On the first approximation of the K-harmonics method[a)]

J. A. Castilho Alcarás

Instituto de Física Teórica, São Paulo, Brazil
(Received 27 June 1977)

Formulas for the calculation of two-body central interactions in the first approximation of the K-harmonics method are presented. These formulas are exact and account for diagonal and off-diagonal matrix elements as well.

1. INTRODUCTION

Simonov and Badalyan[1-4] have developed the so-called K-harmonics method to obtain the energy levels and the corresponding wavefunctions of the bound states of the nuclear many-body problem with two-body interactions. The essence of the method is to expand the wavefunction of the system, in the c.m. frame, in terms of angular functions which form a complete set of functions in the vector space $E_{3(A-1)}$ spanned by the $3(A-1)$ coordinates of the relative vectors of the A particles which constitute the nuclear system. These angular functions, called K harmonics, also contain the spin and isospin coordinates of the nucleons, are totally antisymmetric, and constitute the angular part of harmonic polynomials of degree K in the $3(A-1)$ spatial variables. Due to the antisymmetry of the K harmonics, the label K can assume only the values $K_{\min}, K_{\min}+1, \cdots$, where $K_{\min}$ is a nonnegative integer which depends on Z and $N=A-Z$.

Successive approximations of the method are obtained by considering, in the expansion of the wavefunction, only the K harmonics with K up to $K_{\min}, K_{\min}+1, \cdots$.

Homogeneous polynomials of degree K, totally antisymmetric, can be constructed by filling a Slater determinant with orbitals whose space parts are harmonic polynomials of degree K_i in the relative variables $\rho_i=\mathbf{r}_i-\mathbf{R}$ ($\mathbf{R}$ = center of mass). When $K=\sum_i K_i$ has its minimum value, the polynomial turns out to be harmonic, the corresponding K harmonic being obtained by normalizing it and dividing it by $\rho^{K_{\min}}$ where $\rho=[\sum_{i=1}^{A}\rho_i^2]^{1/2}$ is the hyperdistance in $E_{3(A-1)}$.

Even for $K=K_{\min}$, the calculation of matrix elements of the two-body interactions between these determinantal wavefunctions, for medium and heavy nuclei, is complicated by the fact that the orbitals are not, in general, orthogonal, since the integration is performed in $S_{3(A-1)}$, the unit sphere of $E_{3(A-1)}$. For $K=K_{\min}$, this difficulty was overcome by Baz and Zhukov[5] by introducing a complex contour integral. However, the integrals involved cannot be calculated and they use the saddle point method to obtain approximated expressions for the matrix elements of the two-body interactions. Gorbatov,[6] by introducing a new set of orbitals and a convenient representation of the solid angle element in $E_{3(A-1)}$, obtained an *exact* expression for the diagonal matrix elements of the two-body interactions. The purpose of the present paper is to extend Gorbatov's paper to account for the off-diagonal matrix elements. Introducing generalized Gorbatov coefficients, the resulting formulas that we obtained can be put in a way that shows a structural similarity with the corresponding ones of the nuclear shell model.[7]

2. THE $K_{\min}$ K HARMONIC

Following Gorbatov, we construct the $K_{\min}$ K harmonic by filling a Slater determinant with orbitals

$$\Phi_{abc}^{\mu\tau}(i)=(\rho_{xi})^a(\rho_{yi})^b(\rho_{zi})^c\alpha_{\mu\tau}(i), \quad i=1,2,\ldots,A,$$
$$a,b,c=0,1,2,\cdots, \quad \mu,\tau=+\tfrac{1}{2},-\tfrac{1}{2}, \tag{1}$$

where ρ_{xi}, ρ_{yi}, and ρ_{zi} are the Cartesian components of the relative vector of the particle i and $\alpha_{\mu\tau}(i)$ accounts for its spin-isospin variables. The $K_{\min}$ K harmonic can be conveniently written in an operational form as[8]

$$U_K(\Omega_{3(A-1)})=\frac{i^K B}{\rho^K}D_K(A)_{\mathbf{k}}\exp\{i\prod_{j=1}^{A}\mathbf{k}_j\cdot\boldsymbol{\rho}_j\}, \tag{2}$$

where B is a normalization constant and $D_K(A_{\mathbf{k}})$ is an antisymmetrizer operator given by

$$D_K(A)_{\mathbf{k}}=\lim_{\{\mathbf{k}_j\}\to 0}\sum_P\epsilon_P\prod_{j=1}^{A}\alpha_{\mu_{Pj}\tau_{Pj}}(j)\times\left(\frac{\partial}{\partial k_{xj}}\right)^{a_{Pj}}\left(\frac{\partial}{\partial k_{yj}}\right)^{b_{Pj}}\left(\frac{\partial}{\partial k_{zj}}\right)^{c_{Pj}}, \tag{3}$$

with

$$\sum_{j=1}^{A}(a_j+b_j+c_j)=K, \tag{4}$$

$$\epsilon_P=\epsilon_{(P1)(P2)\cdots(PA)}. \tag{5}$$

In the evaluation of matrix elements in $S_{3(A-1)}$, it is convenient to write the solid angle element as[5]

$$d\Omega_{3(A-1)}=\frac{2A^{3/2}}{(2\pi)^4}\rho^{-(3A-5)}\int d\mathbf{f}\exp[i\mathbf{f}\cdot(\boldsymbol{\rho}_1+\cdots+\boldsymbol{\rho}_A)\times\int_{-\infty}^{\infty}dt\exp[it(\sum_{j=1}^{A}\rho_j^2-\rho^2)]\prod_{s=1}^{A}d\boldsymbol{\rho}_s. \tag{6}$$

Using Eqs. (2) and (6), the matrix element between a K harmonic U'_K with orbitals $\Phi_{a'b'c'}^{\mu'\tau'}$, and U_K with orbitals $\Phi_{abc}^{\mu\tau}$, is given by

$$\langle U'_K|U_K\rangle\equiv\int d\Omega_{3(A-1)}U'^{\dagger}_K(\Omega_{3(A-1)})U_K(\Omega_{3(A-1)})$$
$$=\frac{2A^{3/2}BB'}{(2\pi)^4\rho^{2K+3A-5}}\int_{-\infty}^{\infty}dt\left(\frac{i\pi}{t}\right)^{3A/2}\exp(-it\rho^2)\times\int d\mathbf{f}\exp[-iAf^2/(4t)]D'^{\dagger}_K(A)_{\mathbf{k}'}D_K(A)_{\mathbf{k}}\times\left[\exp\left(-\frac{i}{4t}\sum_{s=1}^{A}[(k_s^2+k_s'^2)-2\mathbf{f}\cdot(\mathbf{k}'_s-\mathbf{k}_s)]\right)\right]\times\exp\left(\frac{i}{2t}\sum_{s=1}^{A}\mathbf{k}_s\cdot\mathbf{k}'_s\right), \tag{7}$$

[a)]Work supported by FINEP, Rio de Janeiro, Brazil.

 0022-2488/78/1901-0001$1.00

where in the last step the integrations in ρ_i were performed.

Gorbatov proved that when K is equal to $K_{\min}$, then, for any symmetric function $\mathcal{S}$ of $\{\mathbf{k}_i\}$ and an arbitrary function P, it holds that

$$D_{K_{\min}}(A)_{\mathbf{k}}\mathcal{S}(\{\mathbf{k}_i\})P=\mathcal{S}(\{0\})D_{K_{\min}}(A)_{\mathbf{k}}P. \tag{8}$$

Using this result, with $\mathcal{S}$ being the quantity between the large square brackets in Eq. (7), the integration in $\mathbf{f}$ is immediate and the result of the antisymmetrizer operators acting over the last exponential in Eq. (7) gives

$$\left(\frac{i}{2t}\right)^K\left(\prod_{j=1}^A a_j!\,b_j!\,c_j!\right)\sum_{P,P'}\epsilon_P\epsilon_{P'}\prod_{i=1}^A\delta_{a'_{P'i}a_{Pi}}\times\delta_{b'_{P'i}b_{Pi}}\delta_{c'_{P'i}c_{Pi}}\delta_{\mu'_{P'i}\mu_{Pi}}\delta_{\tau'_{P'i}\tau_{Pi}}. \tag{9}$$

The product of deltas in Eq. (9) shows that only the P's and P''s, such that

$$(a,b,c,\mu,\tau)_{Pi}=(a',b',c',\mu',\tau')_{P'i},\quad \text{for } i=1,2,\ldots A, \tag{10}$$

contribute to the matrix element. On the other hand, conditions (10) imply that the orbitals of U'_K and U_K must be equal, up to an irrelevant permutation, We then conclude that $K_{\min}$ *K harmonics with a different set of orbitals of the form* (1) *are orthogonal in* $S_{3(A-1)}$. For $U'_K=U_K$, the substitution of Eq. (9) into Eq. (7), the integrations in $\mathbf{f}$ and t and the requirement that U_k be normalized in $S_{3(A-1)}$ yield for B^2 the value

$$B_A^2(K_{\min})=\frac{2^{K_{\min}-1}\Gamma(K_{\min}+3(A-1)/2)}{\pi^{3(A-1)/2}A!\prod_{i=1}^A a_i!\,b_i!\,c_i!}. \tag{11}$$

3. MATRIX ELEMENTS OF TWO-BODY CENTRAL INTERACTIONS

Let us now turn to the evaluation, in $S_{3(A-1)}$, of matrix elements of two-body central interactions which are linear combinations of operators of the form

$$\hat{F}_2=\sum_{i>j=1}^A v(|\rho_i-\rho_j|)O(i,j), \tag{12}$$

where $O(i,j)$ is an operator which acts on the spin-isospin variables of the pair of particles (i,j). Using Eqs. (2), (6), and (12), one obtains

$$\begin{aligned}\langle U'|\hat{F}_2|U\rangle &\equiv\int d\Omega_{3(A-1)}U'^{\dagger}_k(\Omega_{3(A-1)})\hat{F}_2U(\Omega_{3(A-1)})\\ &=\frac{2A^{3/2}BB'}{(2\pi)^4\rho^{2K+3A-5}}\binom{A}{2}\int_{-\infty}^{\infty}dt\left(\frac{i\pi}{t}\right)^{3(A-2)/2}\exp(-it\rho^2)\\ &\times\int d\mathbf{f}\exp[-iAf^2/(4t)]\,D'^{\dagger}_K(A)_{\mathbf{k}'}O(1,2)D_K(A)_{\mathbf{k}}\\ &\times\left[\exp\left(-\frac{i}{4t}\sum_{s=1}^A[(k_s^2+k_s'^2)-2\mathbf{f}\cdot(\mathbf{k}'_s-\mathbf{k}_s)]\right)\right]\exp\left(\frac{i}{2t}\sum_{s=1}^A\mathbf{k}_s\cdot\mathbf{k}'_s\right)\\ &\times\exp\left(\frac{i}{4t}[(\mathbf{f}+\mathbf{k}_1-\mathbf{k}'_1)^2+(\mathbf{f}+\mathbf{k}_2-\mathbf{k}'_2)^2]\right)\\ &\times\int d\rho_1\int d\rho_2\,v(|\rho_1-\rho_2|)\exp\{i[(\mathbf{f}-\mathbf{k}'_1+\mathbf{k}_1)\cdot\rho_1\\ &+(\mathbf{f}-\mathbf{k}'_2+\mathbf{k}_2)\cdot\rho_2+t(\rho_1^2+\rho_2^2)]\},\end{aligned} \tag{13}$$

where we have used the fact that $U'^{\dagger}_K F_2 U_K$ is symmetric under the permutation of all coordinates of the particles to replace $\hat{F}_2$ by $\binom{A}{2}v(|\rho_1-\rho_2|)O(1,2)$ and we have performed the integrations in $\rho_3,\rho_4,\ldots,\rho_A$.

Introducing center of mass coordinates for the pair (1, 2),

$$\mathbf{r}=\rho_1-\rho_2,\quad \lambda=\tfrac{1}{2}(\rho_1+\rho_2), \tag{14}$$

the integration in λ is easily performed, giving

$$\left(\frac{i\pi}{2t}\right)^{3/2}\exp\left(-\frac{i}{8t}(2\mathbf{f}+\mathbf{k}_1-\mathbf{k}'_1+\mathbf{k}_2-\mathbf{k}'_2)^2\right). \tag{15}$$

Using Eqs. (15) and (8) and performing the integration in $\mathbf{f}$, Eq. (13) becomes

$$\begin{aligned}\langle U'|\hat{F}_2|U\rangle=\frac{2A^{3/2}BB'}{(2\pi)^4\rho^{2K+3A-5}}\binom{A}{2}\int_{-\infty}^{\infty}dt\left(\frac{i\pi}{t}\right)^{3(A-2)/2}\left(\frac{2\pi^2}{A}\right)^{3/2}\\ \times\exp(-it\rho^2)D'^{\dagger}_K(A)_{\mathbf{k}'}O(1,2)D_K(A)_{\mathbf{k}}\\ \times\exp\left\{\frac{i}{2t}\sum_{s=3}^A\mathbf{k}_s\cdot\mathbf{k}'_s\right\}\\ \times\exp\left(\frac{i}{8t}[(\mathbf{k}_1-\mathbf{k}_2)^2+(\mathbf{k}'_1-\mathbf{k}'_2)^2+2(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{k}'_1+\mathbf{k}'_2)]\right)\int d\mathbf{r}\,v(r)\exp\left(\frac{i}{2}[(\mathbf{k}_1-\mathbf{k}'_1-\mathbf{k}_2+\mathbf{k}'_2)\cdot\mathbf{r}+tr^2]\right).\end{aligned} \tag{16}$$

The only dependence on $\mathbf{k}_3,\ldots,\mathbf{k}_A$, $\mathbf{k}'_3,\ldots,\mathbf{k}'_A$ is given by

$$\begin{aligned}&D'^{\dagger}_K(A)_{\mathbf{k}'}O(1,2)D_K(A)_{\mathbf{k}}\exp\left(\frac{i}{2t}\sum_{s=3}^A\mathbf{k}_s\cdot\mathbf{k}'_s\right)\\ &=\sum_{P'}\epsilon_{P'}\sum_{P}\epsilon_P\left[\prod_{j=3}^A\left(\frac{i}{2t}\right)^{a_{Pj}+b_{Pj}+c_{Pj}}a_{Pj}!\,b_{Pj}!\,c_{Pj}!\right.\\ &\left.\times\delta_{a_{Pj}a'_{P'j}}\delta_{b_{Pj}b'_{P'j}}\delta_{c_{Pj}c'_{P'j}}\delta_{\mu_{Pj}\mu'_{P'j}}\delta_{\tau_{Pj}\tau'_{P'j}}\right]\\ &\times\alpha^{\dagger}_{\mu'_{P'1}\tau'_{P'1}}(1)\alpha^{\dagger}_{\mu'_{P'2}\tau'_{P'2}}(2)O(1,2)\alpha_{\mu_{P1}\tau_{P1}}(1)\alpha_{\mu_{P2}\tau_{P2}}(2)\\ &\times\lim_{\{\mathbf{k}_i,\mathbf{k}'_i\}\to 0}\prod_{i=1}^2\left(\frac{\partial}{\partial k'_{xi}}\right)^{a'_{P'i}}\left(\frac{\partial}{\partial k'_{yi}}\right)^{b'_{P'i}}\left(\frac{\partial}{\partial k'_{zi}}\right)^{c'_{P'i}}\\ &\times\left(\frac{\partial}{\partial k_{xi}}\right)^{a_{Pi}}\left(\frac{\partial}{\partial k_{yi}}\right)^{b_{Pi}}\left(\frac{\partial}{\partial k_{zi}}\right)^{c_{Pi}}.\end{aligned} \tag{17}$$

The product of the δ's in Eq. (17) gives the result that only the P's and P''s, such that

$$(a,b,c,\mu,\tau)_{Pi}=(a',b',c',\mu',\tau'),\quad i=3,4,\ldots,A, \tag{18}$$

contribute to the matrix element, which fixes P'_i for $i=3,4,\ldots,A$. Therefore, the summation in $P3,\ldots,PA$, $P'3$, $P'4,\ldots,P'A$ simply give $(A-2)!$. Besides that, Eqs. (18) imply that U'_K and U_K must have at least $(A-2)$ orbitals in common in order that the matrix element $\langle U'_K|\hat{F}_2|U_K\rangle$ be nonvanishing.

The integration in the angular part of $\mathbf{r}$ is accomplished by using the plane wave expansion theorem. Finally, we make a power series expansion of $\exp\{(i/8t)[(k_1-k_2)^2+\cdots]\}$ and integrate in t, obtaining

$$\langle U'|\hat{F}_2|U\rangle=\frac{2BB'\nu\nu'\pi^{3A/2}}{\pi^2}A!\sum_{P1,P2}\xi'_{P1P2}\sum_{s=0}^{\infty}\frac{(-)^s}{(2S+1)!}$$

$$\times\sum_{n=0}^{\infty}\frac{2^{-(K'_{P1P2}+n+s)}\rho^{2(n-d'_{P1P2}-s)}}{n!\,\Gamma(K'_{P1P2}+n+3A/2-3)}\sum_{l=0}^{n}\frac{\binom{n}{l}}{2^{2l}}$$

$$\times\int_0^1 dz\, z^{s+1/2}(1-z)^{K'_{P1P2}+n+3A/2-4}\,v(\rho\sqrt{2z})$$

$$\times[J'^{(1)}_{P1P2}\langle P1P2|O(1,2)|P1P2\rangle$$

$$-J'^{(2)}_{P1P2}\langle P2P1|O(1,2)|P1P2\rangle], \tag{19}$$

where

$$\nu=\left[\prod_{i=1}^{A}a_i!\,b_i!\,c_i!\right]^{1/2},\quad \nu'=\left[\prod_{i=1}^{A}a'_i!\,b'_i!\,c'_i!\right]^{1/2}, \tag{20}$$

$$\xi'_{ij}=[a_i!\,b_i!\,c_i!\,a_j!\,b_j!\,c_j!\,a'_i!\,b'_i!\,c'_i!\,a'_j!\,b'_j!\,c'_j!]^{-1/2}, \tag{21}$$

$$d'_{ij}=(a_i+b_i+c_i+a_j+b_j+c_j+a'_i+b'_i+c'_i+a'_j+b'_j+c'_j)/2, \tag{22}$$

$$K'_{ij}=K-d'_{ij}, \tag{23}$$

$$J'^{(1)}_{ij}=\lim_{\{\mathbf{k}_i,\mathbf{k}'_i\}\to 0}\left(\frac{\partial}{\partial k_{x1}}\right)^{a_i}\left(\frac{\partial}{\partial k'_{x1}}\right)^{a'_i}\left(\frac{\partial}{\partial k_{y1}}\right)^{b_i}\left(\frac{\partial}{\partial k'_{y1}}\right)^{b'_i}\left(\frac{\partial}{\partial k_{z1}}\right)^{c_i}$$

$$\times\left(\frac{\partial}{\partial k'_{z1}}\right)^{c'_i}\left(\frac{\partial}{\partial k_{x2}}\right)^{a_j}\left(\frac{\partial}{\partial k'_{x2}}\right)^{a'_j}\left(\frac{\partial}{\partial k_{y2}}\right)^{b_j}\left(\frac{\partial}{\partial k'_{y2}}\right)^{b'_j}\left(\frac{\partial}{\partial k_{z2}}\right)^{c_j}$$

$$\times\left(\frac{\partial}{\partial k'_{z2}}\right)^{c'_j}$$

$$\times(\mathbf{k}_1-\mathbf{k}_2-\mathbf{k}'_1+\mathbf{k}'_2)^{2(l+s)}(\mathbf{k}_1\cdot\mathbf{k}'_1+\mathbf{k}_2\cdot\mathbf{k}'_2)^{n-l}, \tag{24}$$

$$J'^{(2)}_{ij}=\lim_{\{\mathbf{k}_i,\mathbf{k}'_i\}\to 0}$$

$$\times\left(\frac{\partial}{\partial k_{x1}}\right)^{a_i}\left(\frac{\partial}{\partial k'_{x2}}\right)^{a'_i}\left(\frac{\partial}{\partial k_{y1}}\right)^{b_i}\left(\frac{\partial}{\partial k'_{y2}}\right)^{b'_i}\left(\frac{\partial}{\partial k_{z1}}\right)^{c_i}\left(\frac{\partial}{\partial k'_{z2}}\right)^{c'_i}$$

$$\times\left(\frac{\partial}{\partial k_{x2}}\right)^{a_j}\left(\frac{\partial}{\partial k'_{x1}}\right)^{a'_j}\left(\frac{\partial}{\partial k_{y2}}\right)^{b_j}\left(\frac{\partial}{\partial k'_{y1}}\right)^{b'_j}\left(\frac{\partial}{\partial k_{z2}}\right)^{c_j}\left(\frac{\partial}{\partial k'_{z1}}\right)^{c'_j}$$

$$\times(\mathbf{k}_1-\mathbf{k}_2+\mathbf{k}'_2-\mathbf{k}'_1)^{2(l+s)}(\mathbf{k}_1\cdot\mathbf{k}'_1+\mathbf{k}_2\cdot\mathbf{k}'_2)^{n-l}, \tag{25}$$

$$\langle i'j'|O(1,2)|ij\rangle=\alpha^{\dagger}_{\mu'_i,\tau'_i}(1)\alpha^{\dagger}_{\mu'_j,\tau'_j}(2)O(1,2)$$

$$\times\alpha_{\mu_i\tau_i}(1)\alpha_{\mu_j\tau_j}(2). \tag{26}$$

From Eqs. (24) and (25) one sees that only the n's and s's, such that $s+n=d'_{P1P2}$, contribute to the matrix elements. Therefore, in Eq. (19) one may replace the sums in n and s by substituting n by $d'_{P1P2}-s$ and summing in s from 0 to d'_{P1P2}.

Defining the generalized Gorbatov coefficients[9]

$$G^{(q)}(i,j,s)=\frac{\xi_{ij}(-)^s}{2^s s!\,(d'_{ij}-s)!}\sum_{l=0}^{d'_{ij}-s}\binom{d'_{ij}-s}{l}2^{-2l}J'^{(q)}_{ij},\quad q=1,2, \tag{27}$$

one obtains

$$\langle U'_K|\hat{F}_2|U_K\rangle=\frac{\Gamma(3(A-1)/2+K)}{\sqrt{\pi}}\sum_{P1,P2}\sum_{s=0}^{d'_{P1P2}}[\Gamma(K-s+3A/2-3)(2s+1)!!]^{-1}$$

$$\times[\langle P1P2|O(1,2)P1P2\rangle\,G^{(1)}(P1,P2,s)$$

$$-\langle P2P1|O(1,2)|P1P2\rangle\,G^{(2)}(P1,P2,s)]$$

$$\times\int_0^1 dz\, z^{s+1/2}(1-z)^{K-s-3A/2-4}v(\rho\sqrt{2z}). \tag{28}$$

To make the summation in $P1$ and $P2$ explicit, one has to consider the three following possibilities:

(i) when U'_K and U_K have the same orbitals,

$$\langle U'_K|\hat{F}_2|U_K\rangle=\frac{\Gamma(3(A-1)/2+K)}{\sqrt{\pi}}\sum_{j>i=1}^{A}\sum_{s=0}^{d'_{ij}}[\Gamma(K-s+3A/2-3)$$

$$\times(2s+1)!!]^{-1}\int_0^1 dz\, z^{1/2+s}(1-z)^{K+3A/2-s-4}$$

$$\times v(\rho\sqrt{2z})\{G^{(1)}(i,j,s)[\langle ij|O(1,2)|ij\rangle$$

$$+\langle ji|O(1,2)|ji\rangle]$$

$$-G^{(2)}(i,j,s)[\langle ij|O(1,2)|ji\rangle$$

$$+\langle ji|O(1,2)|ij\rangle]\}; \tag{29}$$

(ii) when U'_K and U_K have only $(A-1)$ orbitals in common (numbered $1, 2, \ldots, A-1$), $P1$ or $P2$ must be equal to A, and one has

$$\langle U'_K|\hat{F}_2|U_K\rangle=\frac{\Gamma(K+3A/2-3/2)}{\sqrt{\pi}}\sum_{i=1}^{A-1}\sum_{s=0}^{d'_{iA}}[\Gamma(K-s+3A/2-3)$$

$$\times(2s+1)!!]^{-1}\int_0^1 dz\, z^{s+1/2}(1-z)^{K-s+3A/2-4}$$

$$\times v(\rho\sqrt{2z})\{G^{(1)}(i,A,s)[\langle iA|O(1,2)|iA\rangle$$

$$+\langle Ai|O(1,2)|Ai\rangle]$$

$$-G^{(2)}(i,A,s)[\langle iA|O(1,2)|Ai\rangle$$

$$+\langle Ai|O(1,2)|iA\rangle]\}; \tag{30}$$

(iii) when U'_K and U_K have only $(A-2)$ orbitals in common (numbered $1, 2, \ldots, A-2$), $P1$ and $P2$ can assume only the values A and $A-1$ and one has

$$\langle U'_K|\hat{F}_2|U_K\rangle=\frac{\Gamma(K+3A/2-3/2)}{\sqrt{\pi}}\sum_{s=0}^{d'_{A-1,A}}[\Gamma(K-s+3A/2-3)$$

$$\times(2s+1)!!]^{-1}\int_0^1 dz\, z^{s+1/2}(1-z)^{K-s+3A/2-4}$$

$$\times v(\rho\sqrt{2z})\{G^{(1)}(A-1,A,s)$$

$$\times[\langle A-1A|O(1,2)|A-1A\rangle$$

$$+\langle AA-1|O(1,2)|AA-1\rangle]$$

$$-G^{(2)}(A-1,A,s)[\langle A-1A|O(1,2)|AA-1\rangle$$

$$-\langle AA-1|O(1,2)|A-1A\rangle]\}. \tag{31}$$

For nuclear potentials of the form

$$\hat{F}_2=\sum_{j>i=1}^{A}\{v_{33}(|\rho_i-\rho_j|)P^+_\mu P^+_\tau+v_{31}(|\rho_i-\rho_j|)P^+_\mu P^-_\tau$$

$$+v_{13}(|\rho_i-\rho_j|)P^-_\mu P^+_\tau+v_{11}(|\rho_i-\rho_j|)P^-_\mu P^-_\tau\}, \tag{32}$$

where $P^{\pm}_\mu$ and $P^{\pm}_\tau$ are the triplet and singlet projection operators which act, respectively in the spin and isospin variables of the pair of particles (i,j), one has

$$\langle ij|P^r_\mu P^s_\tau|nl\rangle=\tfrac{1}{4}(\delta_{\mu'_i\mu_n}\delta_{\mu'_j\mu_l}+(-)^r\delta_{\mu'_i\mu_l}\delta_{\mu'_j\mu_n}$$

$$\times(\delta_{\tau'_i\tau_n}\delta_{\tau'_j\tau_l}+(-)^s\delta_{\tau'_i\tau_l}\delta_{\tau'_j\tau_n}), \tag{33}$$

provided one uses the convention that r (or s) is equal to 1 for the singlet and 0 for the triplet.

Substituting Eq. (33) in the diagonal matrix element Eq. (29) one recovers Gorbatov's Eq. (20).

For the Coulomb potential

$$\hat{F}_2 = e^2 \sum_{j>i=1} \frac{Q_i Q_j}{|\rho_i - \rho_j|}, \tag{34}$$

where eQ_i is the charge operator of particle i, one has[10]

$$\langle ij | Q_1 Q_2 | nl \rangle = \tfrac{1}{16}(1-2\tau_i')(1-2\tau_j')(1-2\tau_n) \times (1-2\tau_l)\,\delta_{\mu_i'\mu_n}\delta_{\mu_j'\mu_l}. \tag{35}$$

For the diagonal case, Eqs. (29), (34), and (35) give

$$\langle U_K | F_2 | U_K \rangle = \frac{\Gamma(K+3A/2-3/2)}{2\sqrt{\pi}\,\Gamma(K+3A/2-2)} \frac{e^2}{\rho\sqrt{2}} \sum_{j>i=1}^{A} (1-2\tau_i)(1-2\tau_j) \times \sum_{s=0}^{d_{ij}'} \frac{s!}{(2s+1)!!}[G^{(1)}(i,j,s) - \delta_{\mu_i\mu_j} G^{(2)}(i,j,s)]. \tag{36}$$

[1] Yu. A. Simonov, Sov. J. Nucl. Phys. **3**, 461 (1966).
[2] A. M. Badalyan and Yu. A. Simonov, Sov. J. Nucl. Phys. **3**, 755 (1966).
[3] Yu. A. Simonov, Sov. J. Nucl. Phys. **7**, 722 (1968).
[4] A. M. Badalyan and Yu. A. Simonov, Sov. J. Nucl. Phys. **9**, 42 (1969).
[5] A. I. Baz and M. W. Zhukov, Sov. J. Nucl. Phys. **11**, 435 (1970).
[6] A. M. Gorbatov, Sov. J. Nucl. Phys. **10**, 547 (1970).
[7] H. Bethe, *Intermediate Quantum Mechanics* (Benjamin, New York, 1968).

[8] From now on, we omit the subscript min in $K_{\min}$.
[9] When U_K' and U_K have the same orbitals, the $G^{(q)}(ij,s)$ reduce to the Gorbatov coefficients.
[10] We adopt here the convention that $\tau_i = \frac{1}{2}$ (in units of $\hbar$) if particle i is a neutron and $-\frac{1}{2}$ if it is a proton.

Self-avoiding random walks: Some exactly soluble cases

Deepak Dhar

George W. Downs Laboratory of Physics, California Institute of Technology, Pasadena, California 91125
(Received 12 July 1977)

We use the exact renormalization group equations to determine the asymptotic behavior of long self-avoiding random walks on some pseudolattices. The lattices considered are the truncated 3-simplex, the truncated 4-simplex, and the modified rectangular lattices. The total number of random walks C_n, the number of polygons P_n of perimeter n, and the mean square end to end distance $\langle R_n^2 \rangle$ are assumed to be asymptotically proportional to $\mu^n n^{\gamma-1}$, $\mu^n n^{\alpha-3}$, and $n^{2\nu}$ respectively for large n, where n is the total length of the walk. The exact values of the connectivity constant μ, and the critical exponents λ, α, ν are determined for the three lattices. We give an example of two lattice systems that have the same effective nonintegral dimensionality 3/2 but different values of the critical exponents γ, α, and ν.

I. INTRODUCTION AND OUTLINE

The self avoiding random walk was originally proposed as a model of polymer chains, to study the effect of excluded volume.[1] The properties of such walks are connected with some properties of the Ising model.[2] The study of the configurational problems encountered in this problem may be expected to shed some light on the more general problem of second order phase transitions. The problem has been attached using a variety of analytical and numerical techniques,[3–5] but the number of exact results known is small. It has resisted a complete solution in the physically interesting case of three dimensions, or even in the considerably simpler case of two dimensions.

It is thus of some interest to study the problem for some pseudolattices, where the exact solution may be worked out and its behavior analyzed in detail. An example of such lattices is the Bethe lattice, which has been very important historically in the development of the theory of phase transitions. Detailed study of the Ising model on this lattice has suggested the possibility of a new kind of phase transition (phase transitions of continuous order),[6] which have subsequently been realized on more conventional lattices.[7]

Part of the motivation for the study of pseudolattices springs from the fact that they are very good pedagogical examples of renormalization group techniques at work. Despite enormous progress in the application of renormalization techniques to the field of phase transitions since the pioneering work of Kadanoff and Wilson,[8] the number of cases which show nontrivial phase transitions and where the exact renormalization transformation may be explicitly implemented, has remained rather small. The only other exceptions are the Gaussian model[9] and the hierarchial model.[10] These lattices may also be used to test the validity of new approximation schemes.

The problem of self avoiding random walks on a Bethe lattice is trivial, of course, because of the absence of any closed loops. In this paper we study the self avoiding random walk problem for the truncated tetrahedron lattice, the truncated 4-simplex lattice, and the modified rectangular lattice.

The truncated tetrahedron lattice was defined by Nelson and Fisher.[11] In a previous paper[12] (hereafter referred to as I) we have generalized their construction to define the truncated n-simplex lattice for arbitrary integer n, and shown that the effective dimensionality of this lattice is $2\ln(n)/\ln(n+2)$. We have also defined the modified rectangular lattice, which is planar and is obtained by deleting some bonds from a planar square lattice. For details of the construction of these lattices, the reader is referred to I. In the following discussion familiarity with its contents is assumed.

These lattices are defined recursively, and the exact renormalization equations may be written down for these lattices in terms of only a small number of coupling constants. In I, we used this property to determine the critical behavior of the classical XY model (the results may easily be extended to arbitrary integral spin dimensionality) and the Fortuin–Kasteleyn cluster model on these lattices. No phase transition at a finite (nonzero) temperature is found. The analysis in this paper differs from I mainly in that for the self avoiding random walk problems, the recursion equations have nontrivial fixed points, even for lattices with effective dimensionality less than two. The system shows a phase transition in the sense that the generating functions of the random walk become singular as a function of their argument. We can determine the critical exponents using the standard renormalization group techniques. The recursion equations are coupled algebraic equations and their derivation and analysis is quite straightforward. In the parameter space of the coupling constants, we observe the phenomena of the point specifying the effective interaction approaching the fixed point of renormalization transformation initially. Eventually the point escapes away from the fixed point after a large number of iterations unless the starting system was exactly critical. Linearizing the recursion equations about the fixed point, we determine the critical exponents from the eigenvalues of the linearized renormalization transformation matrix.

II. PRELIMINARIES AND NOTATION

A self avoiding random walk on lattice is a random walk with the constraint that no lattice point is visited more than once. We associate weight factor x with each step of the random walk and define the generating functions

$$C(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{\infty} C_n(N) x^n, \tag{1}$$

 0022-2488/78/1901-0005$1.00

$$P(x)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=2}^{\infty}P_n(N)x^n. \tag{2}$$

Here $C_n(N)$ is the total number of distinct self avoiding randon walks of n steps on a large lattice consisting of N lattice points. $P_n(N)$ is the number of distinct closed simple polygons of perimeter n on the lattice. The random walks may start from any point on the lattice. For large N, the numbers $C_n(N)$ and $P_n(N)$ are asymptotically proportional to N, and the limit exists. For regular lattices, where all the lattice point are equivalent, this limiting procedure is unnecessary because the number of self avoiding walks of length n is independent of the vertex from which the walk starts (so long as the starting vertex is not too close to the boundary of the lattice). This is not the case for the spatially inhomogeneous lattices studied here, and the averaging over all possible positions of the starting point is necessary. We define

$$\bar{P}_n=\lim_{N\to\infty}P_n(N)/N, \tag{3}$$

$$\bar{C}_n=\lim_{N\to\infty}C_n(N)/N. \tag{4}$$

We know that for large n, $\bar{P}_n$ and $\bar{C}_n$ increase geometrically with n. Let us assume that for large n

$$\bar{P}_n\simeq K_1\mu^n n^{\alpha-3}, \tag{5}$$

$$\bar{C}_n\simeq K_2\mu^n n^{\gamma-1}, \tag{6}$$

where K_1 and K_2 are some coefficients of proportionality. In general, we represent a constant of proportionality by K, with or without subscripts. Its numerical value is not necessarily the same in different equations. μ is called the connectivity constant of the lattice and α and γ are the critical indices for the random walk. Substituting the asymptotic behavior of C_n and P_n in Eq. (2) we find that as x tends to $1/\mu$ from below, the asymptotic behavior of $C(x)$ and $P(x)$ is given by

$$C(x)\simeq K_2(1-x\mu)^{-\gamma}+\text{less singular terms}, \tag{7}$$

$$P(x)\simeq K_1(1-x\mu)^{2-\alpha}+\text{less singular terms}. \tag{8}$$

The average number of self avoiding walks per site that return to the origin (polygonal closures) after exactly n steps is given by $2n\bar{P}_n$. We also define the generating function for the mean squared end to end distance by

$$R(x)=\lim_{N\to\infty}\frac{1}{N}\sum_{L}[R(L)]^2x^{n(L)}, \tag{9}$$

where $R(L)$ is the end to end distance for the random walk L with total number of steps given by $n(L)$. The summation extends over all possible self avoiding random walks L on a large lattice of size N. We define the critical exponent ν by the relation

$$\langle R_n^2\rangle\simeq Kn^{2\nu}\quad\text{for large } n, \tag{10}$$

where $\langle R_n^2\rangle$ is the mean squared end to end distance for n-step self avoiding random walks, all walks being weighted equally. Since the number of such walks increases as $\mu^n n^{\gamma-1}$ [Eq. (6)], we find that the asymptotic behavior of $R(x)$ as $x\mu\to 1$ from below is given by

$$R(x)\sim K(1-x\mu)^{-\gamma-2\nu}+\text{less singular terms}. \tag{11}$$

In the next section, we use the renormalization group techniques to determine the values of the constants μ, α, γ, and ν for the truncated 3-simplex, the truncated 4-simplex lattice and the modified rectangular lattices, by determining the singular behavior of their generating functions $C(x)$, $P(x)$, and $R(x)$. We show that for the truncated 3-simplex lattice

$$\mu=1.6180,\quad \alpha=0.7342,\quad \gamma=1.3752,\quad \nu=0.7986. \tag{12a}$$

For the modified rectangular lattice

$$\mu=1.6909,\quad \alpha=0.6699,\quad \gamma=1.4403,\quad \nu=0.6650. \tag{12b}$$

And for the truncated 4-simplex lattice

$$\mu=2.2866,\quad \alpha=0.5413,\quad \gamma=1.4461,\quad \nu=0.7294. \tag{12c}$$

III. CALCULATION OF THE CRITICAL EXPONENTS

In this section we derive the connectivity constants and the critical indices mentioned in the previous section. The analysis of all the three lattices is quite similar and some of the details in the treatment of the truncated 4-simplex lattice and the modified rectangular lattice have been omitted. The treatment may be extended to other recursively defined pseudolattices, but the number of variables that have to be considered to form a closed set of recursion equations soon becomes very large and the analysis becomes difficult.

A. The truncated 3-simplex lattice

We would like to determine the behavior of the generating functions of the random walks $C(x)$, $P(x)$, and $R(x)$ for the truncated 3-simplex lattice. These functions are weighted sums over self avoiding walks. The weight of a walk of length n is x^n. [For $R(x)$, there is an additional multiplicative weight factor depending on the end to end distance of the walk, Eq. (9)]. Instead of assigning a weight x to each step of the walk, we may equivalently assign a weight x to each vertex that the walk passes through, and a weight $\sqrt{x}$ to each of the two vertices that are the starting or the end of the walk. Then, for example, $P(x)$ is the sum over all possible configurations with a single loop.

The renormalization transformation consists of summing over all the internal configurations of the rth order triangles, as was done for the Fortuin–Kasteleyn cluster model in I. We define the rth order restricted partition function as shown in Fig. 1. Here $A^{(r)}$ is the weight of an rth order triangle with one line going in. The end point of the line may be any of the vertices

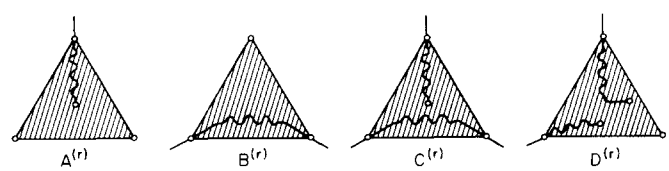

FIG. 1. Restricted partition functions for an rth order triangle. The shaded triangles denote rth order triangles, of which only the corner vertices and the end points of the self-avoiding walks are shown.

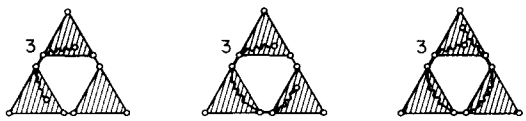

FIG. 2. All possible configurations of an open self avoiding walk of order r. The shaded triangles denote $(r-1)$th order graphs of which only the corner vertices are shown. The factor 3 is for the three possible orientations of the figures.

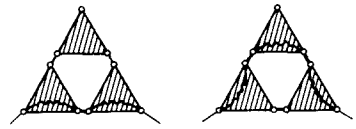

FIG. 3. All possible configurations of rth order triangles (denoted by shaded triangles in the diagram) that contribute to $B^{(r+1)}$.

inside the rth order triangle. We sum over all possible configurations of the rth order triangle consistent with the constraint that one of the end points of the walk lies inside the rth order triangle. Similarly $B^{(r)}$ is the weight of an rth order triangle in which a line goes in from one of the corner vertices and comes out from the other. (The lines are undirected. We use the term going in and coming out rather loosely.) The weights $C^{(r)}$ and $D^{(r)}$ are defined similarly. The starting values of these weights are

$$A^{(0)} = \sqrt{x}\ , \tag{13a}$$

$$B^{(0)} = x, \tag{13b}$$

$$C^{(0)} = D^{(0)} = 0. \tag{13c}$$

We call a closed or open walk L of order r if r is the minimum value of p such that L can be completely described inside a pth order triangle. The sum of weights of all rth order closed loops inside one rth order triangle is clearly $(B^{(r-1)})^3$. Since there are 3^r points in each rth order triangle, the contribution of rth order closed loops per site is $(B^{(r-1)})^3/3^r$. Hence we get

$$P(x) = \sum_{r=1}^{\infty} 3^{-r}(B^{(r-1)})^3. \tag{14}$$

Similarly we get (Fig. 2)

$$C(x) = \sum_{r=1}^{\infty} 3^{-r}\left[3A^{(r-1)^2} + 3B^{(r-1)}(A^{(r-1)})^2 + 3(B^{(r-1)})^2 D^{(r-1)}\right]. \tag{15}$$

It is easy to write down the recursion equations for the weights $A^{(r)}$, $B^{(r)}$, $C^{(r)}$, $D^{(r)}$ by drawing graphically all possible ways a configuration of $(r+1)$th order triangle may arise out of the configurations of (r)th order triangles. Figure 3 shows all the possible configurations that contribute to $B^{(r+1)}$. This shows that

$$B^{(r+1)} = (B^{(r)})^2 + (B^{(r)})^3. \tag{16a}$$

The recursion equations for $A^{(r)}$, $C^{(r)}$, and $D^{(r)}$ are written down similarly and we get

$$A^{(r+1)} = A(1 + 2B + 2B^2) + C(2B^2), \tag{16b}$$

$$C^{(r+1)} = A(B^2) + C(3B^2), \tag{16c}$$

$$D^{(r+1)} = (A^2 + 2A^2B + 4ABC + 6BC^2) + D(2B + 3B^2), \tag{16d}$$

where we have suppressed the superscripts of $A^{(r)}$, $B^{(r)}$, $C^{(r)}$, and $D^{(r)}$ in the right-hand sides of Eqs. (16b)–(16d).

Equations (13)–(16) determine the functions $P(x)$ and $C(x)$ completely. We also notice that the recursion equation for $B^{(r+1)}$ involves only $B^{(r)}$. Also the recursion equations for $A^{(r+1)}$ and $C^{(r+1)}$ are independent of $D^{(r)}$.

From Eq. (16a) and (14), we see that the $P(x)$ satisfies the functional equation

$$P(x) = \frac{x^3}{3} + \frac{1}{3}P(x^2 + x^3). \tag{17}$$

This equation has fixed points given by the equation

$$x^* = x^{*2} + x^{*3} \tag{18}$$

which gives the fixed points $x^* = 0$, $(\sqrt{5}-1)/2$, ∞ in the allowed (real nonnegative) range of x. The fixed points $x^* = 0$ and $x^* = \infty$ are attractive fixed points while the point $x^* = (\sqrt{5}-1)/2$ is repulsive. If the starting value $B^{(0)}$ is less than $(\sqrt{5}-1)/2$, from Eq. (16a) we see that with successive iterations, the value of $B^{(r)}$ decreases to zero. If $B^{(0)}$ is greater than $(\sqrt{5}-1)/2$, for large r, $B^{(r)}$ tends to infinity and $P(x)$ is infinite. This shows that the connective constant of this lattice is given by

$$\mu = 2/(\sqrt{5}-1) \cong 1.6180. \tag{19}$$

Putting $x = 1/\mu$ in Eq. (17) we get

$$P(\mu^{-1}) = \mu^{-3}/2. \tag{20}$$

Consider $x = \mu^{-1} - \delta$ where δ is a small positive number. Then Eq. (17) gives

$$P(\mu^{-1} - \delta) = \tfrac{1}{3}\mu^{-3} - \mu^{-2}\delta + \tfrac{1}{3}P(\mu^{-1} - \delta(2 + \mu^{-2})) + O(\delta^2). \tag{21}$$

We assume that the singular part of $P(\mu^{-1} - \delta)$ varies as $\delta^{2-\alpha}$. This gives us, from Eq. (20),

$$\alpha = 2 - \ln 3/\ln(2 + \mu^{-2}) \simeq 0.7342. \tag{22}$$

Let us define

$$\delta^{(r)} = \mu^{-1} - B^{(r)}\ . \tag{23}$$

Then to lowest order in δ, Eq. (16a) gives the recursion relation

$$\delta^{(r+1)} = (2 + \mu^{-2})\delta^{(r)}\ . \tag{24}$$

We choose a small positive number ϵ, and choose a starting value $\delta^{(0)}$ sufficiently small so that

$$1 \gg \epsilon \gg \delta^{(0)}$$

and

$$r_0 \stackrel{\text{def}}{=} \ln(\epsilon/\delta^{(0)})/\ln(2 + \mu^{-2}) \gg 1. \tag{25}$$

Then for $r < r_0$, $\delta^{(r)}$ is less than ϵ and we may replace $B^{(r)}$ in Eqs. (16b)–(16d) by μ^{-1}. This gives us a set of coupled linear recursion equations for the constants $A^{(r)}$ and $C^{(r)}$,

$$A^{(r+1)} \approx A^{(r)}(1+2\mu^{-1}+2\mu^{-2}) + C^{(r)}2\mu^{-2}\,, \tag{26a}$$

$$C^{(r+1)} \approx A^{(r)}\mu^{-2} + C^{(r)}3\mu^{-2}\,, \tag{26b}$$

which imply that

$$A^{(r)} \approx K_1\lambda^r\,, \tag{27a}$$

$$C^{(r)} \approx K_2\lambda^r \quad \text{for } 1 < r < r_0\,, \tag{27b}$$

where K_1 and K_2 are some constants of proportionality. λ is the larger eigenvalue of the matrix

$$\begin{bmatrix} 1+2\mu^{-1} & +2\mu^{-2} & 2\mu^{-2} \\ & \mu^{-2} & 3\mu^{-2} \end{bmatrix}$$

which gives

$$\lambda = [3+3\mu^{-2}+(9-18\mu^{-2}+17\mu^{-4})^{1/2}]/2. \tag{28}$$

Substituting from Eqs. (27a)—(27b) into Eq. (16d), we see that the recursion equation for $D^{(r+1)}$ has the form

$$D^{(r+1)} \approx K\lambda^{2r} + D^{(r)}(2+\mu^{-2}). \tag{29}$$

Since λ^2 is greater than $(2+\mu^{-2})$, we see that this equation implies

$$D^{(r)} \approx K_1\lambda^{2r} \quad \text{for } 1 < r < r_0\,. \tag{30}$$

For $r > r_0$, the constants $B^{(r)}$ and $C^{(r)}$ rapidly approach zero and the constants $A^{(r)}$ and $D^{(r)}$ tend to finite asymptotic values approximately given by

$$A^{(r)} \approx K_1(\epsilon)\lambda^{r_0} \quad \text{for } r > r_0, \tag{31a}$$

$$D^{(r)} \approx A^{(r)2} \quad \text{for } r > r_0. \tag{31b}$$

Here $K_1(\epsilon)$ is again a constant of proportionality which depends on ϵ, but is independent of δ. We substitute these values from Eqs. (31) and (27) in Eq. (15), and approximating the sum by its largest term we see that

$$C(x) \sim K\frac{\lambda^{2r_0}}{3^{r_0}}\,. \tag{32}$$

Substituting for r_0 from Eq. (25) we get

$$C(x) \sim K\left(\frac{\epsilon}{\delta}\right)^{\gamma} \tag{33}$$

with

$$\gamma = \ln(\lambda^2/3)/\ln(2+\mu^{-2}) \approx 1.3752. \tag{34}$$

In Eq. (33), the constant of proportionality K must vary as $\epsilon^{-\gamma}$; so that $C(x)$ is independent of ϵ as should be obvious from its definition.

The critical exponent ν may be determined similarly. We note that for $r < r_0$, the contribution of the rth order open loops to $R(x)$ is approximately

$$(2^r)^2K\lambda^{2r}/3^r.$$

For $r > r_0$, the coefficient $B^{(r)}$ rapidly become zero. And in a configuration of the type $A^{(r)}$ (Fig. 1) the endpoint of the line stays close to the corner vertex from which it entered the triangle. Thus for $r > r_0$, the contribution of the rth order open loops to $R(x)$ varies as $K(4\lambda^2)^{r_0}3^{-r}$ so that we have

$$R(x) \sim K\{(4\lambda)^2/3\}^{r_0}\,. \tag{35}$$

Substituting for r_0 from Eq. (25) and comparing its dependence on δ with Eq. (11) we get

$$\nu = \frac{\ln 2}{\ln(2+\mu^{-2})} \approx 0.7986. \tag{36}$$

This determines all the critical exponents α, γ, ν. We remark here that though a more complete and rigorous analysis of the recursion equations is certainly possible, it is unnecessary since all the constants μ, α, γ, and ν are determined exactly.

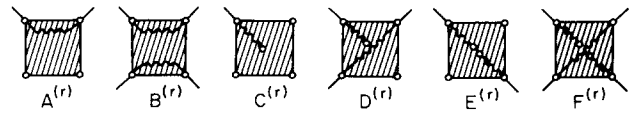

FIG. 4. The restricted partitions for an rth order square. Only the corner vertices of the square and the connections of the walks joining them are shown.

B. The truncated 4-simplex lattice

The analysis for this lattice is very similar to that of the truncated 3-simplex lattice discussion in A. We define the restricted partition functions $A^{(r)}$, $B^{(r)}$, $C^{(r)}$, $D^{(r)}$, $E^{(r)}$, and $F^{(r)}$, for the rth order square, as shown in Fig. 4. Due to the permutation symmetry between the vertices of the rth order square, only six different restricted portion functions are needed. The starting values of these weights are

$$A^{(0)} = x, \tag{37a}$$

$$C^{(0)} = \sqrt{x}, \tag{37b}$$

$$B^{(0)} = D^{(0)} = E^{(0)} = F^{(0)} = 0. \tag{37c}$$

The recursion equations for these weights are written down by constructing graphically all possible ways an $(r+1)$th order square may be constructed out of its constituting rth order squares. We get

$$A^{(r+1)} = A^2 + 2A^3 + 2A^4 + 4A^3B + 6A^2B^2\,, \tag{38a}$$

$$B^{(r+1)} = A^4 + 4A^3B + 22B^4, \tag{38b}$$

$$C^{(r+1)} = C(1+3A+6A^2+6A^3+6A^2B) + D(6A^2+12A^3 + 18A^2B), \tag{38c}$$

$$D^{(r+1)} = C(A^2+2A^3+3A^2B) + D(3A^2+7A^3+16A^2B + 22B^2A+22B^3), \tag{38d}$$

where again the superscripts (r) on each term in the right-hand side of Eq. (38) have been suppressed. We have not written down the explicit expressions for $E^{(r+1)}$ and $F^{(r+1)}$ because, as with Eq. (16d), they are not needed for the determination of critical exponents. The expressions for $E^{(r+1)}$ and $F^{(r+1)}$ involve a sum of terms that are linear in $E^{(r)}$ and $F^{(r)}$ but independent of $C^{(r)}$ and $D^{(r)}$, and terms that are quadratic in $C^{(r)}$ and $D^{(r)}$ but independent of $E^{(r)}$ and $F^{(r)}$. The asymptotic behavior of $E^{(r)}$ and $F^{(r)}$ is determined by the rate at which $C^{(r)}$ and $D^{(r)}$ grow with r.

Explicit expression for $P(x)$ is easily written down,

$$P(x)=\sum_{r=1}^{\infty}4^{-r}\{4A^{(r-1)3}+3A^{(r-1)4}\}, \tag{39}$$

and a similar but more complicated expression for $C(x)$. Again we notice that the recursion equations are quite simple. $A^{(r+1)}$ and $B^{(r+1)}$ depend only on $A^{(r)}$ and $B^{(r)}$. For any given value of x, we can determine $P(x)$ with the help of Eqs. (37)–(39). If x is less than $1/\mu$, where μ is now the connective constant for this lattice, $A^{(r)}$ and $B^{(r)}$ tend to zero for large r, and $P(x)$ is finite. The reverse is the case if $x>\mu^{-1}$. This allows a very easy determination of μ numerically. We start with the obvious bounds $x_{\min}<1/\mu<x_{\max}$ with $x_{\min}=\frac{1}{3}$ and $x_{\max}=1$. Determination of the behavior of $P(x)$ for $x=(x_{\min}+x_{\max})/2$ allows us to reduce the range of uncertainty $(x_{\max}-x_{\min})$ by a factor of 2. The procedure may be repeated till any arbitrary desired accuracy is obtained. Numerically, we find that

$$\mu\cong 2.2866. \tag{40}$$

Equations (38a) and (38b) have a fixed point given by

$$A^*\cong 0.4294 \quad \text{and} \quad B^*\cong 0.04998. \tag{41}$$

There are other fixed points of these recursion equations. The fixed points $A^*=B^*=0$ and $A^*=B^*=\infty$ are the trivial attractive fixed points. Other fixed points of Eqs. (38a) and (38b) are not relevant for the determination of the critical behavior because if we start with $x=1/\mu$, the successive values of $A^{(r)}$ and $B^{(r)}$ tend to the fixed point given by Eq. (41) and hence this is the fixed point which determines the critical behavior of the generating functions $C(x)$, $P(x)$, and $R(x)$.

