{5.3}$$

This identity is the direct generalization for inner plethysms of the identity used by Littlewood as the basis of his third method of calculating outer plethysms.[2] The corresponding method of evaluating inner plethysms involves calculating the right-hand side of (5.3) from a knowledge of plethysms associated with Σ_{m-1}, and then using this for the left-hand side of (5.3) to evaluate the plethysms associated with Σ_m.

For example (5.3) implies that

$$[(21)\otimes\{1^2\}]/(1)=[(2)+(1^2)]\otimes\{1^2\}. \tag{5.4}$$

The use of (3.13) and (3.14) yields for the right-hand side

$$\begin{aligned}[(2)+(1^2)]\otimes\{1^2\}&=(2)\otimes\{1^2\}+(2)\otimes\{1\}\cdot(1^2)\otimes\{1\}\\&\quad+(1^2)\otimes\{1^2\},\\&=(2)\cdot(1^2)=(1^2),\end{aligned} \tag{5.5}$$

so that

$$[(21)\otimes\{1^2\}]/(1)=(1^2). \tag{5.6}$$

Writing

$$(21)\otimes\{1\}=A(3)+B(21)+C(1^3), \tag{5.7}$$

where A, B, C, are nonnegative integers to be determined, we then have

$$[(21)\otimes\{1^2\}]/(1)=(A+B)(2)+(B+C)(1^2). \tag{5.8}$$

Comparison of (5.6) and (5.8) indicates that $A=B=0$ and $C=1$ so that

$$(21)\otimes\{1^2\}=(1^3) \tag{5.9}$$

as is implied, of course, by the more general results (3.9).

More generally, as in Littlewood's third method of evaluating outer plethysms,[17] ambiguities arise. For example they arise in the evaluating of $(31)\otimes\{2\}$. A technique for removing such ambiguities is suggested by a consideration of the subgroup chain $GL(N)\supset\Sigma_m\supset\Sigma_s\otimes\Sigma_t$, where $s+t=m$. The branching rule appropriate to the second link in this chain is

$$(\mu)\to\sum_{\sigma,\tau}m_{\sigma\tau\mu}(\sigma),\ (\tau), \tag{5.10}$$

where the summation is carried out over partitions σ and τ of s and t, respectively. This result implies, through a consideration of the full chain that

$$(\mu)\otimes\{\nu\}\to\left[\sum_{\sigma,\tau}m_{\sigma\tau\mu}(\sigma),\ (\tau)\right]\otimes\{\nu\} \tag{5.11}$$

under the reduction associated with $\Sigma_m\supset\Sigma_s\otimes\Sigma_t$.

The right-hand side of (5.11) may be evaluated for all s and t such that $s<m$ and $t<m$. This information can then be used in the evaluation of the left-hand side. Without loss of generality it is only necessary to consider the cases for which $s\geq t$. The case $s=m-1$, $t=1$ corresponds to the use of (5.3).

For example, under $\Sigma_4\supset\Sigma_3\otimes\Sigma_1$,

$$\begin{aligned}(31)\otimes\{2\}&\to[(3),(1)+(21),(1)]\otimes\{2\}\\&=2[(3),(1)]+2[(21),(1)].\end{aligned} \tag{5.12}$$

Similarly under $\Sigma_4\supset\Sigma_2\otimes\Sigma_2$

$$\begin{aligned}&(31)\otimes\{2\}\to[(2),(2)+(2),(1^2)+(1^2),(2)]\otimes\{2\}\\&=3[(2),(2)]+[(2),(1^2)]+[(1^2),(2)]+[(1^2),(1^2)].\end{aligned} \tag{5.13}$$

Now assuming

$$(31)\otimes\{2\}=A(4)+B(31)+C(2^2)+D(21^2)+E(1^4), \tag{5.14}$$

then under $\Sigma_4\supset\Sigma_3\otimes\Sigma_1$

$$\begin{aligned}(31)\otimes\{2\}\to&(A+B)[(3),(1)]+(B+C+D)[(21),(1)]\\&+(D+E)[(1^3),(1)]\end{aligned} \tag{5.15}$$

while under $\Sigma_4\supset\Sigma_2\otimes\Sigma_2$

$$\begin{aligned}(31)\otimes\{2\}\to&(A+B+C)[(2),(2)]+(B+D)[(2),(1^2)]\\&+(B+D)[(1^2),(2)]+(C+D+E)[(1^2),(1^2)].\end{aligned} \tag{5.16}$$

Comparison of (5.12) with (5.15) and of (5.13) with (5.16) gives

TABLE II. Plethysm coefficients, $p_{\mu\nu,\lambda}$, associated with $GL(N)\supset\Sigma_4/K^{(\mu)}/\{\nu\}\to(\mu)\otimes\{\nu\}=\sum_\lambda p_{\mu\nu,\lambda}(\lambda)$.

(μ); $f^{(\mu)}=N$	(4); 1					(31); 3					(2^2); 2				
$\{\nu\}$ \ (λ)	(4)	(31)	(2^2)	(21^2)	(1^4)	(4)	(31)	(2^2)	(21^2)	(1^4)	(4)	(31)	(2^2)	(21^2)	(1^4)
$\{0\}$	1					1					1				
$\{1\}$	1						1						1		
$\{2\}$	1					1	1	1			1		1		
$\{1^2\}$									1						1
$\{3\}$	1					1	2		1		1		1		1
$\{21\}$							1	1	1				1		
$\{1^3\}$										1					
$\{4\}$	1					2	2	2	1		1		2		
$\{31\}$							2	1	2	1			1		1
$\{2^2\}$						1	1	1		1	1				
$\{21^2\}$									1						
$\{1^4\}$															
$\{5\}$	1					1	4	1	2		1		2		1
$\{41\}$						1	3	2	3	1	1		1		1
$\{31\}$						1	2	1	2				1		
$\{31^2\}$								1	1	1					
$\{2^21\}$							1								
$\{21^3\}$															
$\{1^5\}$															

$$A+B=2,\quad B+C+D=2,\quad D+E=0,$$
$$A+B+C=3,\quad B+D=1,\quad C+D+E=1. \tag{5.17}$$

These equations have the unique solution

$$A=B=C=1,\quad D=E=0,$$

so that in (5.14)

$$(31)\otimes\{2\}=(4)+(31)+(2^2) \tag{5.18}$$

in agreement with the results of Sec. 4.

It should be stressed that the use of (5.3) alone corresponding to the use of (5.12) and (5.15) leads to ambiguities which are eliminated by the use of (5.13) and (5.16). However, in general all such ambiguities may not be eliminated.

This deficiency of this method of calculating inner plethysms using the subgroup chains $GL(N)\supset\Sigma_m\supset\Sigma_s\otimes\Sigma_t$ with all possible distinct pairs of numbers s and t consistent with the condition $s+t=m$ is quite general. Examination of the second link of this chain shows that if

$$\sum_\rho A_\rho(\rho)\to\sum_{\sigma,\tau}B_{\sigma\tau}(\sigma),(\tau), \tag{5.19}$$

where ρ, σ, and τ are partitions of m, s, and t, and the coefficients $B_{\sigma\tau}$ are known for all s and t, then the coefficients A_ρ may be calculated in all cases for which (ρ) is not of the form $(m-k,1^k)$. Furthermore, the sum of each pair of coefficients $A_{(m-k,1^k)}+A_{(m-k-1,1^{k+1})}$ may also be determined, so that a knowledge of any one such coefficient $A_{(m-k,1^k)}$ enables the remainder to be calculated.

Thus in the case of Σ_4 the information available gives $A_{(2^2)}$, $A_{(4)}+A_{(31)}$, $A_{(31)}+A_{(21^2)}$, and $A_{(21^2)}+A_{(1^4)}$. In the example quoted it is the vanishing of the last of these quantities which enables (5.18) to be unambiguously obtained.

Despite the persistence of ambiguities in the general case the method gives a lot of results when used in conjunction with other properties of inner plethysms. For example, the ambiguities may all be removed very easily in the case of plethysms $(\mu)\otimes\{2\}$ and $(\mu)\otimes\{1^2\}$ by virtue of (3.17) and (3.18) which imply that $A_{(m)}=1$ and $A_{(m)}=0$, respectively. From the knowledge of these coefficients all the others may be obtained.