Let us write

$$A^{(r)}=A^*+\delta A^{(r)}, \tag{42a}$$

$$B^{(r)}=B^*+\delta B^{(r)}. \tag{42b}$$

Then to first order in δA and δB, we have the linearized recursion equations

$$\delta A^{(r+1)}=T_{11}\delta A^{(r)}+T_{12}\delta B^{(r)}, \tag{43a}$$

$$\delta B^{(r+1)}=T_{21}\delta A^{(r)}+T_{22}\delta B^{(r)}. \tag{43b}$$

Here T_{ij} are the elements of a (2×2) matrix whose value in terms of A^* and B^* is easily written down, and its eigenvalues and eigenvectors determined. We find that the matrix T has the eigenvalues

$$\lambda_1\cong 2.7965, \tag{44a}$$

$$\lambda_2\cong 0.2538. \tag{44b}$$

We note that only one of the eigenvalues is greater than zero. This implies that

$$\delta A^{(r)}\cong Ka_1\lambda_1^r\delta, \tag{45a}$$

$$\delta B^{(r)}\cong Ka_2\lambda_1^r\delta, \tag{45b}$$

where (a_1, a_2) is the eigenvector corresponding to the eigenvalue λ_1. δ is defined equal to $(\mu^{-1}-x)$ and K is some constant of proportionality same for Eqs. (45a) and (45b). These equations hold when r is not too large (so that the linear approximation is adequate) and not too small (so that the "irrelevant" part of δA and δB, which is proportional to λ_2 is negligible). Let us denote by $P^{\text{sing}}(1/\mu-\delta)$ the singular part of $P(1/\mu-\delta)$. Then from Eq. (39) we get

$$P^{\text{sing}}\left(\frac{1}{\mu}-\delta\right)\approx\tfrac{1}{4}P^{\text{sing}}\left(\frac{1}{\mu}-\lambda_1\delta\right). \tag{46}$$

Assuming that

$$P^{\text{sing}}\left(\frac{1}{\mu}-\delta\right)\sim\delta^{2-\alpha}$$

we get

$$\alpha=2-\ln 4/\ln\lambda_1\simeq 0.5413. \tag{47}$$

Again we choose a small positive number ϵ, and choose a value of x sufficiently close to μ^{-1} so that

$$1\gg\epsilon\gg\delta \tag{48}$$

and

$$r_0\overset{\text{def}}{=}\ln(\epsilon/\delta)/\ln\lambda_1. \tag{49}$$

Then for $r<r_0$ we have $A^{(r)}\approx A^*$, $B^{(r)}\approx B^*$ and the recursion equations for $C^{(r)}$ and $D^{(r)}$ become, from Eqs. (38c) and (38d),

$$C^{(r+1)}\approx C^{(r)}(1+3A^*+6A^{*2}+6A^{*3}+6A^{*2}B^*+D^{(r)}(6A^{*2}+12A^{*3}+18A^{*2}B^*), \tag{50a}$$

$$D^{(r+1)}\approx C^{(r)}(A^{*2}+2A^{*3}+3A^{*2}B^*)+D^{(r)}(3A^{*2}+7A^{*3}+16A^{*2}B^*+22B^{*2}A^*+22B^{*3}). \tag{50b}$$

These are linear recursion equations and show that, for $r<r_0$, $C^{(r)}$ and $D^{(r)}$ increase as λ_+^r, where λ_+ is the larger eigenvalue of the matrix that characterizes the linear transformation of Eq. (50). Numerically, substituting the values of A^* and B^* we find that

$$\lambda_+\cong 4.2069. \tag{51}$$

For $r>r_0$, the coefficients $A^{(r)}$, $B^{(r)}$, and $D^{(r)}$ rapidly approach zero and $C^{(r)}$ tends to its asymptotic value which is proportional to $\lambda_+^{r_0}$. The argument as before shows that

$$C(x)\sim K(\lambda_+^2/4)^{r_0}. \tag{52}$$

Substituting for r_0 from Eq. (49) and comparing with Eq. (7) we get

$$\gamma=\ln(\lambda_+^2/4)/\ln(\lambda_1)\approx 1.4461. \tag{53}$$

And since

$$R(x)\sim K(2^{r_0})^2(\lambda_+^2/4)^{r_0} \tag{54}$$

we get from Eqs. (49) and (9)

$$\nu=\ln 2/\ln\lambda_1\approx 0.7294. \tag{55}$$

C. The modified rectangular lattice

The restricted partition functions for the rth order block[13] of lattice sites are defined in Fig. 5. We have shown only the configuration with no or only one end point of the walk. As before, the restricted partition functions with two end points of restricted walks, may be defined and are necessary to calculate the generating

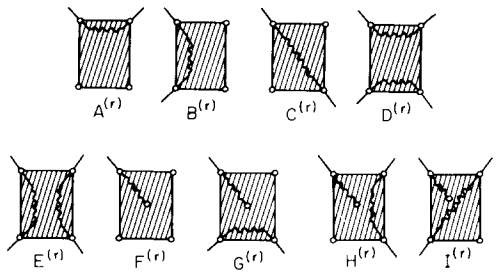

FIG. 5. Restricted partition functions for the rth order block of lattice sites of a modified rectangular lattice. Only the restricted partition functions correspondings to zero or one end points of the self avoiding walk are shown.

function $C(x)$, but do not affect the critical exponents. The recursion equations are complicated by the fact that the number of coupling constants is much larger than in the previous two cases. This is, of course, more representative of real life renormalization calculations where the total number of coupling constants is in principle infinite.

The starting values of the different weights are

$$A^{(1)} = B^{(1)} = x^2 + x^4, \tag{56a}$$

$$C^{(1)} = 2x^3, \tag{56b}$$

$$D^{(1)} = E^{(1)} = x^4, \tag{56c}$$

$$F^{(1)} = x^{1/2}[1 + 2x + 2x^2 + 2x^3], \tag{56d}$$

$$G^{(1)} = H^{(1)} = x^{5/2}(1 + x), \tag{56e}$$

$$I^{(1)} = x^{7/2}. \tag{56f}$$

Again we may write down the recursion equations for these weights by drawing all possible configurations of self avoiding random walks on the $(r+1)$th order block. Figure 6 shows the configurations that contribute to $C^{(r+1)}$. Other recursion equations are similarly written down. We get

$$A^{(r+1)} = B(1 + D), \tag{57a}$$

$$B^{(r+1)} = A^2 + C^2, \tag{57b}$$

$$C^{(r+1)} = 2AC, \tag{57c}$$

$$D^{(r+1)} = B^2 + 2DE\ , \tag{57d}$$

$$E^{(r+1)} = D^2 \tag{57e}$$

$$F^{(r+1)} = F(1 + A + C) + GB + IB, \tag{57f}$$

$$G^{(r+1)} = FB + GE + IE + H(D + A + C), \tag{57g}$$

$$H^{(r+1)} = G(A + D) + IC, \tag{57h}$$

$$I^{(r+1)} = GC + I(A + D), \tag{57i}$$

where again we have suppressed the superscripts (r) in the right-hand sides of Eq. (57). We have quite a simple expression of $P(x)$,

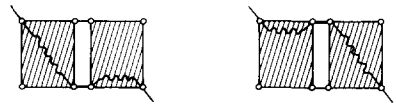

FIG. 6. Configurations of random walks that contribute to $C^{(r+1)}$ for the modified rectangular lattice. The shaded rectangles are the rth order blocks with only the corner vertices and the walk going through them shown.

$$P(x) = \sum_{r=1}^{\infty} (B^{(r)})^2 / 2^{r+2} + \frac{x^4}{4}\ . \tag{58}$$

Numerically, we find the largest value of x for $P(x)$ to converge is

$$x_c \cong 0.5914 \tag{59}$$

which corresponds to the connective constant

$$\mu = 1/x_c \approx 1.6909. \tag{60}$$

If we start with the value $x = \mu^{-1}$, the sequence $(A^{(r)}, B^{(r)}, C^{(r)}, D^{(r)}, E^{(r)})$ converge to the fixed point

$$(0.5000,\ 0.4201,\ 0.4124,\ 0.1902,\ 0.0362). \tag{61}$$

Linearizing the recursion equations about this fixed point, we find the eigenvalues of the transformation matrix. Only one eigenvalue is larger than 1. Numerically, its value is

$$\lambda_1 \cong 1.6839. \tag{62}$$

Thus arguing as before, we find

$$\alpha = 2 - \frac{\ln 2}{\ln \lambda_1} \approx 0.6699, \tag{63}$$

Substituting for $A^{(r)}$, $B^{(r)}$, $C^{(r)}$, $D^{(r)}$, and $E^{(r)}$ the fixed point values given by (61) in the recursion equations (57f)—(57i), we find numerically, the largest eigenvalue of the corresponding transformation matrix. This is found to be

$$\lambda_+ \approx 2.0582. \tag{64}$$

Then analogous to Eq. (53) we find

$$\gamma = \ln(\lambda_+^2/2)/\ln\lambda_1 \approx 1.4403, \tag{65}$$

The diameter of an $(r+2)$th order block is twice the diameter of an rth order block, the diameter being defined as the largest distance between any two points in the block. Thus, arguing as for the truncated 4-simplex lattice, we get

$$\nu = (\ln 2)/2\ln\lambda_1 \approx 0.6650 \tag{66}$$

which determines all the critical exponents for the lattice.

IV. DISCUSSION

We note that in all the cases considered, the critical indices satisfy the relation

$$d\nu = 2 - \alpha, \tag{67}$$

where d is the dimensionality of the lattice as defined by Nelson and Fisher. On the other hand, we know that in other cases (e.g., the XY model discussed in I) the dimensionality of the lattice is more usefully defined by the power law behavior of the cumulative frequency

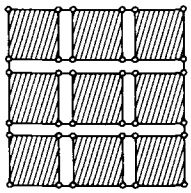

FIG. 7. A generalization of the modified rectangular lattice. The figure shows the graph of an $(r+1)$th order rectangle. Shaded rectangles denote rth order graphs of which only the corner vertices are shown. The case depicted is $p=3$.

distribution function for low frequencies. Perhaps the self avoiding random walks are atypical in that the generating functions whose singularities determine the critical exponents are not given in terms of the partition function of a Hamiltonian. The critical behavior of the self avoiding random walks depends strongly on the connectivity properties of the lattice and not on the dimensionality alone. For example, we expect the critical indices α, γ, ν to be different for the self avoiding walks for planar and nonplanar two-dimensional lattices. This is because planarity determines if the walk can cross itself or not. On the other hand, the critical exponents for the Ising model in two dimensions are expected to remain unchanged if a small next-nearest neighbor interaction is added to the original nearest neighbor Hamiltonian (which makes the lattice nonplanar).

It is possible to construct pseudolattices that have the same effective dimensionality, but different critical exponents for the self avoiding walk problem. Consider, for example, the lattice defined in Fig. 7. This is a simple generalization of the modified rectangular lattice. The first order rectangle is a cyclic graph on four points. An $(r+1)$th order rectangle is formed by taking p^2 rth order rectangles (p is any integer >1) and arranging them in a $p\times p$ array. We connect the rectangles in the same row by horizontal bonds connecting the corner vertices of adjacent rectangles in the same adjacent rectangles in the same row. Finally $2(p-1)$ vertical bonds are added to connect the corner vertices of adjacent rows. In Fig. 7 the construction is illustrated for the case $p=3$.

It is easy to see that the lattice is planar and has coordination number 3. Using the same method as used to determine the effective dimensionality of the modified rectangular lattice in I, it may be shown that the effective dimensionality of the lattice is $\frac{3}{2}$, independent of p. The modified rectangular lattice corresponds to the special case $p=2$. Also, we see that the dimensionality of the lattice is 2 (again independent of p) if we use Nelson and Fisher's definition. Though the coordination number and the dimensionality of the lattice is independent of p (whichever definition of dimensionality is used), it is easy to verify that the critical exponents for the self avoiding walks on these lattices do depend on p. In particular for $p=3$ we find

$$\alpha=0.6589,\quad \gamma=1.4601,\quad \nu=0.6705 \tag{68}$$

which differs from the exponents for $p=2$ (Eq. (12b).

It appears that the connectivity structure of these lattices is quite complicated, and a single value of "effective dimensionality" is not sufficient to completely characterize the critical behavior of self avoiding walks on such lattices. More study in this area is needed to identify the parameters that can be used to completely characterize the critical behavior of different systems on such pseudolattices.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor Jon Mathews for many discusssions and critical comments, and Ms. RLou Norquist for typing the manuscript.

[1]E.W. Montroll, J. Chem. Phys. **18**, 734 (1950).
[2]M.E. Fisher and M.F. Sykes, Phys. Rev. **114**, 45 (1959).
[3]D.S. McKenzie, Phys. Rep. C **27**, 37 (1976); and references cited therein.
[4]S.G. Whittington and J.F. Harris, J. Phys. A **5**, 411 (1972); C. Domb and F.T. Hioe, J. Chem. Phys. **51**, 1920 (1969).
[5]M.F. Sykes *et al.*, J. Phys. A: Gen. Phys. **5**, 653 (1972); F.T. Wall and F. Mandel, J. Chem. Phys. **63**, 4592 (1975).
[6]E. Muller-Hartmann and J. Zittartz, Phys. Rev. Lett. **33**, 893 (1976).
[7]J. Zittartz, Z. Phys. Chem. Chem. Abt. B **23**, 55—63 (1976).
[8]K. Wilson and J. Kogut, Phys. Rep. C **12**, 75 (1974).
[9]This is discussed in Ref. 8.
[10]G. Gallavotti and H. Knopps, Nuovo Cimento **5**, 341 (1975).
[11]D.R. Nelson and M.E. Fisher, Ann. Phys. (N.Y.) **91**, 226 (1975).
[12]D. Dhar, J. Math. Phys. **18**, 577 (1977).
[13]See page 581 of Ref. 12 for the definition of a block of sites.

On projective symmetries of dynamical systems[a)]

Toshihiro Iwai

Department of Applied Mathematics, Gifu University, Kagamigahara, Gifu 504, Japan
(Received 4 August 1977)

The present paper is concerned with symmetry transformations of a dynamical system defined on the tangent bundle of a Riemannian manifold. Of present interest are infinitesimal symmetry transformations of the vector field which defines the dynamical system on the tangent bundle. It is known that a class of such transformations entails infinitesimal projective transformations leaving the vector field invariant. Symmetry algebras formed by such projective transformations are studied. It is shown which dynamical systems admit large symmetry algebras. As a result, two kinds of dynamical systems are determined, which have the base Riemannian manifolds of constant curvature with dimensions $n \geq 4$. The systems are generalizations of the classical harmonic oscillator and Kepler problem usually considered in Euclidean spaces. First integrals quadratic in the velocities are obtained, which are also generalizations of the well-known quadratic integrals for the above classical systems.

1. INTRODUCTION

It is generally recognized that symmetry plays a fundamental role in studying particular aspects of physical systems. A great number of investigations have been carried on into symmetries of mechanical systems. One of the approaches to the subject is based on the symplectic structure (or the Poisson bracket) of the Hamiltonian formalism, in which the relation between dynamical symmetries and constants of the motion is well established.

Recently, there have been many attempts to find symmetries of dynamical systems not restricted within the Hamiltonian formalism, with or without reference to constants of the motion. Katzin and Levine maintain, in their recent works,[1-7] that a dynamical symmetry is a transformation which maps the set of all dynamical paths into itself. Under their view, they set up a "Related integral theorem" which provides a means for deriving new constants of the motion from the old.

In a previous paper,[8] the author reformulated in a more geometric way the idea of symmetry mappings due to Katzin and Levine. That is, a vector field is defined on the tangent bundle of a Riemannian manifold as a geometric object in the place of the equations of motion. A symmetry is then thought of as a certain automorphism for the vector field. In this setting, the related integral theorem is an immediate consequence of the symmetry.[8] In the same paper the conditions for an infinitesimal trajectory collineation were determined by a method different from Katzin's.[2]

This article is a continuation of the previous paper and shows what dynamical systems admit large symmetry algebras (i.e., Lie algebras consisting of infinitesimal trajectory collineations). Katzin[2] was interested in the opposite situation, namely, in the problem of determining the symmetry algebras for given dynamical systems, especially, the familar harmonic oscillator and Kepler problem.

Let M be an n-dimensional Riemannian manifold endowed with a positive-definite metric tensor g and $T(M)$ the tangent bundle of M. Let (x^i) [9] be a local coordinate system in M and (x^i, v^i) the induced coordinate system in $T(M)$. Given a vector field V on M with local components (V^i), one may define a vector field Z on $T(M)$ by

$$Z = v^i \frac{\partial}{\partial x^i} - \begin{Bmatrix} k \\ ij \end{Bmatrix} v^i v^j \frac{\partial}{\partial v^k} + V^i \frac{\partial}{\partial v^i}, \tag{1.1}$$

where the $\{{}^{k}_{ij}\}$'s denote the Christoffel symbols. It is easily verified that (1.1) is independent of the choice of the local coordinate system. Since integral curves of Z determine dynamical trajectories when projected on M, a couple $(T(M), Z)$ may be called a *(Newtonian) dynamical system*. The vector field Z is a fundamental object in place of the equations of motion.

In order to deal with path invariances (or trajectory collineations) of the system, it is convenient to introduce an infinitesimal transformation $\tilde{X}$ of $T(M)$ which is raised from an infinitesimal transformation $X = (\xi^i)$ of M through

$$\tilde{X} = \xi^i \frac{\partial}{\partial x^i} + \frac{\partial \xi^i}{\partial x^j} v^j \frac{\partial}{\partial v^i} - 2\psi v^i \frac{\partial}{\partial v^i},$$
$$\psi = \frac{1}{n+1} \operatorname{div} X = \frac{1}{n+1} \nabla_i \xi^i, \tag{1.2}$$

where ∇_i denotes the covariant derivative with respect to the $\{{}^{k}_{ij}\}$. In this setting, a path invariance can be put in the form

$$[\tilde{X}, Z] \equiv 0 \pmod{Z}, \tag{1.3}$$

where $[\ ,\ ]$ denotes the bracket operation. We may call $\tilde{X}$ an *infinitesimal projective symmetry* of the dynamical system $(T(M), Z)$, when (1.3) is satisfied. Equation (1.3) in turn entails the following equations,[10]

$$\mathcal{L}_X \{{}^{k}_{ij}\} = \psi_i \delta_j{}^k + \psi_j \delta_i{}^k,$$
$$[X, V] = -4\psi V, \tag{1.4}$$

where $\mathcal{L}_X$ denotes the Lie derivative with respect to X and $\psi_i = \partial\psi/\partial x^i$ (see Ref. 8, for details). Since the mapping $X \vdash\!\!\rightarrow \tilde{X}$ is a Lie algebra homomorphism (i.e., $[X, Y] \vdash\!\!\rightarrow [\tilde{X}, \tilde{Y}]$), the Lie algebra of X's satisfying (1.4) may be called a *symmetry algebra* of the dynamical system $(T(M), Z)$.

[a)] The main results of this paper are contained in a doctoral thesis (unpublished) submitted by Kyoto University.

 0022-2488/78/1901-0012$1.00

Equations (1.4) are to be investigated in the following sections. Before working with (1.4), it seems appropriate to point out that (1.4) contains the subordinate conditions

$$\mathcal{L}_X g_{ij}=0, \quad [X,V]=0. \tag{1.5}$$

This has been already discussed in Ref. 11. We shall study (1.4) on the basis of the results obtained there.

Section 2 deals with the integrability conditions of (1.4) and evaluates the maximum possible dimensions of the symmetry algebra determined by (1.4).

Section 3 is concerned with maximal symmetry algebras having maximal subalgebra defined by (1.5). It is shown that if M admits a maximal subalgebra determined by (1.5), then the maximum dimensions of the symmetry algebra evaluated in Sec. 2 are not attained.

Section 4 is devoted to the case where M admits a lower dimensional Lie algebra determined by (1.5). In this case, there is a function U in (an open subset of) M, referred to as a *potential*, such that $V=-\mathrm{grad}\,U$.[11] Furthermore, (1.5) contains the restricted conditions

$$\mathcal{L}_X g_{ij}=0, \quad XU=0. \tag{1.6}$$

Symmetry algebras determined by (1.4) are discussed under the condition that they have subalgebras defined by (1.6). Two kinds of marked symmetry algebras are determined on a space of constant curvature ($n\geq 4$).

In Section 5 dynamical systems admitting the symmetry algebras obtained in Sec. 4 are discussed. These dynamical systems are considered as generalizations of the harmonic oscillator and the Kepler problem usually treated in the Euclidean space. First integrals of the dynamical systems are obtained by the related integral theorem, which are also generalizations of the familiar quadratic integrals. Katzin[2] derived (1.4) for a generic Riemannian space M but applied the equations only to the case of $M=\mathbb{R}^3$. He also applied the related integral theorem to obtain the familiar quadratic first integrals.[2]

It is to be understood throughout this paper that all manifolds, vector fields, functions, etc., introduced are tacitly assumed to be of class C^∞.

2. INTEGRABILITY CONDITIONS

Equations (1.4) can be put into a normal form,

$$\begin{aligned}
&\nabla_i \xi^j=\xi_i{}^j,\\
&\nabla_k \xi_i{}^j=-R_{lki}{}^j\xi^l+\psi_k\delta_i{}^j+\psi_i\delta_k{}^j,\\
&\nabla_j\psi_k=-\frac{1}{n-1}(\xi^l\nabla_l R_{jk}+R_{lk}\xi_j{}^l+R_{jl}\xi_k{}^l,
\end{aligned} \tag{2.1}$$

$$\xi^k\nabla_k V^i-\xi_k{}^l\left(\delta_l{}^i V^k-\frac{4}{n+1}\right)\delta_l{}^k V^i=0, \tag{2.2}$$

where $(R_{lki}{}^j)$ and (R_{jk}) are the curvature and Ricci tensors, respectively. Equations (2.1) are partial differential equations in n^2+2n unknowns $(\xi^i,\xi_j{}^k,\psi_l)$ with the constraints (2.2). If V vanishes identically, the integral curves of the dynamical systems $(T(M),Z)$ reduce to geodesics, and hence (1.4) merely shows that X is an infinitesimal projective transformation without constraints. For this reason we assume that the vector field V does not vanish identically but may have isolated zeros.

The first integrability conditions of (2.1) are written in the form

$$\begin{aligned}
&\xi^l\nabla_l W_{jmk}{}^i+\xi_j{}^l W_{lmk}{}^i+\xi_m{}^l W_{jlk}{}^i+\xi_k{}^l W_{jml}{}^i-\xi_l{}^i W_{jmk}{}^l=0,\\
&\xi^l\nabla_l W_{ijk}+\xi_i{}^l W_{ljk}+\xi_j{}^l W_{ilk}+\xi_k{}^l W_{ijl}+W_{ijk}{}^l\psi_l=0,
\end{aligned} \tag{2.3}$$

where $W_{ijk}{}^l$ and W_{ijk} are, respectively, defined by

$$W_{ijk}{}^l=R_{ijk}{}^l-\frac{1}{n-1}(\delta_i{}^l R_{jk}-\delta_j{}^l R_{ik}), \tag{2.4}$$

$$W_{ijk}=-\frac{1}{n-1}(\nabla_i R_{jk}-\nabla_j R_{ik}). \tag{2.5}$$

The covariant derivatives of (2.2) take the form

$$\begin{aligned}
&\xi^l\nabla_l\nabla_j V^i+\xi_l{}^k(\delta_j{}^l\nabla_k V^i-\delta_k{}^i\nabla_j V^l+\frac{4}{n+1}\delta_k{}^l\nabla_j V^i)\\
&+\psi_k(3\delta_j{}^k V^i-\delta_j{}^i V^k)=0.
\end{aligned} \tag{2.6}$$

We now show that (2.6) is independent of (2.2) and hence gives independent constraints on $(\xi^i,\xi_j{}^k,\psi_l)$. In fact, if (2.6) depends on (2.2), each equation of (2.6) is a linear combination of the equations of (2.2). Then $0=3\delta_j{}^k V^i-\delta_j{}^i V^k$. By contracting this with respect to k and j, we get $(3n-1)V^i=0$, so that $V=0$, contradicting the assumption.

Theorem 2.1: Symmetry algebras determined by (1.4) are at most of dimension n^2.

Proof: Assume that the first integrability conditions (2.3) are valid for $(\xi^i,\xi_j{}^k,\psi_l)$ satisfying (2.2) and (2.6). Define the coefficient of $\xi_k{}^l$ in (2.2) and of ψ_k in (2.6) to be $S^i{}_l{}^k$ and $T^k{}_j{}^i$ respectively,

$$S^i{}_l{}^k=\delta_l{}^i V^k-\frac{4}{n+1}\delta_l{}^k V^i, \quad T^k{}_j{}^i=3\delta_j{}^k V^i-\delta_j{}^i V^k.$$

Let $S^i{}_l{}^k=S^i{}_A$ and $T^k{}_j{}^i=T^k{}_B$, where $A=\binom{k}{l}$ and $B=\binom{i}{j}$ are double indices. By choosing a local coordinate system such that $(V^i)=(1,0,\ldots,0)$ and writting out nonzero elements of the (n,n^2) matrices $(S^i{}_A)$ and $(T^k{}_B)$, we find that these matrices are both of rank n. Thus (2.2) and (2.6) contain $2n$ linearly independent equations at least, and consequently Eqs. (2.1) with (2.2) admit $n^2(=n^2+2n-2n)$ linearly independent solutions at most. This completes the proof.

Remark: Theorem 2.1 also holds good even if the second equation in (1.4) is replaced by $[X,V]=-2\psi V$ (see Theorem 3.2). In this case we have a Lie algebra which does not come from the path invariance of the dynamical system $(T(M),Z)$.

We now refer to a theorem, due to Egorov,[12] with regard to infinitesimal projective transformations.

Proposition 2.2: If an n-dimensional manifold with an affine connection has no torsion and if it admits an effective[13] group of projective transformations of order greater than n^2-2n+5, then it is projectively flat.

The following theorem is an immediate consequence of this proposition.

Theorem 2.3: If an n-dimensional Riemannian manifold M admits the maximal Lie algebra stated in Theorem 2.1 and if $n\geq 3$, then M is of constant curvature.

Proof: By Proposition 2.2 a Riemannian manifold M ($n \geq 3$) admitting the maximal Lie algebra stated in Theorem 2.1 is projectively flat and hence of constant curvature. This is consistent with (2.3), because $W_{ijk}{}^{l}$ and W_{ijk} vanish on M ($n \geq 3$) of constant curvature. This ends the proof.

3. THE CASE WHERE M ADMITS A MAXIMAL SUBALGEBRA OF INFINITESIMAL ISOMETRIES PRESERVING A VECTOR FIELD

As mentioned in Sec. 1, (1.4) contains the condition (1.5) for X to be an infinitesimal isometry which preserves a vector field V. Structures of M's which admit maximal Lie algebras (being of dimension $n(n-1)/2+1$) determined by (1.5) have been specified in Ref. 11. There it was shown that, if $n \geq 4$, M is locally isometric to either a Riemannian manifold of constant negative curvature or the direct product of $\mathbb{R}$ and an $(n-1)$-dimensional manifold of constant curvature. The latter cannot be of constant curvature unless it is flat. In view of this and Theorem 2.3, we are interested in manifolds ($n \geq 4$) of constant nonpositive curvature.

The following proposition is well known about infinitesimal projective transformations (see Ref. 14, for example).

Proposition 3.1: If an Einstein manifold with nonvanishing scalar curvature admits an infinitesimal projective transformation X, then X is uniquely decomposed into

$$X = Y - (1/2K)\mathrm{grad}\psi, \tag{3.1}$$

where Y is a Killing vector field, K is the sectional curvature, and $\mathrm{grad}\psi$ is a gradient vector field defining the infinitesimal projective transformation.[15]

Now we consider spaces of constant nonpositive curvature which admit maximal Lie algebras determined by (1.5).[11] First, let M be a space of constant negative curvature with the sectional curvature $K=-1$. We choose a coordinate system such that

$$ds^2 = \sum_{i=1}^{n} (dx^i)^2/(x^n)^2. \tag{3.2}$$

According to Proposition 3.1, there is such a function ψ that $\frac{1}{2}\mathrm{grad}\psi$ gives a set of infinitesimal projective transformations of M. By substituting $\frac{1}{2}\mathrm{grad}\psi$ for X in the first equations of (1.4) and solving the resulting equations, we can obtain

$$\psi = [1/(x^n)^2](A(x^\lambda) + b_\lambda x^\lambda r^2 + cr^4] + c',^{16} \tag{3.3}$$

where

$$A(x^\lambda) = \tfrac{1}{2}a_{\lambda\mu}x^\lambda x^\mu + a_\lambda x^\lambda + a \quad (a_{\lambda\mu} = a_{\mu\lambda}), \tag{3.4}$$

$r^2 = \sum_{i=1}^{n}(x^i)^2$, and $a_{\lambda\mu}$, a_λ, a, b_λ, c, and c' are constants.

For the space of constant negative curvature (3.2), a basis of the Lie algebra determined by (1.5) consists of

$$x^i \frac{\partial}{\partial x^i}, \quad x^\lambda \frac{\partial}{\partial x^\mu} - x^\mu \frac{\partial}{\partial x^\lambda}, \quad \frac{\partial}{\partial x^\nu}. \tag{3.5}$$

This Lie algebra is isomorphic with the Lie algebra of the Lorentz group $O(n,1)$. The preserved vector field is given by

$$V = x^n \frac{\partial}{\partial x^n} \tag{3.6}$$

within a constant multiple.

In case M is flat $[ds^2 = \sum (dx^\lambda)^2 + (dx^n)^2]$, the Lie algebra determined by (1.5) has a basis consisting of

$$\frac{\partial}{\partial x^\nu}, \quad x^\lambda \frac{\partial}{\partial x^\mu} - x^\mu \frac{\partial}{\partial x^\lambda}, \quad \frac{\partial}{\partial x^n}. \tag{3.7}$$

The preserved vector field is then given by

$$V = \frac{\partial}{\partial x^n}. \tag{3.8}$$

We are now in a position to gain insight into the Lie algebra of the maximum dimension n^2 stated in Theorem 2.1.

Theorem 3.2: If an n-dimensional Riemannian manifold M admits a maximal subalgebra defined by (1.5), then M ($n \geq 4$) admits no symmetry algebras of dimension n^2 determined by (1.4). However, in case the second equation in (1.4) is replaced by $[X, V] = -2\psi V$, M ($n \geq 4$) admits a Lie algebra of dimension n^2.

Proof: As is written above, under the condition of the theorem, we have already obtained preserved vector fields and subalgebras of all Killing vector fields satisfying (1.5) in the both cases where M is of constant negative curvature and where M is flat. Accordingly, we need only to try to look for the infinitesimal projective transformations which satisfy (1.4) but are not Killing vector fields. First we suppose that M is of constant negative curvature, so that, according to (3.1), we devote ourselves to considering gradient vector fields. The function (3.3) with $c' = 0$ satisfies

$$\frac{1}{n+1}\mathrm{div}(\tfrac{1}{2}\mathrm{grad}\psi) = \frac{1}{2(n+1)}\mathrm{tr}(a_{\lambda\mu}) + \psi, \tag{3.9}$$

where $\mathrm{tr}(a_{\lambda\mu}) = \sum a_{\lambda\lambda}$; this may be verified in a straightforward manner. On the other hand, by expanding out the bracket of $\frac{1}{2}\mathrm{grad}\psi$ and V given by (3.6), we see that V is preserved by the infinitesimal transformation $\frac{1}{2}\mathrm{grad}\psi$ if and only if $b_\lambda = c = 0$. That is, under the conditions $b_\lambda = c = 0$, V satisfies

$$[\tfrac{1}{2}\mathrm{grad}\psi, V] = -2\psi V. \tag{3.10}$$

From (3.9) and (3.10) it follows that if $\mathrm{tr}(a_{\lambda\mu}) = b_\lambda = c = 0$ then

$$[\tfrac{1}{2}\mathrm{grad}\psi, V] = -\frac{2}{n+1}[\mathrm{div}(\tfrac{1}{2}\mathrm{grad}\psi)]V. \tag{3.11}$$

Since $\frac{1}{2}\mathrm{grad}\psi$ with $\mathrm{tr}(a_{\lambda\mu}) = b_\lambda = c = 0$ gives rise to $n(n+1)/2 - 1$ linearly independent fields, the linearly independent infinitesimal projective transformations we have obtained amount to n^2 in number; there are $n(n-1)/2+1$ linearly independent Killing vector fields and $n(n+1)/2-1$ linearly independent gradient vector fields. This proves the theorem for M of constant negative curvature. We suppose in turn that M is flat. As is well known, a basis of the infinitesimal projective transformations which are not Killing vector fields consists, in the standard coordinate system, of

$$x^i \frac{\partial}{\partial x^j} + x^j \frac{\partial}{\partial x^i}, \quad x^i x^k \frac{\partial}{\partial x^k}. \tag{3.12}$$

By substituting (3.12) and (3.8) for X and V in (1.4),

respectively, we can easily prove that there are no symmetry algebras of dimension n^2 satisfying (1.4). This completes the proof.

4. THE CASE WHERE M ADMITS A LOWER DIMENSIONAL SUBALGEBRA OF INFINITESIMAL ISOMETRIES PRESERVING A VECTOR FIELD

This section is devoted to a study of the Lie algebra defined by (1.4) under the condition that (1.5) determines a subalgebra of dimension $n(n-1)/2$. We have already analyzed what Riemannian manifolds admit such subalgebras.[11] This section is based on the results obtained in Ref. 11, which we summarize as follows:

Proposition 4.1: If an n-dimensional Riemannian manifold M ($n \geq 5$) admits a Lie algebra of dimension $n(n-1)/2$ determined by (1.5), then there is a function U such that (1.6) holds, and the preserved vector field V is the gradient of U, $V = -\operatorname{grad} U$.[17] The structures of M are classified into two cases,

$$ds^2 = A(r)\sum_{i=1}^{n}(dx^i)^2 + B(r)\left(\sum_{i=1}^{n} x^i dx^i\right)^2,$$
$$U = U(r) \quad \left(r^2 = \sum_{i=1}^{n}(x^i)^2\right), \tag{4.1}$$
$$ds^2 = (dt)^2 + f(t)d\sigma^2,$$
$$U = U(t), \tag{4.2}$$

where $d\sigma^2$ is a metric defining an $(n-1)$-dimensional Riemannian space of constant curvature. In the case (4.1) $\operatorname{grad} U$ vanishes at the point $r=0$, but in the case (4.2) $\operatorname{grad} U$ never vanishes.

On the basis of Proposition 4.1, we investigate Eqs. (1.4) in what follows. First we prove

Theorem 4.2: If symmetry algebras determined by (1.4) have Lie subalgebras of dimension $n(n-1)/2$ ($n \geq 5$) defined by (1.5), the former algebras are of dimension n^2-1 at most.

Proof. The proof proceeds analogously to that of Theorem 2.1. We need only to observe that the second equation in (1.6) is independent both of (2.2) and of (2.6) with (V^i) replaced by (U^i), local components of $\operatorname{grad} U$. This completes the proof.

From Proposition 2.2 and Theorem 4.2, we obtain

Theorem 4.3: If an n-dimensional Riemannian manifold M ($n \geq 5$) admits the symmetry algebra of dimension n^2-1 described in Theorem 4.2, then M is of constant curvature.

Now we proceed to the study of what dynamical systems admit the symmetry algebra of dimension n^2-1 stated above. To do this, we deal with (1.4) in a space of constant curvature. We begin with specifying coordinate systems.

Lemma 4.4: A space of constant curvature ($n \geq 3$) [18] has the following coordinate systems which are consistent with that given in Proposition 4.1:

$$ds^2 = \frac{1}{K(r^2+1/K)^2}\left[(r^2+1/K)\sum_{i=1}^{n}(dx^i)^2 - \left(\sum_{i=1}^{n} x^i dx^i\right)^2\right],$$
$$U = U(r) \quad \left(r^2 = \sum_{i=1}^{n}(x^i)^2\right), \tag{4.3}$$
$$ds^2 = \sum_{i=1}^{n}(dx^i)^2, \quad U = U(r), \tag{4.4}$$
$$ds^2 = \sum_{\lambda=1}^{n-1}(dx^\lambda)^2 + (dt)^2, \quad U = U(t), \tag{4.5}$$

where $K(\neq 0)$ is the sectional curvature. The origin $(x^i)=0$ is included or not according as the metrics (4.3) and (4.4) arise from (4.1) or (4.2). The metric (4.5) arises from (4.2).

Proof: First we suppose that M has the structure (4.1). By setting

$$A(r) = 1/K(r^2+1/K), \quad B(r) = -1/K(r^2+1/K)^2, \tag{4.6}$$

we obtain a metric defining a space of constant curvature with the sectional curvature $K \neq 0$ [i.e., (4.3)]. By setting $A=1$ and $B=0$, we get a metric defining a flat space [i.e., (4.4)]. Next we assume that the structure (4.2) is the case. Then M becomes a space of constant curvature with the sectional curvature K, if $f(t)$ is given by

$$f(t) = (1/K)\sin^2(\sqrt{K}\,t) \quad \text{for } K>0, \tag{4.7a}$$
$$f(t) = (1/-K)\sinh^2(\sqrt{-K}\,t) \quad \text{for } K<0, \tag{4.7b}$$
$$f(t) = t^2 \quad \text{for } K=0, \tag{4.7c}$$

together with $d\sigma^2$ defining an $(n-1)$-dimensional space of constant curvature with the sectional curvature 1. Since the metrics for (4.7) degenerate at $t=0$, we assume that t is positive. By a change of the coordinate system,[19] these metrics can be put into the form

$$ds^2 = (1+\tfrac{1}{4}Kt^2)^{-2}\left(\sum_{i=1}^{n}(d\bar{x}^i)^2\right), \quad \sum_{i=1}^{n}(\bar{x}^i)^2 = t^2 \neq 0. \tag{4.8}$$

By a further change of coordinate systems, this form ($K \neq 0$) can be written in the same form as in (4.3) in the region $(x^i) \neq 0$. If $K=0$, (4.8) is nothing but the metric in (4.4) in the region $(x^i) \neq 0$. In both cases $K=0$ and $\neq 0$, t is expressed as a function of r. Thus we have the same form as in (4.3) and (4.4) in the region $(x^i) \neq 0$. On the other hand, if we set $f(t)=1$ in the case (4.2), we have another structure (4.5). This ends the proof.

In the coordinate systems described in Lemma 4.4, an infinitesimal projective transformation (ξ^i), a solution to the first equation in (1.4), takes the form[20]

$$\xi^i = a_j x^j x^i + b_j{}^i x^j + c^i, \tag{4.9}$$

where a_j, $b_j{}^i$ and c^i constants. A basis of the infinitesimal projective transformations of M is chosen as

$$Y_j = \frac{\partial}{\partial x^j} + K x^j x^k \frac{\partial}{\partial x^k}, \quad L_{ij} = x^i \frac{\partial}{\partial x^j} - x^j \frac{\partial}{\partial x^i}, \tag{4.10}$$
$$A_j = x^j x^k \frac{\partial}{\partial x^k}, \quad S_{ij} = x^i \frac{\partial}{\partial x^j} + x^j \frac{\partial}{\partial x^i}, \tag{4.11}$$

no matter whether K vanishes or not. The linear space spanned by (4.10) is a Lie algebra of all the Killing vector fields. The set of L's generates a Lie subalgebra of dimension $n(n-1)/2$ which preserves U given in (4.3) or (4.4).

In this setting, we proceed with solving (1.4). First

we deal with the case $K \neq 0$. To do this we write out grad$U$ and div$X$ in the coordinate system (4.3). A straightforward calculation shows that grad$U$ and div$X$ take the form

$$\text{grad}U = K^2 \left(r^2 + \frac{1}{K}\right)^2 \frac{U'}{r} x^k \frac{\partial}{\partial x^k}, \tag{4.12}$$

$$\text{div}X = \nabla_i \xi^i = \sum_{k=1}^{n} \frac{\partial \xi^k}{\partial x^k} - (n+1)\left(r^2 + \frac{1}{K}\right)^{-1} \sum_{k=1}^{n} x^k \xi^k, \tag{4.13}$$

where U' denotes the derivative of U with respect to r.

We now prove the following lemma which provides solutions to (1.4).

Lemma 4.5: Under the condition that (1.6) determines a Lie algebra of dimension $n(n-1)/2$,[21] a solution to (1.4) with $V = -\text{grad}U$ takes one of the following forms ($n \geq 4$) in the coordinate system (4.3):

$$X = \sum b_{ij} L_{ij} + \sum c_{ij} S_{ij} \quad \text{with } \text{tr}(c_{ij}) = 0, \tag{4.14a}$$

$$U = \tfrac{1}{2} cr^2 + c' \quad (c \neq 0, c' : \text{consts}), \tag{4.14b}$$

$$X = \sum b_{ij} L_{ij} + \sum a_j A_j, \tag{4.15a}$$

$$U(r) = -c/r + c' \quad (c \neq 0, c' : \text{consts}), \tag{4.15b}$$

where a_j, $b_{ij} = -b_{ji}$ and $c_{ij} = c_{ji}$ are constants. The vector fields A's, L's and S's are given in (4.10) and (4.11).

Proof: Since we have already obtained the Lie subalgebra of dimension $n(n-1)/2$ determined by (1.6), i.e., the Lie algebra spanned by L's, it is sufficient for us to solve the second equation in (1.4) for infinitesimal projective transformations which are not Killing vector fields.

Let X be such a transformation. Then X must be a linear combination of A_j and S_{ij},

$$X = \sum a_j A_j + \sum c_{ij} S_{ij}, \tag{4.16}$$

where a_j and $c_{ij} = c_{ji}$ are constants. Formula (4.13) applied to (4.16) gives

$$\text{div}X = \frac{n+1}{K(r^2 + 1/K)} \sum a_j x^j + 2\text{tr}(c_{ij}) - \frac{2(n+1)}{r^2 + 1/K} \sum c_{ij} x^i x^j. \tag{4.17}$$

Substituting (4.12), (4.16), and (4.17), respectively, for $-V$, X and div$X$ in (1.4) results in

$$2\sum c_{ij} x^i x^j \frac{1}{r} \frac{d}{dr}\left(\frac{U'}{r}\right) + \sum a_j x^j \left[r \frac{d}{dr}\left(\frac{U'}{r}\right) + 3\frac{U'}{r}\right] + \frac{8}{n+1} \text{tr}(c_{ij}) \frac{U'}{r} = 0. \tag{4.18}$$

Let $x^i = rz^i$. Then (4.18) becomes

$$2\sum c_{ij} z^i z^j r \frac{d}{dr}\left(\frac{U'}{r}\right) + \sum a_j z^j r \left[r \frac{d}{dr}\left(\frac{U'}{r}\right) + 3\frac{U'}{r}\right] + \frac{8}{n+1} \text{tr}(c_{ij}) \frac{U'}{r} = 0 \tag{4.19}$$

with the constraint

$$\sum z^i z^i - 1 = 0. \tag{4.20}$$

Denote the left-hand sides of (4.19) and (4.20) by $P(r,z)$ and $Q(z)$, respectively. P and Q are polynomials in (z^i). By the Hilbert zero theorem,[22] there is an integer p and a polynomial $H(r,z)$ in (z^i) such that $P(r,z)^p = H(r,z)Q(z)$. From this, $Q(z)$ must be a factor of P, because Q is irreducible. As P and Q are of the same degree, there is a function $h(r)$ such that

$$P(r,z) = h(r)Q(z). \tag{4.21}$$

From (4.19), (4.20), and (4.21) we obtain the following identities in (z^i),

$$2r \frac{d}{dr}\left(\frac{U'}{r}\right) \sum c_{ij} z^i z^j = h(r) \sum z^i z^i, \tag{4.22a}$$

$$r\left[r \frac{d}{dr}\left(\frac{U'}{r}\right) + 3\frac{U'}{r}\right] \sum a_j z^j = 0, \tag{4.22b}$$

$$\frac{8}{n+1} \text{tr}(c_{ij}) \frac{U'}{r} = -h(r), \tag{4.22c}$$

If $(a_j) \neq 0$, then (4.22b) provides

$$U(r) = -c/r + c' \quad (c \neq 0, c' : \text{consts}). \tag{4.23}$$

The constant c never vanishes, because if so, then grad$U = 0$, contradicting the assumption. If (4.23) is the case, (4.22a), and (4.22c) yield $3c_{ij} = \text{tr}(c_{ij})\delta_{ij}$, so that $\text{tr}(c_{ij}) = 0$ or $n = 3$. Under the condition $n \geq 4$, we have $\text{tr}(c_{ij}) = 0$ and hence $(c_{ij}) = 0$. Conversely, when $(c_{ij}) = 0$, we may suppose that $(a_j) \neq 0$. For otherwise, (4.16) vanishes.

Suppose that $(a_j) = 0$. Equations (4.22) imply that

$$r \frac{d}{dr}\left(\frac{U'}{r}\right) c_{ij} = -\frac{4}{n+1} \text{tr}(c_{ij}) \frac{U'}{r} \delta_{ij}. \tag{4.24}$$

If $c_{ij} \neq 0$ for some distinct i and j, then (4.24) gives

$$U(r) = \tfrac{1}{2} cr^2 + c' \quad (c \neq 0, c' : \text{consts}). \tag{4.25}$$

The reason for $c \neq 0$ is the same as in (4.23). If (4.25) is the case, then $\text{tr}(c_{ij}) = 0$ holds by means of (4.24). Conversely, if the constants $(c_{ij}) \neq 0$ satisfy the condition $\text{tr}(c_{ij}) = 0$, then we have $(a_j) = 0$ from (4.22). This proves the lemma.

We now turn to the case $K = 0$.

Lemma 4.6: Under the same condition as in Lemma 4.5, a solution to (1.4) with $V = -\text{grad}U$ takes the same form as in that lemma also in the coordinate system (4.4).

This may be proved through the same method as above. We now work with (1.4) in the coordinate system (4.5). It is to be noted in advance that grad$U$ never vanishes in this case.

Lemma 4.7: Under the same condition as in Lemma 4.5, a solution to (1.4) with $V = -\text{grad}U$ takes one of the following forms ($n \geq 4$) in the coordinate system (4.5):

$$X = \sum b'_\lambda Y_\lambda + \sum b_{\lambda\mu} L_{\lambda\mu} + \sum a_\lambda A_\lambda + \sum c_{\lambda\mu}\left(S_{\lambda\mu} + \frac{1}{n}\delta_{\lambda\mu} S_{nn}\right), \tag{4.26a}$$

$$U(t) = -c/t^2 + c' \quad (c \neq 0, c' : \text{consts}), \tag{4.26b}$$

$$X = \sum b'_\lambda Y_\lambda + \sum b_{\lambda\mu} L_{\lambda\mu} + \sum c_{\lambda\mu}\left(S_{\lambda\mu} + \frac{4}{n-3}\delta_{\lambda\mu} S_{nn}\right), \tag{4.27a}$$

$$U(t)=\begin{cases} ct+c' \quad (c\neq 0,\ c'\text{: consts}), & \text{if } \mathrm{tr}(c_{\lambda\mu})=0, \\ \text{an arbitrary function}, & \text{if } \mathrm{tr}(c_{\lambda\mu})=0, \end{cases} \tag{4.27b}$$

where a_λ, $b_{\lambda\mu}=-b_{\mu\lambda}$, b'_λ, and $c_{\lambda\mu}=c_{\mu\lambda}$ are constants. The vector fields Y's, L's, A's, and S's are given in (4.10) and (4.11) together with $K=0$.

Proof: It is easily seen that the Lie algebra of dimension $n(n-1)/2$ determined by (1.6) is spanned by $L_{\lambda\mu}$'s and Y_λ's with $K=0$.

Let X be an infinitesimal projective transformation which is not a Killing vector field. It takes the same form as (4.16), if x^n is replaced by t. By substituting this X and $V=-U'\partial/\partial x^n$ into the second equations in (1.4), we have

$$(2c_{n\kappa}+a_n x^\kappa)U'=0, \tag{4.28}$$

$$(x^nU''+3U')\sum a_j x^j+2U''\sum c_{nj}x^j - 2U'\left(c_{nn}-\frac{4}{n+1}\,\mathrm{tr}(c_{ij})\right)-a_n x^n U'=0, \tag{4.29}$$

where U' and U'' denote the derivatives of U with respect to x^n.

Since $U'\neq 0$, (4.28) gives $c_{n\kappa}=a_n=0$. By substituting this into (4.29), we obtain the following identities,

$$(U''+3U')\sum a_\lambda x^\lambda+2c_{nn}U''x^n-2U'\left(c_{nn}-\frac{4}{n+1}\,\mathrm{tr}(c_{ij})\right)=0. \tag{4.30}$$

If $(a_\lambda)\neq 0$, then from (4.30) it follows that

$$U''x^n+3U'=0. \tag{4.31}$$

Since $U'\neq 0$, this implies that $x^n\neq 0$. Equation (4.31) is integrated to give

$$U(t)=-c/t^2+c' \quad (t=x^n;\ c\neq 0,\ c'\text{: consts}) \tag{4.32}$$

Substituting (4.32) into (4.30) results in

$$\sum c_{\lambda\lambda}=nc_{nn}. \tag{4.33}$$

Conversely, if (4.33) holds, (a_λ) does not necessarily vanish because of (4.30). Thus we obtain (4.26).

If $(a_\lambda)=0$, then from (4.30) we have

$$c_{nn}U''x^n-U'\left(c_{nn}-\frac{4}{n+1}\,\mathrm{tr}(c_{ij})\right)=0. \tag{4.34}$$

The first term of (4.34) vanishes on the hypersurface $x^n=0$. However, as U' never vanishes, the coefficient of U' must vanishes, so that $c_{nn}=4\mathrm{tr}(c_{\lambda\mu})/(n-3)$. Consequently, (4.34) gives $c_{nn}U''=0$. If $c_{nn}\neq 0$, U'' vanishes, so that

$$U(t)=ct+c' \quad (t=x^n;\ c\neq 0,\ c'\text{: consts}). \tag{4.35}$$

Conversely, if $c_{nn}=4\mathrm{tr}(c_{\lambda\mu})/(n-3)\neq 0$, then (4.30) implies that $(a_\lambda)=0$. If $c_{nn}=0$, that is, $\mathrm{tr}(c_{\lambda\mu})=0$, then U is arbitrary. Thus we have (4.27). This completes the proof.

We are now in a position to show the existence of the symmetry algebra stated in Theorem 4.3. On the basis of Lemmas 4.5, 4.6, and 4.7, we obtain

Theorem 4.8: Let M be an n-dimensional Riemannian manifold ($n\geq 5$) which admits a Lie algebra of dimension $n(n-1)/2$ defined by (1.5).[23] Then symmetry algebras of dimension n^2-1 determined by (1.4) are attained, when M is a space of constant curvature. If M has nonzero sectional curvature, the symmetry algebra is spanned by the elements (4.14a) in the coordinate system (4.3). When M is flat, two kinds of symmetry algebras are realized. One is a symmetry algebra generated by the elements (4.14a) in the coordinate system (4.4). The other is spanned by the elements (4.26a) in the coordinate system (4.5).

5. DISCUSSION AND APPLICATION

We have studied a class of symmetries of the dynamical system $(T(M), Z)$, which are determined by (1.3). It seems to be a Cartan's idea[24] to formulate symmetries for differential equations in the form of vector fields. Equation (1.3) is an application of his idea. For a study of the symmetry, infinitesimal transformations (1.2) are not the most general allowed in $T(M)$, since they are lifts of transformations of the base space M. However, we chose such transformations for the reason that the vector field Z defining the dynamical system are originally attributable to the equations of motion on M.

Within this class of symmetries, our results obtained in the last section show what dynamical systems admit large symmetry algebras. Before describing this, we recall that such dynamical systems are determined on the assumption that (a) M ($n\geq 5$) admits a Lie algebra of dimension $n(n-1)/2$ defined by (1.5) or (b) M ($n\geq 4$) admits a Lie algebra of dimension $n(n-1)/2$ defined by (1.6) (see Refs. 17 and 23). In the former case (a) the existence of a potential U is a consequence of the assumption, but in the latter case (b) it is assumed against the lower dimensionality $n\geq 4$.

Now, Theorem 4.8 shows what dynamical systems admit maximal symmetry algebras determined by (1.4). We describe them by pointing out Riemannian metrics and potentials. One is a dynamical system which possesses the Riemannian metric (4.4) or (4.3) with constant curvature and the potential (4.14b). Another is given by the flat metric (4.5) and the potential (4.26b).

Aside from the maximality of dimensions of symmetry algebras, we here direct our notice to the fact that Lemma 4.5 provides two kinds of symmetry algebras realized on a space of constant curvature ($n\geq 4$). One is given by (4.14a) and the other by (4.15a). From potential form, the dynamical systems admitting these symmetry algebras are thought of as generalizations of the harmonic oscillator and the Kepler problem usually considered in Euclidean spaces.

Each system has the energy integral

$$E=\frac{1}{2K(r^2+1/K)^2}\sum[-x^ix^j+\delta_{ij}(r^2+1/K)]v^iv^j+U, \tag{5.1}$$

where U is equal to $cr^2/2$ or $-c/r$.

In general, any conservative dynamical system always possesses the energy integral

$$E=\tfrac{1}{2}g_{ij}v^iv^j+U. \tag{5.2}$$

This integral yields a new first integral, when the system admits a projective symmetry. We can easily prove the following theorem by use of (1.3), which was first proved by Katzin and Levine by a different method.[2,4]

Theorem 5.1: Let X be an infinitesimal projective transformation which satisfies (1.4) but is not a Killing vector field, and $\tilde{X}$ be its lift defined by (1.2). Further, assume that the energy integral (5.2) is given. Then $\tilde{X}(E)$, the derivative of the energy integral E with respect to $\tilde{X}$, is a new first integral which is expressed as

$$\tilde{X}(E)=\tfrac{1}{2}\left(\pounds_X g_{ij}-\frac{4}{n+1}(\mathrm{div}X)g_{ij}\right)v^iv^j+XU. \tag{5.3}$$

We now apply this theorem to the energy integral (5.1). Infinitesimal projective transformations which satisfy (1.4) but are not Killing vector fields are already obtained in Lemma 4.5, i.e., A's and S's.

First suppose that we have the energy integral (5.1) with $U=-c/r$. Then Eq. (5.3) applied to $\tilde{A}_k$ provides

$$\tilde{A}_k(E)=\frac{1}{2K^2(r^2+1/K)^2}\sum_{ij}(\delta_{kj}x^i+\delta_{ki}x^j-2\delta_{ij}x^k)v^iv^j + cx^k/r. \tag{5.4}$$

This is an integral corresponding to the so-called Runge—Lenz vector in classical mechanics. Katzin[2] obtained this integral for the Euclidean space.

Suppose in turn that we are given the energy integral (5.1) with $U=cr^2/2$. We then obtain from (5.3)

$$\sum c_{ij}\tilde{S}_{ij}(E) = \frac{2}{K(r^2+1/K)^2}\sum_{i,j,k}c_{ij}(x^iv^k-x^kv^i)(x^jv^k-x^kv^j) + 2\sum_{i,j}c_{ij}\left(\frac{v^iv^j}{K^2(r^2+1/K)^2}\right)+x^ix^j). \tag{5.5}$$

It should be noted that the first terms of the right-hand members of (5.5) is a sum of the products of the linear first integrals. Indeed, for the Killing vector field L_{ij} with local components (ξ^k_{ij}), the function

$$g_{kl}\xi^k_{ij}v^l=\left(\frac{1}{K(r^2+1/K)}\right)(x^iv^j-x^jv^i) \tag{5.6}$$

is a first integral. Subtract the first terms from (5.5), we have new first integrals quadratic in (v^i),

$$\sum_{i,j}c_{ij}\left(\frac{v^iv^j}{K^2(r^2+1/K)^2}+x^ix^j\right),\quad \mathrm{tr}(c_{ij})=0. \tag{5.7}$$

These integrals are generalizations of the well-known quadratic integrals for the harmonic oscillator. Katzin[2] obtained these integrals in the Euclidean space.

We conclude this section with a remark on quadratic first integrals. Nishino[25] studied quadratic first integrals for a space of constant curvature. He treated the dynamical system constituted by the metric ds^2 and the potential U such that

$$ds^2=(1+\tfrac{1}{4}Kr^2)^{-2}\sum(dx^i)^2,\quad U=U(r)\quad\left[r^2=\sum(x^i)^2\right].$$

Though he assumed that the potential is central, that is, depends on only r, the central potential in our case [see (4.3) and (4.4)] is not an assumption but a consequence of the symmetry.

Nishino[25] analyzed the condition for $Q=\tfrac{1}{2}\eta_{ij}v^iv^j+\zeta$ to be a first integral and found out two kinds of potentials to admit a large number of Q's. They are

$$U=cr^2(1-\tfrac{1}{4}Kr^2)^{-2}\quad\text{and}\quad cr^{-1}(1-\tfrac{1}{4}Kr^2).$$

ACKNOWLEDGMENTS

The author wishes to thank Professor M. Ikeda for his encouragement and comments. He is indebted to Scientific Research Fund of Ministry of Education.

[1]G.H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968).
[2]G.H. Katzin, J. Math. Phys. **14**, 1213 (1973).
[3]J. Levine and G.H. Katzin, J. Math. Phys. **14**, 1886 (1973).
[4]G.H. Katzin and J. Levine, J. Math. Phys. **15**, 1460 (1974).
[5]G.H. Katzin and J. Levine, J. Math. Phys. **16**, 548 (1975)
[6]G.H. Katzin and J. Levine, J. Math. Phys. **17**, 1345 (1976).
[7]G.H. Katzin, J. Levine and R.N. Sane, J. Math. Phys. **18**, 424 (1977).
[8]T. Iwai, Tensor N.S. **31**, 98 (1977).
[9]Unless stated otherwise, Latin indices range over 1, 2,..., n.
[10]Eqs. (1.4) were previously derived by G.H. Katzin in Ref. 2 by use of Lie derivatives of the equations of motion.
[11]T. Iwai, to appear in Kodai Math. Sem. Rep.
[12]I.P. Egorov, Doklady Acad. Nauk SSSR (N.S.) **80**, 709 (1951).
[13]A transformation group G is said to act effectively on M, if any transformation that fixes all the points of M must be the identity of G.
[14]K. Yano, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1955), Appendix.
[15]See the first equation of (1.4). A gradient field (ψ^i) is associated with an infinitesimal projective transformation.
[16]Greek indices take the values 1, 2,..., $n-1$.
[17]The condition $n\geq 5$ is necessary for the preserved vector field V to be a gradient field. Under the condition of Proposition 4.1, (1.5) and (1.6) determine the same algebra of dimension $n(n-1)/2$. If we assume that M admits a Lie algebra of dimension $n(n-1)/2$ determined by not (1.5) but (1.6) together with $V=-\mathrm{grad}U$, the conclusion of the theorem holds good for $n\geq 4$.
[18]If the structures (4.1) and (4.2) described in Proposition 4.1 are given for $n\geq 3$ *a priori*, the coordinate systems stated in Lemma 4.4 are valid.
[19]M. Ikeda and Y. Nishino, Math. Japon. **17**, 69 (1972).
[20]L.P. Eisenhart, *Non-Riemannian Geometry* (Amer. Math. Soc., Providence, R.I., 1964), Sec. 44.
[21]See Ref. 17.
[22]S. Lefschetz, *Algebraic Geometry* (Princeton U. P., Princeton, N.J., 1953), Chap. II.
[23]Ref. 17 is available for the present case.
[24]E. Cartan, *Leçon sur les invariants integraux* (Hermann, Paris, 1958), Chap. 9.
[25]Y. Nishino, Math. Japon. **17**, 59 (1972).

"IST-solvable" nonlinear evolution equations and existence—An extension of Lax's method[a)]

I. Miodek

Laboratoire de Physique Mathématique, [b)] *Université des Sciences et Techniques du Languedoc, 34060 Montpellier Cedex, France*

(Received 17 January 1977; revised manuscript received 12 May 1977)

We present here a new and easy method, a natural extension of Lax's method, for obtaining general "IST-solvable" nonlinear evolution equations. These are evolution equations for the potential function(s), v, of a Hamiltonian, H, when the logarithmic t derivatives of H's inverse scattering data are given by a t-dependent ratio of entire functions of E, $\Omega(t,E)$. Here E is the energy variable and Ω is the "dispersion relation" of Abowitz, Kaup, Newell, and Segur (AKNS). We pose the question of existence of the evolution equation's solution. This question is answered completely in the one-dimensional Schrödinger case (first example). In a second example we derive the evolution equation for an $n\times n$ matrix generalization of the Zakharov–Shabat–AKNS equation. Our method displays the central role of analyticity in E in the IST method as a whole.

INTRODUCTION

Inverse methods were introduced into nonlinear evolution equations by Gardner, Green, Kruskal, and Miura (GGKM)[1] in 1967; they showed that the Korteweg–de Vries (KdV) equation and its higher order generalizations could be interrelated by means of the one-dimensional Schrödinger equation and that this family of nonlinear equations could be solved exactly by letting the inverse scattering data of the Schrödinger equation evolve in a manageable way which is determined by the nonlinear equation. Much of the motivation for the study of the KdV equation came from experimental and numerical studies which showed the presence of stable "particlelike" solutions, named solitons. GGKM were able to show that soliton solutions were associated exactly with the presence of proper eigenvalues of the Schrödinger Hamiltonian, and thus their stability was essentially due to the invariance of this Hamiltonian's spectrum.

In 1968 Lax[2] introduced an elegant unitary operator formalism which recovered the GGKM results for the one-dimensional Schrödinger Hamiltonian: Lax showed clearly, in an abstract setting, that what was involved was the "isospectral" evolution of a Hamiltonian $H \equiv H(t) = U(t)H(0)U^*(t)$, induced by the unitary operator $U \equiv U(t)$ $[U(0)=I]$ and that the evolution equation for H was then given by the *commutator equation*

$$H_t = [B_a, H], \quad \text{where } B_a = U_t U^* = -B_a^*.$$

(Notation: $f_s \equiv \partial f/\partial s$ for $s = x$ or t.)

Here B_a, a linear antisymmetric operator, is U's generator. Now if H's operator structure is restricted so that only its potential v can evolve, as is the case when an inverse method is applied, then the commutator equation becomes an evolution equation for v. The great interest is that the evolution equation is in general nonlinear, is exactly solvable, and has soliton type solutions whenever H has proper (discrete and separated) eigenvalues. Given the above, Lax's method consists of somehow finding or constructing antisymmetric linear differential operators, $B^L = -B^{L*}$, such that $[B^L, H]$ has the correct form to be identified with v_t. He explicitly constructed such operators in the Schrödinger case to recover the GGKM results. Lax also looked at several other Hamiltonians.

Zakharov and Shabat[3] applied these methods to a 2×2 matrix equation, whose Hamiltonian wasn't symmetric, but which had proper eigenvalues and a solvable inverse scattering problem: They obtained solvable nonlinear evolution equations which had soliton solutions, some of which they explicitly displayed. The added interest here is that their nonlinear equation which is a kind of nonlinear time-dependent Schrödinger equation, has applications in plasma physics and, more importantly, the IST method worked again.

In 1974 Ablowitz, Kaup, Newell, and Segur[4] (AKNS) presented a comprehensive analysis of the above methods, including a new and more systematic method of deriving even more general evolution equations which could be written in a beautifully concise form. They showed that the process was a nonlinear generalization of the Fourier transform and so they called it the inverse scattering transform (IST). Their method bypasses B_a and derives the evolution equations by using convenient integral representations for the inverse scattering data. Applied to their 2×2 generalization of the Zakharov–Shabat equation they prove that if their two component potential $\mathbf{U}$ satisfies the evolution equa-equation

$$\mathbf{U}_t = \Omega(L^*)\mathbf{U},$$

Then the inverse scattering data evolve according to their "dispersion" relation $\Omega(E)$. Here $\Omega(E)$ is polynomial (or entire or a ratio of entire functions) in E, the energy variable, and L is a linear matrix differential operator whose eigenfunctions are essentially the squared eigenfunctions of the Hamiltonian. They gave similar results for the Schrödinger equation.

Our method, in its present formulation, is a natural extension of Lax's method and is motivated by a general theorem[5] which guarantees the existence of a unitary

[a)] This work was completed under Contract No. 264 of the R. C. P.

[b)] Physique Mathématique et Théorique, Equipe de Recherche Associée au C.N.R.S.

 0022-2488/78/1901-0019$1.00

U (and thus of $B_a = -B_a^*$), given that the evolution of $H = H^*$ has been induced "isospectrally," e.g., by causing H's inverse scattering data to evolve without altering H's spectral structure which includes the proper eigenvalues, continuous spectrum, and their degeneracies. Given $H \equiv H(t)$ (and consequently having $H_t \equiv \partial H/\partial t$) we systematically solve the two following commutator equations, $[B, H]\psi = H_t\psi = [\tilde{B}, H]\psi$ (where $H\psi = E\psi$ and $\tilde{B}\psi = B\psi$), first for the linear E-dependent operator B, which in turn gives us the linear E-independent operator $\tilde{B}$, all modulo the addition of operators which commute with H.

[If $H^* = H$, then $H_t^* = H_t$ and $B_a \equiv \frac{1}{2}(\tilde{B} - \tilde{B}^*)$ must satisfy the same commutator equation as $\tilde{B}$ since $\tilde{B} + \tilde{B}^*$ commutes with H automatically. | The operator equation for B, applied to ψ, takes on a more convenient form which we call the "*reduced commutator equation*," and we call B the *reduced form* of $\tilde{B}$. When the operator structure is introduced into the reduced commutator equation we obtain an ordinary inhomogeneous differential equation in which H_t is the inhomogeneous term, t is a parameter, and the unknown is an "auxilliary function" b, from which B can be recovered. This ordinary differential equation for b becomes the central tool of our method; we refer to it as the "*auxiliary equation.*" In the examples considered the auxiliary equation can be solved by a simple algebraic rearrangement *assuming* that b is polynomial (or entire or a ratio of entire functions) in E. The general evolution equation, in terms of what is essentially $\lim_{x\to\infty} B \equiv \tilde{B}(E)$, is a simultaneous by-product. At this point our results are just those of AKNS in the Schrödinger case, and their 2×2 matrix case is a special case of our second example.

Stopping at this stage the point of our method is not so much that it can generate new evolution equations (which it does) but that it is computationally simpler than and at least as systematic as that of AKNS, and it yields equations which are just as general. We have exploited this simplicity in our second example, using our method to obtain the evolution equation for an $n\times n$ matrix generalization of the (linear) Zakharov–Shabat–A.K.N.S. equation, for which the inverse problem is not yet solved.

The above analyticity assumption is in a sense the second point of our method. If we do not derive this assumption, then our evolution equations, like those of Lax or of AKNS, lack an existence proof for their solutions. We can say, as they can, that if a solution exists, then the IST method constructs it, but existence (or equivalently in our case, analyticity in E) must be proven separately. Because of this it may be possible to obtain evolution equations to which the IST seems to construct solutions, without solutions existing; in such cases "IST solvable" should not be taken to imply existence and the evolution equations obtained should be taken to be candidate evolution equations till existence is proved.

In the following we have five sections. In Sec. 1 we give more details of the inverse method, the IST, and our method as an extension of Lax's method. In Sec. 2 we discuss existence difficulties. Sections 1 and 2 are in an abstract setting. Sections 3 and 4 deal with our first concrete example, the one-dimensional Schrödinger case. Section 3 derives the evolution equation and Sec. 4 proves existence by deriving the analyticity assumption. Section 5 presents our second example, discusses its form briefly, and then derives its evolution equations, finally showing that the AKNS result is our special case $n = 2$. We do not prove existence or solve the inverse problem for the second example.

1. RESUME OF THE IST, LAX'S METHOD, AND ITS EXTENSION

A. The inverse problem

The IST method is based on the solution of the inverse problem for the scattering system of a linear differential equation,

$$H\psi = E\psi, \tag{1.1}$$

where H is the Hamiltonian operator, $E \in \mathbb{C}$ is the energy variable and ψ is the (generalized) eigenfunction. The solution of the inverse problem for say the one-dimensional Schrödinger Hamiltonian, $H = -\partial^2/\partial x^2 + v$, permits us to construct the potential function v (here v is a real valued E-independent function of x, i.e., a multiplication operator in the x representation), given the "inverse scattering data," which we also call the data for short. The data is, essentially, part of the set of asymptotic (i.e., $x \to \infty$) behaviors of the appropriately normalized wavefunctions, $\psi \underset{x\to\infty}{\sim} \psi^\infty$; it is that subset which can be freely specified and from which the rest of the scattering data can be derived. Finding a convenient subset is not a trivial part of the inverse problem.

A given inverse method constructs the unique Hamiltonian, with a given operator structure, from the data. We remark that two Hamiltonians defined on the same space but with different structures (say one with a local potential and the other with a particular kind of nonlocal potential) can both reproduce exactly the same data. The inverse scattering data thus gives us a unique Hamiltonian within the class assumed by the given inverse method—we will use this remark when discussing existence for solutions of evolution equations generated by Lax's method. *The potential is that part of H's structure which can vary.*

B. The inverse scattering transform (IST)

The inverse method can be used to generate an evolution of the potential v by causing the data to evolve with t; v must change if the data changes because the direct scattering problem determines the data uniquely from v. If we can identify the evolution equation satisfied by v when the data evolves in a given way, then v's evolution equation can be solved by[1]

$$v(t=0) \overset{(a)}{\to} \text{data}(t=0) \overset{(b)}{\to} \text{data}(t) \overset{(c)}{\to} v(t).$$

Here:

(a) indicates solving the direct scattering problem at $t = 0$;

(b) indicates integration of the data's evolution equation;

(c) indicates solving the inverse problem for $v(t)$ given data(t).

This is what AKNS[4] later entitled IST.

The evolution equation for the data in step (b) must, of course, be solvable if the IST is to work, and in the cases considered this is trivially true. The nontrivial part is the identification of the evolution equation for v associated with a given (trivially integrable) evolution of the data. The data's evolution will be characterized entirely by what AKNS called the "dispersion relation" Ω; schematically

$$\Omega(t,E)=\frac{\partial}{\partial t}(\ln \text{ data}(t,E)).$$

In the examples to follow, Ω's analyticity in E is the only restriction imposed on the data's evolution.

C. Lax's method and the present generalization

Lax's method[2] for finding the evolution equation for v is based on the existence[5] of unitary operators $U\equiv U(t)$, $UU^*=I$, which generate the isospectral evolution of the Hamiltonian $H\equiv H(t)$ [and of the generalized eigenfunctions $\psi\equiv\psi(t)$] via

$$H=UH(0)U^* \quad [\text{and } \psi=U\psi(0)]$$

or equivalently via (1.2)

$$H_t=[B_a,H] \quad (\text{and } \psi_t=B_a\psi),$$

where $B_a\equiv U_tU^{-1}=-B_a^*$ is U's generator. The evolution of the data is then given by the asymptotic form of B $B_a\sim\overset{\infty}{B}_a$ via the equation

$$\overset{\infty}{\psi}_t=\overset{\infty}{B}_a\overset{\infty}{\psi}. \tag{1.3}$$

Remark: This, and everything else that we do, obviously extends to any number of t-like parameters (cf. results in Ref. 6).

Lax now seeks (somehow) to construct linear antisymmetric operators B^L such that

$$[B^L,H]\equiv H'_t\overset{?}{=}H_t \tag{1.4}$$

has the correct form to be identified with H_t (a multiplication operator for the Schrödinger case).

We, on the other hand, seek to solve the "commutator equation"

$$[B,H]=H_t \tag{1.5}$$

for B, modulo operators which commute with H. Once solved, (1.5) yields the evolution equation for the potential. In comparison, the AKNS method seeks to find (it nearly solves for) the evolution equation directly from the dispersion relation, bypassing B.

Our B need not be antisymmetric. The symmetry of H, constructed by the IST, implies the symmetry of H_t and $[B,H]$ and thus $B_s\equiv B+B^*$ automatically commutes with H; so the antisymmetric part of B, $B_a\equiv\frac{1}{2}(B-B^*)$, also satisfies (1.5).

The explicit form of the antisymmetric part of B is only needed asymptotically, to obtain the data's evolution via Eq. (1.3). In fact the antisymmetry of the generator, i.e., the unitarity of U, should not be overstressed. In the cases considered it suffices to conserve the $\mathbb{L}^2$ norms of the proper eigenfunctions (i.e., bound states) of H and this can be arranged by adding to B an operator which commutes with H. This remark has relevance when H is not symmetric, as, for example, in the Zakharov–Shabat[3] equation and its AKNS generalization.

D. The reduced E-dependent forms and analyticity in E

We are able to solve (1.5) for B in the examples considered thanks to analyticity in E and to completeness of the ψ's. Completeness permits us to "reduce" E-independent linear differential operators $A^\sim$ to more convenient E-dependent forms A. If $A=\sum_n A(n;x,..)E^n$ is a polynomial in E, then, when applied to ψ, it is equivalent to $A^\sim=\sum_n A(n;x,..)H^n$; A is the reduced form of $A^\sim$. In the reverse sense, for example in the case of the ν-dimensional Schrödinger equation

$$H=-\Delta+v, \quad \Delta\equiv\sum_j\partial_j^2, \quad \partial_j\equiv\frac{\partial}{\partial X(j)},$$

∂_1^{2n+2} is equivalent to $\partial_1^{2n}(v-E-\Delta')$, where $\Delta'\equiv\Delta-\partial_1^2$, and so forth till we obtain the "reduced" form $\alpha(E,\cdots)+\beta(E,\cdots)\partial_1$, where the linear operators α,β no longer contain ∂_1. When $\nu=1$, α and β are no longer differential operators.

The reduced form of (1.5), still an operator equation, leads us naturally to a differential equation, the *auxiliary equation*, when the operator form of H_t is incorporated. This auxiliary equation is our central tool.

Since analyticity in the energy variable E underlies the whole inverse method as well as completeness, our method displays the central role played by analyticity (in E) in the IST method as a whole.

E. Obtaining the evolution equation

In the examples we consider here, we show that if we assume that the reduced form of B, as a function of E, is "appropriately analytic" (i.e., polynomial or ratio of polynomials, entire or ratio of entires, with the restriction that B must be regular at the proper eigenvalues of H), then the auxiliary equation can be solved for the reduced form of B by an algebraic rearrangement. This rearrangement simultaneously yields the evolution equation. This method thus generalizes and systematizes that of Lax, and elegantly yields the more general results of AKNS. *Given this analyticity assumption*, our method yields all the evolution equations obtained by AKNS, but our method is computationally simpler, somewhat more systematic[7] and it can be easily generalized[8] to obtain results when the inverse problem is not yet solved.[9] Proving this analyticity assumption is rather long and technical and is equivalent to proving existence of the solution of the evolution equation; but these other methods also require an existence proof. This we now discuss.

2. THE NEED TO PROVE EXISTENCE

Once (1.4) has been satisfied by $B^L=-B^{L*}$, this does not trivially imply that $H'_t=H_t$ where, we repeat, H_t is the partial t-derivative of the $H\equiv H(t)$ constructed by letting the data evolve according to the given dispersion relation $\Omega(t,E)$ which is, in turn, consistent with the asymptotic form of B_L. We know that $H_t=[B_a,H]$ for some linear operator[5] $B_a=-B_a^*$, but we do not know

that $B^L = B_n$ even though both B_n and B^L have asymptotic forms which generate identical evolutions of the inverse scattering data. It is possible that $H^L \equiv H^L(t) = U^L(t)H(0)U^{L*}(t) \neq H$ where $B^L = U_t^L U^{L*} \neq B_n = U_t U^*$ and $H \equiv H(t) = UH(0)U$. If $H^L \neq H$, then H^L and H would be two Hamiltonians defined on the same space, having the same spectral properties and inverse scattering data, but only H would be in the class assumed by the given inverse method. So in this case $H_t^L = [B^L, H^L]$ could not have the same form as H_t, even though $H_t' \equiv [B^L, H]$ does have the same form as H_t. What needs to be done is to show that B^L is indeed equal to B_n and we do this by solving Eq. (1.2) for B_n, given $H(t)$.

If (1.2) is not solved *it may be possible to find linear antisymmetric operators* B^L *which satisfy Lax's criteria without satisfying* $[B^L, H] \equiv H_t' = H_t$. However, given B^L, if this equation as an evolution equation for v, does have a solution, then the v must be the one constructed by the IST; but existence must now be proven separately.

The AKNS method also lacks an existence proof but for a different reason. Schematically they show that if the data evolve according to $\Omega(E)$, then, say for the Schrödinger case,

$$\int \psi\psi\{v_t - \Omega(L^*)v_x\}\,dx = 0 \quad \text{(we will define } L^* \text{ later)}.$$

However AKNS lacked a completeness relation for the squared eigenfunctions $\psi\psi$, similarly for their Zakharov–Shabat results. AKNS proved that if the evolution equation has a solution, then the IST constructs it. More recently Kaup[10] has proved completeness for the squared Zakharov–Shabat eigenfunctions and Jaulent[11] has done so for the squared Shrödinger eigenfunctions. (We are not aware if it is generally realized that these completeness relations are needed to obtain existence by the AKNS method.)

3. THE ONE-DIMENSIONAL SCHRÖDINGER CASE

A. Outline

Here we will carry through a detailed analysis for the one-dimensional Schrödinger equation. We will first show how the form assumed for B by Lax is generalized by our "reduced" (E-dependent) form for B and then derive the reduced form of (1.5). The connection between the reduced form of B and the evolution of the inverse scattering data will then be given, thus introducing the AKNS dispersion relation explicitly. The evolution equation for v will then be derived by a relatively simple algebraic rearrangement of the reduced form of (1.5), assuming analyticity properties in E which will be stated in full but derived separately in Sec. 4. These derivations are rather technical. The one-dimensional Schrödinger case should be taken as an example of the method already described. This section will also serve as a model for a second example, the $n \times n$ matrix generalization of the 2×2 AKNS–Zakharov–Shabat type equations, which will be given in Sec. 5. We note here that the Schrödinger and certain AKNS-type equations are equivalent; see Eqs. (5.5)–(5.7) of the Appendix.

B. Lax's method

The usual one-dimensional Schrödinger equation is

$$H\psi = E\psi \quad \text{with} \quad H = -D^2 + v$$

$$[D \equiv \partial/\partial x,\ E \equiv k^2 \in \mathbb{C},\ v(t,x) \equiv v(x) \in \mathbb{R},\ \psi(t,x,k) \equiv \psi(x,k) \in \mathbb{C}]. \tag{3.1}$$

Following Lax we now look for appropriate linear antisymmetric differential operators B_q, of order $2q+1$ and form

$$B_q = D^{2q+1} + \sum_{j=1}^{q} (b_j D^{2j-1} + D^{2j-1} b_j), \tag{3.2}$$

where $b_j(x) \in \mathrm{R}$. We want the commutator

$$v_t' \equiv [B_q, H]$$

to be a real-valued function of x (i.e., a multiplication operator). This multiplication operator condition does not define B_q uniquely as was thought; it defines B_q modulo the addition of an arbitrary linear combination of B_j's with $j < q$ since, clearly,

$$B_q' \equiv B_q + \sum_{j=1}^{q-1} \omega_j B_j, \quad \omega_j \in \mathbb{R} \tag{3.3}$$

satisfies all the conditions on B_q. (A quick examination of Lax's argument shows that he has overlooked some arbitrary integration constants and interestingly these constants, here characterized by the ω_j's, introduce sufficient extra freedom to recover the AKNS results in the Schrödinger case for polynomial "dispersion" relations.) Instead of extending Lax's results via (3.3) let us examine the multiplication operator condition and Eq. (1.5) when applied to ψ.

C. The reduced form of B

Clearly, when applied to ψ, all operators of the form (3.2) reduce, thanks to (3.1), to the form

$$B = a(x,E) + b(x,E)D, \tag{3.4}$$

where a and b are polynomials in E (and they also depend on t although we will not bother carrying t in all the notation). We now introduce the notation $B^{\sim}$ for the unreduced, E independent, operator which can be recovered from its reduced form, B, by expanding B in powers of E, placing the powers of E on the right in each term and then formally replacing E with H. This recipe defines the relationship between the E-independent form of any finite order differential operator to its reduced form, vis à vis the one-dimensional Schrödinger equation. We will formally extend this definition to entire functions of E and also to the ratios of entire functions of E, however we give no rigorous meaning to the infinite operator sums which occur in the nonpolynomial cases.

D. The reduced commutator equation and the auxiliary equation

When (1.5) is applied to ψ, given (3.4), we obtain

$$H_t\psi = [B^{\sim}, H]\psi = [B, H]\psi$$
$$= (a_{xx} + 2(v-E)b_x + v_x b + (2a_x + b_{xx})D)\psi. \tag{3.5}$$

so

$$v_t = a_{xx} + 2(v-E)b_x + v_x b + (2a_x + b_{xx})D. \tag{3.6}$$

This is the reduced commutor equation, an operator equation. The condition

$$2a_x + b_{xx} = 0 \ (\text{i.e., } a = a^0 - \tfrac{1}{2}b_x,\ a_x^0 = 0) \tag{3.7}$$

is clearly necessary and sufficient to eliminate D and, substituted into (3.6), leaves us with an ordinary inhomogeneous differential equation, the *"auxiliary" equation*, for the *"auxiliary function"* b,

$$v_t = 2(\Lambda - ED)b, \tag{3.8}$$

where

$$\Lambda \equiv \Lambda(v) \equiv -\tfrac{1}{4}D^3 + vD + \tfrac{1}{2}v_x \ [= -\Lambda(v)^* \text{ since } v^* = v\,]. \tag{3.9}$$

Note: The auxiliary equation has thus been obtained from an extremely simple calculation. The evolution equation can now also be obtained by an extremely simple algebraic rearrangement of (3.8) [Eqs. (3.14)—(3.22)], given analyticity in E. We want to stress the simplicity and naturalness of this calculation. Even Λ, whose eigenfunctions happen to be $\psi\psi$ (Lemma 1, Sec. 4) has appeared, as it were, of its own accord [cf. AKNS's Eq. (A3.17)].

E. Evolution of the data and the dispersion relation

The concomitant evolution of the (generalized) eigenfunctions, given by $\psi_t = \tfrac{1}{2}(B^\sim - B^{\sim *})\psi$, now gives us the evolution of the inverse scattering data in terms of $b^\infty(E)$,

$$b^\infty(t, E) \equiv b^\infty(E) \equiv \lim_{x\to\infty} b(t, x, E) \tag{3.10}$$

(the existence of this limit will be proven in Sec. 4).

Now:

$$\begin{aligned}\overset{\infty}{\psi}_t &= [i\,\mathrm{Im}a^0(-D^2) + D\overset{\infty}{b}(-D^2)]\overset{\infty}{\psi} \\ &= i\,\mathrm{Im}a^0(E)\overset{\infty}{\psi} + \overset{\infty}{b}(E)\overset{\infty}{\psi}_x.\end{aligned} \tag{3.11}$$

To obtain the evolution of the inverse scattering data explicitly we consider appropriately normalized scattering wavefunctions (essentially Jost solutions) (3.12a) and bound state wavefunctions (3.12b).

$$\begin{aligned}&\psi_{+\ x\sim\infty}\ c_+(t,k)(\exp(-ikx) + R_+(t,k)\exp(+ikx)) \equiv \overset{\infty}{\psi}_+, \\ &\quad {}_{x\sim-\infty}\ c_+(t,k)\tau(t,k)\exp(-ikx), \quad c_+(t,k) \neq 0\end{aligned} \tag{3.12a}$$

for $\mathrm{Im}k \geq 0$ and $k^2 = E \in \mathbb{C}\backslash\{E_n\}$ where $\{E_n\} \subset \mathbb{R}$ is the set of proper eigenvalues of H [as an operator in $\mathbb{L}^2(\mathbb{R}; dx)$]; while for $E \in \{E_n\}$

$$\psi \equiv \psi_{n\ x\sim\infty}\ c_n(t)\exp(ik_n x), \tag{3.12b}$$

where $ik_n < 0$, $k_n^2 = E_n < 0$; $c_n(t) \in \mathbb{R}$ and $\int_{-\infty}^{\infty} dx\,\psi_n\psi_n = 1$; R_+ is the reflection coefficient to the right; τ is the transmission coefficient; $\{c_n\}$ are the bound state normalization constants. The inverse scattering data is $\{E_n, c_n(t), R_+(t,k)$ (for $k \in \mathbb{R})\}$. Now substituting Eqs. (3.12) into (3.11) gives us

$$\begin{aligned}&(\ln R_+)_t = 2ik\overset{\infty}{b}(t, E), \quad (\ln c_+^2 R_+)_t = 2i\,\mathrm{Im}a^0(E), \\ &(\ln c_n)_t = ik_n\overset{\infty}{b}(t, E_n) + i\,\mathrm{Im}a^0(E_n).\end{aligned} \tag{3.13a}$$

To be consistent with $c_n(t) \in \mathbb{R}$ we must have $\mathrm{Im}a^0(E_n) = 0$. An appropriate choice of normalization, one which simplifies things, is

$$c_+(t,k) = c_+(0,k)/\sqrt{R_+(t,k)} \tag{3.13b}$$

and this leaves us with

$$(\ln R_+)_t = 2ik\overset{\infty}{b}(t, E), \quad (\ln c_n)_t = ik_n\overset{\infty}{b}(t, E_n). \tag{3.13c}$$

Thus we have the logarithmic t derivatives of the inverse scattering data in terms of a *single* polynomial (or entire or meromorphic) function of E, $\overset{\infty}{b}(t, E)$. This is the only restriction which will be imposed on the evolution of the data. Here $\overset{\infty}{b}(t, E) = \Omega(t, E)$, the AKNS dispersion relation.

For completeness, and later use, we also explicitly introduce the second scattering wavefunction (the other Jost solution) for $\mathrm{Im}k \geq 0$, $k^2 = E$:

$$\begin{aligned}&\psi_{-\ x\sim\infty}\ c_-(t,k)\tau_-(t,k)\exp(+ikx), \quad c_-(t,k) \neq 0. \\ &\quad {}_{x\sim-\infty}\ c_-(t,k)(\exp(ikx) + R_-(t,k)\exp(-ikx)).\end{aligned} \tag{3.12c}$$

The Wronskian of ψ_+ and ψ_- is independent of x (from which $\tau_- = \tau$ follows) and is different from zero unless $E \in \{E_n\}$—this can serve as the general definition of the proper eigenvalues of H.

F. The evolution equation for the potential

First we will obtain the results for dispersion relations which are entire in E, $\overset{\infty}{b}(E) \equiv \overset{\infty}{b}(t, E) = \sum_n \overset{\infty}{b}(n; t)E^n$. Assuming that (3.8) has a solution

$$b \in \mathbb{L}^\infty(\mathbb{R}; dx) \cap \mathbb{C}^3(\mathbb{R}) \text{ which satisfies}$$

$$\lim_{x\to\infty}(b, b_x, b_{xx}) = (\overset{\infty}{b}, 0, 0), \tag{3.14}$$

$$b - \overset{\infty}{b}(E) = \int_\infty^x b_x \quad \text{(i.e., that } b \text{ is absolutely continuous)} \tag{3.15}$$

and that $b_x(x, E)$ is also entire in E now enables us to solve (3.8) by algebraic rearrangement. We simultaneously obtain the evolution equation for v. The derivation of these assumptions is standard enough but rather long and technical, so we give it in Sec. 4.

Now substituting (3.15) into (3.8), defining

$$L^* \equiv \Lambda D^{-1} \quad [\text{where } D^{-1}f(x) \equiv \int_\infty^x f(x) \equiv \int_\infty^x dy f(y)\,], \tag{3.16}$$

and noting that

$$(\Lambda 1) = \tfrac{1}{2}v_x \tag{3.17}$$

gives us

$$\begin{aligned}v_t &= 2(L^* - E)b_x + \overset{\infty}{b}(E)v_x \\ &= \overset{\infty}{b}(L^*)v_x + 2(L^* - E)b_x - (\overset{\infty}{b}(L^*) - \overset{\infty}{b}(E))v_x\end{aligned} \tag{3.18}$$

so

$$v_t - \overset{\infty}{b}(L^*)v_x = (L^* - E)\{2b_x - (\Delta\overset{\infty}{b})v_x\}. \tag{3.19a}$$

Here

$$\begin{aligned}(\Delta\overset{\infty}{b})v_x &\equiv (\Delta\overset{\infty}{b})(L^*, E)v_x \\ &\equiv (L^* - E)^{-1}(\overset{\infty}{b}(L^*) - \overset{\infty}{b}(E))v_x \qquad (3.20) \\ &= \sum_{n\geq 1}\sum_{j=0}^{n-1} L^{*n-j-1}E^j\overset{\infty}{b}(n; t). \qquad (3.21)\end{aligned}$$

Now to obtain both the solution of (3.18) for b_x and the evolution equation for v we show that both sides of (3.19a) must be zero. Since $b_x(x; E)$ and $\overset{\infty}{b}(E)$ are entire in E, so is $(\Delta\overset{\infty}{b})(L^*, E)v_x$ as given by (3.21) [modulo questions of convergence of the infinite operator sums when $b(x, E)$ is not polynomial in E]. From (3.19a) and (3.20) we see that the left-hand side of

$$\{2b_x - (\Delta\overset{\infty}{b})v_x\} = (L^* - E)^{-1}(v_t - \overset{\infty}{b}(L^*)v_x) \tag{3.19b}$$

is entire in E; but the other side tends to zero as $|E| \to \infty$. So both sides of (3.19b), and hence of (3.19a), must be zero. That the inverse of the operator $(L^* - E)$ is well defined if $E \notin \{\text{proper eigenvalues of } H\}$, follows from $(\Lambda - ED)\psi\psi = 0 \Longleftrightarrow (H - E)\psi = 0$ (proved in Sec. 4).

We have thus obtained the evolution equation

$$v_t = \overset{\infty}{b}(t, L^*)v_x \tag{3.22}$$

and the solution b of (3.8),

$$b(t, x, E) = \overset{\infty}{b}(t, E) + \tfrac{1}{2}\int_{\infty}^{x}(\Delta\overset{\infty}{b})(L^*, E)v_x, \tag{3.23}$$

where $(\Delta\overset{\infty}{b})$ is given by Eqs. (3.20). Everything is in terms of $\overset{\infty}{b}(t, E)$, where $\overset{\infty}{b}(t, E)$ is *entire* in E and is related to the evolution of the inverse scattering data via (3.13c).

{*Remark*: One can show by induction that

$$L^{*n}v_x \text{ and } \int_{\infty}^{x} L^{*n}v_x \equiv (D^{-1}\Lambda)^n v$$

are both polynomials in v and a finite number of the x derivatives of v. It thus follows from (3.23), from v and its derivatives vanishing as $x \to -\infty$, that

$$\overset{\infty}{b}(t, E) - \lim_{x\to-\infty} b(t, x, E) = \int_{-\infty}^{\infty} dx\, b_x$$

$$= \tfrac{1}{2}\int_{-\infty}^{\infty} dx\,(\Delta b)(L^*, E)v_x = 0.$$

This shows us that the same evolution equation would be obtained using $\overset{\infty}{b}(E)$ instead of $\overset{\infty}{b}(E)$. Applied to ψ_+ as $x \to -\infty$ [from Eq. (3.12a), $\psi_+ \sim c_+\tau\exp(-ikx)$] with Eqs. (3.13b) and (3.13c) taken into account, it also shows that $\tau_t = 0$. Since τ is not an independent part of the scattering data, the t-independence of τ does not concern us directly, but in other approaches[12] this fact plays a more central role.}

G. Dispersion relations which are ratios of entire functions

The results for $\overset{\infty}{b}(t, E)$ being entire in E are easily extended to

$$\overset{\infty}{b}(t, E) = \overset{\infty}{b}_+(E)/\overset{\infty}{b}_-(E) \tag{3.24}$$

being the ratio of entire functions of E, $\overset{\infty}{b}_\pm(E) \equiv \overset{\infty}{b}_\pm(t, E)$, providing that the zeros of the denominator, $\{E_n'\}$, are disjoint from the Schrödinger bound state energies $\{E_n\}$. [We repeat that our results lack rigorous definitions of infinite operator sums unless both $\overset{\infty}{b}_\pm(E)$ are polynomials.] Multiplying (3.18) on the left by $\overset{\infty}{b}_-(L^*)$, then after a little rearrangement we obtain, more or less as before,

$$\overset{\infty}{b}_-(L^*)v_t - \overset{\infty}{b}_+(L^*)v_x$$

$$= (L^* - E)\{2\,\overset{\infty}{b}_-(L^*)b_x - (\Delta\overset{\infty}{b})v_x/\overset{\infty}{b}_-(E)\}, \tag{3.25a}$$

where

$$(\Delta\overset{\infty}{b})v_x \equiv (\Delta\overset{\infty}{b})(L^*, E)v_x \equiv (\overset{\infty}{b}_+(E)(\Delta\overset{\infty}{b}_-) - \overset{\infty}{b}_-(E)(\Delta\overset{\infty}{b}_+))v_x$$

and (3.25b)

$$(\Delta\overset{\infty}{b}_\pm)v_x \equiv (\Delta\overset{\infty}{b}_\pm)(L^*, E)v_x \equiv (L^* - E)^{-1}(\overset{\infty}{b}_\pm(L^*) - \overset{\infty}{b}_\pm(E))v_x.$$

If $\overset{\infty}{b}_\pm(E)$ are polynomial (or entire) in E, so are $(\Delta\overset{\infty}{b}_\pm)v_x$ and $(\Delta\overset{\infty}{b})v_x$ as defined in Eqs. (3.25). It then follows that both sides of

$$\{2\overset{\infty}{b}_-(L^*)b_x - (\Delta\overset{\infty}{b})v_x/\overset{\infty}{b} - (E)\}$$

$$= (L^* - E)^{-1}(\overset{\infty}{b}_-(L^*)v_t - \overset{\infty}{b}_+(L^*)v_x) \tag{3.25c}$$

[and hence both sides of (3.24a)] must be zero because the rhs of (3.24a) tends to zero as $|E| \to \infty$ and can only have polar singularities (as a function of E) at the points $E \in \{E_n\}$ where the lhs of (3.24b) is regular—because of the restriction $\{E_n\} \cap \{E_n' | \overset{\infty}{b}_-(E_n') = 0\} = \phi$. We thus have the evolution equation

$$\overset{\infty}{b}_-(t, L^*)v_t = \overset{\infty}{b}_+(t; L^*)v_x \tag{3.26}$$

and this is how (3.22) is to be interpreted in this more general case.

Thus we have obtained all the AKNS results for the one-dimensional Schrödinger equation. In summary: The reduced form of the commutator equation easily yielded the auxiliary equation for b, $(\Lambda - ED)b = \frac{1}{2}v_t$, and $\overset{\infty}{b}(E) \equiv \lim_{x\to\infty} b$ was seen to be the AKNS dispersion relation; in turn a simple algebraic rearrangement gave us the evolution equation

$$v_t = \overset{\infty}{b}(t, L^*)v_x, \quad L^* \equiv \Lambda D^{-1} \equiv \Lambda\int_{\infty}^{x}$$

as well as the solution for b. All this assumed that b was analytic in E.

If we did not now derive analyticity in E we would only have a candidate for v's evolution equation, $v \equiv v(t)$ being the potential constructed by the IST from the data at t. If this candidate evolution equation has a solution, the IST must construct it, and reversal of the above process constructs b, showing it to be analytic as assumed. Thus existence and analyticity are equivalent and, already at this point, we have recovered the AKNS results. However we now see what else has to be done: Either obtain a completeness relation for the squared eigenfunctions[10,11] or, as we now proceed to do, derive analyticity.

4. PROOF OF EXISTENCE, UNIQUENESS, AND ANALYTICITY OF THE SCHRÖDINGER AUXILIARY FUNCTION

Statement of the problem: We want here to prove the existence, uniqueness, and analytic properties of the "auxiliary function" b which satisfies (3.8), (3.14), and (3.15). Stating this more precisely, we want to show that if *the* (nonarbitrary) rationalized limit function $\overset{\infty}{b}_+(E) \equiv \overset{\infty}{b}_-(E)\overset{\infty}{b}(E)$ is entire in E and if *the* entire function $\overset{\infty}{b}_-(E)$ has all its zeros on $\{E_n'\}$ which is disjoint from $\{E_n\}$ (proper eigenvalues of H), then the rationalized auxiliary equation

$$(\Lambda - ED)b_+ = \tfrac{1}{2}\overset{\infty}{b}_-(E)v_t, \quad b_+ \underset{x\to\infty}{\sim} \overset{\infty}{b}_+(E), \tag{4.1a}$$

where

$$\Lambda \equiv \Lambda(v) \equiv -\tfrac{1}{4}D^3 + \tfrac{1}{2}vD + \tfrac{1}{2}Dv \tag{4.1b}$$

has a unique solution b_+ which is entire in E for fixed x. Clearly Eq. (4.1) is completely equivalent to (3.8) and (3.14), given $b \equiv b_+/\overset{\infty}{b}_-(E)$.

In some respects it is more convenient to work with homogeneous boundary conditions, i.e., with the equivalent equation for $(b - \overset{\infty}{b})$: $\lim_{x\to\infty} D^j(b - \overset{\infty}{b}) = 0$ for $j = 0, 1, 2$ and

$$(\Lambda - ED)(b - \overset{\infty}{b}) = \tfrac{1}{2}v_t - \overset{\infty}{b}(E)v_x. \tag{4.2}$$

[We have used the obvious result $(\Lambda 1) = \frac{1}{2}v_x$.]

Proof: We first consider the relationship of the null space to the Schrödinger equation. In the real Schrödinger case, $H = H^*$, $v = v^*$, and so $\Lambda^* \equiv \Lambda(v)^* = -\Lambda(v^*) = -\Lambda$ is an antisymmetric operator. [Comments on what changes if $v \neq v^*$ will be given enclosed between square brackets.] Thus $i\Lambda(v)$ is essentially self-adjoint, as is $i(\Lambda - ED)$ for $E \in \mathbb{R}$, and both can be diagonalized. If $E_n \in \mathbb{R}$ and $(\Lambda - E_n D)$ (as an operator in $\mathbb{L}^2(\mathbb{R}; dx)$) has a proper null space, then (4.2) has a solution at $E = E_n$ if and only if the rhs of (4.2) is orthogonal to that null space. [If $v \neq v^*$, then $\Lambda \neq -\Lambda^*$ and a necessary condition for the solutions existence is orthogonality to the null space of $(\Lambda^* + E_n^* D)$.]

In Lemma 1 we will prove the following three properties:

(i) For $E \in \mathbb{C} \setminus \{E_n\}$,

$$(\Lambda - ED)h = 0 \tag{4.3a}$$

$$\Longleftrightarrow h = \alpha_1 \psi_+^2 + \alpha_2 \psi_-^2 + \alpha_3 \psi_+ \psi_-, \tag{4.3b}$$

where $(H - E)\psi_\pm = 0$, $\alpha_j \in \mathbb{C}$, ψ_+ and ψ_- being linearly independent. Explicit definitions of ψ_+ and ψ_- were given earlier, ψ^+ in Eq. (3.12a) (the boundary condition as $x \to -\infty$) and ψ^- in Eq. (3.12c) (when $x \to +\infty$).

Note: the linear independence of $\{\psi_+^2, \psi_-^2, \psi_+\psi_-\}$ follows easily from the linear independence of ψ_+ and ψ_-; e.g., use Eq. (4.3b) to consider $(h/\psi_+^2)_x$, then put $h = 0$.

(ii) For $E \in \{E_n\}$,

$$(\Lambda - E_n D)h_n = 0, \tag{4.4a}$$

$$h_n \in \mathbb{L}^2(\mathbb{R}; dx) \cap C^2(\mathbb{R}) \Longleftrightarrow h_n \propto \psi_n \psi_n, \tag{4.4b}$$

where $(H - E_n)\psi_n = 0$, $\psi_n \in \mathbb{L}^2(\mathbb{R}; dx) \cap C^2(\mathbb{R})$. Thus $\psi_n\psi_n \in \mathbb{L}^1(\mathbb{R}; dx) \cap C^2(\mathbb{R}) \subset \mathbb{L}^2(\mathbb{R}; dx)$. [If $v \neq v^*$, then $(\Lambda^* + E_n^* D)h_n^* = 0 \Longleftrightarrow h_n \propto \psi_n\psi_n$.]

(iii) Orthogonality:

$$\int_{-\infty}^{+\infty} dx\, \psi_n \psi_n v_s = 0 \text{ for } s = x \text{ or } t. \tag{4.5}$$

[If $v \neq v^*$, then (4.5) still holds for v_s; but this is not enough—it needs to hold for both $(v^* \pm v)_s$.]

These three properties show that the null space of $(\Lambda - ED)$ is spanned by the squares of the Schrödinger eigenfunctions and that (4.2) has a solution in $\mathbb{L}^2(\mathbb{R}; dx)$ when $E \in \{E_n\}$. [Our proof does not extend to $v \neq v^*$ because v_s is orthogonal to $\psi_n\psi_n$ but not to $(\psi_n\psi_n)^*$.]

Now, in order to construct the solution of (4.1) with the appropriate analytic properties in E, we consider the related initial value problem for "regular" solutions $f \in C^3(\mathbb{R})$,

$$(\Lambda - ED)f = u(E) \equiv \tfrac{1}{2} b_-(E) v_t, \tag{4.6}$$

$$\lim(x \to x_0)(f, f_x, f_{xx}) = (f_1^0, f_2^0, f_3^0) \equiv \mathbf{f}^0(E), \tag{4.7}$$

and we look for conditions on the initial value $\mathbf{f}^0(E)$ so that $f = b_+$ of (4.1).

If v, v_x, and v_t are "smooth enough" locally and if they also decrease "fast enough" as $|x| \to \infty$, standard theory tells us that f is well defined and entire in E if $\mathbf{f}^0(E)$ and $u(E)$ are entire in E. We can also show that the asymptotic behavior as $|x| \to \infty$ can be controlled by imposing analytic conditions on the initial value $\mathbf{f}^0(E)$. More details of this construction are given in proving Lemma 2.

Three linear equations must be satisfied by the three components of $\mathbf{f}^0(E)$ so that $f(x, k^2)$ (with $k^2 = E$, $\mathrm{Im}k > 0$) remains bounded as $x \to \pm\infty$ and so that $\lim_{x\to\infty} f(x, k^2)$ exists uniformly in k as $\mathrm{Im}k \to 0+$. Then $\lim_{x\to\infty} f(x, E) = \overset{\infty}{b}(E)$ gives us the "dispersion relation." These three linear constraints on $\mathbf{f}^0(E)$ are analytic in E, and they can be conveniently put into the 3×3 matrix form

$$M(E)\mathbf{f}^0(E) = \overset{\infty}{\mathbf{f}}(u; E), \tag{4.8}$$

which is obtained in Lemma 2. For our immediate purpose we state that $M(E)$, its determinant $\det M(E)$, and $\overset{\infty}{\mathbf{f}}(u; E)$ are analytic in $k \equiv E^{1/2}$, $\mathrm{Im}k > 0$; that $\overset{\infty}{\mathbf{f}}(u; E)$ is linear in u and that $\overset{\infty}{b}(E)$ is a linear functional of $\mathbf{f}^0(E)$, so for example if $v_t = 0$, then $\overset{\infty}{b}(E) = 0$.

This matrix equation, (4.8), can be solved for the initial value $\mathbf{f}^0(E)$ for all $E \in \mathbb{C} \setminus \{E_n^0 | \det M(E_n^0) = 0\}$. This initial value when used in (4.6) yields an f which satisfies all the conditions on b_+, except possibly when $E \in \{E_n^0\}$. To finish, we now need to show that $\mathbf{f}^0(E)$, and the f it generates, are regular when $E \to E_n$. Since $M(E)$ and its determinant are regular in E, $\mathbf{f}^0(E)$ can only have polar singularities on the point set $\{E_n^0\}$. We show in Lamma 3 that $\{E_n^0\} \subset \{E_n\}$. Now we can make use of the orthogonality property (4.5) with which we showed that when $E = E_n$, Eq. (4.2) has a solution $(b - \overset{\infty}{b})(x, E_n)$ in $\mathbb{L}^2(\mathbb{R}; dx)$. We deduce that when $E = E_n^0$, Eq. (4.8) must also have at least one solution. This solution is the initial value vector of $b(x, E_n)$ at $x = x_0$, noted as $\mathbf{b}^0(E_n)$. So $\overset{\infty}{\mathbf{f}}(v_t; E_n)$ is in the range of $M(E_n)$. This, together with analyticity in E, implies the existence of

$$\lim_{E \to E_n^0} M(E)^{-1} \overset{\infty}{\mathbf{f}}(v_t; E) = \mathbf{b}^0(E_n) \tag{4.9}$$

(see Lemma 4 for details). Thus $\mathbf{f}^0(E)$ is regular on $\{E_n^0\}$ as well.

We have obtained what we wanted. The solution of Eq. (4.8) gives us the appropriate initial value for (4.6) which then generates both the appropriate solution of (4.6) and its asymptotic limit. By construction this solution also satisfies (4.1) and all the analyticity requirements in E. Schematically

$$\begin{aligned} &\mathbf{f}^0(E) = M(E)^{-1}\mathbf{f}(v_t; E), \text{ and then via (4.6)} \\ &\mathbf{f}^0(E) \to f(x, E) = b(x, E), \quad \overset{\infty}{b}(E) = \lim_{x\to\infty} b(x, E) \end{aligned} \tag{4.10}$$

gives us the pathway of this construction. Q.E.D

[If $v \neq v^*$, then $\{E_n^0\} \subset \{E_n\} \subset \mathbb{C}$, but we do not have the appropriate orthogonality property.]

Summary of the proof: The study of $v_s \equiv \partial v/\partial s$ where $s = x$ or t permits us to prove that the squares of the generalized Schrödinger eigenfunctions span the null space of $(\Lambda - ED)$ and that v_s is orthogonal to the squares of the Schrödinger bound states. This orthogonality implies existence of an $\mathbb{L}^\infty(\mathbb{R}; dx)$ solution b of $(\Lambda - ED)b = \frac{1}{2}v_t$ at $E = E_n$. This in turn permits us to show that, when we study the associated initial value problem for a "regular" solution f, we can solve for the initial value

vector $\mathbf{f}^0(E) \equiv (f, f_x, f_{xx})_{x=x_0}$ which is analytic in $E \in \mathbb{C} \setminus \{E_n\} \cup \{E_n\} = \mathbb{C}$. By "regular" we mean (i) regular in E for fixed $x \in \mathbb{R}$, (ii) bounded asymptotically as $|x| \to \pm\infty$ for fixed $E \in \mathbb{C}$, *and* (iii) having an x-independent limit as $x \to +\infty$ uniformly in $k \equiv E^{1/2}$, $\mathrm{Im}k \geq 0$.