VI. TABULATION OF RESULTS AND APPLICATION TO THE SYMMETRY PROPERTIES OF 3j SYMBOLS

The main aim of this paper has been to present some new methods for the evaluation of the coefficients $p_{\mu\nu,\lambda}$ in the inner plethysms

$$(\mu)\otimes\{\nu\}=\sum_\lambda p_{\mu\nu,\lambda}(\lambda). \tag{6.1}$$

Other methods are of course available.[3,4,16,21] These depend principally on the very important embedding of Σ_m in $GL(m-1)$ which is associated with the plethysms $(m-1,1)\otimes\{\nu\}$. Littlewood has given a formula for the evaluation of this class of plethysms[13] and the results have come to play an important role in the study of the nuclear shell model. One particularly important result is that

$$(m-1,1)\otimes(m-k,1^k)\ \text{ for } k=0,1,\ldots,m-1. \tag{6.2}$$

This implies the existence of a subgroup chain $GL(N)\supset GL(m-1)\supset\Sigma_m$ with,

$$N=D_{m-1}\{1^k\}=f^{(m-k,1^k)}=(m-1)!/k!(m-k-1)!. \tag{6.3}$$

TABLE III. Plethysm coefficients $p_{\mu\nu,\lambda}$, associated with $GL(N) \supset \Sigma_5/K^{(\mu)}/\{\nu\} \rightarrow (\mu) \otimes \{\nu\} = \sum_\lambda p_{\mu\nu,\lambda}(\lambda)$.

(μ); $f^{(\mu)} = N$	(5); 1							(41); 4							(32); 5							(31^2); 6						
$\{\nu\}$ \ (λ)	(5)	(41)	(32)	(31^2)	(2^21)	(21^3)	(1^5)	(5)	(41)	(32)	(31^2)	(2^21)	(21^3)	(1^5)	(5)	(41)	(32)	(31^2)	(2^21)	(21^3)	(1^5)	(5)	(41)	(32)	(31^2)	(2^21)	(21^3)	(1^5)
(0)	1							1							1							1						
(1)	1								1								1									1		
(2)	1							1	1	1					1	1	1		1			1	1	2		1		1
(1^2)											1							1		1					1	1	1	
(3)	1							1	2	1	1				1	2	2	1	1	1	1		2	1	5	1	2	
(21)									1	1	1	1				1	2	2	2	1			2	3	4	3	2	
(1^3)													1					1		1		1	1	1		1	1	1
(4)	1							2	3	2	1	1			2	3	4	2	4	1		4	5	7	3	7	3	2
(31)									2	2	3	1	1			3	4	6	4	4	1	1	6	9	10	9	8	3
(2^2)								1	1	2		1			2	3	3	1	2	1	1	3	5	6	3	5	2	1
(21^2)											1	1	1			1	1	3	2	2		3	3	3	4	4	4	
(1^4)														1			1						1	1	1			

Application of the associativity rule to this subgroup chain then yields the identity

$$[(m-1,1) \otimes \{1^k\}] \otimes \{\nu\} = (m-1,1) \otimes [\{1^k\} \otimes \{\nu\}]. \quad (6.4)$$

Hence from (6.2)

$$(m-k,1^k) \otimes \{\nu\} = (m-1,1) \otimes [\{1^k\} \otimes \{\nu\}]. \quad (6.5)$$

This formula may be used in the evaluation of the inner plethysm on the left by using known results for the outer plethysm on the right, followed by the use of Littlewood's technique for the evaluation of plethysm of the form $(m-1,1) \otimes \{\rho\}$.

Butler[21] has gone further by showing that any representation (μ) of Σ_m may be written as a sum of products of representations of of the type $(m-k,1^k)$, so that using the algebra of inner plethysms, followed by (6.5) and Littlewood's theorem, any plethysm $(\mu) \otimes \{\nu\}$ may be evaluated.

These methods have all been used to calculate plethysms but the only tabulation of results is that of Murnaghan[4] for the plethysms $(m-1.1) \otimes \{\nu\}$ for general m, and ν a partition of n with $n \leq 6$. To obtain results for specific values of m from these tables it is necessary to use the rather simple modification rules appropriate to the symmetric group (Ref. 20, p. 98).

On the other hand, the methods described in this paper lend themselves to the evaluation of the more general plethysm $(\mu) \otimes \{\nu\}$ in which μ is any partition of a specific number m. The results obtained using the methods of Secs. 3—5 are given in Tables II, III, and IV in which the coefficients $p_{\mu\nu,\lambda}$ are tabulated for μ a partition of 4, 5, and 6, respectively.