Lemma 1: The squares ψ^2 of the generalized Schrödinger eigenfunctions, $(H - E)\psi = 0$ span the space of homogeneous solutions h of Eq. (4.3a), i.e., h satisfies

$$(\Lambda - ED)h = 0 \quad \text{with} \quad \lim_{x\to\infty} D^j h = 0 \quad \text{for} \quad j = 0, 1, 2. \tag{4.11}$$

Furthermore, the squares ψ_n^2 of proper Schrödinger eigenfunctions are orthogonal to $v_s \equiv \partial v/\partial s$, with $s = x$ or t, in the sense of Eq. (4.5),

$$\int_{-\infty}^{\infty} dx\, \psi_n^2 v_s = 0. \tag{4.12}$$

Proof:

$$v_s = (v - E)_s = (\psi^{-1}\psi_{xx})_s = \psi^{-2}(\psi\psi_{sx} - \psi_s\psi_x), \tag{4.13}$$

$$(4.13) \Longrightarrow \int_{-\infty}^{\infty} dx\, \psi^2 v_s = (\psi\psi_{sx} - \psi_s\psi_x)\Big|_{x=-\infty}^{+\infty}. \tag{4.14}$$

For $E = E_n$, $\psi_n \propto \psi_+ \propto \psi_-$ [defined by Eqs. (3.12a) and (3.12c)] and so

$$\psi_n \underset{|x|\to\infty}{\sim} \exp(-|x|\,|\mathrm{Im}k_n|)$$

($k_n \equiv E_n^{1/2}$; for $v = v^*$, $E_n \leq 0$) and $(\psi_n)_x \to 0$ as $|x| \to \infty$. Thus (4.14) implies (4.12) when $E \in \{E_n\}$.

[For $E \in \mathbb{C} \setminus \{E_n\}$, (4.14) with $\psi = \psi_+$ generates the integral representations for R_+ (when $s = x$) and $(R_+)_t$ (when $s = t$). These are basic to the AKNS method.]

To prove that ψ^2 satisfies $(\Lambda - ED)\psi^2 = 0$ in a way that clearly relates this property to Eq. (4.1) we first note that the ansatz [cf. Eqs. (3.4) and (3.7)]

$$\psi_t = (a^0 - \tfrac{1}{2}b_x)\psi + b\psi_x, \quad a_x^0 = 0 \tag{4.15a}$$

when substituted into Eq. (4.13) with $s = t$ yields Eq. (4.1). Similarly the ansatz

$$\psi_x = -\tfrac{1}{2}g_x\psi + g\psi_x \ (\Longleftrightarrow g - 1 \propto \psi^2) \tag{4.15b}$$

when substituted into Eq. (4.13) with $s = x$ yields

$$\tfrac{1}{2}v_x = (\Lambda - ED)g. \tag{4.16a}$$

But recalling that $\tfrac{1}{2}v_x = \Lambda 1$ and noting the equivalence in (4.15b), we see that

$$(4.16a) \Longleftrightarrow (\Lambda - ED)\psi^2 = 0. \tag{4.16b}$$

If $\{\psi_+, \psi_-\}$ were any pair of linearly independent Schrodinger eigenfunctions of energy E, $\{\psi_+^2, \psi_-^2, \psi_+\psi_-\}$ would also be linearly independent. Thus, noting that $2\psi_+\psi_- = (\psi_+ + \psi_-)^2 - \psi_+^2 - \psi_-^2$, we see that for $E \in \mathbb{C} \setminus \{E_n\}$, an arbitrary homogeneous solution of (4.11) can be expanded as a linear combination of the squares of the Schrödinger (generalized) eigenfunctions defined by (3.12a) and (3.12c) together with $(\psi_+ + \psi_-)^2$. However when $E = E_n$, then $\psi_n \propto \psi_+ \propto \psi_-$, but even so any other linearly independent solution of the Schrödinger equation, ψ_n', must diverge as $x \to \infty$ in such a manner that $\psi_n\psi_n' \not\to 0$ as $x \to \infty$. Thus ψ_n^2 spans the solution space of (4.11), boundary condition included, when $E = E_n$.

This completes the proof of Lemma 1. ∎

[Note: If $v \neq v^*$, Eq. (4.14) still implies Eq. (4.12) because the linear dependence $\psi_n \propto \psi_+ \propto \psi_-$ when $E = E_n$ implies that ψ_n has simple exponential behavior at both $x \to \pm\infty$ and this is enough to ensure that the rhs of (4.14) is zero even when $E_n > 0$.]

Lemma 2: The solution of Eq. (4.6) will be bounded as $x \to \pm\infty$ and will have a constant limit as $x \to +\infty$ uniformly in $k \equiv E^{1/2}$, $\mathrm{Im}k \geq 0$, if and only if the initial value vector at $x = x_0 \in \mathbb{R}$, $\mathbf{f}^0(E) \equiv (f, f_x, f_{xx})|_{x=x_0}$, satisfies three linear constrains which are analytic in E; these constraints can be put in the matrix form of Eq. (4.8).

Proof: We start by sketching a standard method for solving Eq. (4.6). Equation (4.6) can be put into the integral form

$$[1 - A(E)(vD + \tfrac{1}{2}v_x)]f = \mathbf{C}(E)\cdot\mathbf{f}^0(E) - A(E)u(E) \tag{4.17a}$$

where $A(E)$ is a right inverse of $(\frac{1}{4}D^2 + E)D$ $= \frac{1}{4}\exp(\pm 2ikx)D\exp(\mp 2ikx)D\exp(\pm 2ikx)D$. Defining

$$D^{-1} \equiv \int_{x_0}^{x} [\text{so } DD^{-1}g(x) = g(x) \neq D^{-1}Dg(x) = g(x) - g(x_0)],$$

$$A(E) = 4D^{-1}\exp(\mp 2ikx)D^{-1}\exp(\pm 4ikx)D^{-1}\exp(\mp 2ikx). \tag{4.17b}$$

Then $\mathbf{C}(E)\cdot\mathbf{f}^0(E)$ is the usual scalar product for three component vectors and the components of $\mathbf{C} \equiv (C_1, C_2, C_3)$ are given by

$$A(E)(\tfrac{1}{4}D^2 + E)Df(x)$$
$$= f(x) - C_1(E)f(x_0) - C_2(E)f_x(x_0) - C_3(E)f_{xx}(x_0).$$

Using (4.17b) this gives us

$$C_1(E) = 1, \quad C_2(E) = \frac{1}{2k}\sin 2k(x - x_0),$$

$$C_3(E) = \frac{1}{4E}[1 - \cos 2k(x - x_0)]. \tag{4.17c}$$

We now demand: (i) that v and v_x be "smooth enough locally" so that the inverse of $[1 - A(E)(vD + \frac{1}{2}v_x)]$ be well defined in some finite neighborhood of any point $x_0 \in \mathbb{R}$; (ii) that v and v_x "decrease rapidly enough" so that $[1 - A(E)(vD + \frac{1}{2}v_x)]^{-1}$ be well defined whenever $|x| \geq |x_0| \geq r$ for some finite r and the homogeneous solutions of (4.17a) have standard asymptotic ($|x| \to \infty$) behavior [see Eq. (4.19)]; (iii) that v_t be "smooth enough and decrease rapidly enough" so that $[1 - A(E)$ $\times(vD + \frac{1}{2}v_x)]^{-1}A(E)u(E)$ [recall that $u(E) \equiv \frac{1}{2}\overset{\infty}{b}_-(E)v_t$] be well defined and so that it has standard asymptotic behavior with the coefficients that are entire in E. The factor $\overset{\infty}{b}_-(E)$ multiplying v_t indicates that if $u(E) \to \frac{1}{2}v_t$, then the term in v_t will generate poles at $E \in \{E_n' | \overset{\infty}{b}_-(E_n') = 0\}$.

Given that v satisfies these conditions we solve (4.17a) by inverting $[1 - A(E)(vD + \frac{1}{2}v_x)]$ and obtain a sum of a homogeneous solution $f_h(\mathbf{f}^0)$ (linear in $\mathbf{f}^0$) and a particular solution $f_p(-v_t)$ (linear in v_t). What concerns us are the asymptotic behaviors respectively as $x \to \pm\infty$ from which we can obtain boundedness and limit conditions.

Defining

$$f_p(v_t;E)\equiv\tfrac{1}{2}[1-A(E)(vD+\tfrac{1}{2}v_x)]^{-1}A(E)\overset{\infty}{b}_-(E)v_t$$

and (4.18)

$$f_h(\mathbf{f}^0;E)\equiv[1-A(E)(vD+\tfrac{1}{2}v_x)]^{-1}\mathbf{C}(E)\cdot\mathbf{f}^0(E),$$

$$f=f_h(\mathbf{f}^0;E)-f_p(v_t;E)$$

$$\underset{x\sim\pm\infty}{\sim}\exp(-2ikx)(f_{h\pm}^{\pm\infty}-f_{p\pm}^{\pm\infty})+\exp(2ikx)(f_{h\mp}^{\pm\infty}-f_{p\mp}^{\pm\infty})$$

$$+(f_{h0}^{\pm\infty}-f_{p0}^{\pm\infty}). \tag{4.19}$$

[Note: When $\mathrm{Im}k>0$, only the dominant term counts after the terms with null coefficients have been eliminated.]

From (4.19) we see that, for $\mathrm{Im}k>0$, f is bounded as $|x|\to\infty$ if and only if the two following equations hold:

$$f_{h\pm}^{\pm\infty}(\mathbf{f}^0;E)=f_{p\pm}^{\pm\infty}(v_t;E). \tag{4.20a±}$$

The first term is linear in $\mathbf{f}^0$, the second is linear in v_t, and both are analytic in E.

Furthermore from (4.19) we see, that we have eliminated the term in $\exp(-2ikx)$ as $x\to\infty$, that f will take a constant limit as $x\to\infty$ *uniformly* as $\mathrm{Im}k\to 0+$ if and only if a third linear equation holds,

$$f_{h-}^{+\infty}(\mathbf{f}^0;E)=f_{p-}^{+\infty}(v_t;E). \tag{4.20b}$$

We rewrite equations (4.20a±) and (4.20b) in the 3×3 matrix form

$$M(E)\mathbf{f}^0(E)=\overset{\infty}{\mathbf{f}}(v_t;E), \tag{4.20}$$

where $M(E)$ and $\overset{\infty}{\mathbf{f}}(v_t;E)$ are analytic in E and the rhs is linear in v_t.

From equations (4.19) and (4.20) we now have f bounded as $|x|\to\infty$ and the existence of

$$\overset{\infty}{b}_+(E)\equiv\lim_{x\to\infty}f(x,E)=f_{p0}^{+\infty}(v_t;E)-f_{h0}^{+\infty}(\mathbf{f}^0;E) \tag{4.21}$$

uniformly in $k\equiv E^{1/2}$, $\mathrm{Im}k\geq 0$. [Whenever (4.20) is satisfies we can identify f with b_+.] ∎

Lemma 3: $\{E_n^0\,|\,\det M(E_n^0)=0\}=\{E_n\,|\,W(\psi_+,\psi_-)(E_n)=0\}$.

Proof: At $E=E_n^0$ the homogeneous form of (4.20), which is obtained when $v_t=0$, has a nontrivial solution $\mathbf{f}_h^0(E)$. This implies that at $E=E_n^0$, $\mathbf{f}_h^0(E)$ is the initial value of a homogeneous solution f_h of (4.6) satisfying all the conditions of Lemma 2. This f_h is therefore in the null space of $(\Lambda-ED)$ and so, by Lemma 1, f_h can be expanded as a linear combination of $\{\psi_+^2,\ \psi_-^2,\ \psi_+\psi_-\}$. But if $E_n^0\notin\{E_n\}$ it is straightforward to verify that no linear combination of $\{\psi_+^2,\ \psi_-^2,\ \psi_+\psi_-\}$ satisfies all the conditions (boundedness at both $x\to\pm\infty$ and the existence of a constant limit as $x\to+\infty$ uniformly in $k\equiv E^{1/2}$, $\mathrm{Im}k\geq 0$) of Lemma 2. Thus $\{E_n^0\}\subset\{E_n\}$. [This much holds whether or not $v=v^*$.]

If $E=E_n$, then $\psi_n^2(\propto\psi_+^2\propto\psi_-^2)$ is always a homogeneous solution of $(\Lambda-ED)$ which satisfies the conditions of Lemma 2. So, if $E=E_n$, then

$$(\psi_n^2,\ (\psi_n^2)_x,\ (\psi_n^2)_{xx})\big|_{x=x_0}\equiv\mathbf{f}_h^0(E_n)\neq\mathbf{0} \tag{4.22}$$

must be in the null space of $M(E)$ and $\{E_n\}\subset\{E_n^0\}$. ∎

[If $v\neq v^*$, then it is possible for $E_n>0$ and then ψ_n^2 does not have an x-independent limit as $x\to+\infty$.]

Lemma 4: The limit $E\to E_n^0$ of $M(E)^{-1}\overset{\infty}{\mathbf{f}}(v_t;E)\equiv\mathbf{f}^0(E)$ exists.

Proof: We know that $\{E_n^0\}\subset\{E_n\}$ so it suffices to prove that the limit exists when $E\to E_n$. Now if $M_c(E)$ denotes the cofactor matrix transpose of $M(E)$ we have, from (4.20),

$$\mathbf{f}^0(E)=(\det M(E))^{-1}M_c(E)\overset{\infty}{\mathbf{f}}(v_t;E). \tag{4.23}$$

$M_c(E)$ and $\det M(E)$ are analytic in E because $M(E)$ is analytic in E. Let us simplify matters a little by assuming that $\det M(E)$ has only simple zeros. [Multiple zeros can then be treated as the limit where simple zeros coincide or by using the following observation whose proof is quite standard: If a matrix $M(E)$ is analytic in E and if $M(E_n)$ has a null space of dimension j, then $\det M(E)$ must have a zero of order at least j at $E=E_n$.]

Considering (4.23) under these conditions, and remembering that $\overset{\infty}{\mathbf{f}}(v_t;E)$ too is analytic in E, we see that whether or not the limit exists depends on whether or not $M_c(E)\overset{\infty}{\mathbf{f}}(v_t;E)$ has a zero at $E=E_n$ to compensate the zero of $\det M(E)$: It thus suffices for

$$M_c(E_n)\overset{\infty}{\mathbf{f}}(v_t;E_n)=0. \tag{4.24}$$

But since we know from Lemma 1 that (4.1) has a solution at $E=E_n$, $b_+(x,E_n)$, its initial value at $x=x_0$, denoted by $\mathbf{b}_+^0(E_n)$, must satisfy (4.20) at $E=E_n$,

$$M(E_n)\mathbf{b}^0(E_n)=\overset{\infty}{\mathbf{f}}(v_t;E_n). \tag{4.25}$$

Left multiplying by $M_c(E_n)$ now gives us the sought after result,

$$M_c(E_n)\overset{\infty}{\mathbf{f}}(v_t;E_n)=\det M(E_n)\mathbf{b}^0(E_n)=0. \tag{4.26}$$

The existence of a solution of (4.1) at $E=E_n$ has permitted us to show that $\overset{\infty}{\mathbf{f}}(v_t;E)$ was in the range of $M(E_n)$ [Eq. (4.26)] which in turn implied that the analytic function of E, $M_c(E)\overset{\infty}{\mathbf{f}}(v_t;E)$ had a zero at $E=E_n$ to compensate that of $\det M(E)$ (assuming simple zeros) and so the limit of (4.23) exists as $E\to E_n\in\{E_n\}\supset\{E_n^0\}$. Extension to multiple zeros proceeds as indicated above. ∎

5. NONLINEAR EVOLUTION EQUATIONS FOR LINEAR FIRST ORDER $n\times n$ MATRIX EQUATIONS OF THE ZAKHAROV-SHABAT TYPE

A. The linear matrix equation

We will follow the model set up in Sec. 3 for the one-dimensional Schrödinger equation to show that the same method also yields evolution equations for the off-diagonal (i.e., diagonal part is zero) $n\times n$ matrix "potential" $Q\equiv Q^{\mathrm{off}}+Q^{\mathrm{diag}}=Q^{\mathrm{off}}$ of the following first order differential matrix equation for f:

$$(iD-iQ-k\sigma)f=0,\quad \text{i.e.,}\quad f_x=Zf=(Q-ik\sigma)f. \tag{5.1a}$$

The Hamiltonian form is

$$Hf\equiv\sigma^{-1}(iD-iQ)f=kf. \tag{5.1b}$$

Here $D\equiv\partial/\partial x$; $k\in\mathbb{C}$ is the energy variable in this case; σ is a constant (i.e., independent of t,x,k) diagonal matrix (so $\sigma=\sigma^{\mathrm{diag}}$); $\mathrm{tr}\sigma\equiv$ trace $\sigma=0$; the matrix elements of σ satisfy $\sigma_{ij}=\delta_{ij}\sigma_{ii}$ and $\sigma_{ii}\neq0$, so σ^{-1} is well defined; for $n\geq3$ we (eventually) impose the additional condition that the eigenvalues of σ be distinct; finally we assume that $Q\to0$ rapidly as $x\to\pm\infty$ so that asymptotically

$$f\sim\exp(-ikx\sigma)f_0^{\pm\infty},\quad \text{where } (f_0^{\pm\infty})_x=0. \tag{5.1c}$$

In these equations f could be taken to be either a column or a square matrix. We will reserve f for column solutions $\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$, and use F for matrix solutions which are fundamental; i.e., $\tau \equiv \det F \neq 0$ and F's columns form a (fundamental) system of n linearly independent column solutions.

We note that the given conditions imply that $\mathrm{tr}Z = 0$, and since (5.1a) $\Longrightarrow \tau_x = (\mathrm{tr}Z)\tau$, we see that $\mathrm{tr}Z = 0 \Longleftrightarrow \tau_x = 0$. (5.1d)

B. The 2 × 2 cases and self-adjointness

Notice that in general

$$H^* = H \Longleftrightarrow \sigma^* = \sigma \quad \text{and} \quad Q^* = -\sigma^{-1}Q\sigma. \tag{5.2}$$

In the 2×2 case this essentially simplifies to

$$H^* = H \Longleftrightarrow Q = Q^* \tag{5.3}$$

because, without loss of generality, k can be "normalized" so that

$$\sigma^* = \sigma^{-1} = \sigma \equiv \sigma_3 \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{5.4a}$$

If we then introduce the other Pauli matrices σ_j, $j = 1$ or 2,

$$Q \equiv q_1\sigma_1 + q_2\sigma_2 \equiv \begin{bmatrix} 0 & q_1 - iq_2 \\ q_1 + iq_2 & 0 \end{bmatrix} \equiv \begin{bmatrix} 0 & q \\ r & 0 \end{bmatrix}, \tag{5.4b}$$

Eqs. (4.1) reduce exactly to the generalized Zakharov–Shabat equation studied by AKNS (they use q, r and in place of H they use $\not{L}$). The Zakharov–Shabat[3] case itself is then obtained by adding the restriction $r^* = -q$, for which $H^* \neq H$, since $H^* = H \Longleftrightarrow r^* = +q$; however the proper eigenvalues exist and the inverse problem has been solved. As noted by AKNS H has no proper eigenvalues when $H^* = H$ and we add that it is then understood that q and r are "nice" functions of x. What if the AKNS–Zakharov–Shabat sense of "nice" is weakened? We present a partial answer to this question in the Appendix where we show that certain 2×2 cases ($r = q = q_1$, i.e., $q_2 = 0$) are equivalent to the Schrödinger equation. [This discussion has been diverted to an Appendix because it is not directly relevant to obtaining the evolution equations; but it is certainly relevant to relating Secs. 3 and 5. Equations (5.5)–(5.8) are in the Appendix.]

C. The evolution equation and off-diagonalness of Q

The restrictions: that $Q = Q^{\mathrm{off}}$ (i.e., that the diagonal part of Q, Q^{diag}, be zero), that σ be diagonal, of zero trace, and that σ have distinct nonzero eigenvalues, are not all necessary. However they do simplify our results.[13] The potential, Q, is a member of the $N \equiv n^2 - n$ dimensional linear space of off-diagonal matrices. Whereas AKNS obtained an evolution in a two-dimensional linear space (for them $n = 2 \Longrightarrow N = 2^2 - 2 = 2$) we will obtain an evolution equation in an N-dimensional linear space; N-dimensional vectors will be given hats and the $N\times N$ matrices will be underlined to distinguish them from the n-dimensional ones. Notice that the possible values of N are 2, 6, 12, etc. so three-dimensional evolution equations are not naturally obtained from (5.1). The in-between cases ought to be obtainable by relaxing the off-diagonal restriction on Q.

D. Derivation of the reduced commutator equation

Consider the commutator equation

$$H_t F = [B, H]F, \quad H \equiv \sigma^{-1}(iD - iQ) \tag{5.9}$$

for the reduced operator $B \equiv F_t F^{-1}$,

i.e.,

$$F_t = BF \text{ with } F\text{'s right-normalization appropriately fixed.} \tag{5.10}$$

Remark: Although $(FW)_x = F_x W = ZFW$ whenever $W_x = 0$, Eq. (5.10), then becomes $(FW)_t = BFW + FWA_0$ where $A_0 \equiv W^{-1}W_t$; so (5.10) only holds when F is appropriately normalized.

Now

$$\begin{aligned}[B, H] &= [B, \sigma^{-1}(iD - iQ)] \\ &= ([B, \sigma^{-1}]\sigma H + i\sigma^{-1}[B, D - Q]) \\ &= i\sigma^{-1}(-i\sigma[B, \sigma^{-1}]\sigma H + [B, D - Q]) \\ &= i\sigma^{-1}(i[B, \sigma]H + [B, D - Q]).\end{aligned}$$

Applied to F, this can be reduced by replacing the H on the right-hand side with k. Since the operator form of H remains unchanged $H_t = -i\sigma^{-1}Q_t$ and (5.9) now becomes

$$Q_t = [D - Q + ik\sigma, B] = [D - Z, B] \tag{5.11a}$$

$$= B_x - [Q, B] + ik[\sigma, B]. \tag{5.11b}$$

Equation (5.11a) can be obtained with less manipulation by using the form (5.1a) of (5.1), $Q - ik\sigma = Z = F_x F^{-1}$, to consider Z_t as follows[14]:

$$\begin{aligned}Q_t = Z_t = (F_x F^{-1})_t &= F_{xt}F^{-1} - F_x F^{-1}F_t F^{-1} \\ &= (BF)_x F^{-1} - ZB \\ &= B_x + [Z, B] \\ &= [D - Z, B].\end{aligned}$$

{*Notes*: (i) We thus also obtain the precursor of an orthogonality property, using $s = x$ or t, $F^{-1}Q_s F = (F^{-1}F_s)_x$; cf. Eq. (4.13) in the Schrödinger case.

(ii) The homogeneous solutions of (5.11), those when $Q_t = 0$, are obtained by using the operator equality $D - Z = FDF^{-1}$ which gives us

$$[D - Z, B] = F[D, F^{-1}BF]F^{-1} = F(F^{-1}BF)_x F^{-1}.$$

Thus FA_0F^{-1} with $A_{0x} = 0$, is the general homogeneous solution.}

E. The auxiliary equation

Before solving (5.11) for B we have to introduce the properties $Q = Q^{\mathrm{off}}$ and $\sigma = \sigma^{\mathrm{diag}}$. An immediate and obvious consequence of $Q = Q^{\mathrm{off}}$ is that $(\mathrm{tr}B)_x = 0$. Clearly the addition of x-independent multiples of the $n\times n$ identity to B leave $H_t = [B, H]$ invariant: *We can thus choose* $\mathrm{tr}B = 0$ *without loss of generality*.

To introduce the stated properties systematically into (5.11) we will separate the diagonal part, (5.11diag),

and off-diagonal part (5.11off), of Eq. (5.11b). In doing so it is useful to remark that: the product of two off-diagonal matrices has a diagonal part (in general); the product of a diagonal and off-diagonal matrix is off-diagonal while, obviously, the product of two diagonal matrices is diagonal. For example, it follows that $[B^{\rm off}, \sigma]^{\rm diag} = 0 = [B^{\rm diag}, Q]^{\rm diag}$. Thus (5.11b) is equivalent to

$$Q_t = B_x^{\rm off} - [Q, B^{\rm off}]^{\rm off} - [Q, B^{\rm diag}] + ik[\sigma, B^{\rm off}], \tag{5.11off}$$

$$0 = B_x^{\rm diag} - [Q, B^{\rm off}]^{\rm diag}. \tag{5.11diag}$$

Notice that, as a consequence of σ being diagonal, (5.11diag) is independent of k; as a consequence of Q being off-diagonal (5.11diag) is homogeneous and can be immediately integrated,

$$B^{\rm diag} = \overset{\infty}{\gamma} + D^{-1}[Q, B^{\rm off}]^{\rm diag}, \tag{5.12diag}$$

where $D^{-1} = \int_\infty^x$ [i.e. $D^{-1}g(x) = -\int_x^\infty dy\, g(y)$], $\overset{\infty}{\gamma} \equiv \lim_{x\to\infty} B^{\rm diag}$. (5.12diag) explicitly gives us $B^{\rm diag}$ in terms of its limit as $x \to \infty$, $\overset{\infty}{\gamma}$ (*a diagonal $n\times n$ matrix*, $\mathrm{tr}\overset{\infty}{\gamma} = 0 = \overset{\infty}{\gamma}_x$), and in terms of $B^{\rm off}$. Substituting into (5.11off) and defining $b \equiv B^{\rm off}$ (*an off-diagonal $n\times n$ matrix*, thus with $N \equiv n^2 - n$ independent components) gives us

$$Q_t = b_x - [Q, b]^{\rm off} - [Q, D^{-1}[Q, b]^{\rm diag}] - ik[b, \sigma] + [\overset{\infty}{\gamma}, Q]. \tag{5.12off}$$

Equations (5.12) imply (5.11) and also have $Q = Q^{\rm off} \to 0$ as $x \to \infty$ built in. Equation (5.12off) is a form of the reduced auxiliary equation.

F. Vector form of the auxiliary equations

It is convenient to rewrite (5.12off) as a matrix equation in l^N where an off-diagonal matrix $A^{\rm off}$ is represented as a column with N entries, $\hat{A}$, and where a linear operator on $A^{\rm off}$, say $\underline{\sigma}$: $A^{\rm off} \to \underline{\sigma}(A^{\rm off}) \equiv [A^{\rm off}, \sigma]$, is represented by an $N\times N$ matrix $\underline{\sigma}$. So $\underline{\sigma}\hat{A} \equiv \underline{\sigma}(A^{\rm off})$. Matrices operating in l^N will be underlined and vectors in l^N will have hats to distinguish them from those in l^n. Now in vector notation in l^N, (5.12off) becomes

$$\hat{Q}_t = (\underline{I}D - \underline{Q} - ik\,\underline{\sigma})\hat{b} + \underline{\overset{\infty}{\gamma}}(k)\hat{Q}, \tag{5.12vec}$$

i.e.,

$$\hat{Q}_t = (\underline{L} - k\underline{I})i\underline{\sigma}\hat{b} + \underline{\overset{\infty}{b}}(k)i\underline{\sigma}\hat{Q}, \tag{5.13}$$

where

$$\begin{aligned}
&\underline{Q}: A^{\rm off} \to [Q, D^{-1}[Q, A^{\rm off}]^{\rm diag}] + [Q, A^{\rm off}],\\
&\underline{\overset{\infty}{\gamma}}: A^{\rm off} \to [\overset{\infty}{\gamma}, A^{\rm off}],\\
&\underline{\sigma}: A^{\rm off} \to [A^{\rm off}, \sigma],
\end{aligned} \tag{5.14a}$$

$$\underline{\sigma}^{-1} \text{ is the inverse of } \underline{\sigma},$$

$$\underline{\overset{\infty}{b}}(k) \equiv -i\underline{\overset{\infty}{\gamma}}(k)\underline{\sigma}^{-1}, \tag{5.14b}$$

and

$$\underline{L} \equiv -i(\underline{I}D - \underline{Q})\underline{\sigma}^{-1}. \tag{5.14c}$$

Note: $\underline{\sigma}$ is invertable if the eigenvalues of σ are distinct since under this condition σ only commutes with diagonal matrices,

$$[e(l,m), \sigma] = e(l,m)(\sigma_{mm} - \sigma_{ll}), \tag{5.14d}$$

where $e(l,m)_{kj} = \delta_{lk}\delta_{mj}$ is the natural basis for $n\times n$ matrices. [The complexities due to giving up the restriction $Q = Q^{\rm off}$ are similar to those due to permitting say $\sigma_{11} = \sigma_{22}$, because in the latter case the off-diagonal matrix $Q_{12}e(1,2) + Q_{21}e(2,1)$ commutes with σ; so $\underline{\sigma}$ has a null space in both these cases.]

Equation (5.13) is the form of the reduced auxiliary equation which we will now solve for $\hat{b}$, following the model given when solving (3.18) in the Schrödinger case.

G. The evolution equation obtained from the auxiliary equation

We rearrange (5.13) as follows (at least formally):

$$\begin{aligned}
\hat{Q}_t - \underline{\overset{\infty}{\gamma}}(\underline{L})\hat{Q} &= (\underline{L} - k\underline{I})i\underline{\sigma}\hat{b} + (\underline{\overset{\infty}{b}}(\underline{L}) - \underline{\overset{\infty}{b}}(k))i\underline{\sigma}\hat{Q}\\
&= (\underline{L} - k\underline{I})\{i\underline{\sigma}\hat{b} - (\Delta\underline{\overset{\infty}{b}})i\underline{\sigma}\hat{Q}\}.
\end{aligned} \tag{5.15a}$$

Here

$$\begin{aligned}
(\Delta\underline{\overset{\infty}{b}}) &\equiv (\Delta\underline{\overset{\infty}{b}})(\underline{L}, k) \equiv (\underline{L} - k\underline{I})^{-1}(\underline{\overset{\infty}{b}}(\underline{L}) - \underline{\overset{\infty}{b}}(k))\\
&= \sum_{l,m} (\underline{L} - k\underline{I})^{-1}(\underline{\overset{\infty}{b}}_{lm}(\underline{L}) - \overset{\infty}{b}_{lm}(k))e(l,m)\\
&\equiv \sum_{l,m} (\Delta\underline{\overset{\infty}{b}}_{lm})e(l,m)
\end{aligned} \tag{5.16a}$$

and

$$\begin{aligned}
(\Delta\underline{\overset{\infty}{b}}_{lm}) &\equiv (\Delta\underline{\overset{\infty}{b}}_{lm})(\underline{L}, k) \equiv (\underline{L} - k\underline{I})^{-1}(\underline{\overset{\infty}{b}}_{lm}(\underline{L}) - \overset{\infty}{b}_{lm}(k))\\
&= \underline{\overset{\infty}{b}}_{lm}(1; l) + \sum_{j\geq 2} \underline{\overset{\infty}{b}}_{lm}(j; l) \sum_{i=1}^{j-1} \underline{L}^{j-i}k^i,
\end{aligned} \tag{5.16b}$$

given that the matrix entries of $\underline{\overset{\infty}{b}}(k)$ and $\underline{\overset{\infty}{b}}_{lm}(k)$, are polynomial (or entire) in k, with expansions

$$\underline{\overset{\infty}{b}}_{lm}(l,k) = \sum_{j\geq 0} \underline{\overset{\infty}{b}}_{lm}(j; l)k^j \text{ (showing the } l \text{ dependence).} \tag{5.16c}$$

[We apologize for the flood of indices in Eqs. (5.16).] The same method as employed in the Schrödinger case also works here to generalize the result to ratios of polynomial (or entire) functions of k—we will not repeat it here.

Assuming that $\underline{\overset{\infty}{b}}(k)$ is regular in k at all points where $\underline{L}$ has one of its (assumed) finite number of proper eigenvalues, and that $\hat{b} \equiv \hat{b}(t, x, k)$ is regular in k everywhere where $\underline{\overset{\infty}{b}}(k)$ is regular, we conclude that both sides of (5.15a) must be zero for the same reasons as in the Schrödinger case. Namely, (5.15a) is equivalent to

$$\{i\underline{\sigma}\hat{b} - (\Delta\underline{\overset{\infty}{b}})i\underline{\sigma}\hat{Q}\} = (\underline{L} - k\underline{I})^{-1}[\hat{Q}_t - \underline{\overset{\infty}{\gamma}}(\underline{L})\hat{Q}]. \tag{5.15b}$$

Both sides of (5.15b) must be zero because one side or the other is regular in k at all points $k \in \mathbb{C}$, while the right-hand side tends to zero as $|k| \to \infty$. The results are the solution giving $\hat{b}$ in terms of $\underline{\overset{\infty}{\gamma}}$ [i.e., terms of $\overset{\infty}{\gamma} \equiv \overset{\infty}{\gamma}(l,k) \equiv \lim_{x\to\infty} B^{\rm diag}$] and the evolution equation

$$\hat{Q}_t = \underline{\overset{\infty}{\gamma}}(l, \underline{L})\hat{Q}. \tag{5.17}$$

We note that, since $\mathrm{tr}B^{\rm diag} = 0$, our dispersion relation, $\overset{\infty}{\gamma}$, contains $n-1$ scalar functions of k and l.

H. Relation to the scattering data

We have *obtained the evolution equation* (5.16) for Q

on the basis of analyticity assumptions in the k plane, and we have done so *without solving the inverse problem* for (5.1). Since the inverse problem has not been solved we cannot relate $\overset{\infty}{\gamma}(t, k)$ to the inverse scattering data, because the latter is unknown. (Most recent studies[9] which obtain evolution equations for linear operators share this difficulty.)

However it is straightforward to relate $\overset{\infty}{\gamma} \equiv \lim_{x\to\infty} B^{\text{diag}}$ to the asymptotic behavior of F as $x \to \infty$. Equation (5.1c) becomes

$$F \sim \overset{\infty}{F} \equiv \exp(-ikx\sigma)\overset{\infty}{F}_0, \quad \overset{\infty}{F}_{0x} = 0. \tag{5.18a}$$

Now use Eq. (5.10), taking heed of the remark which follows it concerning the right-hand normalization of F,

$$B \equiv F_t F^{-1} - FA_0F^{-1} \Longrightarrow \overset{\infty}{B} = \overset{\infty}{F}_t\overset{\infty}{F}^{-1} - \overset{\infty}{F}A_0\overset{\infty}{F}^{-1}, \tag{5.18b}$$

where $A_{0x} = 0$, and $A_0 = 0$ if and only if E is appropriately right-normalized. The condition $\text{tr}\overset{\infty}{B} = 0$ is equivalent to

$$\text{tr}\overset{\infty}{F}_t F^{-1} = \text{tr}\overset{\infty}{F}_{0t}\overset{\infty}{F}_0^{-1} = \text{tr}A_0 = 0, \tag{5.18c}$$

a condition on the right-normalization of F. We can also adjust the right-normalization so that $\overset{\infty}{B} = \overset{\infty}{B}^{\text{diag}} \equiv \overset{\infty}{\gamma}$, i.e., so that $\lim_{x\to\infty} B^{\text{off}} = 0$. This last condition is equivalent to

$$(\overset{\infty}{F}_{0t}\overset{\infty}{F}_0^{-1})^{\text{off}} = (\overset{\infty}{F}_0 A_0 \overset{\infty}{F}_0^{-1})^{\text{off}} = 0, \tag{5.18d}$$

again a condition on the right-normalization of F. Given these conditions

$$\overset{\infty}{\gamma} = \lim_{x\to\infty} F_t F^{-1}. \tag{5.18e}$$

Recalling (5.1d), its analog with respect to $F_t = BF$ ($\Longrightarrow \tau_t = \tau \text{tr}B$) is

$$\tau_t = 0 \Longleftrightarrow \text{tr}B = 0. \tag{5.18f}$$

This gives us another interpretation, in terms of $\tau \equiv \det F$, of $\text{tr}B = 0$.

Further than this we cannot go without solving, or at least starting to solve, the inverse problem for (5.1). This is not within the scope of this article. However, it may turn out that finding the evolution equation may help in solving the inverse problem—at least in identifying the inverse scattering data. This need not be surprising, if true, because the approach is based on analyticity in the energy variable as are the solutions of inverse problem and the related completeness relations.

I. The AKNS data as the 2 × 2 case

When $n = 2$ and $\sigma = \sigma_3$ the following choice of the right-hand normalization of F,

$$\overset{\infty}{F}_0 = \begin{bmatrix} 1/a & 0 \\ R^+/a & \tau a \end{bmatrix}, \quad a = (R^+)^{1/2}, \tag{5.19a}$$

immediately yields $\overset{\infty}{B} = \lim_{x\to\infty} F_t F^{-1} = \overset{\infty}{F}_{st}\overset{\infty}{F}_0^{-1} = \overset{\infty}{B}^{\text{diag}} = \overset{\infty}{\gamma}(t, k)$,

i.e.

$$\begin{aligned}\overset{\infty}{\gamma}(t, k) &= \exp(-ikx\sigma)\begin{bmatrix} -a_t/a & 0 \\ R_t^+ - 2R^+a_t/a & a_t/a \end{bmatrix}\exp(+ikx\sigma) \\ &= -\tfrac{1}{2}\sigma_3(\ln R^+(t, k))_t \equiv -\sigma_3\Omega(t, k).\end{aligned} \tag{5.19b}$$

Here R^+ is the coefficient of reflection to the right, τ is the transmission coefficient. {The existence of a solution with "normalization" $\overset{\infty}{F}_0$ can be proven by separately imposing appropriate asymptotic conditions on the first and second columns of $F \equiv [\mathbf{F}_1 \vdots \mathbf{F}_2]$, e.g., for $\text{Im}k \geq 0$,

$$\exp(-ikx)\begin{bmatrix}1\\0\end{bmatrix} \underset{x\to-\infty}{\sim} a\tau^{-1}\mathbf{F}_1\left(\underset{x\to+\infty}{\sim}\begin{bmatrix}\exp(-ikx)\tau^{-1}\\ \exp(+ikx)\tau^{-1}R^+\end{bmatrix}\right),$$

$$\exp(+ikx)\begin{bmatrix}0\\1\end{bmatrix} \underset{x\to+\infty}{\sim} (a\tau)^{-1}\mathbf{F}_2\left(\underset{x\to-\infty}{\sim}\begin{bmatrix}\exp(-ikx)\tau^{-1}R^-\\ \exp(+ikx)\tau^-\end{bmatrix}\right),$$

which are linearly independent if and only if $\tau^{-1} \neq 0$; if $\lim_{k\to k_n}\tau^{-1}(k) = 0$ then $\lim_{k\to k_n}\tau^{-1}(k)R^+(k) = C_n^+ \neq 0$ exists. (Cf. The Jost solutions for the Schrödinger case).}

J. The AKNS evolution equation as the 2 × 2 case

We end this section by showing that the 2×2 case of (5.17) is the AKNS evolution equation. We first return to the off-diagonal matrix form of our reduced auxiliary equation, (5.12off). In the 2×2 case this simplicies because when $n = 2$: $\sigma = \sigma_3$; $\overset{\infty}{\gamma}(t, k) = -\Omega(t; k)\sigma_3$;

$$[b, \sigma_3] = -2\sigma_3 b = +2b\sigma_3 \quad \text{(recall that } b \equiv B^{\text{off}}\text{)};$$

$$[\overset{\infty}{\gamma}, Q] = [-\Omega\sigma_3, Q] = [Q, \sigma_3]\Omega = -2\Omega\sigma_3;$$

$$[Q, b] = [Q, b]^{\text{diag}} = (Q_{12}B_{21} - B_{21}Q_{21})\sigma_3.$$

Thus (5.12off) becomes (recall that $Q_{12} = q$, $Q_{21} = r$)

$$Q_t \equiv b_x - 2\sigma_3 QD^{-1}(qb_{12} - rb_{21}) + 2ik\sigma_3 b - 2\Omega\sigma_3 Q. \tag{5.20off}$$

The vector form can in this case (since $N \equiv n^2 - n = 2^2 - 2 = 2 = n$) is easily obtained by applying (5.20off) to $\begin{bmatrix}1\\0\end{bmatrix} \equiv \hat{e}$. Defining $\hat{Q} \equiv Q\hat{e} \equiv \begin{bmatrix}q\\r\end{bmatrix}$, $\hat{b} \equiv b\hat{e}$ (5.20off) takes the "vector" form of (5.12vec) if we also define $\underline{Q}$, $\underline{\sigma}$, and $\underline{\overset{\infty}{\gamma}}(k)$ by

$$\underline{Q} \equiv 2\sigma_3\begin{bmatrix} qD^{-1}q & qD^{-1}r \\ rD^{-1}q & rD^{-1}r \end{bmatrix}\sigma_3, \tag{5.21a}$$

$$\underline{\sigma} \equiv -2\sigma, \quad \underline{\gamma}(t, k) \equiv -2\sigma_3\Omega(t, k).$$

Then Eq. (5.14c) gives us

$$\underline{L} = -\tfrac{1}{2}(\underline{I}D - \underline{Q})i\sigma_3. \tag{5.21b}$$

Finally the evolution equation, (5.17), becomes

$$\hat{Q}_t = \underline{\overset{\infty}{\gamma}}(t, \underline{L})\hat{Q} = -2\sigma_3\Omega(t, \underline{L})\hat{Q}, \tag{5.22}$$

where

$$\Omega(t, k) = \tfrac{1}{2}[\ln R^+(t, k)]_t \tag{5.23}$$

is the AKNS dispersion relation.

6. CONCLUSION

The method outlined abstractly in Sec. 1, which worked for the one-dimensional Schrödinger case in Sec. 3, also works for the $n\times n$ matrix equation and yields the evolution equation (5.17) which, when $n = 2$, reproduces the AKNS evolution equation (5.22). The analyticity assumption and existence have only been proven for the Schrödinger case, not the matrix case.

We feel that the most important general problems with the IST method at the moment are to extend it system-

atically to the cases with several space dimensions and to find a systematic way of going backwards, that is of finding the linear equation associated with a given nonlinear evolution when it can be solved this way. We hope that the method presented here will help to do this.

ACKNOWLEDGMENTS

We would like to sincerely thank Marcel Jaulent and Pierre Sabatier for their help during numerous discussions and during the preparation of this paper.

APPENDIX: CERTAIN 2 × 2 CASES AND THE SCHRÖDINGER EQUATION

In the 2×2 case when $r=q=q_1$ (so $q_2=0$), if one considers the equation satisfied by (cf. Krein,[15] 1953)

$$f^K \equiv T_K f \equiv \frac{1}{\sqrt{2}}(\sigma_1+\sigma_3)f, \tag{5.5}$$

it is easy to show that the components of f^K, $f_\pm^K \equiv (f_1 \pm f_2)/\sqrt{2}$, respectively satisfy one-dimensional Schrödinger equations

$$(-k^2+H_\pm^{\mathrm{Sch}})f_\pm^K \equiv (-D^2+v_\pm^{\mathrm{Sch}}-k^2)f_\pm^K = 0, \tag{5.6a±}$$

where

$$v_\pm^{\mathrm{Sch}} = q^2 \pm q_x = (\psi_0^{\pm1})_{xx}/\psi_0^{\pm1}. \tag{5.6b}$$

The components are related by

$$f_-^K = k^{-1}(iD-iq)f_+^K. \tag{5.6c}$$

The Riccati equations (5.6b) are both solved for q by

$$q = (\ln\psi_0)_x \equiv \pm(\ln\psi_0^{\pm1})_x, \tag{5.7}$$

where ψ_0 is a zero energy (i.e., $k=0$) solution of (5.6a+). It follows that if there exists a bounded zero-energy solution ψ_0 of (5.6a+) which has no zeros (as a function of x), and this can only happen if H_+^{Sch} has no proper eigenvalues (see Martin and Sabatier[16]), then H has no proper eigenvalues either and furthermore (5.6a+) is equivalent to the AKNS case with $q^*=q=r$. It also follows that self-adjoint H_+^{Sch} has proper eigenvalues if and only if H (with $q^*=q=r$) also has proper eigenvalues, but now we must admit potential functions $q=(\ln\psi_0)_x \equiv \psi_{0x}/\psi_0$ where ψ_0 can have zeros as a function of x. If $Q=\left[\begin{smallmatrix}0 & q\\ q & 0\end{smallmatrix}\right]$ is so generalized then this class of AKNS equations is equivalent to the Schrödinger equation.

Without going into details we also state that in the more general case, (4.4b), the AKNS equation can be shown to be equivalent[17] to a Schrödinger equation with an energy dependent potential of the kind

$$v(x,k) = v_1(x) + kv_2(x). \tag{5.8}$$

The inverse problem for this has been extensively studied[18] and the associated evolution equation has been derived by an analog of the AKNS procedure.[12]

[1]C.S. Gardner, J.M. Greene, M.D. Kruskal, and R.M. Miura, Phys. Rev. Lett. **19**, 1095 (1967); Commun. Pure Appl. Math. **XXVII**, 97 (1974).
[2]P.D. Lax, Commun. Pure Appl. Math. **XXI**, 467 (1968); **XXVIII**, 141 (1975).
[3]V.E. Zakharov and A.B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)]; **64**, 1627 (1973) [**37**, 823 (1974)].
[4]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Stud. Appl. Math. **LIII**, 249 (1974). Many physically interesting cases are discussed.
[5]N. Dunford and J.T. Schwartz, *Linear Operators, Part II* (Wiley, New York, 1963), p. 1217, Theorem 16. [See also the remark by P.D. Lax, p. 293, chapter XI in S. Leibovich and A.R. Seebass, Eds., *Nonlinear Waves* (Cornell U.P., Ithaca, New York, 1974).]
[6]F. Calogero and A. Degasperis, Nuovo Cimento B **32**, 201 (1976). (See also preprint submitted in September 1976).
[7]The linear operator Λ, and the associated operators L^* (and $\underline{L}$ in the matrix case) arise here naturally without needing to be separately constructed; cf. the AKNS method.
[8]Apart from the $n\times n$ matrix case in Sec. 5, our method is showing promise for the ν-dimensional Schrödinger case.
[9]Evolution equations of a matrix type are also found and studied, and their physical application discussed in: (a) V.E. Zakharov and S.V. Manakov, Zh. Eksp. Teor. Fiz. Pis. Red. **18**, 413 (1973) [JETP Lett. **18**, 243 (1973)]; (b) M.J. Ablowitz and R. Haberman, J. Math. Phys. **16**, 2301 (1975). With reference to our method we take note of Refs. 16 and 17 of (b).
[10]D.J. Kaup, J. Math. Anal. Appl. **54**, 849 (1976).
[11]M. Jaulent, "Completeness of Squared Schrödinger Eigenfunctions" (unpublished, in preparation).
[12]M. Jaulent and I. Miodek, Lett. Math. Phys. **1**, 243 (1976).
[13]I. Miodek, Preprint (PM/77/1), Lab. de Math-Physique, Université des Sciences et Techniques du Languedoc, Montpellier, France. In this reference we show that a less restricted class of linear first order matrix differential equations can be reduced to Eq. (5.1) by invertible linear transformations; their inverse problems will thus not be independent. It can be expected that the evolution equations of the more general linear equations will not be independent from the ones obtained for Eq. (5.1).
[14]In some respects this other way of obtaining our auxiliary equation is preferable because it seems to involve less manipulation and gives us other important properties in a uniform way, as we have shown. It works in both the Schrödinger and matrix cases and happens to be the way the author first found the auxiliary equation to be useful. Insofar as it is based on direct cross-differentiation it is easily compared to the methods in Ref. 9(b) and the original method of GGKM in Ref. 1. The present presentation has the advantage of abstract generality.
[15]See review by L.D. Faddeev, J. Math. Phys. **1**, 72 (1973), in particular his Eq. (11.19) and his note n° 11.
[16]A. Martin and P. Sabatier, "Impedence, zero energy wave function, and bound states," Preprint (PM/77/2) Physique Mathématique U.S.T.L., Montpellier, France; to be published in J. Math. Phys.
[17]M. Jaulent and I. Miodek, Lett. Nuovo Cimento (to be published).
[18](a) M. Jaulent, Ann. Inst. Henri Poincaré **XVII**, 363 (1972); (b) M. Jaulent and C. Jean, Commun. Math. Phys. **28**, 177 (1972); Ann. Inst. Henri Poincaré **XXV**, 105, 119 (1976).

Dense electron-gas response at any degeneracy

C. Gouedard and C. Deutsch

Laboratoire de Physique des Plasmas, Université Paris XI, 91405 Orsay, France
(Received 28 March 1977)

An exact expression for the linear response function of the dense electron gas valid at any temperature is worked out in the ring (RPA) approximation. The $T=0$ and $T=\infty$ limits reproduce the already known results. It is used to explain the longitudinal oscillations and the screening around a test charge. The latter is either Thomas–Fermi-like or Friedel-like according to the values of the parameters.

I. INTRODUCTION

As a prerequisite to the study of the thermal conductivity in the dense partially degenerate electron gas encountered in the final stage of laser fusion experiments,[1] we have to work out the properties of the dynamic dielectric function $\epsilon(\mathbf{q}, \omega)$ given by

$$\epsilon(\mathbf{q}, \omega) = 1 - V(q)F(\mathbf{q}, \omega) \tag{I. 1}$$

with $V(q) = 4\pi e^2 q^{-2}$, in terms of the dynamic linear response function of the electron gas[2]

$$F(\mathbf{q}, \omega) \equiv F^0(\mathbf{q}, \omega) = -(2S+1)\int \frac{d^3\mathbf{k}}{(2\pi)^3} \cdot \frac{n^0(\mathbf{k}+\mathbf{q}) - n^0(\mathbf{k})}{\hbar\omega + i\eta - (\epsilon^0_{\mathbf{k}+\mathbf{q}} - \epsilon^0_{\mathbf{k}})} \tag{I. 2}$$

taken in the dense random phase approximation (RPA). η is a small positive convergence parameter. Henceforth, we neglect any spin dependence (except for an overall factor of two) in the electron–electron interaction. $\epsilon(\mathbf{q}, \omega)$ is the cornerstone of any Boltzmann-like approach through the electron–ion interaction to the time-independent transport coefficients. The main purpose of the present work is to investigate for any degree of degeneracy the quantities $\epsilon(\mathbf{q}, \omega)$ and $F^0(\mathbf{q}, \omega)$ with the usual fermionic distribution ($\beta^{-1} = k_B T$, μ is the grand canonical chemical potential)

$$n^0(k) = \frac{1}{\exp[\beta(\epsilon^0_k - \mu)] + 1}, \quad \epsilon^0_k = \frac{\hbar^2 k^2}{2m_e}. \tag{I. 3}$$

At this preliminary stage, the Coulomb interaction does not play any role, as far as the linear response is concerned. Equation (I. 2) has already been considered for any degree of degeneracy[3] in a condensed matter framework, where it defines the dynamic susceptibility in connection with inelastic neutron scattering from liquid[3] He. Here, our basic motivation is to bridge the gap between the dilute high-temperature nondegenerate electron plasma worked out by Vlasov and the well known $T=0$ dense liquid electron approximation.[2] Both pictures delineate the extreme limits of the random phase approximation when applied to the one-component plasma with a fixed neutralizing continuum background. From the start, we postulate that this very simple model provides the dominant contribution to the linear response in the real dense electron gas up to any nonzero temperature. Furthermore, it will remain to investigate in a future work the temperature dependence of the exchange corrections. In Eq. (I. 2), $\hbar\mathbf{q}$ is the momentum and $\hbar\omega$ is the energy transferred in a given scattering process (neutron, light, etc, $\cdots$) probing the electron plasma fluctuations. In the sequel, it will prove useful to introduce the dimensionless variables $Q = q/k_F$ and $\nu = \hbar\omega/\epsilon^0_F$, in terms of the corresponding $T = 0\,^\circ$K Fermi quantities. This way of picturing the linear response function may be given further support through the well-known relationship

$$\frac{1}{\epsilon(\mathbf{q}, \omega)} = 1 + V(q)\int_{-\infty}^{\infty} d\omega' S(\mathbf{q}, \omega') \times \left(\frac{1}{\omega - \omega' + i\eta} - \frac{1}{\omega + \omega' + i\eta}\right), \tag{I. 4}$$

between the dielectric function and the Van Hove structure factor[4] $S(\mathbf{q}, \omega)$. As is well known, Eq. (I. 4) provides, through its Fourier transform, a direct insight into the equilibrium properties of the partially degenerate electron gas. Our ultimate goal is to explain the real and the imaginary parts of the response function, through a kind of Padé-like interpolation procedure between the well-known classical results (Fried–Conte[5]) and the $T=0$ RPA (Lindhard[6]) limit.

The present paper is organized as follows:

In Sec. II we rederive the real and the imaginary parts of the RPA response function with techniques quite similar, although differing in some important respect, to those already used recently by Khanna and Glyde.[3] The present treatment leads easily to a convergent real part series. These results are numerically displayed in Sec. III and their static limit detailed in Sec. IV. The high temperature limit is considered in Sec. V, and the longitudinal plasma modes are explained in Sec. VI. The last sections are devoted to the screening property around a test change.

II. DYNAMIC LINEAR RESPONSE

Our present task is now to explain Eq. (I. 2) for any finite temperature. A simple change of variable makes it appear as

$$F^0(\mathbf{q}, \omega) = -\frac{2}{(2\pi)^3}\int d^3\mathbf{k}\, n^0(k) \times \left(\frac{1}{\hbar\omega + i\eta - (\epsilon^0_{\mathbf{k}} - \epsilon^0_{\mathbf{k}-\mathbf{q}})} - \frac{1}{\hbar\omega + i\eta - (\epsilon^0_{\mathbf{k}+\mathbf{q}} - \epsilon^0_{\mathbf{k}})}\right). \tag{II. 1}$$

Introducing in the rhs the dimensionless momentum $\mathbf{k}' = \mathbf{k}/k_F$ together with $\gamma = \mu/\epsilon^0_F$, $Z = k_B T/\epsilon^0_F$, the Fermi distribution function (I. 3) becomes $n^0(k') = \{\exp[(k'^2 - \gamma)Z^{-1}] + 1\}^{-1}$, and we obtain

$$F^0(\mathbf{q}, \omega) \equiv \frac{-k_F^3}{(2\pi)^2 \epsilon^0_F} Q^{-1}[f^0(p_+) - f^0(p_-)], \tag{II. 2}$$

 0022-2488/78/1901-0032$1.00

where

$$p_\pm = \frac{\nu}{2Q} \pm \frac{Q}{2},$$

and

$$f^0(p) = \int_0^\infty dk\, k^2 n^0(k) \int_{-1}^{+1} \frac{du}{p + i\eta - ku}, \tag{II. 3}$$

with

$$\mathrm{Im}F^0(\mathbf{q}, \omega) = \frac{-k_F^3 r_s \alpha Z}{4\pi^2 Q}\{\ln(1 + \exp[(\gamma - p_+^2)/Z]) - \ln(1 + \exp[(\gamma - p_-^2)/Z])\}. \tag{II. 4}$$

r_s is the usual dimensionless interelectronic distance in number of a_0, $\alpha = (9\pi/4)^{-1/3}$. As already noticed,[3] $\mathrm{Re}F^0(\mathbf{q}, \omega)$ is a much more involved quantity. To evaluate it, let us start from Eq. (II. 3) and compute

$$\frac{\partial}{\partial p}[\mathrm{Re}f^0(p)] = \int_{-\infty}^{\infty} dk\, n^0(k) - p^2 PP \int_{-\infty}^{\infty} \frac{dk n^0(k)}{p^2 - k^2}. \tag{II. 5}$$

The last quadrature may be interpreted as a contour integral, because its integrand behaves as k^{-2} when $|k| \to \infty$. Only the $n^0(k)$ poles are to be taken into account. They are an infinite number.[3] The n^{th} one is given by

$$k_n^2 = \gamma + i(2n + 1)Z, \quad n = -\infty, \ldots, \infty. \tag{II. 6}$$

The location of the poles in the complex k plane, is deduced from

$$k_n = a_n + ib_n \tag{II. 7a}$$

$$a_n^2 - b_n^2 = \gamma, \tag{II. 7b}$$

$$2a_n b_n = (2n + 1)\pi Z, \tag{II. 7c}$$

$$r_n^2 = a_n^2 + b_n^2, \tag{II. 7d}$$

through the intersection of the two hyperbolas $x^2 - y^2 = \gamma$ and $2xy = (2n + 1)\pi Z$. In the upper half-plane, we retain only $k_n = a_n + ib_n$, with a_n and $b_n > 0$, where

$$a_n = \frac{1}{\sqrt{2}}\{\gamma + [\gamma^2 + (2n + 1)^2\pi^2 Z^2]^{1/2}\}^{1/2},$$

$$b_n = \frac{1}{\sqrt{2}}\{-\gamma + [\gamma^2 + (2n + 1)^2\pi^2 Z^2]^{1/2}\}^{1/2}. \tag{II. 8}$$

The single contour integral may be closed up by a semicircle in the upper k half plane, yielding

$$\int_{-\infty}^{\infty} \frac{dk n^0(k)}{p^2 - k^2} = \frac{-2\pi Z}{p^2} \cdot \sum_{n=0}^{\infty} \left(\frac{b_n}{r_n^2} + \frac{b_n(p^2 + r_n^2)}{(p^2 - r_n^2)^2 + 4p^2 b_n^2} \right). \tag{II. 9}$$

Equation (II. 5) then becomes

$$\frac{\partial}{\partial p}[\mathrm{Re}f^0(p)] = 2\int_0^\infty dk\, n^0(k) + 2\pi Z \sum_{n=0}^{\infty} \times \left(\frac{b_n}{r_n^2} + \frac{b_n(p^2 + r_n^2)}{(p^2 - r_n^2)^2 + 4p^2 b_n^2} \right). \tag{II. 10}$$

It must be pointed out that the two discrete sums on the last line cannot converge separately, since a_n, b_n, and r_n grow up as $n^{1/2}$. Finally, we thus obtain

$$F^0(\mathbf{q}, \omega) = -\frac{\alpha r_s k_F^2}{\pi^2 e^2} g(\mathbf{q}, \omega),$$

where

$$\mathrm{Im}g(\mathbf{q}, \omega) = \frac{\pi Z}{Q} \ln\left(\frac{1 + \exp[(\gamma - p_+^2)/Z]}{1 + \exp[(\gamma - p_-^2)/Z]} \right) \tag{II. 11}$$

and

$$\mathrm{Re}g(\mathbf{q}, \omega) = \int_0^\infty dk\, n^0(k) + \pi Z \sum_{n=0}^{\infty} \left\{ \frac{b_n}{r_n^2} - \frac{1}{2Q} \left[\mathrm{Arctan}\left(\frac{p_+ + a_n}{b_n} \right) + \mathrm{Arctan}\left(\frac{p_+ - a_n}{b_n} \right) - \mathrm{Arctan}\left(\frac{p_- + a_n}{b_n} \right) - \mathrm{Arctan}\left(\frac{p_- - a_n}{b_n} \right) \right] \right\}, \tag{II. 12}$$

a convergent quantity.

III. NUMERICAL ANALYSIS

The numerical analysis of $F^0(\mathbf{q}, \omega)$ may be easily performed through the Kramers–Kronig expression

$$\mathrm{Re}g(\mathbf{q}, \omega) = \pi^{-1} PP \int_{-\infty}^{\infty} \frac{d\omega\, \mathrm{Im}g(\mathbf{q}, \omega')}{\omega' - \omega}, \tag{III. 1}$$

a rather intricate expression to deal with. This assertion is made clear by the introduction of

$$PP\frac{1}{x} = \lim_{\epsilon \to 0} \frac{x}{x^2 + \epsilon^2}, \tag{III. 2}$$

in Eq. (III. 1), followed by the limit $\epsilon \to 0$. Such a procedure already used by Khanna and Glyde yields results of a rather low accuracy.[3] A very simple way to circumvent this difficulty is to inject in Eq. (III. 1) the rhs of

$$PP\int_{-\infty}^{\infty} \frac{dx f(x)}{t - x} = \int_{-\infty}^{\infty} dx \frac{f(x) - f(t)}{t - x}, \tag{III. 3}$$

since

$$PP\int_{-\infty}^{\infty} \frac{dx}{t - x} = 0.$$

With a simple transformation, Eq. (III. 1) reads

$$\mathrm{Re}g(\mathbf{q}, \omega) = \frac{T}{4Q}[g(p_+) - g(p_-)], \tag{III. 4}$$

where

$$g(p) = \int_{-\infty}^{\infty} \frac{dp'(\ln[1 + \exp[(\gamma - p'^2)/T]] - \ln[1 + \exp[(\gamma - p^2)/T]])}{[p' - p]^{-1}}. \tag{III. 5}$$

In order to compute $g(\mathbf{q}, \omega)$ for different Z values, we need the reduced chemical potential $\gamma = T/\epsilon_F^0$. For further applications to the electronic transport properties in a partially degenerate electron gas, it appears of interest to parametrize the quantities of relevance with the degeneracy factor $\alpha = \beta\mu$, obtained by inverting

$$\tfrac{2}{3}Z^{-3/2} = \int_0^\infty \frac{dx\, x^{1/2}}{\exp(x - \alpha) + 1}. \tag{III. 6}$$

γ and α are depicted numerically in Fig. 1.

The variations of $g(\nu, Q)$ as a function of $\nu/2Q$ are reported in Fig. 2 for various values of the degeneracy parameter.

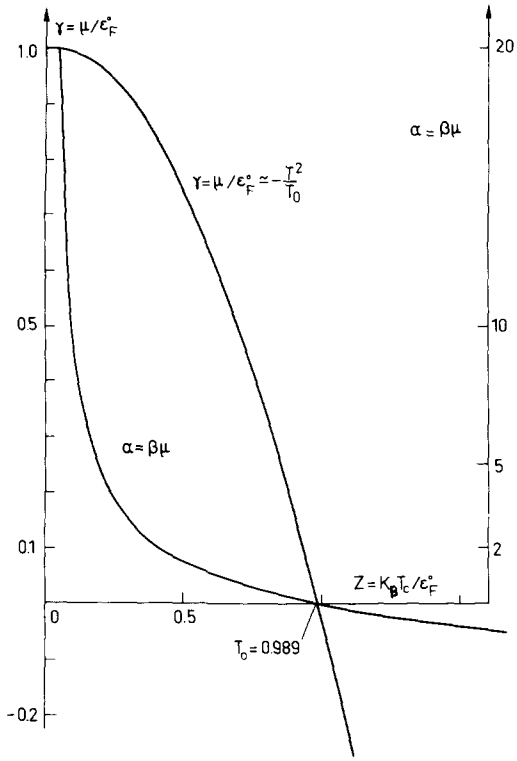

FIG. 1. Reduced chemical potential γ and degeneracy factor α as a function of T.

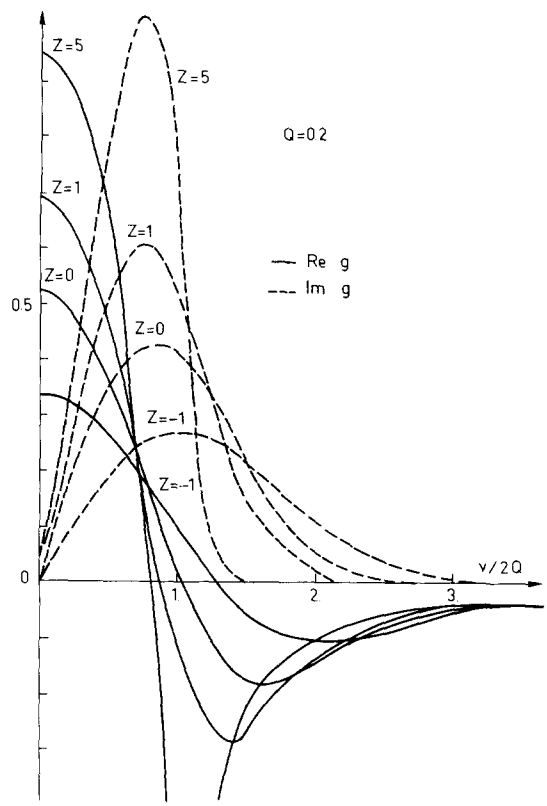

FIG. 2. Variations of $g(Q, \nu)$ as a function of $\nu/2Q$ for $Q = 0.2$ and several degeneracy degrees.

They are steeper around $\nu/2Q = 1$, when the degeneracy increases. Also, the imaginary part reaches its maximum near the same value and vanishes quicker when the degeneracy is the largest. $g(\nu, Q)$ is only weakly Q dependent for $Q \leq 1$.

IV. STATIC RESPONSE FUNCTION $F^0(\mathbf{q}, 0)$

In the $\omega = 0$ limit, we simply get $p_+ = p_- = Q/2 = p$. $\mathrm{Im}g(\mathbf{q}, \omega)$ is an odd function of ω, and $\mathrm{Im}g(\mathbf{q}, 0) = 0$. $\mathrm{Re}g(\mathbf{q}, \omega)$ is even in ω, so

$$g(\mathbf{q}, 0) = \mathrm{Re}g(\mathbf{q}, 0) = g(0, 0) + \pi Z \sum_{n=0}^{\infty} \frac{b_n}{\gamma_n^2} \left[-\frac{1}{2p} \left(\mathrm{Arctan}\ \frac{p + a_n}{b_n} + \mathrm{Arctan}\ \frac{p - a_n}{b_n} \right) \right] \tag{IV.1}$$

where $g(0, 0) = \int_0^\infty dk\, n^0(k)$ and

$$g(\mathbf{q}, 0) \sim \frac{Z^{3/2}}{p} \sqrt{\pi}\, I_{1/2}\left(\frac{\gamma}{Z}\right), \quad p \to \infty \tag{IV.2}$$

with

$$I_{1/2}(y) = \frac{2}{\Gamma(\frac{3}{2})} \cdot \int_0^\infty \frac{dx\, x^2}{1 + \exp(x^2 - y)}.$$

The presence of the Arctan makes $g(\mathbf{q}, 0)$ analytic in the complex q plane minus a number of branch cuts. Let us introduce $g(p) = g(\mathbf{q}, 0)$. The ratio of the discrete sum in Eq. (IV.1) is $(2n + 1)Z$, vanishing in the $Z \to 0$ limit. Therefore, the zero temperature limit has to be integrated through a continuous sum. For this purpose, we introduce

$$g(u) = \frac{b}{\gamma^2} - \frac{1}{2p} \left(\tan^{-1} \frac{p + a}{b} + \tan^{-1} \frac{p - a}{b} \right), \tag{IV.3}$$

$$F(0) = a_0 + \frac{a_0 b_0}{p} \left(\tan^{-1} \frac{p + a_0}{b_0} + \tan^{-1} \frac{p - a_0}{b_0} \right) + \frac{\gamma - p^2}{4p} \ln \frac{b_0^2 + (p + a_0)^2}{b_0^2 + (p - a_0)^2}. \tag{IV.4}$$

So, the discrete sum in the rhs of Eq. (IV.1) reads

$$\pi T \sum_{n=0}^{\infty} g[(2n + 1)\pi Z] = \frac{\pi T}{2} g(\pi Z) + \frac{F(0)}{2} + O(Z^2). \tag{IV.5}$$

The $\tan^{-1}$ terms cancel out, and

$$g(p) = g(0) - \frac{\gamma a_0}{2\gamma_0^2} + \frac{\gamma - p^2}{8p} \ln \frac{b_0^2 + (p + a_0)^2}{b_0^2 + (p - a_0)^2} + O(Z^2) \tag{IV.6}$$

in the $T \to 0$ limit. In order to preserve the equality

$$g(p) = O(p^{-2}), \quad p \to \infty \tag{IV.7}$$

in the zero-temperature range, we put

$$g(0) = \frac{a_0^3}{\gamma_0^2} + O(Z^2), \tag{IV.8}$$

whence

$$g(p) \approx \frac{a_0}{2} + \frac{\gamma - p^2}{8p} \ln \frac{b_0^2 + (p + a_0)^2}{b_0^2 + (p - a_0)^2} + O(Z^2). \tag{IV.9}$$

Recalling that in the $Z \to 0$ limit $a_0 \sim 1$, $\gamma \sim 1$, and $b_0 \sim 0$, we recover the well-known Lindhard result[6]

$$g(p) = \frac{1}{2} + \frac{1 - p^2}{4p} \ln \frac{p + 1}{p - 1}. \tag{IV.10}$$

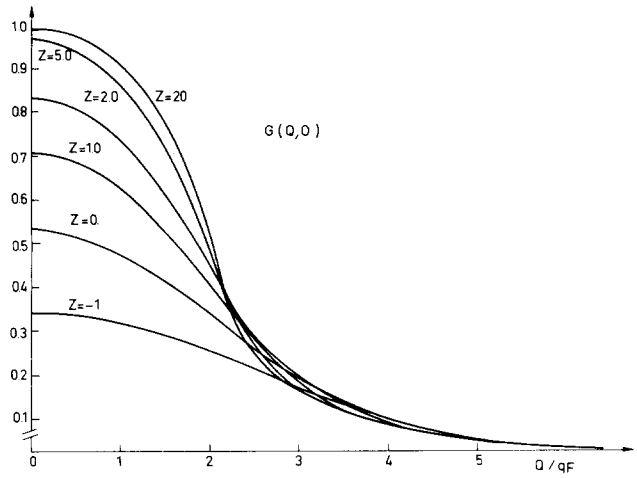

FIG. 3. Q variations of the static response $g(Q,0)$ for several degeneracy degrees.

Equation (IV. 1) is pictured in Fig. 3. Its Q variations get strongly damped in the $Z\to-\infty$ limit. $g(0)$ is given in Table I as a function of Z, and compared with its asymptotic estimates at low and high temperature respectively.

TABLE I. $g(0)$ with its low and high temperature asymptotic estimates.

α	$T=\frac{1}{\beta\epsilon_F^0}$	$g(0)$ exact	$g(0)=a_0^3/r_0^2$, $T\to0$	$g(0)=2/3T$, $T\to\infty$
20	0.0499	0.997941		
10	0.0992	0.991515	0.984654	
5	0.1934	0.963852	0.948239	
4	0.2372	0.943481		
3	0.3040	0.905109		
2	0.4140	0.8350	0.836609	
1	0.6109	0.7114	0.746575	
0.5	0.7682	0.6274	0.68802	
0.1	0.9383	0.552434		
0.01	0.9836	0.535003		
0	0.9888	0.533054	0.623123	0.674250
−0.1	1.0427	0.513556	0.609759	0.639363
−0.5	1.3000	0.435738	0.556315	0.512824
−1.0	1.7399	0.3437		
−1.5	2.3607	0.2632		
−2.0	3.2349	0.197117	0.386572	0.206087
−2.5	4.4624	0.145301		
−5.0	23.2233	0.028544	0.248429	0.028707

V. HIGH TEMPERATURE LIMIT

Although Eqs. (II. 11) and (II. 12) are exact RPA quantities, the latter is more suited to extract in a convenient way the low temperature limit than the higher one. In order to recover from below the well-known classical $Z\to-\infty$ limit,[5] it is useful to represent $F^0(\mathbf{q},\omega)$ as a series expansion in $\exp(\beta\mu)$ for large and negative μ values, through

$$n^0_{\mathbf{p}+\mathbf{q}/2}\approx\exp[\beta\mu-\beta\hbar^2(\mathbf{p}-\mathbf{q}/2)^2/2m]\times\left(1+\sum_{n=1}^{\infty}\{-\exp[\beta\mu-\beta\hbar^2(\mathbf{p}+\mathbf{q}/2)^2/2m]\}^n\right). \quad (V.1)$$

The general term of order n in Eq. (I. 2) is easily integrated using cylindrical polar coordinates $\mathbf{p}=\mathbf{p}_\parallel+\hat{q}\mathbf{p}_\perp$ with $\hat{q}$ as the polar axis.[2] So, we get

$$F^0(\mathbf{q},\omega)=-2\sum_{n=0}^{\infty}(-)^n\exp[(n+1)\beta\mu)]\times\int\frac{d^2\mathbf{p}}{(2\pi)^2}\exp[-(n+1)\beta p_\perp^2\hbar^2/2m]\times\int_{-\infty}^{\infty}\frac{dp_\parallel}{2\pi}\exp[-(n+1)\beta(p_\parallel+q/2)^2\hbar^2/2m]\times\left(\frac{1}{\hbar\omega+i\eta-\hbar^2p_\parallel q/m}-\frac{1}{\hbar\omega+i\eta+\hbar^2p_\parallel q/m}\right). \quad (V.2)$$

Separating out the real and the imaginary parts, we have

$$F(\mathbf{q},\omega)=F_1^0(\mathbf{q},\omega)+iF_2^0(\mathbf{q},\omega),$$

where

$$F_2^0(\mathbf{q},\omega)=2\sum_{n=0}^{\infty}A_n\frac{m}{\hbar^2q}\left(\exp\left[-\frac{(n+1)\beta m\omega^2}{2q^2}-\frac{(n+1)\beta\hbar^2q^2}{8m}\right]\right)\times\sinh\left(\frac{\beta(n+1)\hbar\omega}{2}\right),$$
$$A_n=\frac{(-)^n\exp[\beta\mu(n+1)]m\pi^{-1}}{(n+1)\beta\hbar^2}, \quad (V.3)$$

an odd function of ω, vanishing at $\omega=0$.

It should be noticed that in the $T\to\infty$ limit, Eqs. (II. 11) and the first term in Eq. (V. 3) are both proportional to ν/Q. However the first one does not vanish with $\beta\to0$, thus making it clear that even in this limit the correct behavior of the real part can only be obtained from the complete sum (V. 3). Actually, these $T\to\infty$ expansions are asymptotic to the exact quantities. In contrast, the real part F_1^0 cannot be expressed in terms of elementary functions, but a change of variables[2] yields

$$F_1^0(\mathbf{q},\omega)=-\frac{m}{\pi^{1/2}q\hbar^2}\sum_{n=0}^{\infty}A_n\left\{\Phi\left[\left(\frac{m\beta(n+1)}{2}\right)^{1/2}\left(\frac{\omega}{q}+\frac{\hbar q}{m}\right)\right]-\Phi\left[\left(\frac{m\beta(n+1)}{2}\right)^{1/2}\left(-\frac{\omega}{q}+\frac{\hbar q}{m}\right)\right]\right\}, \quad (V.4)$$

an even function of ω, in agreement with the requirements of the fluctuation-dissipation theorem. $\Phi(x)$ denotes the real part of the classical plasma dispersion function[5]

$$\Phi(x)=\frac{1}{\sqrt{\pi}}PP\int_{-\infty}^{\infty}\frac{dy\exp(-y^2)}{x-y},\quad \Phi(x)=-\Phi(-x). \quad (V.5)$$

VI. PLASMA OSCILLATIONS

As a first application of the ring response function $F^0(\mathbf{q},\omega)$ we may determine the electron plasma modes in thermodynamic equilibrium for any degeneracy. They are obtained by setting the real and the imaginary parts of Eq. (I. 1) equal to zero with $F(\mathbf{q},\omega)\equiv F^0(\mathbf{q},\omega)$. The natural oscillation frequencies are determined by the poles of the retarded density correlation function,[7] which occur at the solutions $\Omega_q-i\gamma_q$ of the equation

$$1-\frac{4\pi e^2}{q^2}F_1^0(\mathbf{q},\Omega_q-i\gamma_q)-\frac{i4\pi e^2}{q^2}F_2^0(\mathbf{q},\Omega_q-i\gamma_q)=0. \quad \text{(VI. 1)}$$

This description is entirely general. Furthermore the damping is small ($\gamma_q \ll \Omega_q$), so the real and imaginary parts separate easily with the results

$$1=\frac{4\pi e^2}{q^2}F_1^0(\mathbf{q},\Omega_q), \quad \text{(VI. 2)}$$

$$\gamma_q=F_2^0(\mathbf{q},\Omega_q)\left(\frac{\partial F_1^0(\mathbf{q},\omega)}{\partial\omega}\bigg|_{\Omega_q}\right)^{-1}, \quad \text{(VI. 3)}$$

obtained from $\omega=\Omega_q+i\gamma_q$ and assuming that both γ_q and $\partial F_2^0/\partial\omega$ are small quantities whose product is negligible. As in the $T\to 0$ limit, this approximate separation of real and imaginary parts will be shown to be valid at long wavelengths, so we consider small expansions of $F_1^0(\mathbf{q},\omega)$ and $F_2^0(\mathbf{q},\omega)$. A simple change of variables in the first term of Eq. (I. 2) reduces it to[8]

$$F^0(\mathbf{q},\omega)=\frac{2q^2}{m}\int\frac{d^3\mathbf{k}}{(2\pi)^3}n^0(k)\frac{1}{[\omega-(\hbar\mathbf{q}\cdot\mathbf{k}/m)+i\eta]^2-(\hbar q^2/2m)^2}. \quad \text{(VI. 4)}$$

As in the $T=0$ limit, the well-known limiting relation (Plemelj)

$$\frac{1}{X+i\eta}\underset{\eta\to 0+}{\to}PP\frac{1}{X}-i\pi\delta(X), \quad \text{(VI. 5)}$$

shows that $F_2^0(\mathbf{q},\omega)=0$ if

$$|\omega|>\frac{\hbar k_F q}{m}+\frac{\hbar q^2}{2m}.$$

In this region of the $q\omega$ plane, $F_1^0(\mathbf{q},\omega)$ may then be evaluated as an ascending series in q. To order q^4, we have

$$F_1^0(\mathbf{q},\omega)=\frac{2q^2}{m\omega^2}\int\frac{d^3\mathbf{k}}{(2\pi)^3}n^0(k)\left[1+2\frac{\hbar\mathbf{k}\cdot\mathbf{q}}{m\omega}+3\left(\frac{\hbar\mathbf{k}\cdot\mathbf{q}}{m\omega}\right)^2+-\right] \quad \text{(VI. 6)}$$

Using the relationships

$$2\int\frac{d^3k}{(2\pi)^3}n^0(k)=\frac{N}{V}\equiv\rho \quad \text{(VI. 7)}$$

and

$$\int\frac{d^3\mathbf{k}}{(2\pi)^3}\left(\exp\left[\beta\left(\frac{\hbar^2k^2-\mu}{2m}\right)\right]+1\right)^{-1}\left(\frac{\hbar\mathbf{k}\cdot\mathbf{q}}{m\omega}\right)^2=\tfrac{1}{3}\hbar^2\cdot\frac{q^2\pi}{\lambda^5m^2\omega^2}f_{5/2}(z), \quad \text{(VI. 8)}$$

where

$$f_n(z)=\frac{1}{\Gamma(n)}\int_0^\infty\frac{dx\,x^{n-1}}{z^{-1}e^x+1}. \quad \text{(VI. 9)}$$

Therefore, the dispersion relation [Eq. (V. 1)] now reads

$$1=\frac{4\pi e^2\rho}{m\Omega^2q}\left(1+\frac{12\pi\hbar^2q^2}{m^2\Omega^2q}\frac{f_{5/2}(z)}{\lambda^5\rho}+-\right) \quad \text{(VI. 10)}$$

$$\underset{T\to\infty}{\approx}\frac{4\pi e^2\rho}{m\Omega^2q}\left(1+\frac{3q^2k_BT}{\Omega^2q}\right).$$

The last expression on the rhs is obtained from Eq. (VI. 10) with $z=n\lambda^3/2$. Solving iteratively, one obtains

$$\Omega_q=\pm\,\Omega_{pl}[1+q^2G'(z)],\quad G'(z)\equiv\frac{G(z)}{2\Omega_{pl}^2}, \quad \text{(VI. 11)}$$

where $\Omega_{pl}^2=4\pi e^2\rho/m$ is the square of the plasma frequency. One sees immediately that whatever the degeneracy degree may be, the resonant frequency at zero wavelength is the classical frequency, and it is therefore independent of $\hbar$, as obtained from the classical derivation of plasma oscillations.[8] In the $T=0$ limit, the rhs of Eq. (VI. 8) simplifies to $\frac{3}{5}k_F^2$, so that[9] Eq. (VI. 10) becomes

$$1=\frac{4\pi e^2}{m\Omega_q^2}\left[1+\tfrac{3}{5}\left(\frac{\hbar k_Fq}{m\Omega q}\right)^2+-\right] \quad \text{(VI. 12)}$$

with

$$\Omega_q=\pm\,\Omega pl\left(1+\frac{9}{10}\frac{q^2}{q_{TF}^2}+-\right). \quad \text{(VI. 13)}$$

$q_{TF}=(6\pi\rho e^2/\epsilon_F^0)^{1/2}$ is the Thomas–Fermi (Debye–Hückel when $T\to\infty$) wavenumber. It is worthwhile to notice that upon the replacement $\epsilon_F^0\to k_BT$, Eq. (VI. 13) gives back at once the classical Bohm–Gross-like dispersion relation ($\bar V_{th}^2=2k_BT/m_e$)

$$\Omega^2\mathbf{q}=\Omega^2pl+\tfrac{3}{5}\bar V_{th}^2q^2,\quad T\to\infty, \quad \text{(VI. 14)}$$

while the exact classical result obtained from (VI. 12) is the expected Bohm–Gross expression

$$\Omega^2\mathbf{q}=\Omega^2p+\tfrac{3}{2}\bar V_{th}^2q^2,\quad T\to\infty. \quad \text{(VI. 15)}$$

The slight discrepancy between both results is a quantitative estimate of the brute force replacement[7(b)] of the Boltzmann velocity distribution by a $T=0$ Fermi-like in the classical dielectric function yielding Eq. (VI. 15). The numerical results given in Fig. 3 for $F_2^0(\mathbf{q},\omega)$ show that for any $T\neq 0$, Eq. (VI. 3) will produce a nonzero damping (Landau). Therefore, the collective modes are now damped, even in the lowest order approximation. Working out the $q\to 0$ limit of $F_2^0(\mathbf{q},\omega)$ from the first term on the rhs of Eq. (V. 3), one gets from Eq. (VI. 3) the classical result[2]

$$\gamma_q=q^{-3}\Omega^4pl(m\beta)^{3/2}\left(\frac{\pi}{8}\right)^{1/2}\exp\left(\frac{-\Omega_{pl}^2m\beta}{2q^2}\right),\quad T\to\infty. \quad \text{(VI. 16)}$$

In general γ_q is positive. Results related to Eq. (VI. 16) are obtained for intermediate T values from the complete expansion (VI. 6), so the approximation of small damping appears fully justified at long wavelengths, because $|\gamma_q/\Omega_q|$ vanishes exponentially. The temperature affects both the damping and the q^2 correction to the dispersion relation, but does not alter the fundamental plasma frequency.

VII. SCREENING AROUND A TEST CHARGE

In the foregoing sections we have obtained an exact expression for the linear response at any degeneracy in the ring approximation. As a second application, let us consider the electron gas screening, i.e., the specific response to an external perturbation $\phi^{ex}(\mathbf{r},t)$. Following the standard folklore in this field, we take it as produced by an extra point charge so $\phi^{ex}(\mathbf{q},\omega)=(4\pi ze/$

$q^2)2\pi\delta(\omega)$, and we need only the zero frequency component of $F^0(\mathbf{q}, \omega)$. The main purpose of the present section is to show that the asymptotic limit of the fluctuation density function $\delta n(\mathbf{r})$ exhibits at all temperatures both a Thomas–Fermi-like screening and damped Friedel-like oscillations. The induced density fluctuation by $\phi^{ex}(\mathbf{q}, \omega)$ in the uniform electron gas is[2]

$$\delta n(\mathbf{q}, \omega) = -\pi^R(\mathbf{q}, \omega)\, e\phi^{ex}(\mathbf{q}, \omega), \tag{VII. 1}$$

with

$$\pi^R(\mathbf{q}, \omega) = \frac{F(\mathbf{q}, \omega + i\eta)}{[1 - V(q)F(\mathbf{q}, \omega + i\eta)]}. \tag{VII. 2}$$

In the ring approximation $F(\mathbf{q}, \omega_0^+) \sim F^0(\mathbf{q}, \omega_0^+)$, so ($z$ = charge number)

$$\delta\rho(\mathbf{x}) = -e\delta n(\mathbf{x}) = ze\int \frac{d^3\mathbf{q}}{(2\pi)^3} \exp(i\mathbf{q}\cdot\mathbf{x}) \frac{V(q)F^0(\mathbf{q}, 0)}{1 - V(q)F^0(\mathbf{q}, 0)}. \tag{VII. 3}$$

We already know that $[\alpha r_s = (k_F a_0)^{-1}]$

$$F^0(\mathbf{q}, 0) = -\frac{4\alpha r_s}{\pi} \frac{k_F^2}{4\pi e^2} g\left(\frac{q}{q_F}\right), \tag{VII. 4}$$

where ($p = Q/2$)

$$g(Q) = \int_0^\infty dk\, n^0(k) + \pi Z \sum_{n=0}^{\infty} \frac{b_n}{r_n^2} + \frac{1}{2p} \tan^{-1} \frac{2pb_n}{p^2 - r_n^2}, \tag{VII. 5}$$

$$\equiv g(0) + \sum_{n=0}^{\infty} g_n(Q).$$

Equation (VII. 3) may then be rewritten as

$$\delta\rho(\mathbf{x}) = -ze(2k_F)^3 \int \frac{d^3\mathbf{p}}{(2\pi)^3} \frac{\exp(2i\mathbf{p}\cdot\mathbf{r})(\alpha r_s/\pi)g(p)}{p^2 + (\alpha r_s/\pi)g(p)}, \tag{VII. 6}$$

with $p = q/2k_F$ and $\sigma = k_F\mathbf{x}$. The angular average is straightforward, so

$$\delta\rho(\mathbf{x}) = \delta\rho(\sigma) = \frac{-ze}{i\pi^2} k_F^3 \frac{1}{r} \int_{-\infty}^{\infty} dp\, p \frac{\exp(2ipr)(\alpha r_s/\pi)g(p)}{p^2 + (\alpha r_s/\pi)g(p)}. \tag{VII. 7}$$

In view of the complex analytic structure of $g(p)$ in the complex p plane, we find it better to expand[4(b),10] the integrand in powers of r_s. The only remaining singularities in the denominator are the $p = \pm ip_D$ poles. More precisely, we set $g(p) = g_0(1 - \phi(p))$ with $|\phi(p)| < 1$, for all p, and also

$p_D^2 = \frac{\alpha r_s}{\pi} g_0 = (4\lambda_{TF}^2 k_F^2)^{-1}$. Thus, we have

$$\left(p^2 + \frac{\alpha r_s}{\pi} g(p)\right)^{-1} = \frac{1}{p^2 + p_D^2}\left(1 + \frac{p_D^2\phi(p)}{p^2 + p_D^2} + O(r_s^2)\right) \tag{VII. 8}$$

and also

$$\frac{\delta\rho(x)}{-ze k_F^3/i\pi^2} = I_0 + I_1 + \cdots \tag{VII. 9}$$

with

$$I_0 = \frac{1}{Z} \int_{-\infty}^{\infty} \frac{dp\, p \exp(2ipr)}{p^2 + p_D^2} \frac{\alpha r_s}{\pi} g_0(1 - \phi(p)). \tag{VII. 10}$$

The contribution of the $p = ip_D$ pole makes the Thomas–Fermi (Debye–Hückel) term appear,

$$I_0^{TF} = \frac{i\pi p_D^2}{r} \exp(-2p_D r)[1 - \phi(ip_D)]. \tag{VII. 11}$$

The logarithmic branch cuts of the $\tan^{-1}$ terms yield the Friedel-like term in the form

$$I_0^{\text{Friedel}} = -\frac{p_D^2}{r} \sum_{n=0}^{\infty} \int_{c_n} \frac{dp\, p \exp(2ipr)}{p^2 + p_D^2} \cdot \frac{\pi Z}{g_0} \cdot \frac{1}{2p}\left(\tan^{-1}\frac{p + a_n}{b_n} + \tan^{-1}\frac{p - a_n}{b_n}\right), \tag{VII. 12}$$

where c_n is the appropriate contour around the logarithmic branch cut. Taking into account that the left contour contribution in Fig. 4 is complex conjugate to the right one, we can rewrite

$$I_0^{\text{Friedel}} = \frac{i\pi^2 Z p_D^2}{2r g_0}\left(\sum_{n=0}^{\infty} \exp[-2r(b_n + ia_n)] \times \int_0^{\infty} \frac{dy \exp(-2ry)}{(b_n + ia_n + y)^2 - p_D^2} + \text{c.c.}\right), \tag{VII. 13}$$

where we used $\Delta[\tan^{-1}(p - a_n)/b_n] = \pi$ and $p = a_n + ib_n + iy$ along c_n right. Let us introduce $\beta_n = b_n + ia_n$, so that the sum in the last line becomes

$$\frac{1}{2p_D} \sum_{n=0}^{\infty} \exp(-2r\beta_n)\{\exp[2r(\beta_n - p_D)]E_1[2r(\beta_n + p_D)] - \exp[2r(\beta_n + p_D)]E_1[2r(\beta_n + p_D)]\}$$

with

$E_1(z) = \int_z^{\infty} \frac{dt\, e^{-t}}{t}$. This gives

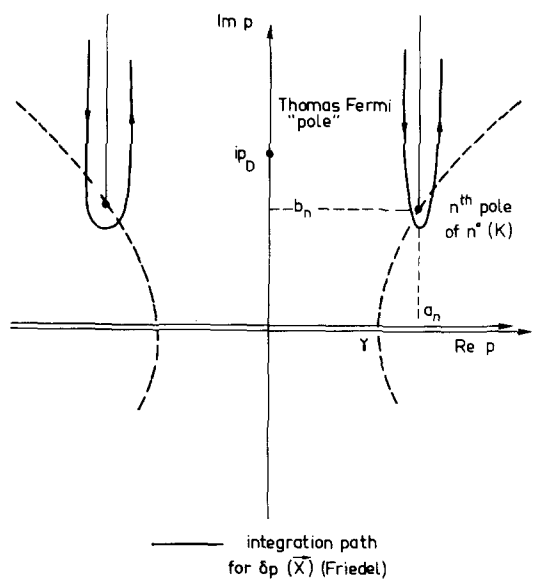

FIG. 4. $g(p)$ singularities in the p-plane. The $p = ip_D$ pole gives rise to the Thomas–Fermi (Debye) screening, while two of the symmetric contours produce the long ranged Friedel oscillations.

$$I_0^{\text{Friedel}} \underset{r\to\infty}{\sim} \frac{i\pi p_D}{4rg_0}\frac{\pi Z}{2r}\left[\sum_{n=0}^{\infty}\exp(-2\beta_n r)\frac{1}{\beta_n - p_D} - \frac{1}{\beta_n + p_D} + \text{c.c.}\right]. \tag{VII. 14}$$

As a provisional conclusion, we have

$$\frac{\delta\rho^{TF}(\mathbf{r})}{-zek_F^3/\pi} = \frac{p_D^2}{r}\exp(-2p_D r) \times\left(1 + \frac{\pi Z}{2g_0}\left(\sum_{n=0}^{\infty}\frac{1}{\beta_n} - \frac{1}{2p_D}\ln\frac{\beta_n + p_D}{\beta_n - p_D} + \text{c.c.}\right)\right), \tag{VII. 15}$$

$$\frac{\delta\rho(\mathbf{r})^{\text{Friedel}}}{-zek_F^3/\pi} = \frac{p_D^2}{4rg_0}\frac{\pi Z}{2p_D r}\left[\sum_{n=0}^{\infty}\exp(-2\beta_n r) \times\left(\frac{1}{\beta_n - p_D} - \frac{1}{\beta_n + p_D}\right) + \text{c.c.}\right], \tag{VII. 16}$$

which shows that in the asymptotic range, and for any temperature, the dominant term will be Friedel-like when $b_0 < p_D$ or Thomas–Fermi- (Debye–Hückel-) like for $b_0 > p_D$. $b_0 = p_D$ thus define a border between a gas-like and a Fermi liquidlike behavior with long range order. This is explained in the T_e-n_e plane through the reduced temperature $Z = k_B T/\epsilon_F^0$ with $\epsilon_F^0 = (\hbar^2/2m_e)k_F^2$, the r_s of equivalence between the two terms when $r = k_F x \to \infty$, the corresponding $n_e = 1.613\times 10^{24} r_s^{-3}$ and the equivalent temperature $Z_{\text{eq}} = 6.5\times 10^5\, T r_s^{-2}$. In this connection, it should be noted that the exponential screening of the Friedel oscillations at $T \neq 0$, has been also discussed by Fetter.[11]

VIII. SCREENING AT $T \simeq 0$

The Friedel sum becomes in the $T \to 0$ limit

$$\pi T\sum_{n=0}^{\infty}\exp(-2\beta_n r)\left(\frac{1}{\beta_n - p_D} - \frac{1}{\beta_n + p_D}\right) \underset{r\to\infty}{\sim} \frac{\pi z}{2}\exp(-2\beta_0 r)\left(\frac{1}{\beta_0 - p_D} - \frac{1}{\beta_0 + p_D}\right) - \frac{ip_D}{2r}\exp(-\beta_0 r)\left(\frac{1}{\beta_0 - p_D} - \frac{1}{\beta_0 + p_D}\right). \tag{VIII. 1}$$

In the present case we have $\pi z/2 < p_D/2r$, so the first term is negligible and

$$\frac{\delta\rho(\mathbf{r})^{\text{Friedel}}}{-zek_F^3/\pi} = \frac{-ip_D^2}{8g_0}\cdot\frac{1}{2r^3}\cdot\left[\exp(-2\beta_0 r)\left(\frac{1}{\beta_0 - p_D} + \frac{1}{\beta_0 + p_D}\right) - \text{c.c.}\right]. \tag{VIII. 2}$$

Upon introducing $b_0 \sim \pi Z/2$, $a_0 \sim 1$, and $\beta_0 \sim 1$, it becomes

$$\frac{\delta\rho(\mathbf{r})^{\text{Friedel}}}{zek_F^3/\pi} = \frac{p_D^2}{4g_0}\cdot\frac{1}{1+p_D^2}\cdot\exp(-2b_0 r)\cos\frac{2a_0 r}{r^3}. \tag{VIII. 2'}$$

The Thomas–Fermi-like term behaves as

$$\frac{\delta\rho(r)^{TF}}{-zek_F^3/\pi} = \frac{p_D^2}{rg_0}\exp(-2p_D r)g(ip_D). \tag{VIII. 3}$$

In the $T \to 0$ limit

$$g(ip_D) \sim \frac{a_0}{2} + \frac{1+p_D^2}{4p_D}\tan^{-1}\frac{2a_0 p_D}{r_0^2 - p_D^2}. \tag{VIII. 4}$$

The latter becomes

$$g(ip_D) = 1 + \frac{p_D^2}{2}, \tag{VIII. 4'}$$

with $a_0 \sim 1$ and $b_0 \ll p_D \ll a_0$. The relative importance of both contributions is measured by

$$\frac{\delta\rho(\mathbf{r})^{\text{Friedel}}}{\delta\rho^{TF}(\mathbf{r})} \approx \frac{\exp[-2r(b_0 - p_D)]}{4(1+p_D^2)r^2}, \tag{VIII. 5}$$

an increasing function of $r_s(p_D^2)$. Further remarks are in order. At nonzero T, the Friedel sum behaves asymptotically as $r^{-2}\exp(-2\beta_0 r)$, while it becomes equivalent to $r^{-3}\exp(-2\beta_0 r)$ in the $T \to 0$ limit. Moreover, our $T \to 0$ limit differs very slightly from the Langer–Vosko result[10]

$$\frac{\delta\rho(\mathbf{r})}{-zek_F^3/\pi} = \frac{p_D^2}{5(1+p_D/2)^2}\exp(-2b_0 r)\cdot\frac{\cos 2a_0 r}{r^3}. \tag{VIII. 6}$$

[1] See for instance, H. Minoo, C. Deutsch, and J. P. Hansen, Phys. Rev. A **14**, 840 (1976) and H. Brysk, P. M. Campbell, and P. Hammerling, Plasma Phys. **17**, 473 (1975).
[2] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971), p. 305.
[3] F. C. Khanna and H. R. Glyde, Can. J. Phys. **54**, 648 (1976).
[4] (a) U. Fano, Phys. Rev. **103**, 1202 (1956), also (b) A. Ishihara, Physica **71**, 83 (1974).
[5] B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).
[6] J. Lindhard, K. Dan. Vidensk. Selsk. Mat. Fis. Medd. **28**, 8 (1954).
[7] (a) Ref. 2, p. 307. Also (b) E. G. Harris, *A Pedestrian Approach to Quantum Field Theory* (Wiley-Interscience, New York, 1972), Chap. 9.
[8] L. Tonks and I. Langmuir, Phys. Rev. **33**, 195 (1929).
[9] Ref. 2, p. 182.
[10] J. S. Langer and S. H. Vosko, J. Phys. Chem. Solids **12**, 196 (1960).
[11] A. L. Fetter, Phys. Rev. **140**, A1921 (1965).

Dynamics and phase transitions for a continuous system of quantum particles in a box [a)]

Guy A. Battle

Duke University, Durham, North Carolina 27706
(Received 15 December 1976)

A particular type of continuous quantum system with infinitely many particles is analyzed, and the existence of dynamics is proven in the GNS representations of certain states. The dynamics is not a group of automorphisms on the original algebra, so equilibrium states are defined in terms of the KMS condition in the representations of the states. The basic theorems about KMS states do not apply here. Nevertheless, for a special class of interactions it is proven that the central decomposition of an equilibrium state is concentrated on a Borel set of equilibrium factor states and that such factor states are precisely the extremal equilibrium states. Furthermore, the equilibrium factor states are in one-to-one correspondence with sets of functions satisfying a certain system of trace equations. This explicit correspondence is then used to show that there are no phase transitions for high temperature, and an example of a phase transition is constructed for low temperature. The phase transition also provides an example of continuous symmetry breaking.

1. INTRODUCTION

We wish to study the equilibrium states of an infinite collection of distinguishable, quantum-mechanical particles in a one-dimensional box with periodic boundary conditions, where the interaction is infinitely weak in a sense that will be made precise. The first step is to prove the existence of dynamics for the system, and we do so by abandoning the usual notion that the desired time-evolution must be defined on the algebra of observables. Instead, we single out a certain kind of state, whose properties guarantee the existence of dynamics in the GNS representation of that state. The need for this approach is very strongly indicated in the study of a certain quantum lattice interaction in Ref. 1; the authors demonstrate that there is a degree of arbitrariness in the time-evolution for each state that disappears for equilibrium states. The second step is to prove that the central decomposition of an equilibrium state yields equilibrium factor states, and this is done for a special class of interactions. Such a result is not immediately clear because the standard theory of KMS states does not apply to this situation, as we shall presently see. The third step is to characterize the equilibrium factor states as finite sets of functions satisfying the system (1.2) of equations, and, again, this is done for a special class of interactions. Using this explicit characterization we prove that phase transitions are absent for high temperatures and we construct an example of a phase transition for low temperatures.

We begin by indexing the particles with Z^+ and letting $[-\pi, \pi]$ be the interval of space to which the system is confined. Let X_j be the jth copy of $[-\pi, \pi]$, x_j the usual multiplication operator on $L^2(X_j)$, and $H_j = L^2(X_j)$. The interaction is defined in the following operational way.

Introduce the particles into the system one after another. When the first N particles are present, the interaction is a pair interaction, and the potential energy operator for the pair (j, k) of particles is $\mu_{jk}(N) f(x_j - x_k)$, where $\mu_{jk}(N)$ is the coupling coefficient and f is the functional form of the potential. In this case the energy of the system is given by

$$H_N = \sum_{k=1}^{N} (-\Delta_k) + \tfrac{1}{2} \sum_{j,k=1}^{N} \mu_{jk}(N) f(x_j - x_k) - \tfrac{1}{2} f(0) \sum_{k=1}^{N} \mu_{kk}(N). \tag{1.1}$$

Our insistence on periodic boundary conditions is contained in the requirement that the kinetic energy operator $-\Delta_k$ for the kth particle be the periodic self-adjoint extension of $-\Delta$ on $C_0^\infty(X_k)$. Our assumptions on the $\mu_{jk}(N)$ will include the assumption that $\lim_{N\to\infty} \mu_{jk}(N) = 0$, so the limiting situation as $N \to \infty$ may be thought of as an infinite system governed by an "infinitely weak" interaction.

We emphasize that the main obstacle in our program is the handling of Hamiltonians which are not "classical," i.e., the kinetic energy does not commute with the interaction. The idea for overcoming this difficulty is that an infinitely weak interaction provides just enough "Abelianness" to make the computation of explicit conditions possible.

We take $\mathcal{A}_j = \mathcal{B}(H_j)$, the algebra of all bounded operators on H_j, to be the algebra of observables for the jth particle. For a finite $\Omega \subset Z^+$, let $H_\Omega = \otimes_{j\in\Omega} H_j$ and $\mathcal{A}_\Omega = \mathcal{B}(H_\Omega)$, which is isomorphic to $\otimes_{j\in\Omega} \mathcal{B}(H_j) = \otimes_{j\in\Omega} \mathcal{A}_j$. Since the particles are distinguishable, $\mathcal{A}_\Omega$ is the algebra of observables for the set Ω of particles. The family $\{\mathcal{A}_\Omega\}$ is a directed system of C^*-algebras in the usual sense. We take the C^*-closure of $\cup_\Omega \mathcal{A}_\Omega$ to be the C^*-algebra of the infinite system. We will refer to $\mathcal{A}$ as a quasilocal algebra and an element of $\mathcal{A}$ that happens to lie in some $\mathcal{A}_\Omega$ as a local observable. Notice that "local" refers to the number of particles and has no spatial meaning.

In the case where only the first N particles are in the system the dynamics is the one-parameter group of automorphisms of $\mathcal{A}$ implemented by the one-parameter unitary group on $H_{\{1,\ldots,N\}}$ generated by H_N. It is not at all clear that the one-parameter group of automorphisms converges in any sense as $N \to \infty$, and physical intuition suggests that it does not. As we have

[a)] Since submission for publication, this paper has been incorporated into the author's dissertation, which has been submitted to the Mathematics Department at Duke University in partial fulfillment of the requirements for the Ph. D. degree.

 0022-2488/78/1901-0039$1.00

already stated, this difficulty is circumvented by taking only some states to be physically meaningful.

In Sec. 2 we single out those states which are locally normal on $\mathcal{A}$. In contrast to the situation studied in Ref. 2, where local normality is derived from the KMS condition, the situation here compels us to assume local normality at the very outset. We also define a notion of asymptotic state and show that the set of all such states is a convex, Borel set. We conclude the section with a result concerning integral decompositions of asymptotic states.

In Sec. 3 we demonstrate the existence of dynamics for an asymptotic state. More specifically, if p is an asymptotic state and $(\mathcal{H}_p, \pi_p, \Phi_p)$ is the GNS representation of p, we have a time-evolution $\{\alpha_p^t : \pi_p(\mathcal{A}) \to \pi_p(\mathcal{A})'' \mid -\infty < t < \infty\}$ which is a reasonable dynamics from a physical point of view. Notice that α_p^t may not lift to an automorphism of $\pi_p(\mathcal{A})''$. A similar time-evolution was proposed in Ref. 1, where a version of the Weiss theory was studied for the quantum lattice.

In Sec. 4 we give a definition of β-equilibrium state which is essentially the same condition that was considered in Ref. 1. It is an obvious extension of the β-KMS condition to the kind of dynamics we have. We show that the set of β-equilibrium states is a convex, Borel set for each inverse temperature β. We prove in Sec. 5 a preliminary result concerning the support of the central measure of a β-equilibrium state. We develop a Fourier analysis of the interaction in Sec. 6, and in Sec. 7 we apply this analysis in our examination of β-equilibrium factor states.

In Sec. 8 we restrict ourselves to a special class of interactions (interactions of the Pth kind) and prove that:

(a) The central measure of a β-equilibrium state is concentrated on a Borel set of β-equilibrium factor states.

(b) The β-equilibrium factor states are exactly the extremal β-equilibrium states. This is proven directly since the standard theory of KMS states does not apply here.

(c) There is a one-to-one correspondence between β-equilibrium factor states and P-tuples $(g_1, \ldots, g_P)$ of continuous functions satisfying the system of equations

$$g_j(x) = \sum_{k=1}^{P} \lambda_{jk} \frac{\mathrm{tr}\{\exp[-\beta(-\Delta + g_k(y))] f(y-x)\}}{\mathrm{tr}\{\exp[-\beta(-\Delta + g_k(y))]\}}, \tag{1.2}$$

where the λ_{jk} are given numbers related to the interaction, $-\Delta + g_k(y) = -d^2/dy^2 + g_k(y)$ operates on $L^2(-\pi, \pi)$, and for each fixed x, $f(y-x)$ denotes the operator

$$\phi(y) \to f(y-x)\phi(y)$$

on $L^2(-\pi, \pi)$.

This explicit characterization of extremal β-equilibrium states allows us to show in Sec. 9 that for an interaction of the Pth kind there is a temperature above which there is only one equilibrium state. In Sec. 10 we give an example of a phase transition at low temperature and show that it is also an example where the central decomposition breaks symmetry.

Our results generalize to higher-dimensional systems. We confine our attention to a one-dimensional system because the generalization does not alter the results in any significant way.

2. ASYMPTOTIC STATES

Throughout we assume that f is an even, real-valued, continuous function on $\mathbf{R}$ which is periodic with period 2π. Let $\{a_n \mid n \in \mathbb{Z}\}$ be the sequence of Fourier coefficients of f, and note that the a_n are real and $a_{-n} = a_n$. We assume that $\sum_{n=-\infty}^{\infty} |a_n| < \infty$. Let μ_{jk} denote the function $N \to \mu_{jk}(N)$ on $\mathbb{Z}^+$. The double sequence $\{\mu_{jk}\}$ is assumed to have the following properties:

$$\mu_{jk} = \mu_{kj}, \tag{2.1}$$

$$\lim_{N \to \infty} \mu_{jk}(N) = 0, \tag{2.2}$$

$$\sup_{N,j} \sum_{k=1}^{N} |\mu_{jk}(N)| < \infty. \tag{2.3}$$

We also introduce the following notation:

$$W_{j,N} = \sum_{k=1}^{N} \mu_{jk}(N) f(x_j - x_k), \tag{2.4}$$

$$V_{j,N}(x) = \sum_{k=1}^{N} \mu_{jk}(N) f(x - x_k) \tag{2.5}$$

for $-\infty < x < \infty$. Clearly, $V_{j,N}(x)$ is a multiplication operator on $\mathcal{H}_{\{1,\ldots,N\}}$ for fixed x, and it is periodic in x with period 2π. $W_{j,N}$ is a multiplication operator on $\mathcal{H}_{\{1,\ldots,N\}\cup\{j\}}$. Also, the $V_{j,N}(x)$ and $W_{j,N}$ are uniformly bounded in j, N, and x. If Ω is a finite subset of $\mathbb{Z}^+$, we denote by $\alpha_{\Omega,N}$ the one-parameter group of automorphisms of $\mathcal{A}$ implemented by the one-parameter unitary group on $\mathcal{H}_{\{1,\ldots,N\}\cup\Omega}$ generated by the self-adjoint operator $\sum_{j\in\Omega}(-\Delta_j + W_{j,N})$. These time-evolutions are not the same as those considered in the Introduction.

The states of the physical system are states on the C^*-algebra $\mathcal{A}$. We wish to define a notion of equilibrium state for the situation we have described, and we begin by confining ourselves to a certain class of states. Let $\mathcal{N}$ be the set of states which are locally normal on the quasilocal structure of $\mathcal{A}$, i.e., normal on $\mathcal{A}_\Omega$ for each finite $\Omega \subset \mathbb{Z}^+$. Because local normality will be one of the conditions needed for defining our dynamics, we cannot take advantage of the results of Takesaki and Winnink in Ref. 2 in order to *prove* local normality of our equilibrium states. Let $\mathcal{J}$ be the set of all p in $\mathcal{N}$ such that the strong limits

$$s\text{-}\lim_{N \to \infty} \pi_p(V_{j,N}(x)) \equiv V_{j,p}(x)$$

and

$$s\text{-}\lim_{N \to \infty} \pi_p(W_{j,N}) \equiv W_{j,p}$$

exist for $-\infty < x < \infty$, where $(\mathcal{H}_p, \pi_p, \Phi_p)$ will always denote the canonical cyclic representation of $\mathcal{A}$ with respect to p. We refer to the states in $\mathcal{J}$ as *asymptotic states*.