These results give the reduction of every inner plethysm associated with the embedding of Σ_4, Σ_5, and Σ_6, respectively, in $GL(N)$ provided that the embedding is defined in terms of a mapping of the form $\{1\} \rightarrow (\mu)$ with (μ) irreducible and faithful. Of course the conjugacy theorem (3.16) is used to extend the range of the tables.

All the results pertaining to Σ_2 are, of course, trivial, since the only representations (2) and (1^2) are covered by the general results (3.13) and (3.14). These also cover the inner plethysms associated with the representations (3) and (1^3) of Σ_3. The plethysms associated with (21) are given by (4.1) and (4.4). They may also be obtained directly from Table II by noting that the relevant subgroup chain, $GL(2) \supset \Sigma_4/K^{(2^2)} = \Sigma_3$, is defined by the mapping

$$\{1\} \rightarrow (21). \quad (6.6)$$

TABLE IV. Plethysm coefficients $p_{\mu\nu,\lambda}$, associated with $GL(N) \supset \Sigma_6/K^{(\mu)}/\{\nu\} \rightarrow (\mu) \otimes \{\nu\} = \sum_\lambda p_{\mu\nu,\lambda}(\lambda)$.

(μ); $f^{(\mu)} = N$	(6); 1											(51); 5											(42); 9										
$\{\nu\}$ \ (λ)	(6)	(51)	(42)	(41^2)	(3^2)	(321)	(31^3)	(2^3)	(2^21^2)	(21^4)	(1^6)	(6)	(51)	(42)	(41^2)	(3^2)	(321)	(31^3)	(2^3)	(2^21^2)	(21^4)	(1^6)	(6)	(51)	(42)	(41^2)	(3^2)	(321)	(31^3)	(2^3)	(2^21^2)	(21^4)	(1^6)
(0)	1											1											1										
(1)	1												1											1									
(2)	1											1	1	1									1	1	2			1		1			
(1^2)															1											1		1	1				
(3)	1											1	2	1	1	1							2	2	5	2	1	3	2	2		1	
(21)													1	1	1		1							2	4	3	1	6	3	2	2	1	
(1^3)																		1								2	1	1	2		2	1	
(μ); $f^{(\mu)} = N$								(41^2); 10											(3^2); 5										(321); 16				
(0)	1											1											1										
(1)				1												1												1					
(2)	1	1	2		1	1		1		1		1		1					1				1	2	3	1	1	3	1	2	1	1	1
(1^2)				1		1	1		1								1									3	1	2	3		2	1	
(3)		2	2	6	2	4	4		2				1		1	2				1		1	2	6	10	12	6	17	12	6	10	6	2
(21)		2	4	5	2	8	4	2	4	2					1	1	1			1			1	9	17	18	9	32	18	9	17	9	1
(1^3)	1	1	2		1	2	2	2	2	2	1							1					1	4	7	8	4	12	8	4	7	4	1

Interpreting inner plethysms as symmetrized products it is then clear that under $\Sigma_4\supset\Sigma_3$

$$(2^2)\otimes\{\nu\}\to(21)\otimes\{\nu\}. \tag{6.7}$$

However, in general

$$(2^2)\otimes\{\nu\}=A(4)+B(2^2)+C(1^4) \tag{6.8}$$

since the nth Kronecker power of (2^2) only contains the representations (4), (2^2), and (1^4). Decomposing these representations in the usual way into the representations (3), (21), and (1^3), it then follows from (6.7) that

$$(21)\otimes\{\nu\}=A(3)+B(21)+C(1^3). \tag{6.9}$$

Thus the plethysm (6.9) associated with (21) may be read off from the tabulation of (6.8) associated with (2^2).

It should be stressed that in drawing up the tables the well known identity (Ref. 16, p. 43, Ref. 20, p. 89)

$$\{\nu\}=\left|\{1^{\tilde{\nu}_i-i+j}\}\right|, \tag{6.10}$$

has been used in conjunction with (2.5) and (2.6) to reduce the amount of work involved. Thus it is only necessary to calculate inner plethysms of the type $(\mu)\otimes\{1^k\}$, from which all others may be calculated.

A number of interesting facts may be established very easily from the results of the tables. For example, the only irreducible representations of Σ_m which define an embedding in $GL(3)$ are (31) and its conjugate (21^2). Both representations are orthogonal since $(31)\otimes\{2\}$ and $(21^2)\otimes\{2\}$ both contain the representation (4). However, only (21^2) defines an embedding in SO(3) since $(31)\otimes\{1^3\}=(1^4)$ whilst $(21^2)\otimes\{1^3\}=(4)$. This is sufficient to identify the embedding of Σ_4 in SO(3) with the embedding of the point symmetry group, O, the octahedral group, in SO(3) as discussed in Sec. 4.

It is then easy to read off from Table II the branching rules associated with $SO(3)\supset O$. Adopting the notation of Griffith[6] defined in Table I and making use of the fact that a state of integral angular momentum J is defined by $[J]=\{J\}-\{J-2\}$, Table II gives

$$\begin{aligned}
&[0]\to A_1,\\
&[1]\to T_1,\\
&[2]\to T_2+E,\\
&[3]\to T_2+T_1+A_2,\\
&[4]\to A_1+T_2+E+T_1,\\
&[5]\to T_2+E+2T_1,
\end{aligned} \tag{6.11}$$

in agreement with the tabulation of Griffith (Ref. 6, p. 389) and (4.8).

Of course results other than those of Sec. 4 may also be obtained from the tables. For example it is easy to see that the symmetrized and antisymmetrized squares of the irreducible representations of O are given by (Ref. 6, p. 405)

$$\begin{aligned}
&A_1\otimes\{2\}=A_1, && A_1\otimes\{1^2\}=0,\\
&T_2\otimes\{2\}=A_1+T_2+E, && T_2\otimes\{1^2\}=T_1,\\
&E\otimes\{2\}=A_1+E, && E\otimes\{1^2\}=A_2,\\
&T_1\otimes\{2\}=A_1+T_2+E, && T_1\otimes\{1^2\}=T_1,\\
&A_2\otimes\{2\}=A_1, && A_2\otimes\{1^2\}=0.
\end{aligned} \tag{6.12}$$

Such results are important in establishing selection rules for certain physical processes. They may also be derived without the explicit use of the character table for the group O by the methods of Smith and Wybourne[10] who make use of the subgroup decomposition $SO(3)\supset Y$, and the symmetrized and antisymmetrized Kronecker squares of representations of $SO(3)$ to derive the corresponding results for Y. In the case of the group O for example, (6.11) and (4.15) imply that

$$(T_2+E)\otimes\{2\}=2A_1+2T_2+2E+T_1.$$

However, $T_2\cdot E=T_2+T_1$, whilst any power of E only contains the representations A_1, A_2, and E. Hence

$$T_2\otimes\{2\}=A_1+T_2+E \quad\text{and}\quad E\otimes\{2\}=A_1+E$$

in agreement with (6.12). Clearly the tabulation given here provides a more direct way of obtaining these results.

These results on symmetrized squares are also relevant to the determination of the symmetry properties of the $3j$ symbols.

Quite generally the physical states associated with the representations of a symmetry group labelled by μ, ν, and λ couple together in an invariant way if and only if the product $\mu\cdot\nu\cdot\lambda$ contains the identity, 1-dimensional representation of the group which it is convenient to label by η. In general

$$\mu\cdot\nu\cdot\lambda=M\eta+\cdots \tag{6.13}$$

and M is the number of distinct couplings invariant under the symmetry operations of the group in question. The physical states associated with the representations μ, ν, and λ may be specified by indices i, j, and k. With this notation it is convenient to denote the corresponding $3j$ symbols[27] by $(\mu\nu\lambda)^r_{ijk}$, where r is a multiplicity label taking on the values $1,2,\ldots,M$.

If all the representations μ, ν, and λ are distinct, the $3j$ symbols may be chosen so that the pairs of labels μi, νj, and λk may be permuted without altering the value of the coefficients for every value of r.

If, however, $\mu=\nu\neq\lambda$ and $M=M_2+M_{1^2}$ where

$$[\mu\otimes\{2\}]\cdot\lambda=M_2\eta+\cdots \tag{6.14}$$

and

$$[\mu\otimes\{1^2\}]\cdot\lambda=M_{1^2}\eta+\cdots,$$

then (Ref. 7, p. 264)

$$(\mu\mu\lambda)^r_{jik}=\delta_r(\mu\mu\lambda)^r_{ijk}, \tag{6.15}$$

where $\delta_r=+1$ or -1 according as r is associated with the set of M_2 symmetric couplings or the set of M_{1^2} antisymmetric couplings.