We restrict ourselves to $\mathcal{N}$ because, when one is considering a finite system of quantum-mechanical

particles, a physically meaningful state is always assumed to be a density matrix, i.e., normal. We further stipulate that our equilibrium states must lie in $\mathcal{G}$, since it is reasonable to expect the corresponding representations of $V_{j,N}(x)$ and $W_{j,N}$ to converge in some sense as $N\to\infty$, i.e., the limiting behavior of the quantities should not be "blurred" in the representation corresponding to an equilibrium state. Requiring convergence in the operator norm is too strong because, for a locally normal state p, π_p is a C^*-isomorphism between $\mathcal{A}$ and $\pi_p(\mathcal{A})$, so if $\mathcal{G}\neq\phi$ it would follow that $V_{j,N}(x)$ and $W_{j,N}$ converge in norm as $N\to\infty$. This need not be the case.

$\mathcal{G}$ is obviously convex. However, it is not clear that $\mathcal{G}$ is nonempty. We will prove this for a special class of interactions in Sec. 8. For the present, we will prove a topological property of $\mathcal{G}$ as a subset of the set $\mathcal{S}$ of all states on $\mathcal{A}$, where $\mathcal{S}$ will always be equipped with the weak * topology. Every assertion we will make with regard to $\mathcal{N}$ both here and in the future are either proven in Ref. 3 (pp. 158–60) or can be proven by adaptations of the arguments given in Ref. 3.

2.1. *Theorem*: $\mathcal{G}$ is a Borel subset of $\mathcal{S}$.

Proof: $\mathcal{N}$ is a G_δ set in $\mathcal{S}$, so we need only show that $\mathcal{G}$ is a Borel set in $\mathcal{N}$. Consider the condition that $W_{j,p}$ and $V_{j,p}(x)$ exist for j in $\mathbb{Z}^+$ and $-\infty<x<\infty$ for an arbitrary locally normal state p.

Let $\{A_m\}$ be a sequence in $\mathcal{A}$ such that $\{A_m\in\mathcal{A}_\Omega \mid m\in\mathbb{Z}^+\}$ is strongly dense in $\mathcal{A}_\Omega=\mathcal{B}(\mathcal{H}_\Omega)$ for all finite $\Omega\subset\mathbb{Z}^+$. There exists such a sequence because there are only a countable number of finite $\Omega\subset\mathbb{Z}^+$ and $\mathcal{B}(\mathcal{H}_\Omega)$ is strongly separable for each such Ω. Thus the choice of $\{A_m\}$ is independent of p. We have immediately that $\{\pi_p(A_m)\mid m\in\mathbb{Z}^+$ and $A_m\in\mathcal{A}_\Omega\}$ is strongly dense in $\pi_p(\mathcal{A}_\Omega)$ by the strong continuity of π_p on $\mathcal{A}_\Omega$. As π_p is norm-continuous on $\mathcal{A}$, $\cup_\Omega\mathcal{A}_\Omega$ is norm-dense in $\mathcal{A}$, and Φ_p is cyclic in $\mathcal{H}_p$ under the action of $\pi_p(\mathcal{A})$, we have in particular that $\{\pi_p(A_m)\Phi_p\}$ is dense in $\mathcal{H}_p$. Since the operators $\pi_p(V_{j,N}(x))$ and $\pi_p(W_{j,N})$ are bounded uniformly in N, the existence of $W_{j,p}$ and $V_{j,p}(x)$ follows from the existence of

$$\lim_{N\to\infty}\pi_p(V_{j,N}(x))\pi_p(A_m)\Phi_p \qquad \text{(l. j. x. m)}$$

and

$$\lim_{N\to\infty}\pi_p(W_{j,N})\pi_p(A_m)\Phi_p \qquad \text{(l. j. m.)}$$

for $m\in\mathbb{Z}^+$. Now let $\{b_k\}$ be a dense sequence on the real line. Since the vector-valued functions $x\to\pi_p(V_{j,N}(x))\pi_p(A_m)\Phi_p$ are equicontinuous for fixed m, the existence of (l. j. x. m) for $-\infty<x<\infty$ follows from the existence of

$$\lim_{N\to\infty}\pi_p(V_{j,N}(b_k))\pi_p(A_m)\Phi_p \qquad \text{(l′. j. k. m)}$$

for $k\in\mathbb{Z}^+$. Thus p lies in $\mathcal{G}$ if and only if a countable number of limits involving p exist, namely (l. j. m) and (l′. j. k. m). It is obvious that the functions $p\to\|\pi_p(W_{j,N}-W_{j,M})\pi_p(A_m)\Phi_p\|^2$ and $p\to\|\pi_p(V_{j,N}(b_k)-V_{j,M}(b_k))\pi_p(A_m)\Phi_p\|^2$ are continuous, so $\mathcal{G}$ must therefore be a countable intersection of countable unions of G_δ sets in $\mathcal{N}$, and the proof is complete.

The next step is to examine the states we obtain in an integral decomposition of a state in $\mathcal{G}$. To this end, let μ be a probability measure on $\mathcal{S}$ whose resultant p is in $\mathcal{G}$. It follows from an argument given by Ruelle in Ref. 3 that an integral decomposition of a state is concentrated on $\mathcal{N}$ if and only if the state lies in $\mathcal{N}$. Thus μ is concentrated on $\mathcal{N}$, i.e., $\mu(\mathcal{N})=1$.

Let S be an increasing sequence of positive integers. We denote by J_S the set of all states p in $\mathcal{N}$ such that the strong limits

$$s\text{-}\lim_{N\to\infty}\pi_p(V_{j,s_N}(x))\equiv V_{j,s,p}(x)$$

and

$$s\text{-}\lim_{N\to\infty}\pi_p(W_{j,s_N})\equiv W_{j,s,p}$$

exist for $-\infty<x<\infty$. By reasoning identical to that in the proof of Theorem 2.1, J_S is a Borel set in $\mathcal{S}$. It is obvious that $\mathcal{G}\subset J_S$ for every increasing sequence S of positive integers, and that the J_S are convex sets. For a given S, we refer to the elements of J_S as *S-asymptotic states*.

2.2. *Theorem*: There exists an increasing sequence S of positive integers such that μ is concentrated on J_S.

Proof: Let $\{A_m\}$ and $\{b_k\}$ be as in the proof of Theorem 2.1 and $\tilde{A}_m=A_m/\|A_m\|$.

Since p is an asymptotic state, (l, j, m) and (l′, j, k, m) exist for $j, k, m\in\mathbb{Z}^+$. As the $V_{j,N}(b_k)$ and $W_{j,N}$ are uniformly bounded with respect to j and k,

$$\sum_{j,k,m}\frac{1}{2^{j+k+m}}\|\pi_p(V_{j,M}(b_k)-V_{j,N}(b_k))\pi_p(\tilde{A}_m)\Phi_p\|^2$$

$$+\sum_{j,m}\frac{1}{2^{j+m}}\|\pi_p(W_{j,M}-W_{j,N})\pi_p(\tilde{A}_m)\Phi_p\|^2$$

converges to zero as $M,N\to\infty$. Since $p=\int_{\mathcal{N}}\sigma\,d\mu(\sigma)$ in the weak * sense, it follows that $\|\pi_p(A)\Phi_p\|^2=\int_{\mathcal{N}}\|\pi_\sigma(A)\times\Phi_\sigma\|^2\,d\mu(\sigma)$ for all $A\in\mathcal{A}$. Hence

$$\sum_{j,k,m}\frac{1}{2^{j+k+m}}\int_{\mathcal{N}}\|\pi_\sigma(V_{j,M}(b_k)-V_{j,N}(b_k))\pi_\sigma(\tilde{A}_m)\Phi_\sigma\|^2\,d\mu(\sigma)$$

$$+\sum_{j,m}\frac{1}{2^{j+m}}\int_{\mathcal{N}}\|\pi_\sigma(W_{j,M}-W_{j,N})\pi_\sigma(\tilde{A}_m)\Phi_\sigma\|^2\,d\mu(\sigma)$$

converges to zero as $M,N\to\infty$. Now consider the topological space $(\mathbb{Z}^+\times\mathbb{Z}^+\times\mathbb{Z}^+\times\mathcal{N})\cup(\mathbb{Z}^+\times\mathbb{Z}^+\times\mathcal{N})$ with the measure ν defined on it such that

$$\nu(\{j\}\times\{k\}\times\{m\}\times\mathcal{J})=\frac{\mu(\mathcal{J})}{2^{j+k+m}}$$

and

$$\mu(\{j\}\times\{m\}\times\mathcal{J})=\frac{\mu(\mathcal{J})}{2^{j+m}}$$

for all Borel $\mathcal{J}\subset\mathcal{N}$. Thus we have a sequence of vector functions which is Cauchy in the norm implicitly defined for the vector functions. The functions $\|\pi_\sigma(W_{j,M}-W_{j,N})\pi_\sigma(\tilde{A}_m)\Phi_\sigma\|^2$ and $\|\pi_\sigma(V_{j,M}(b_k)-V_{j,N}(b_k))\pi_\sigma(A_m)\Phi_\sigma\|^2$ are continuous functions on our measure space, and

this fact is all that is needed to carry out the same argument as the one used for proving that if a sequence of L^1 functions on some measure space is Cauchy in the L^1 norm, then there is a subsequence that converges almost everywhere.

Thus we have an increasing sequence S of positive integers such that $\lim_{N\to\infty}\pi_\sigma(V_{j,s_N}(b_k))\pi_\sigma(\tilde{A}_m)\Phi_\sigma$ and $\lim_{N\to\infty}\pi_\sigma(W_{j,s_N})\,\pi_\sigma(\tilde{A}_m)\Phi_\sigma$ exist almost everywhere on the measure space, so they exist μ-almost everywhere for all j, k, m. As we have already seen, this is enough to show that σ is an S-asymptotic state for μ-almost all σ in $\mathcal{N}$. This completes the proof.

Let $J=\cup_S J_S$; J may not be convex, and, since there are an uncountable number of increasing sequences of positive integers, J may not be Borel.

2.3. *Definition*: Let $p\in J$ and let S be an increasing sequence of positive integers. Then S is a *workable sequence for* p if and only if $p\in J_S$.

3. DYNAMICS OF THE SYSTEM

As we have already explained in the Introduction, our approach to proving the existence of dynamics for asymptotic states is to prove that there exists a limiting time-evolution in the GNS representation of each such state.

Let $p\in I$ and $j\in\mathbb{Z}^+$. Since π_p is strongly continuous on $\mathcal{A}_j=\mathcal{B}(\mathcal{H}_j)$ and $t\to\exp(it\Delta_j)$ is a one-parameter, strongly continuous, unitary group on $\mathcal{H}_j$, it follows that $t\to\pi_p(\exp(it\Delta_j))$ is a one-parameter, strongly continuous, unitary group on $\mathcal{H}_p$. Let $\Delta_{j,p}$ be the infinitesimal generator of this group. Thus $\Delta_{j,p}$ is a self-adjoint operator on $\mathcal{H}_p$, and

$$\exp(it\Delta_{j,p})=\pi_p(\exp(it\Delta_j)) \tag{3.1}$$

for all real t.

3.1. *Lemma*: $\exp[it(-\Delta_{j,p}+\pi_p(W_{j,N}))]=\pi_p(\exp[it(-\Delta_j+W_{j,N})])$ for all real t.

Proof: Since $\pi_p(\exp(itW_{j,N}))=\exp(it\pi)p(W_{j,N})$ for all real t, and π_p is strongly continuous on $\mathcal{A}_{\{1,\ldots,N\}\cup\{j\}}=\mathcal{B}(\mathcal{H}_{\{1,\ldots,N\}\cup\{j\}})$, the conclusion follows from Eq. (3.1) together with an application of the Trotter product formula.

3.2. *Lemma*: $-\Delta_{j,p}+W_{j,p}$ is a self-adjoint operator on $\mathcal{H}_p$ and

$$\exp[it(-\Delta_{j,p}+W_{j,p})]=\operatorname{s-lim}_{N\to\infty}\exp[it(-\Delta_{j,p}+\pi_p(W_{j,N}))] \tag{3.2}$$

for all real t.

Proof: Since $W_{j,p}$ is a bounded, self-adjoint operator, the first part of the statement follows from the Kato–Rellich theorem.

Let $\Psi\in\mathcal{H}_p$. By Duhamel's formula we have

$$\exp[it(-\Delta_{j,p}+\pi_p(W_{j,N}))]\Psi-\exp[it(-\Delta_{j,p}+W_{j,p})]\Psi$$
$$=i\int_0^t\exp[i(t-s)(-\Delta_{j,p}+\pi_p(W_{j,N}))]$$
$$\times(\pi_p(W_{j,N})-W_{j,p})\exp[is(-\Delta_{j,p}+W_{j,p})]\Psi\,ds.$$

Therefore, the norm of the left-hand side of the equation is dominated by

$$\int_0^t\|(\pi_p(W_{j,N})-W_{j,p})\exp[is(-\Delta_{j,p}+W_{j,p})]\Psi\|ds.$$

As $W_{j,p}=\text{s-lim}_{N\to\infty}\pi_p(W_{j,N})$, the integrand converges pointwise to zero. Since $W_{j,N}$ is uniformly bounded in N, we have that the integral tends to zero. This completes the proof.

3.3. *Lemma*: $\exp[it(-\Delta_{j,p}+W_{j,p})]$ commutes with $\exp[it(-\Delta_{r,p}+W_{r,p})]$.

Proof: We may assume that $j\neq r$. $W_{j,p}$ commutes with $\pi_p(\mathcal{A})$ because $\mu_{jr}(N)\to 0$ as $N\to\infty$. In particular, $W_{j,p}$ commutes with $\exp(it\Delta_{r,p})$; $W_{j,p}$ obviously commutes with $W_{r,p}$. It follows from the Trotter product formula that $W_{j,p}$ commutes with $\exp[it(-\Delta_{r,p}+W_{r,p})]$.

Since $W_{r,p}$ commutes with $\pi_p(\mathcal{A}_j)$, we have that $\exp(it\Delta_{j,p})$ commutes with $W_{r,p}$; $\exp(it\Delta_{j,p})$ certainly commutes with $\exp(it\Delta_{r,p})$, so a second application of the Trotter product formula shows that $\exp(it\Delta_{j,p})$ commutes with $\exp[it(-\Delta_{r,p}+W_{r,p})]$. Applying the Trotter product formula to the two results, the conclusion follows.

Let α_p^t be defined on the algebra $\cup_\Omega\pi_p(\mathcal{A}_\Omega)$ by

$$\alpha_p^t(\pi_p(A))=\exp(itH_{\Omega,p})\pi_p(A)\exp(itH_{\Omega,p}) \tag{3.3}$$

for all $A\in\mathcal{A}_\Omega$, where

$$H_{\Omega,p}=\sum_{j\in\Omega}(-\Delta_{j,p}+W_{j,p}). \tag{3.4}$$

It is easy to demonstrate that α_p^t is a well-defined * homomorphism. Indeed, suppose $\Omega\subset\Omega'$ and $A\in\mathcal{A}_\Omega$. Then by Lemma 3.3,

$$\exp(itH_{\Omega',p})\pi_p(A)\exp(-itH_{\Omega',p})$$
$$=\exp(itH_{\Omega,p})\pi_p(A)\exp(-itH_{\Omega,p}).$$

α_p^t clearly has range in $\pi_p(\mathcal{A})''$. Since α_p^t is an isometry, it can be extended uniquely to $\pi_p(\mathcal{A})$. Thus α_p is a one-parameter family of C^*-embeddings of $\pi_p(\mathcal{A})$ in $\pi_p(\mathcal{A})''$. Obviously α_p is not a group. Not only is it defined in the canonical representation of p, but it is also irreversible in the sense that it "smears" local and quasi-local observables into $\pi_p(\mathcal{A})''$. Of course, one would have a group structure if the α_p^t could be lifted to automorphisms of $\pi_p(\mathcal{A})''$, but whether this can be done or not is an open question. We will presently see, however, that a great deal can be said about this peculiar dynamics as it stands. A similar dynamics is proposed by Emch and Knops in Ref. 1, where an infinitely weak, infinitely long-range interaction for the quantum lattice is studied. That α_p is indeed the limiting dynamics in the GNS representation of p is stated as follows:

3.4. *Proposition*: Let Ω be a finite subset of $\mathbb{Z}^+$ and $A\in\mathcal{A}_\Omega$. Then

$$\alpha_p^t(\pi_p(A))=\operatorname{s-lim}_{N\to\infty}\pi_p(\alpha_{\Omega,N}^t(A)). \tag{3.5}$$

The proof is an easy combination of a revision of the above lemmas. One has to revise Lemma 3.2 to read:

$$\exp(itH_{\Omega,p}) = \text{s-}\lim_{N\to\infty} \exp\left(it \sum_{j\in\Omega} (-\Delta_{j,p} + \pi_p(W_{j,N}))\right)$$

but this is easy enough to prove because Ω is fixed. Moreover, one has to revise Lemma 3.1 to read

$$\exp\left(it \sum_{j\in\Omega} (-\Delta_{j,p} + \pi_p(W_{j,N}))\right) = \pi_p\left(\exp\left(it \sum_{j\in\Omega} (-\Delta_j + W_{j,N})\right)\right),$$

but this can also be handled. Notice that (3.5) involves the $\alpha_{\Omega,N}$ rather than the time evolutions mentioned in the Introduction.

Now consider the more general case where $p \in J$, and let S be a workable sequence for p. Then all of the above discussion can be duplicated for this situation. We obtain a C^*-embedding $\alpha^t_{p,S}$ of $\pi_p(\mathcal{A})$ in $\pi_p(\mathcal{A})''$ defined by $\alpha^t_{p,S}(\pi_p(A)) = \exp(itH_{S,\Omega,p})\pi_p(A)\exp(-itH_{S,\Omega,p})$ for $A \in \mathcal{A}_\Omega$, where

$$H_{S,\Omega,p} = \sum_{j\in\Omega} (-\Delta_{j,p} + W_{j,S,p}), \tag{3.6}$$

and $\alpha^t_{p,S}$ has the property that

$$\alpha^t_{p,S}(\pi_p(A)) = \text{s-}\lim_{N\to\infty} \pi_p(\alpha^t_{\Omega,S_N}(A)) \tag{3.7}$$

for all $A \in \mathcal{A}_\Omega$. Such time evolutions are introduced for technical reasons that will soon become apparent.

4. DEFINITION OF EQUILIBRIUM

We are now ready to propose a definition of equilibrium with respect to the interaction.

4.1. *Definition*: Let $\beta > 0$ and $p \in I$. Then p is a β-*equilibrium state* if and only if

$$\int_{-\infty}^{\infty} \hat{\phi}(t)(\alpha^t_p(\pi_p(A))\pi_p(B)\Phi_p, \Phi_p)\,dt = \int_{-\infty}^{\infty} \hat{\phi}(t+i\beta)(\pi_p(B)\alpha^t_p(\pi_p(A))\Phi_p, \Phi_p)\,dt \tag{4.1}$$

for all $A, B \in \mathcal{A}$ and $\phi \in C_0^\infty(\mathbb{R})$, where $\hat{\phi}$ is the Fourier transform of ϕ. We denote the set of such states by $\mathcal{E}_\beta$.

This definition of equilibrium was proposed by Emch and Knops in Ref. 1. It is an obvious extension of the KMS condition forced upon us by the fact that the α^t_p are *not* automorphisms of $\mathcal{A}$, but map $\pi_p(\mathcal{A})$ into $\pi_p(\mathcal{A})''$.

As in the case of ordinary KMS states, every equilibrium state is time invariant (with respect to the dynamics defined in terms of it). $\mathcal{E}_\beta$ is obviously a convex set. It is no longer clear, however, that the extremal states in $\mathcal{E}_\beta$ are exactly the factor states in $\mathcal{E}_\beta$. Nevertheless, when we consider certain classes of interactions, we will see that this is indeed the case. Furthermore, the factor states in $\mathcal{E}_\beta$ will prove to be very interesting, so it will be important to examine the central decomposition of an arbitrary β-equilibrium state in any case. The existence of β-equilibrium states will not be established until we examine our special classes of interactions in Sec. 8.

Let $\{A_m\}$ be as in the proof of Theorem 2.1. We have already seen that $\{\pi_p(A_m) | A_m \in \mathcal{A}_\Omega\}$ is strongly dense in $\pi_p(\mathcal{A}_\Omega)$ for all finite $\Omega \subset \mathbb{Z}^+$. Since α^t_p is strongly continuous on $\pi_p(\mathcal{A}_\Omega)$ for all finite $\Omega \subset \mathbb{Z}^+$, it is not hard to verify that (4.1) holds for all $A, B \in \mathcal{A}_\Omega$ and all $\phi \in C_0^\infty(\mathbb{R})$ if it holds for $A = A_m$, $B = A_n$ for all $A_m, A_n \in \mathcal{A}_\Omega$ and $\phi \in C_0^\infty(\mathbb{R})$. (Dominated convergence is applicable, since we may arrange $\{A_m \in \mathcal{A}_\Omega\}$ to be strongly dense in $\mathcal{A}_\Omega = \mathcal{B}(\mathcal{H}_\Omega)$ in the sequential sense, i.e., such that a given element in $\mathcal{A}_\Omega$ has a sequence of elements in $\{A_m \in \mathcal{A}_\Omega\}$ converging to it.) Since α^t_p is norm-continuous on $\pi_p(\mathcal{A})$, $p \in \mathcal{E}_\beta$ if (4.1) holds for $A = A_m$, $B = A_n$ for all $m, n \in \mathbb{Z}^+$ and $\phi \in C_0^\infty(\mathbb{R})$.

Also, there is a sequence (ϕ_k) in $C_0^\infty(\mathbb{R})$ such that if (3.1) holds for all $\phi = \phi_k$ then it holds for all $\phi \in C_0^\infty(\mathbb{R})$ (where A and B are fixed). In summary, $p \in \mathcal{E}_\beta$ if and only if (4.1) holds for $\phi = \phi_k$, $A = A_m$, and $B = A_n$ for all triples (k, m, n) of positive integers. Thus the set $\mathcal{E}_\beta$ is defined by a countable number of equations.

4.2. *Theorem*: $\mathcal{E}_\beta$ is a Borel set in $\mathcal{S}$.

Proof: It is sufficient to show that $\mathcal{E}_\beta$ is a Borel set in $\mathcal{G}$. By the above remarks we are finished if we show that the left- and right-hand sides of (4.1) are Borel functions of p for fixed $A, B \in \mathcal{A}_\Omega$ and $\phi \in C_0^\infty(\mathbb{R})$. For the sake of definiteness, consider the left-hand side. $\pi_p(\alpha^t_{\Omega,N}(A)) \to \alpha^t_p(\pi_p(A))$ strongly as $N \to \infty$ for all $p \in \mathcal{G}$, so $p(\alpha^t_{\Omega,N}(A)B) \to (\alpha^t_p(\pi_p(A))\pi_p(B)\Phi_p, \Phi_p)$ as $N \to \infty$. By dominated convergence, $\int_{-\infty}^{\infty} \hat{\phi}(t)p(\alpha^t_{\Omega,N}(A)B)\,dt$ converges to the left-hand side of (4.1). For each N, it is obvious that the function $p \to \int_{-\infty}^{\infty} \hat{\phi}(t)p(\alpha^t_{\Omega,N}(A)B)\,dt$ is continuous on $\mathcal{N}$ (since $\mathcal{N}$ is metrizable, sequential continuity is enough, so dominated convergence is applicable). Thus we have a sequence of continuous functions on $\mathcal{N}$ which converges pointwise on the Borel set $\mathcal{G}$ to the left-hand side of (4.1). Hence the left-hand side of (4.1) is a Borel function on $\mathcal{G}$.

4.3. *Definition*: Let $p \in J$ and S be a workable sequence for p. p is an $S-\beta$-*equilibrium state* if and only if

$$\int_{-\infty}^{\infty} \hat{\phi}(t)(\alpha^t_{p,S}(\pi_p(A))\pi_p(B)\Phi_p, \Phi_p)\,dt = \int_{-\infty}^{\infty} \hat{\phi}(t+i\beta)(\pi_p(B)\alpha^t_{p,S}(\pi_p(A))\Phi_p, \Phi_p)\,dt \tag{4.2}$$

for all $A, B \in \mathcal{A}$ and $\phi \in C_0^\infty(\mathbb{R})$.

We denote the set of $S-\beta$-equilibrium states by $\mathcal{E}^S_\beta$. By an argument similar to the one above, $\mathcal{E}^S_\beta$ is a Borel set in $\mathcal{S}$.

5. CENTRAL DECOMPOSITION OF AN EQUILIBRIUM STATE

Our next step is to examine the central measure of a β-equilibrium state. By Ref. 4, the central algebra of a locally normal state on $\mathcal{A}$ is also the algebra at infinity of that state. By Ref. 5, the decomposition at infinity of a locally normal state is concentrated on a Borel set of locally normal states with trivial algebra at infinity. Hence, the central measure is concentrated

on a Borel set of locally normal factor states. We wish to answer the following question: Is the central measure of a β-equilibrium state concentrated on a Borel set of β-*equilibrium* factor states?

5.1. *Lemma*: Let $p \in \mathcal{E}_\beta$ and let μ be the central measure of p. Let $\mathcal{J}$ be a Borel set of states such that $\mu(\mathcal{J}) > 0$. Then the state

$$\lambda(A) = \frac{1}{\mu(\mathcal{J})} \int_{\mathcal{J}} \sigma(A)\, d\mu(\sigma)$$

is also a β-equilibrium state.

Proof: By the nature of the central decomposition,[6] there is an orthogonal projection $E \in \pi_p(\mathcal{A})' \cap \pi_p(\mathcal{A})''$ such that

$$(E\pi_p(A)\Phi_p, \Phi_p) = \int_{\mathcal{J}} \sigma(A)\, d\mu(\sigma)$$

for all $A \in \mathcal{A}$, so

$$\lambda(A) = \frac{1}{\|E\Phi_p\|^2} (E\pi_p(A)\Phi_p, \Phi_p).$$

Since p is locally normal, $\mu(N) = 1$. Thus $\lambda \in N$.

Notice that $(E\mathcal{H}_p,\ A \to \pi_p(A)|_{E\mathcal{H}_p},\ E\Phi_p/\|E\Phi_p\|)$ is the canonical cyclic representation of $\mathcal{A}$ with respect to λ because E is a central projection, so the fact that λ lies in I follows immediately from the fact that p does. It is easy to verify that $\alpha_\lambda^t(\pi_p(A)|_{E\mathcal{H}_p}) = \alpha_p^t(\pi_p(A))|_{E\mathcal{H}_p}$ for all $A \in \mathcal{A}$, so in order to show that λ is a β-equilibrium state, it is sufficient to show that

$$\int_{-\infty}^{\infty} \hat{\phi}(t)(\alpha_p^t(\pi_p(A))E\pi_p(B)\Phi_p, \Phi_p)\, dt = \int_{-\infty}^{\infty} \hat{\phi}(t + i\beta)(E\pi_p(B)\alpha_p^t(\pi_p(A))\Phi_p, \Phi_p)\, dt \qquad (5.1)$$

for $A, B \in \mathcal{A}$ and $\phi \in C_0^\infty(\mathbb{R})$. Fix A, B, and ϕ and let $\{A_\delta\}$ be a net of elements such that $\|A_\delta\| \leq 1$ and $\pi_p(A_\delta) \to E$ weakly (using Kaplansky's theorem). We certainly have

$$\int_{-\infty}^{\infty} \hat{\phi}(t)(\alpha_p^t(\pi_p(A))\pi_p(A_\delta B)\Phi_p, \Phi_p)\, dt = \int_{-\infty}^{\infty} \hat{\phi}(t + i\beta)(\pi_p(A_\delta B)\alpha_p^t(\pi_p(A))\Phi_p, \Phi_p)\, dt$$

for all δ since $p \in \mathcal{E}_\beta$. We wish to show that each side of the equation converges to the corresponding side of (5.1). For the left-hand side, the net of integrands converges pointwise. The strong continuity of $\alpha_p^t(\pi_p(A))$ together with the uniform boundedness of $\{A_\delta\}$ implies equicontinuity of the net of integrands. Hence, the net of integrands converges uniformly on compact intervals, so that the net of integrals converges to the left-hand side of (4.1) (since the integrands also vanish uniformly at infinity). The argument for the right-hand side is similar, except that we use the strong continuity of $\alpha_p^t(\pi_p(A^*))$. This completes the proof of the lemma.

We are now ready to give a partial answer to our question.

5.2. *Proposition*. Let $\lambda \in \mathcal{E}_\beta$ and let μ be the central measure of λ. Then $\mu(I \setminus \mathcal{E}_\beta) = 0$.

Proof. Suppose $\mu(I \setminus \mathcal{E}_\beta) > 0$. Let $\{\phi_k\}$ and $\{A_m\}$ be as in the preceding section and let I_{kmn} be the set of all states in I such that (4.1) does *not* hold for $A = A_m$, $B = A_n$, and $\phi = \phi_k$. Since each side of (4.1) is a Borel function of p, I_{kmn} is a Borel set. Furthermore, $I \setminus \mathcal{E}_\beta = \cup_{k,m,n} I_{kmn}$, so there exist k_0, m_0, n_0 such that $\mu(I_{k_0 m_0 n_0}) > 0$.

Let K_N be the set of all states p in $I_{k_0 m_0 n_0}$ such that the number

$$c(p) = \int_{-\infty}^{\infty} \phi_{k_0}(t)(\alpha_p^t(\pi_p(A_{m_0}))\pi_p(A_{n_0})\Phi_p, \Phi_p)\, dt - \int_{-\infty}^{\infty} \phi_{k_0}(t + i\beta)(\pi_p(A_{n_0})\alpha_p^t(\pi_p(A_{m_0}))\Phi_p, \Phi_p)\, dt$$

satisfies $|c(p)| > 1/N$ and $2\pi(r-1)/17 < \arg c(p) \leq 2\pi(r/17)$. Then K_N is a Borel set and $I_{k_0 m_0 n_0} = \cup_{r=1}^{17} \cup_{N=1}^{\infty} K_N$. Hence $\mu(K_{r_0 N_0}) > 0$ for some r_0, N_0. Therefore, $\int_{K_{r_0 N_0}} c\, d\mu \neq 0$, so by the Fubini theorem we have

$$\int_{-\infty}^{\infty} \phi_{k_0}(t) \int_{K_{r_0 N_0}} (\alpha_p^t(\pi_p(A_{m_0}))\pi_p(A_{n_0})\Phi_p, \Phi_p)\, d\mu(p)\, dt \neq \int_{-\infty}^{\infty} \phi_{k_0}(t + i\beta) \int_{K_{r_0 N_0}} (\pi_p(A_{n_0})\alpha_p^t(\pi_p(A_{m_0}))\Phi_p, \Phi_p)\, d\mu(p)\, dt.$$

Let

$$\lambda'(A) = \frac{1}{\mu(K_{r_0 N_0})} \int_{K_{r_0 N_0}} p(A)\, d\mu(p)$$

for all A in $\mathcal{A}$. Using Proposition 3.4, it is easy to verify that

$$(\alpha_{\lambda'}^t(\pi_{\lambda'}(A_{m_0}))\pi_{\lambda'}(A_{n_0})\Phi_{\lambda'}, \Phi_{\lambda'}) = \frac{1}{\mu(K_{r_0 N_0})} \int_{K_{r_0 N_0}} (\alpha_p^t(\pi_p(A_{m_0}))\, \pi_p(A_{n_0})\Phi_p, \Phi_p)\, d\mu(p).$$

By Lemma 5.1, $\lambda' \in I$, so $\alpha_{\lambda'}^t$ makes sense. Hence λ' is *not* a β-equilibrium state. This contradicts Lemma 5.1, and the proof is complete.

This result does not rule out the possibility of "spilling" into $N \setminus I$; it claims that the part of μ concentrated on I is concentrated on $\mathcal{E}_\beta$. However, we know from Sec. 2 that μ is concentrated on J_S for some increasing sequence S of positive integers.

5.3. *Theorem*: Let $\lambda \in \mathcal{E}_\beta$ and μ be the central measure of λ. Then there exists an increasing sequence S of positive integers such that μ is concentrated on $\mathcal{E}_{S,\beta}$.

The proof of this theorem differs in no essential way from the proof of the proposition, so we omit it. As we mentioned at the beginning of this section, μ is also concentrated on a Borel set of factor states, so μ is concentrated on a Borel set of factor states in $\mathcal{E}_{S,\beta}$. Obviously, our next step is to study such states. Before doing so, however, we must analyze the interaction more carefully.

6. FOURIER ANALYSIS OF THE INTERACTION

On the space of all bounded, complex sequences let λ be an arbitrary Banach limit (see Ref. 7). Ultimately

there will be no dependence on our choice of λ. Notice that λ is a bounded, self-adjoint, linear functional on a commutative C^*-algebra. In order to avoid confusion with regard to indices, we will denote a sequence $\{b_N\}$ by $N \to b_N$.

6.1. *Definition*: Let p be a state on $\mathcal{A}$. Then $B_{j,n,p}$ is the operator on $\mathcal{H}_p$ defined by

$$(B_{j,n,p}\Psi_1, \Psi_2) = \lambda\left(N \to \sum_{k=1}^{N} \mu_{jk}(N)(\pi_p[\exp(inx_k)]\Psi_1, \Psi_2)\right) \tag{6.1}$$

for all $\Psi_1, \Psi_2 \in \mathcal{H}_p$.

It follows from (2.3) that the sequence in the argument of λ is indeed bounded, so (6.1) makes sense. Note that the $B_{j,n,p}$ are uniformly bounded, and recall that $\{a_n | n \in \mathbb{Z}\}$ is the sequence of Fourier coefficients of f.

6.2. *Proposition*: $B_{j,n,p} \in \pi_p(\mathcal{A})' \cap \pi_p(\mathcal{A})''$ and $\sum_{n=-\infty}^{\infty} a_n B_{j,n,p}\pi_p[\exp(-inx_j)]$ converges absolutely in norm.

Proof: The second part of the statement follows from the fact that $\sum_{n=-\infty}^{\infty} |a_n| < \infty$.

Let $A \in \mathcal{A}_r$ and $\Psi_1, \Psi_2 \in \mathcal{H}_p$. Letting $\pi_p(A)\Psi_1, \Psi_2$ be the vectors in (6.1), we obtain

$$(B_{j,n,p}\pi_p(A)\Psi_1, \Psi_2) = \lambda\left(N \to \sum_{k=1}^{N} \mu_{jk}(N)(\pi_p[\exp(inx_k)A]\Psi_1, \Psi_2)\right).$$

Letting $\Psi_1, \pi_p(A)^*\Psi_2$ be the vectors, we get

$$(\pi_p(A)B_{j,n,p}\Psi_1, \Psi_2) = \lambda\left(N \to \sum_{k=1}^{N} \mu_{jk}(N)(\pi_p(A\exp[inx_k])\Psi_1, \Psi_2)\right).$$

Since $A \in \mathcal{A}_r$, it follows from (1.2) that

$$\begin{aligned}([B_{j,n,p}, \pi_p(A)]\Psi_1, \Psi_2) &= \lambda(N \to \mu_{jr}(N)(\pi_p([\exp(inx_r), A])\Psi_1, \Psi_2)) \\ &= \lim_{N\to\infty} \mu_{jr}(N)(\pi_p([\exp(inx_r), A])\Psi_1, \Psi_2) \\ &= 0.\end{aligned}$$

Thus $B_{j,n,p} \in \pi_p(\mathcal{A})'$.

Let $C \in \pi_p(\mathcal{A})'$ and $\Psi_1, \Psi_2 \in \mathcal{H}_p$. Using the vectors $C\Psi_1, \Psi_2$ and also the vectors $\Psi_1, C^*\Psi_2$, we obtain

$$\begin{aligned}([B_{j,n,p}, C]\Psi_1, \Psi_2) &= \lambda\left(N \to \sum_{k=1}^{N} \mu_{jk}(N)([\pi_p(\exp[inx_k]), C]\Psi_1, \Psi_2)\right) \\ &= 0.\end{aligned}$$

Thus $B_{j,n,p} \in \pi_p(\mathcal{A})''$.

This result will play a key role in the characterization of β-equilibrium factor states for the special class of interactions to be considered. The commutation result proven above depends upon the fact that the interaction is infinitely weak, and it is in this sense that such an interaction provides a certain amount of "Abelianness" as remarked upon in the Introduction.

6.3. *Proposition*: If $p \in \mathcal{S}$, then $W_{j,p} = \sum_{n=-\infty}^{\infty} a_n B_{j,n,p} \times \pi_p(\exp(-inx_j))$ and $V_{j,p}(x) = \sum_{n=-\infty}^{\infty} a_n B_{j,n,p} \exp(-inx)$.

Proof: Observe that $\sum_{k=1}^{N} \mu_{jk}(N) f(x_k - x_j) = \sum_{n=-\infty}^{\infty} a_n \times \exp(-inx_j) \sum_{k=1}^{N} \mu_{jk}(N) \exp(inx_k)$ and that the convergence of the series is uniform in N. Since λ is a bounded linear functional on the space of all bounded sequences, we have

$$\begin{aligned}(W_{j,p}\Psi_1, \Psi_2) &= \lim_{N\to\infty}\left[\pi_p\left(\sum_{n=-\infty}^{\infty} a_n \exp(-inx_j)\sum_{k=1}^{N} \mu_{jk}(N)\exp(inx_k)\right)\Psi_1, \Psi_2\right] \\ &= \lambda\left(N \to \left[\pi_p\left(\sum_{n=-\infty}^{\infty} a_n \exp(-inx_j)\sum_{k=1}^{N} \mu_{jk}(N)\exp(inx_k)\right)\Psi_1, \Psi_2\right]\right) \\ &= \sum_{n=-\infty}^{\infty} a_n(\pi_p[\exp(-inx_j)])B_{j,n,p}\Psi_1, \Psi_2).\end{aligned}$$

The proof of the other equation is similar.

6.4. *Definition*: Let $p \in \mathcal{S}$ and let S be an increasing sequence of positive integers. $B_{S,j,n,p}$ is the operator on $\mathcal{H}_p$ such that

$$(B_{S,j,n,p}\Psi_1, \Psi_2) = \lambda\left(N \to \sum_{k=1}^{S_N} \mu_{jk}(S_N)(\pi_p[\exp(inx_k)]\Psi_1, \Psi_2)\right) \tag{6.2}$$

for all $\Psi_1, \Psi_2 \in \mathcal{H}_p$.

It is clear that $B_{S,j,n,p}$ has the properties stated in Proposition 6.2 for $B_{j,n,p}$. Furthermore, if $p \in \mathcal{J}_S$, then $V_{S,j,p}(x) = \sum_{n=-\infty}^{\infty} a_n B_{S,j,n,p} \exp(-inx)$ and $W_{S,j,p} = \sum_{n=-\infty}^{\infty} a_n B_{S,j,n,p}\pi_p[\exp(-inx_j)]$.

7. EQUILIBRIUM FACTOR STATES

Let p be a factor state in $\mathcal{E}_\beta$. Since $\pi_p(\mathcal{A})' \cap \pi_p(\mathcal{A})'' = \mathbb{C}1_{\mathcal{H}_p}$, the operators $B_{j,n,p}$ of the preceding section are scalar multiples of the identity operator. By the expressions we have in Proposition 6.3 for $V_{j,p}$ and $W_{j,p}$, it follows that $V_{j,p}$ is the identity operator times a real-valued, continuous function, and, identifying $V_{j,p}$ with that function, it also follows that $W_{j,p} = \pi_p(V_{j,p}(x_j))$. In physical terms, $V_{j,p}$ is the potential field in which the jth particle is moving when the system is in state p.

7.1. *Lemma*: $\pi_p(\exp[it(-\Delta_j + V_{j,p}(x_j))]) = \exp[it(-\Delta_{j,p} + W_{j,p})]$ for all real t.

The proof of this lemma differs in no essential way from the proof of Lemma 3.1, so we omit it.

Thus

$$\exp(itH_{\Omega,p}) = \pi_p\left(\exp\left(it\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))\right)\right)$$

for all finite $\Omega \subset \mathbb{Z}^+$. Since p satisfies (4.1), it follows that p satisfies the β-KMS condition with respect to the group of automorphisms of $\mathcal{A}_\Omega$ implemented by the unitary group $\exp[it\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))]$ on $\mathcal{H}_\Omega$. But p is also normal on $\mathcal{A}_\Omega = \mathcal{B}(\mathcal{H}_\Omega)$.

Now, every normal state on $\mathcal{B}(\mathcal{H}_\Omega)$ is clearly a factor state on $\mathcal{B}(\mathcal{H}_\Omega)$ by the strong continuity of the corre-

sponding faithful representation. Also, the set of normal β-KMS states on $\mathcal{B}(\mathcal{H}_\Omega)$ is obviously convex. By Ref. 8, the β-KMS factor states are exactly the extremal β-KMS states. Hence there is at most one normal β-KMS state on $\mathcal{B}(\mathcal{H}_\Omega)$.

On the other hand, $\exp[-s\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))]$ is a trace-class operator for all $s > 0$ because $-\Delta_j$ has the discrete spectrum $\{m^2 \mid m \in \mathbb{Z}\}$ and $V_{j,p}$ is a bounded function (see Ref. 9). Thus there is exactly one normal β-KMS state on $\mathcal{B}(\mathcal{H}_\Omega)$, namely the one whose density matrix is

$$\frac{1}{\operatorname{tr}\exp[-\beta\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))]} \times \exp\left[-\beta\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))\right].$$

Thus we have proven the following.

7.2. *Theorem*:

$$p(A) = \frac{1}{\operatorname{tr}\exp[-\beta\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))]} \times \operatorname{tr}\left(A\exp\left[-\beta\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))\right]\right)$$

for all $A \in \mathcal{A}_\Omega$.

Since Ω was arbitrary, we now have a preliminary form for p on the whole algebra. Of course, the expressions are not explicit.

7.3. *Corollary*: p is a product state.

Proof: If $A \in \mathcal{A}_\Omega$ and $B \in \mathcal{A}_{\Omega'}$, where $\Omega \cap \Omega' = \phi$, then

$$\begin{aligned} p(AB) &= \frac{1}{\operatorname{tr}\exp[-\beta\sum_{j\in\Omega\cup\Omega'}(\Delta_j + V_{j,p}(x_j))]} \times \operatorname{tr}\left[AB\exp\left[-\beta\sum_{j\in\Omega\cup\Omega'}(-\Delta_j + V_{j,p}(x_j))\right]\right] \\ &= \operatorname{tr}\left[A\exp\left[-\beta\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))\right]\right] \times \operatorname{tr}\left[B\exp\left[-\beta\sum_{j\in\Omega'}(-\Delta_j + V_{j,p}(x_j))\right]\right] \Big/ \\ &\quad \operatorname{tr}\exp\left[-\beta\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))\right] \times \operatorname{tr}\exp\left[-\beta\sum_{j\in\Omega'}(-\Delta_j + V_{j,p}(x_j))\right] \\ &= p(A)p(B), \end{aligned}$$

since $\exp[-\beta\sum_{j\in\Omega}(-\Delta_j + V_{j,p}(x_j))]$ and $\exp[-\beta\sum_{j\in\Omega'}(-\Delta_j + V_{j,p}(x_j))]$ lie in $\mathcal{A}_\Omega$ and $\mathcal{A}_{\Omega'}$, respectively. This completes the proof.

Now recall that $V_{j,p}(x) = \text{s-}\lim_{N\to\infty}\pi_p(\sum_{k=1}^N \mu_{jk}(N)f(x_k - x))$. Since $V_{j,p}(x)$ is a scalar, we have in particular

$$\begin{aligned} V_{j,p}(x) &= (V_{j,p}(x)\Phi_p, \Phi_p) \\ &= \lim_{N\to\infty}\left(\pi_p\left(\sum_{k=1}^N \mu_{jk}(N)f(x_k - x)\right)\Phi_p, \Phi_p\right) \\ &= \lim_{N\to\infty}\sum_{k=1}^N \mu_{jk}(N)p(f(x_k - x)). \end{aligned}$$

Hence, we have the following compatibility conditions:

$$\begin{aligned} V_{j,p}(x) &= \lim_{N\to\infty}\sum_{k=1}^N \mu_{jk}(N) \times \frac{\operatorname{tr}[\exp[-\beta(-\Delta_k + V_{k,p}(x_k))]f(x_k - x)]}{\operatorname{tr}\exp[-\beta(-\Delta_k + V_{k,p}(x_k))]} \\ &= \lim_{N\to\infty}\sum_{k=1}^N \mu_{jk}(N)\frac{\operatorname{tr}[\exp[-\beta(-\Delta + V_{k,p}(y))]f(y - x)]}{\operatorname{tr}\exp[-\beta(-\Delta + V_{k,p}(y))]}. \end{aligned} \tag{7.1}$$

These equations hold pointwise in x. $-\Delta + V_{k,p}(y) = -d^2/dy^2 + V_{k,p}(y)$ operates on $L^2(-\pi, \pi)$, and for each fixed x, $f(y - x)$ denotes the operator

$$\phi(y) \to f(y - x)\phi(y),$$

where the translation $y - x$ is modulo 2π.

These equations for the functions $V_{j,p}$ are *necessary* conditions for the state p to be a β-equilibrium factor state. This fact alone, of course, is not enough to make the equations themselves interesting from the standpoint of computing β-equilibrium states. There will be more to say about this problem when we consider interactions of a special kind.

Suppose that S is an increasing sequence of positive integers and that p is an S-β-equilibrium factor state. Then by reasoning identical to that given above, we obtain that p is a product state,

$$p(A) = \frac{\operatorname{tr}[\exp[-\beta(-\Delta_j + V_{S,j,p}(x_j))]A]}{\operatorname{tr}\exp[-\beta(-\Delta_j + V_{S,j,p}(x_j))]} \tag{7.2}$$

for $A \in \mathcal{A}_j$, and

$$V_{S,j,p}(x) = \lim_{N\to\infty}\sum_{k=1}^{S_N} \mu_{jk}(S_N) \times \frac{\operatorname{tr}[\exp[-\beta(-\Delta + V_{S,k,p}(y))]f(y - x)]}{\operatorname{tr}\exp[-\beta(-\Delta + V_{S,k,p}(y))]}. \tag{7.3}$$

8. INTERACTIONS OF THE Pth KIND

We are now ready to consider special classes of interactions for which more complete results can be obtained.

8.1. *Definition*: $\{\mu_{jk}\}$ is said to be *of the Pth kind* if and only if $\{\mu_{jk}\}$ has period P with respect to j (and therefore with respect to k) and

$$\lim_{N\to\infty}\mu_{jk}(N)\left|\{n \in \mathbb{Z}^+ \mid nP + k \le N\}\right| \tag{8.1}$$

exists for $1 \le k \le P$, where $|\ |$ denotes the cardinality of the set.

Suppose $\{\mu_{jk}\}$ is of the Pth kind, and let p be a factor state in $\mathcal{E}_{S,\beta}$ for an increasing sequence S in $\mathbb{Z}^+$. By the preceding section we have

$$V_{S,j,p}(x) = \lim_{N\to\infty}\sum_{k=1}^{S_N} \mu_{jk}(S_N) \times \frac{\operatorname{tr}[\exp(-\beta(-\Delta + V_{S,k,p}(y))]f(y - x)]}{\operatorname{tr}\exp[-\beta(-\Delta + V_{S,k,p}(y))]}.$$

Since $\{\mu_{jk}\}$ is periodic in j, the sequence of functions $\{V_{S,j,p}\}$ is periodic in j. It follows that the sequence

$$\frac{\operatorname{tr}[\exp[-\beta(-\Delta + V_{S,j,p}(y))]f(y - x)]}{\operatorname{tr}\exp[-\beta(-\Delta + V_{S,j,p}(y))]}$$

is periodic in j. Since $\{\mu_{jk}\}$ is of the Pth kind,

$$\lim_{N\to\infty}\sum_{k=1}^{N}\mu_{jk}(N)\frac{\mathrm{tr}[\exp[-\beta(-\Delta+V_{S,k,p}(y))]f(y-x)]}{\mathrm{tr}\exp[-\beta(-\Delta+V_{S,k,p}(y))]}$$

exists and is equal to

$$\sum_{k=1}^{P}\lambda_{jk}\frac{\mathrm{tr}[\exp[-\beta(-\Delta+V_{S,k,p}(y))]f(y-x)]}{\mathrm{tr}\exp[-\beta(-\Delta+V_{S,k,p}(y))]}$$

where

$$\lambda_{jk}=\lim_{N\to\infty}\mu_{jk}(N)\left|\{n\in\mathbf{Z}^{+}\mid nP+k\le N\}\right|.$$

In summary, we have

$$V_{S,j,p}(x)=\sum_{k=1}^{P}\lambda_{jk}\times\frac{\mathrm{tr}[\exp[-\beta(-\Delta+V_{S,k,p}(y))]f(y-x)]}{\mathrm{tr}\exp[-\beta(-\Delta+V_{S,k,p}(y))]}.$$

Before proceeding further, we need a lemma, which also happens to demonstrate the abundance of asymptotic states for an interaction of the Pth kind. The proof is an adaptation of a proof in Ref. 1 that the time-evolution is unambiguous for equilibrium states of a quantum lattice.

8.2. *Lemma*: Suppose $\{\mu_{jk}\}$ is of the Pth kind, and let p be a locally normal product state such that the sequence $\{p(\exp(inx_j))\}$ is periodic in j with period P for all n. Then p is an asymptotic state.

Proof: Let $A_{j,n,p}=\sum_{k=1}^{P}\lambda_{jk}p(\exp(inx_k))$ and $A\in\mathcal{A}_\Omega$ for some finite $\Omega\subset\mathbf{Z}^{+}$. We wish to show first that

$$\left\|\sum_{j=1}^{N}\mu_{rj}(N)\pi_p(\exp(inx_j))\pi_p(A)\Phi_p-A_{r,n,p}\pi_p(A)\Phi_p\right\|$$

converges to zero as $N\to\infty$. The expression is certainly dominated by

$$\left\|\sum_{\substack{j=1\\ j\in\Omega}}^{N}\mu_{rj}(N)\pi_p(\exp(inx_j))\pi_p(A)\Phi_p\right\|+\left\|\sum_{\substack{j=1\\ j\notin\Omega}}^{N}\mu_{rj}(N)\pi_p(\exp(inx_j))\pi_p(A)\Phi_p-A_{r,n,p}\pi_p(A)\Phi_p\right\|.$$

Since Ω is fixed, the first term tends to zero. Squaring the second term, we get the expression

$$\sum_{\substack{j=1\\ j\notin\Omega}}^{N}\sum_{\substack{j'=1\\ j'\notin\Omega}}^{N}\mu_{rj}(N)\mu_{rj'}(N)p(A^{*}\exp(-inx_{j'})\exp(inx_j)A)-\overline{A_{r,n,p}}\sum_{\substack{j'=1\\ j'\notin\Omega}}^{N}\mu_{rj'}(N)p(A^{*}\exp(-inx_{j'})A)-A_{r,n,p}\times\sum_{\substack{j=1\\ j\notin\Omega}}^{N}\mu_{rj}(N)p(A^{*}\exp(inx_j)A)+|A_{r,n,p}|^{2}p(A^{*}A).\qquad(*)$$

Since p is a product state and $A\in\mathcal{A}_\Omega$, (*) may be rewritten as

$$p(A^{*}A)\left|\sum_{\substack{j=1\\ j\notin\Omega}}^{N}\mu_{rj}(N)p(\exp(inx_j))-A_{r,n,p}\right|^{2}+p(A^{*}A)\sum_{\substack{j=1\\ j\notin\Omega}}^{N}\mu_{rj}(N)^{2}(1-|p(\exp(inx_j))|^{2}).$$

Using the fact that $p(\exp(inx_j))$ is periodic in j with period p and that $\mu_{rj}(N)\to 0$ as $N\to\infty$ for each j in Ω, we see that (*) becomes in the limit

$$p(A^{*}A)\left|\sum_{k=1}^{P}\lambda_k p(\exp(inx_k))-A_{r,n,p}\right|^{2}+p(A^{*}A)\lim_{N\to\infty}\sum_{k=1}^{P}\mu_{rk}(N)^{2}\left|\{n\in\mathbf{Z}^{+}\mid np+k\le N\}\right|\times|(1-|p(\exp(inx_k))|^{2}).$$

The first term is certainly zero, and the second term is zero because $\lim_{N\to\infty}\mu_{rk}(N)=0$ and $\lambda_{rk}=\lim_{N\to\infty}\mu_k(N)\times|\{n\in\mathbf{Z}^{+}\mid nP+k\le N\}|$.

Thus we have shown that $\sum_{j=1}^{N}\mu_{rj}(N)\pi_p(\exp(inx_j))-A_{r,n,p}1_{H_p}$ as $N\to\infty$ on a dense subspace of H_p. Since the operators $\sum_{j=1}^{N}\mu_{rj}(N)\,\pi_p(\exp(inx_j))$ are uniformly bounded in N, we have pointwise convergence everywhere in H_p. Now recall that

$$V_{r,N}(x)=\sum_{j=1}^{N}\mu_{rj}(N)f(x_j-x)=\sum_{n=-\infty}^{\infty}a_n\exp(-inx)\sum_{j=1}^{N}\mu_{rj}(N)\exp(inx_j).$$

Given $\Psi\in H_p$, it follows that

$$\left\|\pi_p(V_{r,N}(x))\Psi-\sum_{n=-\infty}^{\infty}a_nA_{r,n,p}\exp(-inx)\Psi\right\|$$

tends to zero as $N\to\infty$, since $\sum_{n=-\infty}^{\infty}|a_n|<\infty$. Thus $V_{r,p}(x)$ exists. By similar reasoning, $W_{r,p}$ exists. This completes the proof that p is an asymptotic state.