This result may be generalized in such a way that if $\mu=\nu=\lambda$ and $M=M_3+2M_{21}+M_{1^3}$ with

$$\begin{aligned}
&\mu\otimes\{3\}=M_3\eta+\cdots,\\
&\mu\otimes\{21\}=M_{21}\eta+\cdots,\\
&\mu\otimes\{1^3\}=M_{1^3}\eta+\cdots,
\end{aligned} \tag{6.16}$$

then

$$(\mu\mu\mu)^{\tau}_{\pi(ijk)}=\sum_s D^{\nu_\tau}_{\tau s}(\pi)(\mu\mu\mu)^{s}_{ijk}, \tag{6.17}$$

where π is an element of the permutation group Σ_3 acting on the letters ijk, and $D^{\nu_\tau}(\pi)$ is the matrix representing this permutation.[28–32] The particular representation, ν_τ, is (3), (21), or (1^3) according as τ belongs to the set of M_3 totally symmetric couplings, the set of $2M_{21}$ mixed symmetry couplings or the set of M_{1^3} totally antisymmetric couplings, respectively. Thus the symmetry properties of the $3j$ symbols involve more than a phase factor if and only if M_{21} is nonzero.[29] A group is said to be a simple phase group[33] if for every representation μ of the group

$$(\mu\mu\mu)^{\tau}_{\pi(ijk)}=\epsilon_\tau(\mu\mu\mu)^{\tau}_{ijk}, \tag{6.18}$$

with

$$|\epsilon_\tau|^2=1.$$

Only the values $\epsilon_\tau=+1$ and -1 occur for the couplings associated with M_3 and M_{1^3}. Clearly (6.18) will be invalid and the group will not be a simple phase group if there exists μ such that M_{21} is nonzero since $D^{21}(\pi)$ is a 2×2 matrix.

Examination of Tables II, III, and IV and the results pertaining to Σ_2 and Σ_3 indicates that Σ_m is a simple phase group if $m\leq5$. However, for $m=6$,

$$(321)\otimes\{21\}=(6)+\cdots,$$

so that Σ_6 is not a simple phase group. This result has been noted elsewhere[29,33] but the connection between the symmetry properties of $3j$ symbols and inner plethysms, although hinted at,[33] has been neither explicitly stated nor used before. Butler has independently noted this connection.[32]

Returning to the octahedral group, isomorphic with Σ_4, it follows from Table II, or equivalently (6.12), that the only couplings $(\mu\mu\lambda)$ with $\mu\neq\lambda$ for which δ_τ is not $+1$ are those denoted by $(T_2T_2T_1)$ and (EEA_2). In these cases $\delta_\tau=-1$. In the same way it follows from Table II that the couplings $(\mu\mu\mu)$ have the symmetry properties of (6.18) since in all cases $M_{21}=0$. Moreover, $\epsilon_\tau=+1$ except in the case $(T_1T_1T_1)$ for which $\epsilon_\tau=+1$ or -1 according as the permutation π of (6.18) is even or odd. This is in agreement with the work of Griffith (Ref. 6, p. 446, Ref. 28).

From these remarks it should be clear that the rather obscure operation of inner plethysm has an important role to play in understanding the properties of finite subgroups of continuous groups and in physical applications.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge a private communication from Dr. P.H. Butler confirming that he had independently arrived at the same conclusions as the author regarding the relationship between inner plethysms and the symmetry properties of $3j$ symbols, and to thank Dr. A.J. van Zanten and Dr. E. deVries for communicating their work to me prior to publication. The author is extremely grateful to Professor B.G. Wybourne for pointing out several errors in the original manuscript version of Table III as well as a number of typing errors. The author also wishes to express his appreciation of the kind hospitality extended to him by Professor G. de B. Robinson and Professor G. F. D. Duff of the Mathematics Department at the University of Toronto, where most of this work was carried out.

*Permanent address: The Mathematics Department, The University, Southampton, England.

[1]D. E. Littlewood, J. Lond. Math. Soc. **11**, 49 (1936).
[2]D. E. Littlewood, Philos. Trans. R. Soc. Lond. A **239**, 305 (1944).
[3]D. E. Littlewood, Can. J. Math. **10**, 1 (1958).
[4]F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **41**, 514 (1955); Proc. Natl. Acad. Sci. USA **41**, 1096 (1955).
[5]G. Ya. Lyubarskii, *The Application of Group Theory in Physics* (Pergamon, New York, 1960).
[6]J. S. Griffith, *The Theory of Transition-Metal Ions* (Cambridge U.P., Cambridge, 1961).
[7]M. Hamermesh, *Group Theory and Its Application To Physical Problems* (Addison-Wesley, Reading, Mass., 1962).
[8]J. L. Birman, Phys. Rev. **127**, 1093 (1962).
[9]B. R. Judd and H. T. Wadzinski, J. Math. Phys. **8**, 2125 (1967).
[10]P. R. Smith and B. G. Wybourne, J. Math. Phys. **8**, 2434 (1967).
[11]C. J. Bradley and B. L. Davis, J. Math. Phys. **11**, 1536 (1970).
[12]M. Kretzschmar, Z. Phys. **157**, 433 (1960); Z. Phys. **158**, 284 (1960).
[13]D. E. Littlewood, Can. J. Math. **10**, 17 (1958).
[14]L. Sabliauskas, Liet. Fiz. Rinkinys **9**, 477 (1969); V. Vanagas, *Algebraic Methods in Nuclear Theory* (Vilnius, 1971-in Russian).
[15]B. G. Wybourne and P. H. Butler, J. Phys. (Paris) **30**, 181 (1969); B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New York, 1970).
[16]G. de B. Robinson, *Representation Theory of the Symmetric Group* (Edinburgh U. P., Edinburgh, 1961).
[17]P. H. Butler and R. C. King, J. Math. Phys. **14**, 741 (1973).
[18]C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).
[19]D. E. Littlewood and A. R. Richardson, Philos. Trans. R. Soc. Lond. A **233**, 99 (1934).
[20]D. E. Littlewood, *The Theory of Group Characters* (Oxford U. P., Oxford, 1950), 2nd ed.
[21]P. H. Butler, J. Phys. (Paris) Suppl. **31**, 47 (1970).
[22]F. G. Frobenius, Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl. 303 (1901).
[23]P. H. Butler and R. C. King, J. Math. Phys. **14**, 1176 (1973).
[24]B. M. Puttaswamaiah and G. de B. Robinson, Can. J. Math. **16**, 587 (1964).
[25]J. S. Lomont, *Applications of Finite Groups* (Acadamic, New York, 1959).
[26]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
[27]J. R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[28]J. S. Griffith, Mol. Phys. **3**, 285 (1960).
[29]J. R. Derome, J. Math. Phys. **7**, 612 (1966).
[30]J. R. Derome, J. Math. Phys. **8**, 714 (1967).
[31]J. A. R. Coope, J. Math. Phys. **11**, 1591 (1970).
[32]P. H. Butler (private communication).
[33]A. J. van Zanten and E. deVries, J. Algebra 25, 475 (1973).

Some comments on a recent paper of Glasser [J. Math. Phys. 14, 409 (1973)]

André Hautot*

Institut de Physique, Université de Liège, Sart Tilman par 4000 Liège 1, Belgium
(Received 5 November 1973)

The calculations which lead to results recently established by Glasser are simplified.

In a recent issue of the journal [J. Math. Phys. 14, 409 (1973)], Glasser published a very interesting paper dealing with the possibility of evaluating exactly certain double series. The final results are expressed by him in terms of Riemann zeta and beta functions; they are listed in Table I of his paper. The aim of this short note simply lies in the two following point:

(a) An error has been committed in the evaluation of S_2; we correct it.
(b) it is possible to simplify greatly a part of the calculations, especially those which lead to the value of S_5.