8.3. *Theorem*: Suppose that $\{\mu_{jk}\}$ is of the Pth kind, and let p be a factor state in $\mathcal{E}_{S,\beta}$ for some increasing S in $\mathbf{Z}^{+}$. Then p is a factor state in $\mathcal{E}_\beta$.

Proof: It is clear that we need only show that $p\in I$. By the preceding lemma, it is sufficient to show that p is a locally normal product state such that $p(\exp(inx_j))$ is periodic in j with period P.

By the preceding section, p is certainly a locally normal product state. Furthermore, we have in particular that

$$p(\exp(inx_j))=\frac{\mathrm{tr}[\exp[-\beta(-\Delta+V_{S,j,p}(y))]\exp(iny)]}{\mathrm{tr}\exp[-\beta(-\Delta+V_{S,j,p}(y))]}.$$

As we have pointed out at the beginning of this section, $V_{S,j,p}$ is periodic in j with period P. This completes the proof.

8.4. *Corollary*: The central measure of a β-equilibrium state is concentrated on a Borel set of β-equilibrium factor states when the interaction is of the Pth kind.

We finally have the decomposition theorem we want. Our next step is to characterize the factor states in $\mathcal{E}_\beta$ for interactions of the Pth kind, and we begin by stating what has been accomplished so far. If p is a factor state in $\mathcal{E}_\beta$, then p is a product state such that

$$p(A)=\frac{\mathrm{tr}[\exp[-\beta(-\Delta_j+V_{j,p}(x_j))]A]}{\mathrm{tr}\exp[-\beta(-\Delta_j+V_{j,p}(x_j))]} \tag{8.2}$$

for all A in $\mathcal{A}_j$, and

$$V_{j,p}(x)=\sum_{k=1}^{P}\lambda_{jk}\frac{\mathrm{tr}[\exp[-\beta(-\Delta+V_{k,p}(y))]f(y-x)]}{\mathrm{tr}\exp[-\beta(-\Delta+V_{k,p}(y))]} \tag{8.3}$$

for all $j\in\mathbb{Z}^+$.

These conditions are *necessary* for p to be a factor state in $\mathcal{E}_\beta$. The following result establishes that the consistency equations (8.3) are *sufficient* conditions for the state corresponding to the functions via (8.2) to be a factor state in $\mathcal{E}_\beta$.

8.5. *Theorem*: Suppose $\{\mu_{jk}\}$ is of the Pth kind, and let $\{g_j\}$ be a sequence of real-valued, bounded, measurable functions such that

$$g_j(x)=\sum_{k=1}^{P}\lambda_{jk}\frac{\mathrm{tr}[\exp[-\beta(-\Delta+g_k(y))]f(y-x)]}{\mathrm{tr}\exp[-\beta(-\Delta+g_k(y))]}. \tag{8.4}$$

Let p be the product state such that

$$p(A)=\frac{\mathrm{tr}[\exp[-\beta(-\Delta_j+g_j(x_j))]A]}{\mathrm{tr}\exp[-\beta(-\Delta_j+g_j(x_j))]}$$

for all $A\in\mathcal{A}_j$ and all j. Then p is a factor state in $\mathcal{E}_\beta$. Therefore, $V_{j,p}$ exists and is equal to g_j.

Proof: Obviously, p is a locally normal product state. Moreover,

$$p(\exp(inx_j))=\frac{\mathrm{tr}[\exp[-\beta(-\Delta+g_j(y))]\exp(iny)]}{\mathrm{tr}\exp[-\beta(-\Delta+g_j(y))]}.$$

Since g_j is periodic in j with period P, so is $p(\exp(inx_j))$. By Lemma 8.2, it follows that p is an asymptotic state.

The next thing to notice is that p is a β-KMS state on $\mathcal{A}_j$ with respect to the automorphism group of $\mathcal{A}_j=\mathcal{B}(\mathcal{H}_j)$ implemented by the unitary group $\exp[it(-\Delta_j+g_j(x_j))]$ on $\mathcal{H}_j$, for all j. Hence p is a β-KMS state with respect to the automorphism group of $\mathcal{A}$ which is locally defined in this way. We wish to show that $p\in\mathcal{E}_\beta$.

Since p is a locally normal product state, it follows from Ref. 4 that p is a factor state. But recall that for an asymptotic state which is also a factor state, $V_{j,p}$ is an ordinary function and $W_{j,p}=\pi_p(V_{j,p}(x_j))$. In fact, by Proposition 6.2, the $B_{j,n,p}$ are scalars and

$$\begin{aligned}B_{j,n,p}&=\lambda\left(N\to\sum_{k=1}^{N}\mu_{jk}(N)p[\exp(inx_k)]\right)\\&=\lim_{N\to\infty}\sum_{k=1}^{N}\mu_{jk}(N)p[\exp(inx_k)]\\&=\sum_{k=1}^{P}\lambda_{jk}p[\exp(inx_k)],\end{aligned}$$

since $p[\exp(inx_k)]$ is periodic in k with period P and $\{\mu_{jk}\}$ is of the Pth kind. We know that

$$V_{j,p}(x)=\sum_{n=-\infty}^{\infty}a_nB_{j,n,p}\exp(-inx),$$

and we have by assumption that

$$\begin{aligned}g_j(x)&=\sum_{k=1}^{P}\lambda_{jk}p(f(x_k-x))\\&=\sum_{k=1}^{P}\lambda_{jk}\sum_{n=-\infty}^{\infty}a_n\exp(-inx)p[\exp(inx_k)].\end{aligned}$$

Thus $V_{j,p}=g_j$. It follows that the dynamics α_p of p induces the automorphism group of $\mathcal{A}$ mentioned above. Since p is β-KMS with respect to this automorphism group, the proof is complete.

Thus, combining this theorem with the remark preceding it, we have

8.6. *Theorem*: Suppose the interaction is of the Pth kind. Then there is a one-to-one correspondence between factor states in $\mathcal{E}_\beta$ and p-tuples $(g_1,\ldots,g_p)$ of bounded, real-valued, measurable functions satisfying (8.4).

Thus, the computation of β-equilibrium factor states is reduced to a very concrete problem. Notice that there is one and only one solution $(g_1,\ldots,g_p)$ whose entries are constant, namely $g_j=a_0\sum_{k=1}^{P}\lambda_{jk}$, as the following calculation shows:

$$\begin{aligned}g_j&=\sum_{k=1}^{P}\lambda_{jk}\frac{\mathrm{tr}[\exp[-\beta(-\Delta+g_k)]f(y-x)]}{\mathrm{tr}\exp[-\beta(-\Delta+g_k)]}\\&=\sum_{k=1}^{P}\lambda_{jk}\frac{\mathrm{tr}[\exp(\beta\Delta)f(y-x)]}{\mathrm{tr}\exp(\beta\Delta)}.\end{aligned}$$

Taking the trace with respect to the eigenfunctions $\exp(imy)$ of Δ, we get

$$\begin{aligned}g_j&=\sum_{k=1}^{P}\lambda_{jk}\frac{1}{\sum_{m=-\infty}^{\infty}\exp(-\beta m^2)}\sum_{m=-\infty}^{\infty}\exp(-\beta m^2)\\&\quad\times\frac{1}{2\pi}\int_{-\pi}^{\pi}f(y-x)\exp(imy)\exp(-imy)\,dy\\&=\sum_{k=1}^{P}\lambda_{jk}\frac{1}{2\pi}\int_{-\pi}^{\pi}f(y-x)\,dy\\&=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(y)\,dy\sum_{k=1}^{P}\lambda_{jk}.\end{aligned}$$

Since $a_0=(1/2\pi)\int_{-\pi}^{\pi}f(y)\,dy$, the calculation is complete. Thus the constant potential field in which the jth particle moves when the system is in this state is computed by averaging the underlying interaction function f and multiplying by the "total coupling strength" $\sum_{k=1}^{P}\lambda_{jk}$. This state is a "free" state, and it should not be surprising that there is one, since the interaction we are considering is "infinitely weak." We note that the constant solution $(g_1,\ldots,g_p)$ is independent of β, while the corresponding product state

$$p(A)=\frac{\mathrm{tr}[\exp(\beta\Delta_j)A]}{\mathrm{tr}\exp(\beta\Delta_j)},\quad A\in\mathcal{A}_j,$$

is dependent on β. In particular, $\mathcal{E}_\beta\neq\phi$ for all $\beta>0$.

In Sec. 9 we will given an example where the free solution is not the only solution. We note at this point that if $(g_1,\ldots,g_P)$ is a solution with nonconstant entries, then all of its spatial translates modulo 2π are also solutions because $-\Delta$ has periodic boundary conditions.

Our characterization of β-equilibrium factor states for interactions of the Pth kind enables us to prove the following theorem:

8.7. *Theorem*: Suppose that the interaction is of the Pth kind and $p \in \mathcal{E}_\beta$. Then p is an extreme point of $\mathcal{E}_\beta$ if and only if p is a factor state.

Proof: That every extreme point of $\mathcal{E}_\beta$ is a factor state follows immediately from Corollary 8.4. Now suppose that p is a factor state and that p is not an extreme point of $\mathcal{E}_\beta$. Then $p = \frac{1}{2}(p_1 + p_2)$ for some $p_1, p_2 \in \mathcal{E}_\beta$ such that $p_1 \neq p_2$. Let μ_1 and μ_2 be the central measures of p_1 and p_2, respectively. By Corollary 8.4 μ_1 and μ_2 are concentrated on Borel sets of β-equilibrium factor states. Let $\mu = \frac{1}{2}(\mu_1 + \mu_2)$ and note that p is the resultant of μ and μ is concentrated on a Borel set of β-equilibrium factor states. In particular,

$$p(f(x - x_k) f(x - x_{k+P})) = \int \sigma(f(x - x_k) f(x - x_{k+P}))\, d\mu(\sigma)$$

for $-\infty < x < \infty$ and $k \in \mathbb{Z}^+$. Since p is a factor state in $\mathcal{E}_\beta$, p is a product state and

$$p(f(x - x_{k+P})) = p(f(x - x_k)).$$

As μ is concentrated on a Borel set of states with the same properties, we have

$$\begin{aligned} p(f(x - x_k))^2 &= p(f(x - x_k)) p(f(x - x_{k+P})) \\ &= p(f(x - x_k) f(x - x_{k+P})) \\ &= \int \sigma(f(x - x_k) f(x - x_{k+P}))\, d\mu(\sigma) \\ &= \int \sigma(f(x - x_k))(f(x - x_{k+P}))\, d\mu(\sigma) \\ &= \int \sigma(f(x - x_k))^2\, d\mu(\sigma). \end{aligned}$$

Hence

$$\left[\int \sigma(f(x - x_k))\, d\mu(\sigma)\right]^2 = \int \sigma(f(x - x_k))^2\, d\mu(\sigma).$$

Therefore, $\sigma(f(x - x_k))$ must be a constant μ-almost everywhere for fixed x, and, since

$$p(f(x - x_k)) = \int \sigma(f(x - x_k))\, d\mu(\sigma),$$

that constant is $p(f(x - x_k))$.

Let $\{b_m\}$ be a sequence that is dense in the real line and let $\mathcal{J}_{mk}$ be the set of all states σ in K such that $\sigma(f(b_m - x_k)) = p(f(b_m - x_k))$, where K is a Borel set of factor states in $\mathcal{E}_\beta$ such that $\mu(K) = 1$. Then $\mu(\cap_{m,k=1}^{\infty} \mathcal{J}_{mk}) = 1$ and $\cap_{m,k=1}^{\infty} \mathcal{J}_{mk} = \{\sigma \in K \mid \sigma(f(x - x_k)) = p(f(x - x_k))$ for $-\infty < x < \infty$ and $k \in \mathbb{Z}^+\}$. For each factor state σ in $\mathcal{E}_\beta$, we denote its corresponding P-tuples of functions by $(g_1^\sigma, \ldots, g_P^\sigma)$. Since

$$g_j^\sigma(x) = \sum_{k=1}^{P} \lambda_{jk} \sigma(f(x - x_k)),$$

it follows that $g_j^\sigma(x) = g_j^\rho(x)$ for all $\sigma \in \cap_{m,k=1}^{\infty} \mathcal{J}_{mk}$, $-\infty < x < \infty$, and $j \in \mathbb{Z}^+$. As $(g_1^\sigma, \ldots, g_P^\sigma)$ characterizes σ, we have $\cap_{m,k=1}^{\infty} \mathcal{J}_{mk} = \{\rho\}$, so $\mu(\{\rho\}) = 1$. Hence $\mu_1(\{\rho\}) = \mu_2(\{\rho\}) = 1$, so $p_1 = p_2 = p$. This is a contradiction.

Thus the central decomposition of a β-equilibrium state is also an extremal decomposition. Whether $\mathcal{E}_\beta$ is a simplex or not is an open question. Of course, $\mathcal{E}_\beta$ is not a Choquet simplex because it is not compact, but whether the extremal decomposition of a β-equilibrium state is unique or not is still a pertinent question.

We close this section with the statement of a result which is stronger than the converse of Corollary 8.4.

8.8. *Proposition*: Let μ be a probability measure concentrated on $\mathcal{E}_\beta$. Then the resultant of μ lies in $\mathcal{E}_\beta$.

The proof does not depend on the theory developed in this section. One need only use the definition of asymptotic state together with a dominated convergence argument. Proving the compatibility of the dynamics is straightforward.

9. ABSENCE OF PHASE TRANSITIONS AT HIGH TEMPERATURES

Our next step is to examine the bifurcation theory with respect to the inverse temperature β. We have the following result:

9.1. *Theorem*: Suppose that the interaction is of the Pth kind. Then there is only one β-equilibrium state for β sufficiently small.

Proof: By the decomposition theorem, it is sufficient to show that there is only one β-equilibrium factor state for β small enough. By our characterization of β-equilibrium factor states and the remarks of the preceding section, it is enough to prove that the free solution is the only solution to our system of equations for β sufficiently small.

Notice that if $(g_1, \ldots, g_P)$ solves the equations for f, then for a real constant C, $(g_1 + C\sum_{k=1}^{P} \lambda_{1k}, \ldots, g_P + C\sum_{k=1}^{P} \lambda_{Pk})$ solves the equations for $f + C$, so the abundance of solutions is unaffected by the assumption that

$$\int_{-\pi}^{\pi} f(x)\, dx = 0.$$

In this case $(0, \ldots, 0)$ is the free solution. Let $(g_1, \ldots, g_P)$ be an arbitrary β-solution. We need to find a $\beta_0 > 0$, independent of $(g_1, \ldots, g_P)$, such that if $\beta \leq \beta_0$, then $g_j = 0$ for $1 \leq j \leq P$. Since

$$\operatorname{tr}[\exp[-\beta(-\Delta)] f(y - x)] = 0,$$

we have

$$\begin{aligned} |g_j(x)| &\leq \sum_{k=1}^{P} |\lambda_{jk}| \frac{|\operatorname{tr}[\exp[-\beta(-\Delta + g_k(y))] f(y - x)]|}{\operatorname{tr}\exp[-\beta(-\Delta + g_k(y))]} \\ &= \sum_{k=1}^{P} |\lambda_{jk}| \times \frac{|\operatorname{tr}\{[\exp[-\beta(-\Delta + g_k(y))] - \exp[-\beta(-\Delta)]] f(y - x)\}|}{\operatorname{tr}\exp[-\beta(-\Delta + g_k(y))]}. \end{aligned}$$

By Duhamel's formula, we have

$$\exp[-\beta(-\Delta + g_k(y))] - \exp[-\beta(-\Delta)] = -\int_0^\beta \exp[-(\beta - s)(-\Delta + g_k(y))] g_k(y) \exp[-s(-\Delta)]\, ds,$$

where the integral is a Riemann integral taken in the strong sense. By the Fubini theorem, we may interchange the integral and trace to obtain

$$|g_j(x)| \leq \sum_{k=1}^{P} |\lambda_{jk}| \int_0^\beta \frac{|\mathrm{tr}\{\exp[-(\beta-s)(-\Delta+g_k(y))]g_k(y)\exp[-s(-\Delta)]f(y-x)\}|}{\mathrm{tr}\exp[-\beta(-\Delta+g_k(y))]}$$

Recall that if A is a positive, trace-class operator on a Hilbert space, then $|\mathrm{tr}(AB)| \leq \|B\|\,\mathrm{tr}A$. Since $\|\exp[-s(-\Delta)]\| \leq 1$, $\|g_k(y)\| = \|g_k\|_\infty$, and $\|f(y-x\| = \|f\|_\infty$,

$$\begin{aligned}
&|g_j(x)| \\
&\leq \|f\|_\infty \sum_{k=1}^{P} |\lambda_{jk}| \|g_k\|_\infty \\
&\quad \times \int_0^\beta \frac{\mathrm{tr}\exp[-(\beta-s)(-\Delta+g_k(y))]}{\mathrm{tr}\exp[-\beta(-\Delta+g_k(y))]}\,ds \\
&= \|f\|_\infty \sum_{k=1}^{P} |\lambda_{jk}|\ \|g_k\|_\infty \frac{1}{\sum_{m=1}^{\infty}\exp(-\beta E_{km})} \\
&\quad \times \sum_{m=1}^{\infty} \int_0^\beta \exp[-(\beta-s)E_{km}]\,ds,
\end{aligned}$$

where the E_{km} are the eigenvalues of $-\Delta+g_k(y)$, indexed in increasing order, counting multiplicity. Now notice that we also have the very crude estimate

$$\|g_k\|_\infty \leq \|f\|_\infty \sum_{r=1}^{P} |\lambda_{kr}| \equiv b_k.$$

Hence

$$-\Delta - b_k \leq -\Delta + g_k(y) \leq -\Delta + b_k$$

on their common domain of self-adjointness. If E_m is the mth eigenvalue of $-\Delta$ in increasing order, counting multiplicity, it follows that

$$E_m - b_k \leq E_{km} \leq E_m + b_k,$$

from which we obtain

$$\begin{aligned}
&|g_j(x)| \\
&\leq \|f\|_\infty \sum_{k=1}^{P} |\lambda_{jk}| \|g_k\|_\infty \frac{1}{\exp(-\beta b_k)\sum_{m=1}^{\infty}\exp(-\beta E_m)} \\
&\quad \times \sum_{m=1}^{\infty} \int_0^\beta \exp[-(\beta-s)(E_m-b_k)]\,ds \\
&= \|f\|_\infty \sum_{k=1}^{P} |\lambda_{jk}|\ \|g_k\|_\infty h_k(\beta), \qquad (9.1)
\end{aligned}$$

where

$$h_k(\beta) = \frac{1}{\exp(-\beta b_k)\sum_{m=1}^{\infty}\exp(-\beta E_m)} \times \left(\sum_{\substack{m=1 \\ E_m \neq b_k}}^{\infty} \frac{1-\exp[-\beta(E_m-b_k)]}{E_m-b_k} + \beta\,|\{m \mid E_m = b_k\}| \right).$$

Now, $h_k(\beta)$ is independent of $(g_1, \ldots, g_P)$ and is *monotonically increasing* in β. The denominator obviously blows up as $\beta \downarrow 0$, and the numerator remains bounded because $E_{2m-1} = E_{2m} = (m-1)^2$ (i.e., $|\{m \mid E_m = b_k\}| \leq 2$, there are only a finite number of m such that $E_m < b_k$, and

$$\sum_{\substack{m=1 \\ E_m > b_k}}^{\infty} \frac{1}{E_m - b_k} < \infty.)$$

Choosing β_0 such that $h_k(\beta) \leq [2P\|f\|_\infty(1+|\lambda_{jk}|)]^{-1}$ for $\beta \leq \beta_0$, $1 \leq j,\ k \leq P$, (9.1) implies

$$\|g_j\|_\infty \leq \frac{1}{2P} \sum_{k=1}^{P} \|g_k\|_\infty$$

for $\beta \leq \beta_0$, $1 \leq j \leq P$. Summing over j, we obtain

$$\sum_{j=1}^{P} \|g_j\|_\infty \leq \tfrac{1}{2} \sum_{k=1}^{P} \|g_k\|_\infty,$$

so $g_k = 0$ for all k.

10. PHASE TRANSITIONS AND SYMMETRY BREAKING AT LOW TEMPERATURES

In considering interactions of the first kind, the β-equilibrium factor states correspond to functions g satisfying the equation

$$g(x) = \lambda \frac{\mathrm{tr}[\exp[-\beta(-\Delta+g(y))f(y-x)]}{\mathrm{tr}\exp[-\beta(-\Delta+g(y))]}, \qquad (10.1)$$

where λ is a constant depending upon the strength of the "infinitely weak" interaction. We wish to find an interaction and a temperature for which a phase transition occurs—that is, a function f with the required properties and a number $\beta > 0$ for which (10.1) has at least two solutions. For each $\beta > 0$, (10.1) has a constant solution $g = \lambda a_0$, so it is only necessary to find an f and β such that (10.1) has a nonconstant solution.

Let $\{\psi_m\}$ be a complete, orthonormal sequence of eigenvectors of $-\Delta+g(y)$ such that the corresponding sequence $\{E_m\}$ of eigenvalues is increasing. In terms of this basis, (10.1) becomes

$$g(x) = \frac{1}{2\pi\sum_{m=1}^{\infty}\exp(-\beta E_m)} \sum_{m=1}^{\infty} \exp(-\beta E_m) \times \int_{-\pi}^{\pi} f(\mu-x)\,|\psi_m(\mu)|^2\,d\mu. \qquad (10.2)$$

Since f is a continuous, real-valued, periodic function, g is also. If $\{b_n\}$ is the sequence of Fourier coefficients of g, we have

$$b_n = \frac{\lambda a_{-n}}{2\pi\sum_{m=1}^{\infty}\exp(-\beta E_m)} \sum_{m=1}^{\infty} \exp(-\beta E_m) \times \int_{-\pi}^{\pi} \exp(-in\mu)\,|\psi_m(\mu)|^2\,d\mu. \qquad (10.3)$$

Since $\{\psi_m\}$ and $\{E_m\}$ are indirectly related to $\{b_n\}$, this system of equations appears more hopeless than ever. However, all we need is to *find* a summable sequence $\{a_n\}$ such that (10.3) holds and $b_k \neq 0$ for some $k \neq 0$, so we embark on the simple-minded strategem of *picking* a convenient $\{b_n\}$, from which $\{\psi_m\}$ and $\{E_m\}$ arise, and solving for $\{a_n\}$,

$$a_n = \frac{2\pi b_{-n}}{\lambda \sum_{m=1}^{\infty} \exp(-\beta E_m) \int_{-\pi}^{\pi} \exp(in\mu) |\psi_m(\mu)|^2 d\mu} \times \sum_{m=1}^{\infty} \exp(-\beta E_m). \tag{10.4}$$

The only remaining obstacle is the possibility that the denominator may vanish for some n where $b_{-n} \neq 0$. We must find a $\beta > 0$ such that this does not happen for any such n. In this way we will have found an f and a β which yields as a solution the nonconstant function g that was picked to begin with.

To this end, let $b_1 = b_{-1} = -\frac{1}{6}$ and $b_n = 0$ for all other n. Then $g(x) = -\frac{1}{3}\cos x$. We need not compute $\{\psi_m\}$ and $\{E_m\}$ for $-\Delta - \frac{1}{3}\cos y$, but just show that there exists a $\beta > 0$ such that neither $\sum_{m=1}^{\infty} \exp(-\beta E_m) \int_{-\pi}^{\pi} \exp(-i\mu) \times |\psi_m(\mu)|^2 d\mu$ nor $\sum_{m=1}^{\infty} \exp(-\beta E_m) \int_{-\pi}^{\pi} \exp(i\mu) |\psi_m(\mu)|^2 d\mu$ vanish. Since one expression is the complex conjugate of the other, it is enough to show that the sum does not vanish. Hence the problem is reduced to showing that

$$\sum_{m=1}^{\infty} \exp(-\beta E_m) \int_{-\pi}^{\pi} \cos\mu |\psi_m(\mu)|^2 d\mu \neq 0 \tag{10.5}$$

for some $\beta > 0$. (10.5) holds for large β if

$$\sum_{E_m = E_1} \int_{-\pi}^{\pi} \cos\mu |\psi_m(\mu)|^2 d\mu \neq 0.$$

By the min–max principle discussed in Ref. 9, $-\frac{1}{3} \leq E_1 \leq \frac{1}{3}$ and $\frac{2}{3} \leq E_2 \leq \frac{4}{3}$; hence E_1 has no multiplicity, so in order to prove (10.5) for some β, it is sufficient to show that

$$\int_{-\pi}^{\pi} \cos\mu |\psi_1(\mu)|^2 d\mu \neq 0. \tag{10.6}$$

We must therefore investigate the behavior of the ground-state eigenfunction of the Hamiltonian $-\Delta - \frac{1}{3}\cos y$. Thus we have the equation

$$-\psi_1''(x) = (\tfrac{1}{3}\cos x + E_1)\psi_1(x)$$

with periodic boundary conditions. Since E_1 has no multiplicity, ψ_1 must be a complex multiple of a real-valued function, so we may assume that ψ_1 is real-valued. Since the law of nodes applies to periodic boundary conditions (see Ref. 10), we also have the information that $\psi_1(x) \neq 0$ for $-\pi < x < \pi$, so we may assume that $\psi_1(x) > 0$ for $-\pi < x < \pi$. Let c be the zero of $\frac{1}{3}\cos x + E_1$ such that $0 \leq c \leq \pi$. Then the graph of ψ_1 is concave down between $-c$ and c, and concave up in the other two regions. Since ψ_1 must also be an even function, it follows that $\psi_1'(-\pi) = -\psi_1'(\pi)$; but ψ_1 also satisfies the periodic boundary condition: $\psi_1'(-\pi) = \psi_1'(\pi)$. Hence $\psi_1'(-\pi) = \psi_1'(\pi) = 0$. All of these facts taken together imply that ψ_1 is monotonically increasing from $-\pi$ to 0 and monotonically decreasing from 0 to π, and is a positive, even function. Hence, the function $|\psi_1|^2$ has the same properties. Therefore, any value assumed by $|\psi_1|^2$ between $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$ is greater than any value assumed by $|\psi_1|^2$ on the other two regions. By the nature of the cosine function, (10.6) therefore holds. Thus there exists an f and a $\beta > 0$ such that $g(x) = -\frac{1}{3}\cos x$ satisfies (10.1), and we have the desired example.

Now consider a function f and a number $\beta > 0$ such that (10.1) is satisfied by $g(x) = -\frac{1}{3}\cos x$. By our remarks in Sec. 8, we know that (10.1) is satisfied by every spatial translate of g. Let p_x be the state corresponding to the translation of g by x, and let p be the state defined by

$$p(A) = \frac{1}{2\pi} \int_{-\pi}^{\pi} p_x(A)\, dx$$

for all A in $\mathcal{A}$. By Proposition 8.9, $p \in \mathcal{E}_\beta$. Also, p is translation invariant by construction, so we certainly have an example of symmetry breaking in the sense that we have a translation-invariant equilibrium state expressed as a mixture of equilibrium states that are not translation invariant.

Although this decomposition of p may not be the central decomposition, we can still show that the central decomposition of p breaks symmetry. To this end, let μ be the central measure of p and let K be a Borel set of factor states in $\mathcal{E}_\beta$ such that $\mu(K) = 1$. Suppose μ does not break symmetry. Then $\mu(\{\sigma \in K | \sigma \text{ is translation invariant}\}) = 1$. But as we already know, there is only one factor state in $\mathcal{E}_\beta$ that is translation invariant, i.e., there is only one constant solution to (10.1). Hence μ is concentrated on a set consisting of one state so p is that state. But if p is a factor state, then p is an extreme point of $\mathcal{E}_\beta$, so we have the desired contradiction.

ACKNOWLEDGMENTS

The author would like to thank Michael Reed for many useful and stimulating discussions, which led the author to many of these results. His criticism of the manuscript was also very helpful. Thanks are due to Richard Herman for reading the manuscript and making valuable suggestions.

[1] G. G. Emch and H. J. F. Knops, J. Math. Phys. 11, 3008–18 (1970).
[2] M. Takesaki and M. Winnink, Commun. Math. Phys. 30, 129–52 (1973).
[3] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[4] R. Haag, R. V. Kadison, and D. Kastler, Commun. Math. Phys. 16, 81–104 (1970).
[5] D. Ruelle, *Cargese Lectures in Physics*, edited by D. Kastler (Gordon and Breach, New York, 1970), Vol. 4, pp. 169–94.
[6] S. Sakai, *C*-Algebras and W*-Algebras* (Springer-Verlag, Berlin, Hiedelberg, New York, 1971).
[7] N. Dunford and J. T. Schwartz, *Linear Operators*, Part I (Interscience, New York, 1958).
[8] H. Araki, Commun. Math. Phys. 14, 120–157 (1969).
[9] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, to be published), Vol. III.
[10] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I.

Inverse Gaussian transforms: General properties and application to Slater-type orbitals with noninteger and integer n in the coordinate and momentum representations

William J. Taylor

Department of Chemistry, Ohio State University, Columbus, Ohio 43210
(Received 3 January 1977)

The use of Gaussian-type orbitals (GTO) facilitates the evaluation of the multicenter integrals encountered in quantum chemistry by reducing all integrals of more than two centers to two-center integrals. On the other hand Slater-type orbitals (STO), while leading to more time-consuming integral evaluations, provide a better approximation to variationally determined atomic orbitals. Thus, for a basis set of given size, STO's generally give better accuracy than GTO's. Kikuchi proposed the representation of STO's as integral Gaussian transforms, or in effect by *continuous* expansions in GTO's, and Shavitt, Karplus, and Kern have applied this technique to the evaluation of multicenter integrals over STO's. If these procedures are to be extended, it is desirable to develop a more systematic approach to the representation of a given basis function, $\psi(r)$, as a Gaussian transform, $\psi(r) = \mathcal{G}[f(t);r] = \int_0^\infty f(t)\exp(-r^2 t)dt$; what this reduces to is the problem of calculating the *inverse* Gaussian transform, $f(t) = \mathcal{G}^{-1}[\psi(r);t]$. In the present investigation it is pointed out that $f(t) = \mathcal{L}^{-1}[\psi(s^{1/2});t]$, where $\mathcal{L}^{-1}$ represents the inverse *Laplace* transformation. On this basis conditions on $\psi(r)$ necessary for the existence of a unique continuous Gaussian inverse, $f(t)$, are formulated, and general rules for the manipulation of inverse Gaussian transforms are developed. Finally, the formulas for the inverse Gaussian transforms of STO's obtained previously by Kikuchi and Wright are generalized to noninteger principal quantum number, and angle-dependent STO's, in both the coordinate and momentum representations.

INTRODUCTION

The use of Gaussian-type orbitals (GTO) greatly simplifies the evaluation of multicenter molecular quantum integrals, as a result of the theorem that a product of any number of Gaussian functions of the same argument, but having arbitrary centers, is reducible to a single Gaussian function referred to an intermediate center. Thus, all three- and four-center two-electron repulsion integrals, and three-center nuclear-electron attraction integrals, reduce to two-center integrals; the latter are further reducible to one-dimensional integrals which are evaluable in terms of the incomplete gamma function.[1,2] On the other hand, a *single* GTO provides a poor representation of an atomic orbital, because of the incorrect behavior of the GTO at both small and large distances from its center. This is, of course, a reflection of the proportionality of the exponent in a GTO to the *square*, rather than the first power, of the distance from its center. One solution to the latter difficulty is to approximate a single atomic orbital by a linear combination of a relatively small number of Gaussian orbitals having different exponents, and in some cases different centers (as in the Gaussian-lobe method). The coefficients and exponents in such *finite* Gaussian expansions can be determined by least-squares[3] or variational[4] methods. An alternative approach is the representation of each atomic orbital as an integral *Gaussian transform*, or effectively by a *continuous* expansion in Gaussian functions. Such representations are exact, when they exist, and were first suggested by Kikuchi,[5] and extensively applied to the evaluation of many-center integrals over Slater-type orbitals (STO) by Shavitt, Karplus, and Kern.[6,7,2] In the present article we develop a systematic approach to the derivation and properties of such transforms, based on the theory of Laplace transforms. The method is then applied to STO's in both the coordinate and momentum representations, and for general and integer values of n, leading to some results which appear to be new.

1. LAPLACE AND GAUSSIAN TRANSFORMS

We restrict the present discussion to spherically symmetric functions in the coordinate or momentum space, e.g., s-type atomic orbitals, the *radial* parts of angle-dependent atomic orbitals, or spherically symmetric potential functions. Thus, we write for the functions of interest $\psi(r)$ in coordinate space, and $\varphi(p)$ in momentum space, where $r = |\mathbf{r}|$ and $p = |\mathbf{p}|$. To avoid repetition the *general* part of the discussion will be directed toward functions $\psi(r)$ in coordinate space; to obtain the corresponding momentum space results simply replace r by p, and $\psi(r)$ by $\varphi(p)$.

Since a Gaussian transform is a modified Laplace transform, and a rigorous theory exists for the latter,[8-10] we first review briefly the existence and uniqueness of Laplace transforms in order to establish the same properties for Gaussian transforms. Let $f(t)$ be a real-valued function of a real variable, t, such that: (1) $f(t)$ is integrable over every finite interval $[a,b]$, $0 < a < b$; (2) the limit of one such integral as $a \to 0+$ exists; and (3) $f(t)$ is of *exponential order*, or $O(e^{\alpha t})$ (i.e., real α, M, and t_0 exist such that $|f(t)| < Me^{\alpha t}$ for $t > t_0$). Then

$$g(s) = \mathcal{L}[f(t);s] = \int_0^\infty e^{-st} f(t)dt \tag{1}$$

exists, and defines a unique analytic function, $g(s)$, of the complex variable s in the half-plane $\mathrm{Re}(s) > \alpha_0$ where α_0 is the greatest lower bound of the set of numbers $\{\alpha\}$ (Ref. 9, Sec. 59; Ref. 10, Sec. 2.1). The function $g(s)$ can usually be defined over the entire complex s plane by analytic continuation, except possibly at isolated singularities (poles or branch points). The necessity for

 0022-2488/78/1901-0052$1.00

permitting s and $g(s)$ to be complex-valued will become clear when the *inverse* of the preceding transformation is discussed in the next section. Equation (1) may be regarded as a linear integral Laplace transformation of the *object* function, $f(t)$, to obtain the *image* function, or *direct* transform, $g(s)$, as indicated by the notation $\mathcal{L}$ for the Laplace operator in (1).

We now define the Gaussian transform of an object function $f(t)$ (which satisfies the previously stated conditions) as

$$\psi(r) = \mathcal{G}[f(t); r] = \int_0^\infty f(t)\exp(-r^2 t)dt. \tag{2}$$

Comparison of (1) and (2) shows that

$$\psi(r) = \mathcal{G}[f(t); r] = \mathcal{L}[f(t); r^2] = g(r^2). \tag{3}$$

It follows that $\psi(r)$ exists and is a unique analytic function of r for $\mathrm{Re}(r^2) > \alpha_0$ (i.e., for a *hyperbolic* region in the complex r plane). It will be convenient to refer to $f(t)$ in the context of the Gaussian transform, (2), as the *shape* function.[11] For given $f(t)$, one can obtain $\psi(r)$ by evaluation of the integral in (2), or use of (3) and a table of direct Laplace transforms (Ref. 2, p. 15). However, in the context of the present article it is usually the atomic orbital or other function, $\psi(r)$, which is given, and the shape function, $f(t)$, which is sought. There are then disadvantages associated with the use of (3), as it stands. Since (3) is equivalent to (1), the *existence* of such a transform pair is made to hinge on the properties of the *unknown* object function. Further, the search for a transform pair satisfying (3) in a table of *direct* transforms is a hit or miss proposition, because it is the unknown object function, $f(t)$, which is ordered systematically in such a table, while the order of the image functions is largely unpredictable; this aspect should not be minimized since the more extensive tables[12,13] contain $\sim 10^3$ transform pairs.

2. INVERSE LAPLACE AND GAUSSIAN TRANSFORMS

The difficulties discussed in the preceding paragraph are obviated if one introduces the *inverses* of the Laplace and Gaussian tranformations, (1) and (2), denoted formally by

$$f(t) = \mathcal{L}^{-1}[g(s); t] \tag{4}$$

and

$$f(t) = \mathcal{G}^{-1}[\psi(r); t], \tag{5}$$

respectively. Using (3), (4) can be rewritten as

$$f(t) = \mathcal{L}^{-1}[\psi(s^{1/2}); t], \tag{6}$$

and from (5) and (6),

$$\mathcal{G}^{-1}[\psi(r); t] = \mathcal{L}^{-1}[\psi(s^{1/2}); t]. \tag{7}$$

Thus, the calculation of inverse Gaussian transforms is reduced to the evaluation of inverse Laplace transforms. Bishop and Somorjai[11] also suggested use of inverse Laplace transforms, but the emphasis in their work is on the choice of particular shape functions and the evaluation of the corresponding direct transforms as potentially useful basis functions, rather than development of a systematic procedure for obtaining the shape function corresponding to a *given* basis function as in the present study. With the latter objective in mind we next summarize the relevant portions of the theory of inverse Laplace transformations, again with emphasis on existence and uniqueness.

By definition, any function, $f(t)$, for which the (unique) Laplace transform $g(s)$ of (1) exists is an inverse Laplace transform of $g(s)$. However, not every function $g(s)$ possesses such an inverse transform, and conditions for the existence of an inverse are therefore of interest. The only generally useful representation of the inverse Laplace transform is the *complex inversion integral* (CII),

$$f(t) = \lim_{b\to\infty} (2\pi i)^{-1} \int_{c-ib}^{c+ib} e^{ts} g(s) ds, \tag{8}$$

where s is the complex variable of integration, and $b > 0$, c, and t, are real numbers. The CII is thus an integral along a line parallel to the imaginary axis and intersecting the real axis at $x = c$. The CII converges absolutely and also uniformly in t provided: (i) $g(s)$ is an analytic function of s of order s^{-k} ($k > 1$) in some half-plane, $\mathrm{Re}(s) > \gamma$, and is real on the part of the real axis lying in this half-plane, and (ii) $c > \gamma$ [which insures that the path of integration in (8) lies to the right of all singularities of $g(s)$]. When these conditions are met the indicated limit of the integral on the rhs of (8) defines a real-valued function of t, $f(t)$, which is *independent* of the value of c, and has the Laplace transform $g(s)$. Further, $f(t)$ is continuous for $-\infty < t < \infty$, vanishes for $t \le 0$, and is $O(e^{\gamma t})$ as $t \to \infty$ (Ref. 9, Sec. 63, Theorem 5).

The preceding conditions on $g(s)$ can be relaxed if $f(t)$ is required to be only *sectionally* continuous in the preceding interval. In this case $f(t)$ is not unique at its points of discontinuity, because its values at such isolated points make no contribution to the Laplace integral, (1) (see Ref. 9, Sec. 64, or Ref. 14, pp. 140, 141, including Lerch's theorem). The simplest example of such a case is $g(s) = s^{-n}$, $0 < n \le 1$ which is analytic in the half-plane $\mathrm{Re}(s) > 0$, but is not of order s^{-k} there, with $k > 1$. Nevertheless, the CII, (8), converges for this case (provided $c > 0$) to yield the inverse transform $f(t) = t^{n-1}/\Gamma(n)$, $t > 0$, and $f(t) = 0$, $t < 0$, which is discontinuous at $t = 0$ (in fact, singular there except for $n = 1$).

We consider next the inversion of the Gaussian transformation, (2). It is clear from (6) that if $\psi(s^{1/2})$ satisfies the conditions stated for $g(s)$ following (8), the substitution of $\psi(s^{1/2})$ for $g(s)$ in (8) yields a function, $f(t)$, having $\psi(r)$ as its Gaussian transform, as well as the other properties listed at the end of the next preceding paragraph [as discussed in the final paragraph of this section the fact that $\psi(s^{1/2})$ may be a multiple-valued function in the complex s plane does not present a problem]. We have thus obtained necessary and sufficient conditions for a given basis function, $\psi(r)$, to have a unique and continuous Gaussian inverse or shape function, $f(t)$, and have also found an integral formula for this shape function.

Realistic atomic orbitals in the coordinate representation (e.g., STO's, or those determined variationally), are of exponential order, $O(e^{-\zeta r})$, as $|r| \to \infty$, with $\zeta > 0$, so

that $\psi(s^{1/2})$ is of order $\exp(-\zeta s^{1/2})$. Setting $s = |s|e^{i\theta}$, $s^{1/2} = |s^{1/2}|(\cos\frac{1}{2}\theta + i\sin\frac{1}{2}\theta)$, one finds that $|\psi(s^{1/2})|$ is of order $\exp(-\zeta|s^{1/2}|\cos\frac{1}{2}\theta)$ as $|s| \to \infty$. The singularities of $\psi(s^{1/2})$ include at least a branch point at $s = 0$, so $\gamma \geq 0$, and $\cos\frac{1}{2}\theta > 2^{-1/2}$ in the half-plane $\mathrm{Re}(s) > \gamma$. Thus, $\psi(s^{1/2})$ is of order $\exp[-\zeta|(s/2)^{1/2}|]$ as $|s| \to \infty$ in the half-plane, which is more than sufficient to fulfill the order condition on $g(s)$ stated following (8). Therefore, such basis functions can be represented as Gaussian transforms of *continuous* shape functions. This is not true of the Gaussian function itself, $\psi(r) = \exp(-\beta r^2)$, for which $\psi(s^{1/2}) = e^{-\beta s}$. Proceeding in the same manner one finds that $|\psi(s^{1/2})|$ is now of order $\exp(-\beta|s|\cos\theta) = \exp[-\beta\mathrm{Re}(s)]$ which is not sufficient. This is not surprising since the inverse Laplace transform of $e^{-\beta s}$ is the Dirac delta function, $\delta(t-\beta)$, which is a *generalized function* or *distribution*. It *is* perhaps unexpected that a similar conclusion holds for *any* function which is of *Gaussian order*, or $O[\exp(-\beta r^2)]$, as $|r| \to \infty$.

It will seldom be necessary to actually calculate the CII on the rhs of (8), because of the existence of extensive tables of inverse Laplace transforms;[12,13,15] we therefore discuss only briefly the evaluation of the CII. Although the CII can be converted to a real integral (Ref. 9, Sec. 62), it is most readily evaluated by the methods of contour integration in the complex s plane. Since e^{ts} is an *entire* function of s, the singular points of the integrand of the CII occur at the positions of the singularities (if any) of $g(s)$, all of which lie in the *left* half-plane, $\mathrm{Re}(s) \leq \gamma$, by hypothesis. If the only singularities of $g(s)$ are a finite number of *poles* (Case I), the contour of integration in (8) may be completed with a circular arc, $\Gamma(R)$, of radius $R = (b^2 + c^2)^{1/2}$, centered at the origin and lying in the left half-plane. The integral along $\Gamma(R)$ will approach zero as b (and R) $\to \infty$, provided $|g(s)|$ is of order $|s|^{-k}$ as $|s| \to \infty$ in the left half-plane, for some $k > 0$. In this case, from Cauchy's residue theorem, the limit of the CII in (8), and therefore the inverse transform, $f(t)$, equals the sum of the residues of $e^{ts}g(s)$ at its poles. This result can be extended to the case in which the number of poles is infinite, under appropriate conditions (see Ref. 9, Secs. 66 and 67, especially Theorem 10, p. 193).

The procedure must be modified if $g(s)$ has one or more *branch points* (Case II). The contour must then be deformed to exclude the branch points, in order that $g(s)$ be single-valued on and within the contour, and analytic there except at its poles (if these also exist). The integrals along the deformed segments of the contour do not vanish, in general, and must be subtracted from the preceding sum of residues (if any) to obtain $f(t)$ (Ref. 9, Sec. 68; Ref. 10, Sec. 3.11; Ref 15a, pp. 599–603). It will be seen in Sec. 4 that it is Case II which is applicable to the calculation of the shape functions for STO's in both the coordinate and momentum representations.

3. RULES FOR MANIPULATION OF INVERSE GAUSSIAN TRANSFORMS

In working with any given type of integral transform it is useful to have available rules giving the effect on the image function of certain general operations on the object function, or the converse. A few such rules for Gaussian transforms are formulated in the present section. We limit the discussion to the effect on the inverse Gaussian transform (or shape function), $f(t)$, of operations on the image (or basis) function, $\psi(r)$, which are likely to be of interest in the context set forth in the Introduction. In view of (7), we expect these rules to be related to the corresponding rules for inverse Laplace transforms (Ref. 9, Appendix 2; Ref. 16, Secs. 4–9).

It follows from (7) and the linearity of the inverse Laplace transformation, $\mathcal{L}^{-1}$, that the inverse Gaussian transformation, $\mathcal{G}^{-1}$, is also *linear*, i.e., if a_1 and a_2 are arbitrary constants,

$$\mathcal{G}^{-1}[a_1\psi_1(r) + a_2\psi_2(r); t] = a_1\mathcal{G}^{-1}[\psi_1(r); t] + a_2\mathcal{G}^{-1}[\psi_2(r); t], \tag{9}$$

provided the two inverses on the rhs exist (throughout the present section $\mathcal{G}^{-1}[\psi(r); t]$ may also be read as "the shape function for $\psi(r)$, with argument t"). Clementi, Roothaan, and Yoshimine[17] were able to approximate numerical Hartree–Fock orbitals for atoms sufficiently closely to give seven-figure accuracy in the total energy using linear combinations of ~10 STO's. In view of (9), the same linear combinations of the shape functions for STO's given in the next section will yield equivalent approximations to the shape functions for the Hartree–Fock orbitals, and thus render the Gaussian transform method[6,7,12] applicable to molecular quantum integrals containing Hartree–Fock atomic orbitals.

For $a_2 = 0$, (9) reduces to the rule for *multiplication by a constant*. In the next section we obtain the shape functions for *unnormalized* STO's in the coordinate and momentum representations; multiplication of these shape functions by the appropriate normalization constants yields the shape functions for normalized STO's.

A *change of scale* of r (or p) may occur as a result of a change of the *units* in which these quantities are expressed or, in the quantum-chemistry context, from the *scaling procedure* needed to bring the ratio of kinetic and potential energies calculated from an approximate molecular wavefunction into agreement with the *virial theorem*.[18] The appropriate rule is

$$\mathcal{G}^{-1}[\psi(ar); t] = (1/a^2)\mathcal{G}^{-1}[\psi(r); (t/a^2)], \tag{10}$$

where a is any constant. The analogs for Gaussian transforms of the two translation theorems for Laplace transforms (Ref. 16, pp. 23, 24) do not appear to be of interest in the present context, and we omit them.

The inverse Gaussian transform of a *product* of two functions, $\psi_1(r)$ and $\psi_2(r)$, is evaluable as a *finite convolution* of the inverse transforms of $\psi_1(r)$ and $\psi_2(r)$, i.e., (Ref. 16, p. 28),

$$\mathcal{G}^{-1}[\psi_1(r)\psi_2(r); t] = \int_0^t \mathcal{G}^{-1}[\psi_1(r); t']\mathcal{G}^{-1}[\psi_2(r); t - t']dt'. \tag{11}$$

The convolution operation is *commutative*, so that ψ_1 and ψ_2 may be interchanged on the rhs of (11); it is also *associative*, so that if (11) is extended to a product of n functions the $n-1$ convolutions needed may be performed

in any order. The *complex convolution* of two image (basis) functions which corresponds to the product of two object (shape) functions (Ref. 16, p. 29) seens to be of little interest in the present context and is omitted.

The general formula for the inverse Laplace transform of a *composite* function is of basic importance, as it is applicable to numerous special cases, including one of particular interest in the present discussion. It is rather surprising that this formula is omitted from most lists (an exception is Ref. 13, p. 174, No. 56), and the writer is not aware of any published derivation of the formula. Let $g(s)$ and $h(s)$ be two functions such that inverse Laplace transforms exist for $g(s)$ as well as the composite functions $g[h(s)]$ and $\exp[-uh(s)]$, for all real positive u. Substitution of the identity,

$$g[h(s)] = \int_0^\infty \mathcal{L}^{-1}[g(s); u]e^{-uh(s)}du, \tag{12}$$

for $g(s)$ in (8), and inversion of the order of integrations (justified by uniform convergence of the CII), then yields

$$\mathcal{L}^{-1}\{g[h(s)]; t\} = \int_0^\infty \mathcal{L}^{-1}[g(s); u]\mathcal{L}^{-1}[e^{-uh(s)}; t]du. \tag{13}$$

In view of (7) the corresponding formulas for the inverse Gaussian transform of a composite function are

$$\mathcal{G}^{-1}\{\psi[\omega(r)]; t\} = \int_0^\infty \mathcal{L}^{-1}[\psi(s); u]\mathcal{L}^{-1}[e^{-u\omega(s^{1/2})}; t]du \tag{14a}$$

$$= \int_0^\infty \mathcal{G}^{-1}[\psi(r^2); u]\mathcal{G}^{-1}[e^{-u\omega(r)}; t]du. \tag{14b}$$

On setting $\omega(r) = r$ in (14a) one also obtains

$$\mathcal{G}^{-1}[\psi(r); t] = \int_0^\infty \mathcal{L}^{-1}[\psi(s); u]\mathcal{L}^{-1}[e^{-us^{1/2}}; t]du$$

$$= (4\pi t^3)^{-1/2}\int_0^\infty u\exp(-u^2/4t)\mathcal{L}^{-1}[\psi(s); u]du, \tag{15}$$

which shows that the inverse Gaussian transform of $\psi(r)$ can be expressed in terms of $\mathcal{L}^{-1}[\psi(s); t]$, as well as in terms of $\mathcal{L}^{-1}[\psi(s^{1/2}); t]$ as in (7). Although it will generally be more convenient to evaluate $\mathcal{G}^{-1}[\psi(r); t]$ from (7) than from (15), the latter formula will be useful for the derivation of general rules covering the differentiation and integration of Gaussian transforms, as now discussed.

The object function corresponding to the nth derivative of a Laplace transform is (Ref. 9, p. 47, Theorem 6)

$$\mathcal{L}^{-1}[\psi^{(n)}(s); t] = (-t)^n\mathcal{L}^{-1}[\psi(s); t]; \tag{16}$$

similarly for the integral indicated (Ref. 9, p. 53, Theorem 7),

$$\mathcal{L}^{-1}[\int_s^\infty \psi(s')ds'; t] = t^{-1}\mathcal{L}^{-1}[\psi(s); t]. \tag{17}$$

The simplicity of these results is a reflection of the properties of the *kernel*, e^{-st}, of the Laplace transform (1), and is lost when the latter is replaced by the kernel, $\exp(-r^2t)$, of the Gaussian transform, (2). However, by using (16) and (17) in combination with (15) and (7) the following formulas are obtained for Gaussian transforms:

$$\mathcal{G}^{-1}[\psi^{(n)}(r); t]$$

$$= (-1)^n(4\pi t^3)^{-1/2}\int_0^\infty u^{n+1}\exp(-u^2/4t)\mathcal{G}^{-1}[\psi(r^2); u]du, \tag{18}$$

$$\mathcal{G}^{-1}[\int_r^\infty \psi(r')dr'; t]$$

$$= (4\pi t^3)^{-1/2}\int_0^\infty \exp(-u^2/4t)\mathcal{G}^{-1}[\psi(r^2); u]du. \tag{19}$$

In *specific* cases it will usually be simpler to evaluate the inverse Gaussian transforms on the lhs of (18) and (19) by substitution of $\psi^{(n)}(r)$ and $\int_r^\infty \psi(r')dr'$, respectively, for $\psi(r)$ in (7). However, this procedure does not yield general formulas involving only inverse Gaussian transforms, such as are of interest in this section.

Finally, we take up the case in which the image function, $\psi(r, \zeta)$, depends upon a *parameter* ζ (as is the case, for example, with the STO, $r^{n-1}e^{-\zeta r}$), and consider differentiation and integration with respect to this parameter. The following formulas are usually valid (although the validity of interchanging the order of operations must be established for each case):

$$\mathcal{G}^{-1}\left[\frac{\partial^n\psi(r, \zeta)}{\partial\zeta^n}; t\right] = \left(\frac{\partial^n}{\partial\zeta^n}\right)\mathcal{G}^{-1}[\psi(r, \zeta); t], \tag{20}$$

$$\mathcal{G}^{-1}[\int_a^b \psi(r, \zeta)d\zeta; t] = \int_a^b \mathcal{G}^{-1}[\psi(r, \zeta); t]d\zeta. \tag{21}$$

Equation (20) has an application to the generation of an ns STO, $n > 1$, by $n-1$ differentiations of the $1s$ STO, $e^{-\zeta r}$, while (21) is relevant to the Hulthén-type basis function,[19,11]

$$\int_a^b e^{-\zeta r}d\zeta = (e^{-ar} - e^{-br})/r, \tag{22}$$

which may be regarded as a linear combination of $1s$ STO's for which ζ varies continuously over the interval $[a, b]$.

4. SHAPE FUNCTIONS FOR SLATER-TYPE ORBITALS

In this section we obtain explicit shape functions for the general ns STO in the coordinate and momentum representations. It will be convenient to write $\psi_n(r, \zeta)$ and $\varphi_n(p, \zeta)$, respectively, for the STO's in these two cases, to indicate their dependence on the parameter ζ (i.e., the *orbital exponent* in the coordinate representation). For each representation we consider first the case of noninteger n, and then the simplifications which occur for n a nonnegative integer. Following common practice we use *unnormalized* STO's [in this connection see the paragraph following (9)]. Finally, we discuss the extension of the present method to angle-dependent STO's.