Let us first recall the notations used: m and n run over all positive integers; p and q run over all positive even integers; and k and l run over all positive odd integers. We wish to evaluate the following sums:

$$S_1 = \sum_{m,n} (m^2 + n^2)^{-s}, \qquad S_2 = \sum_{m,n} (-1)^{m+n}(m^2 + n^2)^{-s},$$

$$S_3 = \sum_{m,n} (-1)^{m-1}(m^2 + n^2)^{-s}, \qquad S_4 = \sum_{k,l} (k^2 + l^2)^{-s},$$

$$S_5 = \sum_{m,p} (m^2 + p^2)^{-s}, \qquad S_6 = \sum_{k,p} (k^2 + p^2)^{-s},$$

$$S_7 = \sum_{m,k} (m^2 + k^2)^{-s}.$$

As we shall see later it is only necessary to evaluate two of these sums since all the others will follow through elementary arithmetical deductions. It is possible to simplify the procedure indicated by Glasser in the following way. To evaluate S_1 and S_3 we start with Jacobi's identities ($|q| < 1$):

$$\sum_0^\infty \sum_0^\infty (-1)^r q^{(2r+1)(t+1)} = \sum_0^\infty \sum_1^\infty q^{m^2+n^2},$$

$$\sum_0^\infty \sum_0^\infty (-1)^{r+t} q^{(4r+2)(t+1)} = \sum_1^\infty \sum_1^\infty (-1)^{m+1} q^{m^2+n^2} - \sum_1^\infty q^{4n^2}.$$

The reader who is not familiar with the theory of the theta functions can verify these identities by equating in the two members the terms of equal power in q. Put $q = e^{-x}$, multiply the two members by x^{s-1} and integrate both sides between $x = 0$ and $x = \infty$. The results announced by Glasser are immediate. The evaluation of the other sums can, of course, be performed through the same procedure; but there is a simpler way. It is evident that the set of positive integers can be split into two subsets: the set of positive odd integers and the set of positive even integers. Our conclusion is that one has: $2S_7 = S_1 + S_3$ and $2S_5 = S_1 - S_3$ through a simple arithmetical device.

The reader will prove without difficulties that

$$S_2 = (2^{2-2s} - 1)S_1 + 2S_3,$$

$$S_4 = 2^{-2s}S_1 + S_3,$$

$$2S_6 = (1 - 2^{1-2s})S_1 - S_3.$$

Thus we have proved that all the sums $S_1 \cdots S_7$ are deduced linearly from S_1 and S_3. Finally, one has the following table (with the corrected value of S_2):

$$S_1 = \zeta(s)\beta(s) - \zeta(2s),$$

$$S_2 = (1 - 2^{1-2s})\zeta(2s) - (1 - 2^{1-s})\zeta(s)\beta(s),$$

$$S_3 = 2^{-s}[2^{-s}\zeta(2s) + (1 - 2^{1-s})\zeta(s)\beta(s)],$$

$$S_4 = 2^{-s}(1 - 2^{-s})\zeta(s)\beta(s),$$

$$S_5 = \tfrac{1}{2}(1 - 2^{-s} + 2^{1-2s})\zeta(s)\beta(s) - \tfrac{1}{2}(1 + 2^{-2s})\zeta(2s),$$

$$S_6 = \tfrac{1}{2}(1 - 2^{-s})\zeta(s)\beta(s) - \tfrac{1}{2}(1 - 2^{-2s})\zeta(2s),$$

$$S_7 = \tfrac{1}{2}(1 + 2^{-s} - 2^{1-2s})\zeta(s)\beta(s) - \tfrac{1}{2}(1 - 2^{-2s})\zeta(2s).$$

*Presently Professor at the National University of Zaïre, Kinshasa.

Erratum: Modified Lippmann–Schwinger equations for two-body scattering theory with long-range interactions [J. Math. Phys. 14, 1398 (1973)]

E. Prugovečki and J. Zorbas

Department of Mathematics, University of Toronto, Toronto, Canada M5S 1A1
(Received 19 October 1973)

The definition of D in (4.5) should read

$$(D\chi_{[a,b)})(\lambda) = \left| \lambda^{1/2}\left(1 - \cos\frac{q_0}{\sqrt{\lambda}} \log\left|\frac{\lambda - b}{\lambda - a}\right|\right)\right|^{1/2}.$$

On the right-hand side of the inequality (4.8) there should be x instead of x^2. With these corrections Lemma 4.1 still shows that $\| Z_\mp^\Delta \chi_{[a,b)} \| \leq \text{const} \| \chi_{[a,b)} \|$ for $[a, b) \subset \Delta$, but it is not sufficient for establishing boundedness for the general case when $Z_\mp^\Delta$ is applied to an arbitrary element of the form (4.6). In order to prove boundedness in this case, improved estimates taking into account the highly oscillatory behaviour of the kernel of $Z_{\mp\epsilon}^\Delta$ are required.