A. *ns* STO's in the coordinate representation

The usual form of the ns STO is

$$\psi_n(r, \zeta) = r^{n-1}e^{-\zeta r}, \tag{23}$$

where n is restricted to positive integer values. Parr and Joy[20] suggested a generalization of STO's to positive fractional values of n, and integrals involving such orbitals have been studied by Geller,[21] Silverstone,[22] and Allouche.[22a] From (6), the shape function corresponding to (23) for general n is

$$f_n(t,\zeta) = \mathcal{L}^{-1}[s^{(n-1)/2}\exp(-\zeta s^{1/2});\, t]$$
$$= (2^n\pi)^{-1/2} t^{-(n+1)/2}\exp(-\zeta^2/8t) D_n(\zeta/2^{1/2}t^{1/2}),$$
$$[\mathrm{Re}(\zeta^2)>0,\ \mathrm{Re}(s)>0] \tag{24}$$

Ref. 12, p. 246, No. 9; Ref. 13, p. 247, No. 24). Here $D_n(z)$ is the *parabolic cylinder* (or *Weber–Hermite*) *function*, which is expressible as a confluent hypergeometric function, and is an entire function of z.[23]

Equation (24), which has apparently not been given previously, is valid for all real and complex values of n. However, for n a nonnegative integer,

$$D_n(z) = 2^{-n/2}\exp(-z^2/4) H_n(z/2^{1/2}), \tag{25}$$

where $H_n(z)$ is the nth Hermite polynomial (not to be confused with the alternative form He_n preferred in Refs. 12 and 13; see Ref. 12, p. 369). Thus, in this case (24) reduces to

$$f_n(t,\zeta) = (2^n\pi^{1/2})^{-1} t^{-(n+1)/2}\exp(-\zeta^2/4t) H_n(\zeta/2t^{1/2})$$
$$[n=0,1,2,\ldots,\ \mathrm{Re}(\zeta^2)>0,\ \mathrm{Re}(s)>0]. \tag{26}$$

Substitution of (26) for $f(t)$ in (2) reproduces the Gaussian transform of the ns STO, for n a nonnegative integer, as first given by Wright [Ref. 24, Eq. (20); there is a misprint in this transform in Eq. (66) of the second paper of Ref. 6]. The Gaussian transform of the $1s$ STO was given still earlier by Kikuchi.[5] For $n=0$, (26) yields the shape function for the $0s$ STO, $e^{-\zeta r}/r$, introduced for convenience in connection with certain recursion formulas for matrix elements relative to STO's by Roothaan.[25] Equation (26) is not valid for $\zeta=0$, except for the particular case $n=0$ for which it reduces to the shape function, $(\pi t)^{-1/2}$, for the Coulomb potential, $1/r$.

Finally, we remark that the explicit derivation of (24), through evaluation of the complex inversion integral (8), provides an example of the previously defined Case II, since $\psi_n(s^{1/2},\zeta)$ has a *branch point* at $s=0$ (and no other singularities). The evaluation of the CII for $n=1$, to obtain the shape function, $f_1(t,\zeta)$, for a $1s$ STO is discussed in Ref. 14, pp. 178–80. As shown there, integration of $f_1(t,\zeta)$ with respect to ζ, and use of (21), yields $f_0(t,\zeta)$. Similarly, differentiation of $f_1(t,\zeta)$ n times with respect to ζ, and use of (20), yields $(-1)^n f_{n+1}(t,\zeta)$, $n=1,2,\ldots$.

B. *ns* STO's in the momentum representation

The three-dimensional complex exponential Fourier transform of the general (unnormalized) STO,

$$r^{n-1}e^{-\zeta r}P_l^{|m|}(\cos\theta)e^{im\Phi}, \tag{27}$$

where $P_l^{|m|}(x)$ is an associated Legendre function, and n may be a noninteger, has been calculated by Silverstone [see Ref. 22, Eqs. (12) and (13)]. In order to convert Silverstone's results to the accepted form of the momentum amplitude function,[26] one must replace his wavenumber, k, by $p/\hbar$, and multiply the complex conjugate of the resulting function by $(2\pi\hbar)^{-3/2}$.[27] However, we assume atomic units, so that $\hbar=1$, $k=p$, and the preceding factor becomes $(2\pi)^{-3/2}$. On also setting $l=m=0$, we obtain for the momentum amplitude function corresponding to the ns STO, (23),

$$\varphi_n(p,\zeta) = (2\pi)^{-1/2}\Gamma(n+1)(i/p)[(\zeta+ip)^{-n-1} - (\zeta-ip)^{-n-1}]. \tag{28}$$

From (2), (5), (9), and (7), $\varphi_n(p,\zeta)$ may be represented as a Gaussian transform,

$$\varphi_n(p,\zeta) = \int_0^\infty f_n(t,\zeta)\exp(-p^2t)dt, \tag{29}$$

with shape function [this is another example of Case II, since $\varphi_n(s^{1/2},\zeta)$ has a branch point at $s=0$, at least]

$$f_n(t,\zeta) = (2\pi)^{-1/2}\Gamma(n+1)\{i^{-n}\mathcal{L}^{-1}[(s^{1/2}-i\zeta)^{-n-1}/s^{1/2};\, t]$$
$$+ i^n\mathcal{L}^{-1}[(s^{1/2}+i\zeta)^{-n-1}/s^{1/2};\, t]\}. \tag{30}$$

Equation (30) may be reduced as follows (see Ref. 13, p. 212, No. 54, and Ref. 23, p. 117, Eq. 6):

$$f_n(t,\zeta) = \pi^{-1}\Gamma(n+1)(2t)^{n/2}\exp(-\zeta^2t/2)$$
$$\times[i^nD_{-n-1}(i2^{1/2}\zeta t^{1/2}) + i^{-n}D_{-n-1}(-i2^{1/2}\zeta t^{1/2})]$$
$$= (2/\pi)^{1/2}(2t)^{n/2}\exp(-\zeta^2t/2)D_n(2^{1/2}\zeta t^{1/2}),$$
$$[\mathrm{Re}(n)>-2,\ \mathrm{Re}(s)>0]. \tag{31}$$

Further, if n is a nonnegative integer, substitution of (25) yields

$$f_n(t,\zeta) = (2/\pi)^{1/2}t^{n/2}\exp(-\zeta^2t)H_n(\zeta t^{1/2}),$$
$$(n=0,1,2,\ldots). \tag{32}$$

C. Angle-dependent Slater-type orbitals

As our final topic we discuss the extension of the preceding treatment to the *radial* part of the general angle-dependent STO in the momentum representation. We again assume this orbital has the form (27) in the coordinate representation; in this case the radial part, $r^{n-1}e^{-\zeta r}$, is independent of the quantum numbers l and m, so that the corresponding shape function remains of the form (24), or for integer n, (26). On the other hand, the three-dimensional Fourier transform of (27), or general STO in the momentum representation, has a "radial" part,

$$\varphi_{nl}(p,\zeta) = (2\pi)^{-1/2}\Gamma(n-l+1)(ip)^l\left(p^{-1}\frac{\partial}{\partial p}\right)^l(i/p)$$
$$[(\zeta+ip)^{l-n-1} - (\zeta-ip)^{l-n-1}], \tag{33}$$

which is dependent on l (but not on m); the complete angle-dependent orbital in momentum space has the form

$$\varphi_{nlm}(p,\Theta,\Phi;\zeta) = \varphi_{nl}(p,\zeta)P_l^{|m|}(\cos\Theta)e^{im\Phi}, \tag{34}$$

where Θ and Φ are spherical polar angles in momentum space [Ref. 22, Eqs. (12) and (13)].

Generalizing (30), the shape function corresponding to (33) is

$$f_{nl}(l,\zeta)$$
$$= (2\pi)^{-1/2}(2i)^l\Gamma(n-l+1)\mathcal{L}^{-1}\left[s^{l/2}\left(\frac{\partial}{\partial s}\right)^l s^{-1/2}\right.$$
$$\left.\times\{i^{l-n}(s^{1/2}-i\zeta)^{l-n-1} + i^{n-l}(s^{1/2}+i\zeta)^{l-n-1}\};\, l\right]$$
$$= 2^{(n+l+1)/2}i^l\pi^{-1/2}\mathcal{L}^{-1}\left\{s^{l/2}\left(\frac{\partial}{\partial s}\right)^l\mathcal{L}[u^{(n-l)/2}\exp(-\zeta^2u/2)\right.$$
$$\left.\times D_{n-l}(2^{1/2}\zeta u^{1/2});\, s];\, t\right\}$$
$$= 2^{(n+l+1)/2}i^{-l}\pi^{-1/2}\mathcal{L}^{-1}\{s^{l/2}\mathcal{L}[u^{(n+l)/2}\exp(-\zeta^2u/2)$$
$$\times D_{n-l}(2^{1/2}\zeta u^{1/2});\, s];\, t\}, \tag{35}$$

where we have used Ref. 13, p. 212, No. 54; Ref. 23, p. 117, Eq. (6); and (16), in that order.

For $l=0$ the expression on the final line of (35) reduces to $f_n(t,\zeta)$ of (31), or to (32) for n a nonnegative integer. The further reduction of (35) when $l>0$ proceeds somewhat differently for the cases of *even* and *odd* integer l. For l even, $s^{l/2}$ is a positive integer power of s, and by Theorem 8, p. 184 of Ref. 9, (35) reduces to the form

$$f_{nl}(t,\zeta)=(-1)^{l/2}2^{(n+l+1)/2}\pi^{-1/2}\left(\frac{\partial}{\partial t}\right)^{l/2}\times[t^{(n+l)/2}\exp(-\zeta^2 t/2)D_{n-l}(2^{1/2}\zeta t^{1/2})],\quad (l=2,4,6,\ldots) \tag{36a}$$

$$=(-1)^{l/2}(2/\pi)^{1/2}2^{l}\left(\frac{\partial}{\partial t}\right)^{l/2}\times[t^{(n+l)/2}\exp(-\zeta^2 t)H_{n-l}(\zeta t^{1/2})]$$

$$(n=0,1,2,\ldots,\quad l=2,4,6,\ldots). \tag{36b}$$

On the other hand, for l odd, $s^{l/2}$ is a positive half-integer power of s which we rewrite as $s^{-1/2}s^{(l+1)/2}$ [the alternative decomposition, $s^{1/2}s^{(l-1)/2}$, is not useful because $\mathcal{L}^{-1}(s^{1/2})$ does not exist]. Then from Theorem 3, p. 37 of Ref. 9, the final line of (35) (exclusive of constant factors) reduces to the Laplace convolution of $\mathcal{L}^{-1}(s^{-1/2})=(\pi t)^{-1/2}$ and $\mathcal{L}^{-1}\{s^{(l+1)/2}\mathcal{L}[F(u);s];t\}$, where $F(u)$ is the function within the square brackets in the final line of (35). But by the first theorem referred to in this paragraph the latter inverse transform equals the derivative $F^{(l/2+1/2)}(t)$. Hence

$$f_{nl}(t,\zeta)=i(-1)^{(l+1)/2}2^{(n+l+1)/2}\pi^{-1}\times\int_0^t[F^{(l/2+1/2)}(u)/(t-u)^{1/2}]du,\quad (l=1,3,5,\ldots), \tag{37}$$

where

$$F^{(l/2+1/2)}(t)=\left(\frac{\partial}{\partial t}\right)^{(l+1)/2}[t^{(n+l)/2}\exp(-\zeta^2 t/2)D_{n-l}(2^{1/2}\zeta t^{1/2})], \tag{38a}$$

or for n a nonnegative integer,

$$F^{(l/2+1/2)}(t)=2^{(l-n)/2}\left(\frac{\partial}{\partial t}\right)^{(l+1)/2}[t^{(n+l)/2}\exp(-\zeta^2 t)H_{n-l}(\zeta t^{1/2})],\quad (n=0,1,2,\ldots). \tag{38b}$$

5. CONCLUDING REMARKS

Among the principal objectives of the present article have been the determination of existence and uniqueness conditions for inverse Gaussian transforms, and development of a systematic procedure for their calculation when these conditions are fulfilled (Sec. 2). Basic rules for the manipulation of inverse Gaussian transforms have also been presented (Sec. 3). Finally, we have applied these methods to obtain the inverse Gaussian transforms (or shape functions) for Slater-type orbitals, for noninteger and integer n, in both the coordinate and momentum representations (Sec. 4). The last results are in large part new, and are certainly of some intrinsic interest in view of the important role played by STO's in quantum chemistry. It remains to be seen whether they will be useful in extending previous Gaussian transform methods[6,7] for the evaluation of multicenter quantum integrals over STO's. However, the derivation of these generalized shape functions for STO's must obviously precede, and may stimulate, investigations of their usefulness.

ACKNOWLEDGMENT

The desirability of a more systematic approach to the calculation of inverse Gaussian transforms became apparent during discussions with Raymond Thompson, in the course of his preparation of a review of integral transform methods for the evaluation of multicenter quantum integrals.[29]

[1] S. F. Boys, Proc. Roy. Soc. **A200**, 542 (1950).
[2] I. Shavitt, in *Methods in Computational Physics*, edited by B. Alder *et al.* (Academic, New York, 1963), Vol. 2, pp. 1–46.
[3] S. F. Boys and I. Shavitt, Proc. Roy. Soc. **A254**, 487 (1960).
[4] R. McWeeny, Acta Cryst. **6**, 631 (1953).
[5] R. Kikuchi, J. Chem. Phys. **22**, 148 (1954).
[6] I. Shavitt and M. Karplus, J. Chem. Phys. **36**, 550 (1962); **43**, 398 (1965).
[7] C. W. Kern and M. Karplus, J. Chem. Phys. **43**, 415 (1965).
[8] G. Doetsch, *Theorie und Anwendung der Laplace-Transformation* (Springer-Verlag, Berlin, 1937).
[9] R. V. Churchill, *Operational Mathematics* (McGraw-Hill, New York, 1958), 2nd ed.
[10] M. G. Smith, *Laplace Transform Theory* (Van Nostrand, London, 1966).
[11] D. M. Bishop and R. L. Somorjai, J. Math. Phys. **11**, 1150 (1970).
[12] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. 1.
[13] G. E. Roberts and H. Kaufman, *Table of Laplace Transforms* (Saunders, Philadelphia, 1966).
[14] I. N. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972).
[15] Reference 12 lists about 650 inverse Laplace transforms, as well as about 800 direct transforms and (especially with the inclusion of volume 2) a number of other kinds of integral transforms; Ref. 13 is the most extensive table of Laplace transforms known to the writer, with over 1550 inverse transforms, and has complete statements of validity conditions on parameters and the variable s. Where appropriate we will give specific references to both of these tables, because Ref. 12 is more widely available than Ref. 13 (even though less complete).
[15a] J. Irving and N. Mullineux, *Mathematics in Physics and Engineering* (Academic, New York, 1959).
[16] G. Doetsch, *Guide to the Applications of the Laplace and Z-Transforms* (Van Nostrand Reinhold, London, 1971).
[17] E. Clementi, C. C. J. Roothaan, and M. Yoshimine, Phys. Rev. **127**, 1618 (1962).
[18] P.-O. Löwdin, in *Advances in Chemical Physics*, edited by I. Prigogine (Interscience, New York, 1959), Vol. 2, pp. 219–223.
[19] L. Hulthén, Ark. Fys. **28A**, 5 (1942).
[20] R. G. Parr and H. W. Joy, J. Chem. Phys. **26**, 424 (1957).
[21] M. Geller, J. Chem. Phys. **36**, 2424 (1962).
[22] H. J. Silverstone, J. Chem. Phys. **45**, 4337 (1966).
[22a] A. Allouche, Theoret. Chim. Acta **34**, 79 (1974); **42**, 325

(1976).
[23]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, Chap. 8.
[24]J. P. Wright, in Solid-State and Molecular Theory Group Quarterly Progress Report No. 50 (Massachusetts Institute of Technology, October 15, 1963), p. 35.
[25]C. C. J. Roothaan, J. Chem. Phys. **19**, 1445 (1951).
[26]W. Pauli, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1958), Vol. 5, Part 1, p. 13, Eq. (3.4).
[27]The factor $(2\pi\hbar)^{-3/2}$ is required if the amplitude functions in coordinate and momentum space are to satisfy Parseval's formula, and thus have the *same normalization constant*[28] (even though this constant may be omitted as is the case here). The necessity for using the complex conjugate function (in general) is connected with the sign of the exponent in Silverstone's definition of the Fourier transform; however, for the STO's, (27), under consideration this represents at most multiplication by a phase factor, as the momentum amplitude is real for even l and pure imaginary for odd l.
[28]W. J. Taylor, Int. J. Quantum Chem. **2**, 735 (1968).
[29]R. L. Thompson, M.S. thesis, Ohio State University (1977).

Nontranslationally covariant currents and associated symmetry generators

W. D. Garber and H. Reeh

Institut für Theoretische Physik, Universität Göttingen, Bunsenstr. 9, D-3400 Göttingen, Germany
(Received 21 March 1977)

Within the framework of axiomatic field theory, the general case of a translationally noncovariant conserved local current is investigated. It is shown that the associated symmetry does not change the particle number nor the mass or the momentum of one-particle states. There is an integer N such that the N-fold commutators of the generator with the momentum as well as with the mass operator vanish.

1. INTRODUCTION

There is a series of investigations in the literature concerning the interplay of continuous symmetry transformations in general with the symmetries of the Poincaré group. Important contributions are given, e.g., in Refs. 1–3. The first two have the character of no-go theorems and lead to the result that there are no nontrivial combinations of those symmetries. In particular it is shown that the symmetries do not change the mass of one-particle states or their momentum. In Ref. 3 these results are used (along with other results of Ref. 2) to investigate the situation if supersymmetries are included.

In Ref. 1, group-theoretic assumptions and methods are used to obtain the results whereas Ref. 2 heavily relies on assumptions concerning the scattering matrix. In the present work, we investigate the situation in a pure field-theoretic framework using the Wightman formulation and the assumption that the symmetries in question can be obtained from conserved local currents.

The case of currents transforming covariantly under translations is completely investigated if there is a mass gap and the theory is asymptotically complete, see, e.g., Refs. 4 and 5. (The mass gap assumption in this case excludes spontaneous symmetry breaking.) The associated symmetry generators are even known to be self-adjoint.[6] There are also results for the case of nontranslationally covariant currents of a specific structure as expected for the generators of the Poincaré and conformal group, see, e.g., Refs. 7–10.

If one is interested in the basic question whether there exist symmetry transformations changing mass or momentum, the class of currents just mentioned is obviously too restrictive. We therefore consider, in the following, the general case of translationally noncovariant conserved local currents. In addition to the usual assumptions of the Wightman framework we need: (i) existence of a mass gap (for the construction of the generator); (ii) isolated one-particle hyperboloids of finite multiplicity (for the construction of asymptotic states); (iii) invariance of the vacuum [in general not implied by (i), compare Sec. 3].

Furthermore, we assume, in Sec. 7, asymptotic completeness and in Sec. 8 we assume the existence of a local interpolating field for every particle. We then prove in Sec. 8 that also under these assumptions mass and momentum of one-particle states are not changed by the symmetry. In addition, higher multiple commutators of the symmetry generator with the momentum and mass operator have to vanish.

Sections 2–5 contain the discussion of the relation between the current and the generator. In Secs. 6 and 7 we give a natural definition of the generator on asymptotic states which is consistent with the usual definition in case the current is covariant. We recover the result (known in the translationally covariant case) that the matrix elements of the generator between different numbers of particles vanish. This then leads to the results in Sec. 8. The questions whether the generator is self-adjoint and commutes with the scattering matrix will be discussed in a subsequent paper.

2. ASSUMPTIONS ON THE CURRENTS

We consider a quantum field theory in the Wightman framework[11] given by a finite set of fields $\{\phi_i(x)\}$. By $\mathcal{P}_{SL}$ we denote the polynomial algebra spanned by the $\phi_i(f)$ with $f \in \mathcal{D}(\mathbb{R}^4)$ (test functions on $\mathbb{R}^4$ with compact support), by Ω the vacuum state, by D_0 the domain $\mathcal{P}_{SL}\Omega$, by D the common dense domain of the elements of $\mathcal{P}_{SL}$, and by $U(a)$, $a \in \mathbb{R}^4$, the unitary representation of the translations.

Besides the covariant fields above, we consider an operator-valued tempered distribution $k(x)$ commuting with itself and with the elements of $\mathcal{P}_{SL}$ for spacelike separation, fulfilling $k(f)D_0 \subset D$ and $(k(g))^* D_0 \subset D$, $f, g \in \mathcal{D}(\mathbb{R}^4)$, * denotes the adjoint operator. However, $k(x)$ need not be covariant under translations. The translates $k(x,a) := U(a)k(x)U^{-1}(a)$ have the following properties:

Lemma 2.1.: Let $g \in \mathcal{D}(\mathbb{R}^4)$, $\varphi \in D_0$. Then

(i) $k(g,a)\varphi$ is strongly continuous in a;

(ii) $\|k(g,a)\varphi\|$ is polynomially bounded as $|a| \to \infty$ ($|a|^2 := |a^0|^2 + |\mathbf{a}|^2$).

Proof of (ii): $\varphi \in D_0$ is a finite sum of vectors $\phi_{i_1}(f_{i_1}) \cdots \phi_{i_n}(f_{i_n})\Omega$. Hence

 0022-2488/78/1901-0059$1.00

$$\|k(g,a)\varphi\|^2 = (U^*(a)\varphi \,|\, k^*(g)k(g)U^{-1}(a)\varphi)$$
$$= \sum (U(-a)\phi_{l_1}(f_{l_1}) \cdots |$$
$$\times k^*(g)k(g)U(-a)\phi_{l_1'}(f_{l_1'}) \cdots \Omega)$$
$$\leq \text{const} \sum \|f_{l_1,a}\|_{r_1,s_1} \cdots \|g\|_{r,s}$$
$$\times \|g\|_{r,s} \|f_{l_1',a}\|_{r_1',s_1'} \cdots ,$$

where $\|\cdot\|_{rs}$ denotes the r,s norm on $\mathcal{S}^{11}$ which, for the translated function $f_{l,a}$ is bounded by a polynomial in the components of a.

Proof of (i):

$$(k(g,a) - k(g,a'))\varphi$$
$$= [U(a)k(g)U^{-1}(a) - U(a')k(g)U^{-1}(a')]\varphi$$
$$= [U(a) - U(a')]k(g)U^{-1}(a)\varphi$$
$$+ U(a')k(g)[U^{-1}(a) - U^{-1}(a')]\varphi .$$

Keep a fixed and let $a' \to a$. The first term vanishes strongly. The norm of the second can be estimated by

$$\|U^{-1}(a) - U^{-1}(a')\varphi\|^{1/2} \|k^*(g)k(g)\{U^{-1}(a) - U^{-1}(a')\}\varphi\|^{1/2} ,$$

where the second factor is bounded as in (ii).□

By the regular transformation of variables $(x,a) \to (x+a=y,x)$ we can rewrite

$$k(x,a) = j(x+a,x) ,$$

where j transforms covariantly under translations with respect to the first variable while the second remains unchanged.

We now take four such nontranslationally covariant fields $j^\mu(y,x)$, $\mu = 0,1,2,3$, as components of a current density for which we assume

(i) $j^\mu(y,x)$ is Hermitian;

(ii) $j^\mu(y,x)$ is local and relatively local in y for every x;

(iii) $\partial_\mu j^\mu(y,y) := \frac{\partial}{\partial y^\mu} j^\mu(y,y) = 0$.

By translation one gets from (iii) $\partial_\mu j^\mu(y+a,y) = 0$ for any a.

3. DEFINITION AND SOME PROPERTIES OF THE GENERATOR

A conserved current defines an infinitesimal symmetry transformation via a generator Q. Accordingly, we define a linear operator $Q(a)$ by

$$Q(a)\Omega := 0 , \tag{3.1}$$
$$Q(a)A\Omega := \lim_{r\to\infty} [Q_r(a), A]\Omega , \quad A \in \mathcal{P}_{SL} \tag{3.2}$$

with

$$Q_r(a) := \int j^0(y+a,y)\vartheta_r(\mathbf{y})\eta(y^0)d^4y ,$$
$$\vartheta_r(\mathbf{y}) = \vartheta\left(\left|\frac{\mathbf{y}}{r}\right|\right), \quad \vartheta \in \mathcal{D}(\mathbf{R}^1), \quad \vartheta(s) = \begin{cases} 1, & \text{for } 0 \leq s \leq 1, \\ 0, & \text{for } s \geq 2, \end{cases}$$
$$\eta(y^0) \in \mathcal{D}(\mathbf{R}^1), \quad \int \eta(y^0)dy^0 = 1 .$$

$Q(a)$ is well defined since Ω is separating for local operators. As in the translationally covariant case, $Q(a)$ does not depend on the unspecified details of ϑ and η.

We now explicitly assume invariance of the vacuum under the symmetry transformations [not to be confused with (3.1)]:

$$\lim_{r\to\infty} (\Omega \,|\, [Q_r(a), A]\Omega) = 0 \quad \text{for all} \quad A \in \mathcal{P}_{SL} . \tag{3.3}$$

[It would be sufficient to assume (3.3) for one vector a only.] In case of a nontranslationally covariant current, (3.3) is not implied by the mass gap assumption.[12]

(3.3) implies that $Q(a)$ is symmetric and hence closable (for $A, B \in \mathcal{P}_{SL}$,

$$(A\Omega \,|\, QB\Omega) - (QA\Omega \,|\, B\Omega) = \lim_{r\to\infty} (\Omega \,|\, [Q_r, A^*B]\Omega) = 0$$

see, e.g., Ref. 6).

Since the infinitesimal symmetry tranformation and Q are not affected by adding a c-number to j^μ, we may redefine j^μ by $j'^\mu(y,x) = j^\mu(y,x) - (\Omega \,|\, j^\mu(y,x)\Omega)$ so that $(\Omega \,|\, j'^\mu(y,x)\Omega) = 0$ and $\partial_\mu j'^\mu = 0$, just as in the translationally covariant case. Henceforth we always assume that this redefinition has been done.

Lemma 3.1.: Let $\overline{Q}(a)$ denote the closure of $Q(a)$ and let $\mathcal{D}_{\overline{Q}(a)}$ denote its domain. Then

$$\mathcal{D}_{\overline{Q}(a)} \supset \{A_1(f_1) \cdots A_n(f_n)\Omega\}_{f_i \in \mathcal{S}(\mathbf{R}^4),\ A_i \in \mathcal{P}_{SL},\ n \in \mathbf{N}}$$
$$[A(x) := U(x)AU^{-1}(x)] .$$

Proof: The norm of

$$QA_1(x_1) \cdots A_n(x_n)\Omega$$
$$= \lim_{r\to\infty} \sum_{i=1}^{n} A_1(x_1) \cdots [Q_r, A_i(x_i)] \cdots A_n(x_n)\Omega \tag{3.4}$$

is polynomially bounded in $x_1, \ldots, x_n$: Consider a typical term

$$\phi_1(g_{1,x_1}) \cdots [Q_r, \phi(g_{l,x_l})] \cdots \phi(g_{m,x_m})\Omega , \tag{3.5}$$

where g_{l,x_l} denotes the test function $g_l \in \mathcal{D}(\mathbf{R}^4)$ shifted by x_l. Because of locality, the limit in (3.5) is attained already for finite r, $r \geq r_0$ (suppg_l, suppη) + $|x_l|$. The norm of (3.5) is bounded by Schwartz norms

$$2 \cdot \text{const} \|g_{1,x_1}\|_{t_1,s_1} \cdots \|\vartheta_r \otimes \eta\|_{t,s} \cdots \|g_{m,x_m}\|_{t_m,s_m} .$$

The $\|g_{l,x_l}\|_{t_l,s_l}$ are bounded by polynomials in x_l and

$$\|\vartheta_r \otimes \eta\|_{t,s} \leq \text{const}(1 + r^{t-s}) .$$

Since we may choose $r = r_0 + |x_l|$, boundedness follows.

The scalar product of (3.4) with any vector in D_0 is continuous, weak continuity follows and we may integrate. For step functions $\chi_i(x_i)$ we get

$$\int \chi_1(x_1) \cdots \chi_n(x_n) QA_1(x_1) \cdots A_n(x_n)\Omega dx_1 \cdots dx_n \tag{3.6}$$
$$= QA_1(\chi_1) \cdots A_n(\chi_n)\Omega .$$

$f_i \in \mathcal{S}(\mathbf{R}^4)$ can be approximated in the supremum norm by sequences of step functions $\chi_{i,\nu}$, uniformly in x_i. The induced sequence of the left-hand side of (3.6) converges weakly. $A_1(\chi_{1,\nu_1}) \cdots A_n(\chi_{n,\nu_n})\Omega$ converges to

$A_1(f_1)\cdots A_n(f_n)\Omega$. Now (5.12), p. 165 of Ref. 13 implies the statement. □

In the following, we write Q instead of $\overline{Q}$.

We note that, as in the translationally covariant case, $[Q(a), \phi_i(x)]$ is again local with respect to x, however it is not translationally covariant in x.

4. MATRIX ELEMENTS OF Q

We next discuss the convergence of matrix elements of Q_r as $r \to \infty$ under the assumptions of the existence of a mass gap and the invariance of the vacuum (3.3). By $\mathcal{P}_1$ we denote the polynomial algebra generated by the $A(f)$ with $A \in \mathcal{P}_{SL}$, $f \in \mathcal{S}(\mathbb{R}^4)$.

Lemma 4.1: For all $B \in \mathcal{P}_1$, $\lim_{r\to\infty}(\Omega | Q_r(a)B\Omega) = 0$.

For the proof we need the following lemma:

Lemma 4.2: Let $L > 0$ and $0_L := \{x : |\mathbf{x}| + |x^0| < L\}$, and denote by $\mathcal{P}_{SL}(0_L)$ the subalgebra of $\mathcal{P}_{SL}$ spanned by the $\phi_i(f)$, $f \in \mathcal{D}(\mathbb{R}^4)$, $\mathrm{supp} f \subset 0_L$. Let B be quasilocal, i.e., to every $L > 0$ there is a decomposition

$$B = B_L + B - B_L$$

with $B_L \in \mathcal{P}_{SL}(0_L)$ and $L^N \|(B - B_L)\Omega\| \to 0$, $L^N \|(B^* - B_L^*)\Omega\| \to 0$ for $L \to \infty$ and any $N \in \mathbb{N}$. Then

$$\lim_{r\to\infty}(\Omega | [Q_r(a), B]\Omega) = 0.$$

Proof: $(\Omega | [Q_r(a), B]\Omega) = (\Omega | [Q_r(a), B_L] + (\Omega | [Q_r(a)(B - B_L)]\Omega)$. The first term is independent of r for $r \geq L + d(\mathrm{supp}\eta)$ where $d(\mathrm{supp}\eta)$ denotes the diameter of $\mathrm{supp}\eta$, and vanishes by (3.3) for $L \leq r - d$. The second is bounded by

$$\|Q_r(a)\Omega\|(\|(B - B_L)\Omega\| + \|(B^* - B_L^*)\Omega\|)$$

which tends to zero for the choice $L = r - d$ since $\|Q_r(a)\Omega\|$ is polynomially bounded in r. □

Proof of Lemma 4.1: It suffices to consider Hermitian B. Then by Lemma 4.2

$$0 = \lim_{r\to\infty}(\Omega | [Q_r(a), B]\Omega) = 2i \lim_{r\to\infty} \mathrm{Im}(\Omega | Q_r(a)B\Omega). \qquad (4.1)$$

Because of the gap there exists[14] a B', Hermitian and quasilocal, such that

$$B\Omega - (\Omega | B\Omega) = P^0 B'\Omega$$

with the infinitesimal generator of time translations P^0. Hence [remember $(\Omega | Q_r(a)\Omega) = 0$!]

$$(\Omega | Q_r(a)B\Omega) = (\Omega | Q_r(a)P^0 B'\Omega)$$
$$= -(\Omega | [P^0, Q_r(a)]B'\Omega) = -i(\Omega | Q_r^\cdot(a)B'\Omega),$$

where $Q_r^\cdot(a)$ belongs to the current $(\partial/\partial a^0)j^\mu(y + a, y)$. Therefore,

$$\mathrm{Im}(\Omega | Q(a)B\Omega) = -\mathrm{Re}(\Omega | Q_r^\cdot(a)B'\Omega)$$

and by (4.1)

$$\lim_{r\to\infty} \mathrm{Re}(\Omega | Q_r^\cdot(a)B'\Omega) = 0.$$

On the other hand, the symmetry induced by $Q_r^\cdot(a)$ also leaves the vacuum invariant [because this is the case for $Q_r(a)$ and all a]. Hence (4.1) applies to $Q_r^\cdot(a)$ and B' and yields

$$\lim_{r\to\infty} \mathrm{Im}(\Omega | Q_r^\cdot(a)B'\Omega) = 0.$$

Thus,

$$\lim_{r\to\infty}(\Omega | Q_r^\cdot(a)B'\Omega) = 0 = \lim_{r\to\infty}(\Omega | Q_r(a)B\Omega). \ \square$$

An important consequence is

Theorem 4.3: Let $B, B' \in \mathcal{P}_1$. Then

$$\lim_{r\to\infty}(B\Omega | Q_r(a)B'\Omega) = (B\Omega | Q(a)B'\Omega).$$

Proof: By Lemma 3.1 we have

$$(B\Omega | Q(a)B'\Omega) = \lim_{r\to\infty}(B\Omega | [Q_r(a), B']\Omega)$$
$$= \lim_{r\to\infty}((B\Omega | Q_r(a)B'\Omega) - (B\Omega | B'Q_r(a)\Omega)).$$

The last term vanishes by Lemma 4.2. □

5. RELATED CURRENT DENSITIES

For the next section it will be useful to consider the following objects: The Fourier transform

$$\tilde{j}^\mu(y, p) := \frac{1}{(2\pi)^2}\int e^{-ipx} j^\mu(y, x)d^4x$$

(which, of course, needs smearing in y and p) is a covariant field with parameter p,

$$U(a)\tilde{j}^\mu(y, p)U^{-1}(a) = \tilde{j}^\mu(y + a, p).$$

$\tilde{j}^\mu(y, p)$ is not conserved,

$$\frac{\partial}{\partial y^\mu}\tilde{j}^\mu(y, p) = -ip_\mu \tilde{j}^\mu(y, p).$$

However, the translationally noncovariant density

$$\tilde{J}^\mu(y, p) := e^{ipy}\tilde{j}^\mu(y, p)$$

is conserved, and $\int \tilde{J}^\mu(y, p)\tilde{f}(p)d^4p = \int j^\mu(y, y + x)f(x)d^4x$, $f \in \mathcal{S}(\mathbb{R}^4)$, generates the same symmetry transformation as

$$\int j^\mu(y - x, y)f(x)d^4x$$

with the generator

$$Q^f = \int Q(x)f(-x)d^4x.$$

In the following, it will sometimes be convenient to use Q^f instead of $Q(a)$. [We will also use $Q^f(y)$: $= U(y)Q^f U^{-1}(y)$.] Then we shall consider the following instead of Theorem 4.3.

Theorem 5.1: Let $B, B' \in \mathcal{P}_1$. Then

$$\lim_{r\to\infty}(B\Omega | Q_r^f B'\Omega) = (B\Omega | Q^f B'\Omega)$$

with

$$Q_r^f := \int j^0(y, y + x)f(x)\vartheta_r(y)\eta(y^0)d^4x d^4y.$$

Proof: It is sufficient to replace everywhere the local current density $j^\mu(y + a, y)$ by the local current density $\int j^\mu(y, y + x)f(x)d^4x$. □

Now, $Q(a)$ or Q^f resp. are defined on strictly local and some quasilocal states including one-particle asymptotic states. In view of Theorem 4.3 or Theorem 5.1 resp.

we give in the next sections a natural extension of Q^f to multiparticle asymptotic states.

6. ASYMPTOTIC ESTIMATES

To get $Q(a)$ on asymptotic multiparticle states one has to apply an asymptotic condition. For this, we slightly generalize the Haag–Ruelle construction[15-17] to include a field with an additional parameter and nontranslationally covariant x dependence, respectively. (If the noncovariance results simply from multiplying a Wightman field by a c-number function this generalization has already been considered.[18])

Lemma 6.1: Let $f\in\mathcal{D}(\mathbb{R}^4)$, $\eta\in\mathcal{D}(\mathbb{R}^1)$, $l\in\mathbb{N}$, $B_i\in\rho_1$. Then there exists $m\in\mathbb{N}$ such that for all $N\in\mathbb{N}$ the truncated vacuum expectation value

$$W(\mathbf{a}_1,\dots,\mathbf{a}_{l+n};y,p):$$
$$=\langle B_1(\mathbf{a}_1)\cdots B_l(\mathbf{a}_l)\tilde{j}^0(y,p)B_{l+1}(\mathbf{a}_{l+1})\cdots B_{l+n}(\mathbf{a}_{l+n})\rangle^T$$

fulfills

$$\left|\int W(\mathbf{a}_1,\dots,\mathbf{a}_{l+n};y,p)e^{ipy}f(p)\eta(y^0)d^4pd^4y\right|$$
$$\leq C_{l,n,m,N}(1+|\mathbf{a}_l|^2)^m\prod_{\nu=1}^{l+n-1}\frac{1}{(1+|\mathbf{a}_\nu-\mathbf{a}_{\nu+1}|^2)^N}$$

with a number $C_{l,n,m,N}$ not depending on $\mathbf{a}_1,\dots,\mathbf{a}_n$.

Proof: According to one of the standard results of the Haag–Ruelle theory, one has for $\eta\in\mathcal{D}(\mathbb{R}^1)$ that

$$\int W(\mathbf{a}_1,\dots,\mathbf{a}_{l+n};y,p)f(p)\eta(y^0)dy^0d^4p$$

is strongly decreasing in the difference variables

$$\mathbf{a}_1-\mathbf{a}_2,\cdots,\mathbf{a}_l-\mathbf{y},\mathbf{y}-\mathbf{a}_{l+1},\cdots,\mathbf{a}_{l+n-1}-\mathbf{a}_{l+n},$$

and a test function from $\mathcal{S}$ in these variables. We now keep supp$f$ and suppη fixed. Then outside supp$f$, suppη we may put $W=0$. The so changed W is in $0'_c$ with respect to the variables

$$\mathbf{a}_1-\mathbf{a}_2,\cdots,\mathbf{a}_l-\mathbf{y},\mathbf{y}-\mathbf{a}_{l+1},\cdots,\mathbf{a}_{l+n-1}-\mathbf{a}_{l+n},y^0,p,$$

and has a representation as a finite sum[19]

$$W=\sum_{\underline{\nu}}D_p^{\underline{\nu}}w_{\underline{\nu}}(\mathbf{a}_1-\mathbf{a}_2,\dots,\mathbf{a}_{l+n-1}-\mathbf{a}_{l+n};y^0,p)$$

where $w_{\underline{\nu}}$ are continuous, strongly decreasing functions and $D_p^{\underline{\nu}}$ is a differential monomial in p of degree $|\underline{\nu}|$ ($\underline{\nu}$ is a multi-index). $\int Wf(p)dp$ is a continuous function in y; hence, f may be replaced by $e^{ipy}f(p)$ and

$$\int We^{ipy}f(p)\eta(y^0)d^4yd^4p$$
$$=\sum_{\underline{\nu}}\int w_{\underline{\nu}}(-1)^{|\underline{\nu}|}[D_p^{\underline{\nu}}e^{ipy}f(p)]\eta(y^0)d^4yd^4p.$$

For $y^0\in$ suppη and $p\in$ supp$f$ one has an estimate

$$|D_p^{\underline{\nu}}e^{ipy}f(p)|\leq C_f(1+|\mathbf{y}|^2)^{|\underline{\nu}|} \tag{6.1}$$

leading for all $M\in\mathbb{N}$ to

$$\left|We^{ipy}f(p)\eta(y^0)d^4yd^4p\right|$$
$$\leq C_M\int d^3\mathbf{y}(1+|\mathbf{y}|^2)^{|\underline{\nu}|}\frac{1}{(1+|\mathbf{a}_1-\mathbf{a}_2|^2)^M}\cdots$$
$$\times\frac{1}{(1+|\mathbf{a}_l-\mathbf{y}|^2)^M}\frac{1}{(1+|\mathbf{y}-\mathbf{a}_{l+1}|^2)^M}\cdots.$$

The integral over $\mathbf{y}$ can be estimated as

$$\int d^3\mathbf{y}\frac{(1+|\mathbf{y}|^2)^{|\underline{\nu}|}}{(1+|\mathbf{y}-\mathbf{a}_l|^2)^M(1+|\mathbf{y}-\mathbf{a}_{l+1}|^2)^M}$$
$$=\int d^3\mathbf{x}\frac{(1+|\mathbf{x}+\mathbf{a}_l|^2)^{|\underline{\nu}|}}{(1+|\mathbf{x}|^2)^M(1+|\mathbf{x}-\mathbf{a}_{l+1}-\mathbf{a}_l|^2)^M}$$
$$\leq\text{const}\int d^3\mathbf{x}\frac{(1+2|\mathbf{a}_l|^2)^{|\underline{\nu}|}(1+2|\mathbf{x}|^2)^{|\underline{\nu}|}}{(1+|x|^2+|\mathbf{x}-\boldsymbol{\alpha}|^2)^M}$$

with $\mathbf{a}_{l+1}-\mathbf{a}_l=\boldsymbol{\alpha}$

$$\leq\text{const}(1+|\mathbf{a}_l|^2)^{|\underline{\nu}|}|\boldsymbol{\alpha}|^{-2M+3+2|\underline{\nu}|}$$
$$\times\int d^3\mathbf{u}\frac{(1+|\mathbf{u}|^2)^{|\underline{\nu}|}}{(|\mathbf{u}|^2+(|\mathbf{u}|-|\boldsymbol{\alpha}/|\boldsymbol{\alpha}||)^2)^M}$$

for $|\boldsymbol{\alpha}|\geq 1$. This implies the lemma. □

An immediate consequence is

Corollary 6.2: Let $f\in\mathcal{D}(\mathbb{R}^4)$, $\eta\in\mathcal{D}(\mathbb{R}^1)$, $l\in\mathbb{N}$, $B_i\in\rho_1$. Then there exist $m,m'\in\mathbb{N}$ such that for all $N\in\mathbb{N}$ the truncated vacuum expectation value

$$W_t(\mathbf{a}_1,\dots,\mathbf{a}_{l+n};y,p):$$
$$=\langle B_1(\mathbf{a}_1,t)\cdots B_l(\mathbf{a}_l,t)\tilde{j}^0(y,p)B_{l+1}(\mathbf{a}_{l+1},t)\cdots B_{l+n}(\mathbf{a}_{l+n},t)\rangle^T$$

fulfills

$$\left|\int W_t(\mathbf{a}_1,\dots,\mathbf{a}_{l+n};y,p)e^{ipy}f(p)\eta(y^0)d^4pd^4y\right|$$
$$\leq C_{l,m,m',n,N}(1+|t|)^{m'}(1+|\mathbf{a}_l|^2)^m\prod_{\nu=1}^{l+n-1}\frac{1}{(1+|\mathbf{a}_\nu-\mathbf{a}_{\nu+1}|^2)^N}$$

with a number $C_{l,m,m',n,N}$ not depending on $\mathbf{a}_1,\dots\mathbf{a}_{l+n},t$.

Proof: Translational covariance and

$$\partial_\mu[\tilde{j}^\mu(y,p)e^{ipy}]=0$$

yield for the left-hand side

$$\int\langle B_1(\mathbf{a}_1,0)\cdots B_l(\mathbf{a}_l,0)\tilde{j}^0(y,p)e^{ipy}B_{l+1}(\mathbf{a}_{l+1},0)\cdots\rangle^T$$
$$\times e^{ip^0t}f(p)\eta(y^0)d^4pd^4y.$$

Taking into account the additional t dependence of (6.1), one gets the result. □

For the construction of scattering states one needs the following standard results on solutions of Klein–Gordon equations:

Lemma 6.3: Let

$$f(\mathbf{x},t)=\frac{1}{(2\pi)^{3/2}}\int\exp[-i(\omega_{\mathbf{p}}t-\mathbf{px})]\check{f}(\mathbf{p})\frac{d^3\mathbf{p}}{2\omega_{\mathbf{p}}},$$
$$\check{f}\in\mathcal{S}(\mathbb{R}^3),\ \omega_{\mathbf{p}}=(\mathbf{p}^2+m^2)^{1/2},\ m>0.$$

(i) Define the velocity $\mathbf{v}:=\mathbf{p}/\omega_{\mathbf{p}}$, $\Sigma:=\{\mathbf{v}(\mathbf{p}):\mathbf{p}\in\text{supp}\check{f}\}$ and let U be any open set containing Σ. Then

$$|f(\mathbf{v}t,t)|\leq A\cdot(1+|t|)^{-3/2}\ \text{ for }\ \mathbf{v}\in U, \tag{6.2}$$
$$|f(\mathbf{v}t,t)|\leq A_N(1+|\mathbf{v}|)^{-N}(1+|t|)^{-N}\ \text{ for }\ \mathbf{v}\notin U, \tag{6.3}$$

with constants A,A_N not depending on $\mathbf{v}$.

(ii) For any polynomial P_n of degree n in $\mathbf{x}$,

$$\int|P_n(\mathbf{x})f(\mathbf{x},t)|d^3\mathbf{x}\leq C(1+|t|)^{3/2+n} \tag{6.4}$$

with a constant C.

Proof: (i) and (ii) for $P=$ const are due to Ruelle.[20] By rewriting the factor $P_n(x)$ in p-space as derivatives

and applying partial integration (ii) is reduced to the case $P = \text{const}$. □

Lemma 6.4: Let f_i, $i = 1, 2$ be as in Lemma 6.3 with disjoint support in velocity space, $l \in \mathbb{N}$ and $N > l + \frac{3}{2}$. Then there is a $C_{l,N}$ such that

$$\int \frac{|f_1(\mathbf{a}_1, t)|\,|f_2(\mathbf{a}_2, t)|\,|\mathbf{a}_2|^{2l}}{(1+|\mathbf{a}_1 - \mathbf{a}_2|^2)^N} d^3\mathbf{a}_1 d^3\mathbf{a}_2 \leq C_{l,N}(1+|t|)^{3+2l-2N}. \tag{6.5}$$

Proof: The left-hand side can be written as

$$|t|^{6+2l} \int \frac{|f_1(\mathbf{v}_1 t, t)|\,|f_2(\mathbf{v}_2 t, t)|\,|\mathbf{v}_2|^{2l}}{(1+t^2|\mathbf{v}_1 - \mathbf{v}_2|^2)^N} d^3\mathbf{v}_1 d^3\mathbf{v}_2 .$$

Denote by U_i, $i = 1, 2$ two disjoint open sets containing Σ_i, respectively. Splitting the integrals one gets

(i) $|t|^{6+2l} \int_{\mathbf{v}_2 \notin U_2} \cdots \leq C_{L_2}(1+|t|)^{-L_2} \times \int d^3\mathbf{a}_1 d^3\mathbf{a}_2 \frac{|f_1(\mathbf{a}_1, t)|\,|\mathbf{a}_2|^{2l}}{(1+|\mathbf{a}_1 - \mathbf{a}_2|^2)^N}$.

The $\mathbf{a}_2$ integration leads to a polynomial in $\mathbf{a}_1$ and (6.4) implies a bound of the form (6.5),

(ii) $|t|^{6+2l} \int_{\substack{\mathbf{v}_2 \in U_2 \\ \mathbf{v}_1 \notin U_1}} \cdots \leq C_{L_1}(1+|t|)^{-L_1} \times \int d^3\mathbf{a}_1 d^3\mathbf{a}_2 \frac{|f_2(\mathbf{a}_2, t)|\,|\mathbf{a}_2|^{2l}}{(1+|\mathbf{a}_1 - \mathbf{a}_2|^2)^N}$.

Integrating over $\mathbf{a}_1$ and applying (6.4) again implies the bound.

(iii) $|t|^{6+2l} \int_{\substack{\mathbf{v}_2 \in U_2 \\ \mathbf{v}_1 \notin U_1}} \leq \frac{\text{const}}{(1+l^2 \cdot \inf|\mathbf{v}_1 - \mathbf{v}_2|^2)^N} |t|^{3+2l}$. □

7. Q ON SCATTERING STATES

A first step for the definition of Q on scattering states is

Lemma 7.1: Let f_ν be smooth solutions of Klein–Gordon equations with mass $m > 0$ as in Sec. 6 and with pairwise disjoint support in velocity space, and

$$B_\nu^{f,t} = \int B_\nu(\mathbf{a}, t) f(\mathbf{a}, t) d^3\mathbf{a}, \quad B_\nu \in \mathcal{P}_1 .$$

Let $l \geq 2$ or $n \geq 2$. Then for all $N \in \mathbb{N}$ there is a C_N such that for $f \in \mathcal{S}(\mathbb{R}^4)$

$$\Big| \int \langle B_1^{f_1,t} \cdots B_l^{f_l,t} e^{ipy} \tilde{j}^0(y,p) B_{l+1}^{f_{l+1},t} \cdots B_{l+n}^{f_{l+n},t} \rangle^T \times f(p)\eta(y^0) d^4p d^4y \Big| \leq C_N(1+|t|)^{-N}.$$

Proof: Assume $l \geq 2$. Then

$$\Big| \int \langle B_1(\mathbf{a}_1, t) \cdots B_l(\mathbf{a}_l, t) e^{ipy} \tilde{j}^0(y,p) B_{l+1}(\mathbf{a}_{l+1}, t) \cdots \rangle^T \times f(p)\eta(y^0) \prod_{\nu=1}^{l+n} f_\nu(\mathbf{a}_\nu, t) \prod_{\nu=1}^{l+n} d^3\mathbf{a}_\nu d^4p d^4y \Big|$$
$$\leq C_{l,m,m',n,M}(1+|t|)^{m'} \int \frac{(1+|\mathbf{a}_l|^2)^m}{(1+|\mathbf{a}_1 - \mathbf{a}_2|^2)^M} \times \prod_{\nu=1}^{l+n} |f_\nu(\mathbf{a}_\nu, t)| \prod_\nu d^3\mathbf{a}_\nu$$

by Corollary 6.2 with $M > m + \frac{3}{2}$. In case $l = 2$, the left-hand side of the above equation equals

$$C_{2,m,m',n,M}(1+|t|)^{m'} \int \frac{(1+|\mathbf{a}_2|^2)^m |f_1(\mathbf{a}_1)|\,|f_2(\mathbf{a}_2)|}{(1+|\mathbf{a}_1 - \mathbf{a}_2|^2)^M} \times d^3\mathbf{a}_1 d^3\mathbf{a}_2 \prod_{\nu \geq 3} \int |f_\nu(\mathbf{a}, t)| d^3\mathbf{a}$$

which yields the statement by Lemmas 6.4 and 6.3. In case $l > 2$, the above equals

$$C_{l,m,m',M}(1+|t|)^{m'} \int \frac{|f_1|\,|f_2|}{(1+|\mathbf{a}_1 - \mathbf{a}_2|^2)^M} d^3\mathbf{a}_1 d^3\mathbf{a}_2 \times \int (1+|\mathbf{a}_l|^2)^m |f_l(\mathbf{a}_l, t)| d^3\mathbf{a}_l \times \prod_{\substack{\nu \geq 3 \\ \nu \neq l}} \int |f_\nu(\mathbf{a}, t)| d^3\mathbf{a} .$$

Again the result follows by the same lemmas. □

To get proper scattering states, we adopt the following *normalization convention*: The two-point function can be written as

$$(B_\nu^{f_\nu,t}\Omega | B_{\nu'}^{f_{\nu'},t}\Omega) = (2\pi)^{-3} \int \overline{\check{f}_\nu(\mathbf{q})} \check{f}_{\nu'}(\mathbf{q}) (B_\nu \Omega | dE(q) B_{\nu'}\Omega)$$

for $B_\nu, B_{\nu'}$ such that only one-particle states of mass $q^2 = m_\nu^2 = m_{\nu'}^2$ contribute. By a suitable choice of the B_ν one can achieve that

$$(B_\nu \Omega | dE(q) B_{\nu'}\Omega) = \delta(q^2 - m_\nu^2)\delta_{\nu\nu'}\sigma_\nu(\mathbf{q}) d^4q$$

with an infinitely often differentiable $\sigma_\nu(\mathbf{q}) > 0$ for almost all $\mathbf{q}$.[17,21]

We now put

$$\hat{f}_\nu(\mathbf{q}) := (2\pi)^{-3/2} \check{f}_\nu(\mathbf{q}) \sqrt{\sigma_\nu(\mathbf{q})} .$$

Then

$$(B_\nu^{f,t}\Omega | B_{\nu'}^{g,t}\Omega) = \delta_{\nu\nu'} \int \overline{\hat{f}(\mathbf{p})} \hat{g}(\mathbf{p}) \frac{d^3\mathbf{p}}{2(\mathbf{p}^2 + m^2)^{1/2}} .$$

In this notation, a result of the Haag-Ruelle theory is

$$\varphi^{\text{out}}_{\text{in}\,\nu_1 \cdots \nu_n}(\hat{f}_1, \ldots, \hat{f}_n) = \text{s-lim}_{t \to \pm\infty} B_{\nu_1}^{f_1,t} \cdots B_{\nu_n}^{f_n,t}\Omega . \tag{7.1}$$

Here, the left-hand side denotes the outgoing or incoming free n-particle state characterized by quantum numbers ν_i, masses m_{ν_i} and wavefunctions $\hat{f}_i$. It is assumed that the $\hat{f}_i$ have compact support in p and pairwise disjoint support in velocity space. The limit in (7.1) is then attained faster than any inverse power in t.[16,17]

The subsets $D^{\text{out}}_{\text{in}}$ of states of the form (7.1) are dense in the Hilbert spaces of asymptotic configurations.

Theorem 7.2: Consider two states of the form (7.1). Then

$$\lim_{t \to \pm\infty} (B_{\nu_1}^{f_1,t} \cdots B_{\nu_l}^{f_l,t}\Omega | Q^f B_{\mu_1}^{g_1,t} \cdots B_{\mu_m}^{g_m,t}\Omega)$$
$$= \int \sum_{i,k} (\varphi^{\text{out}}_{\text{in}\,\nu_1 \cdots \nu_l}(\hat{f}_1, \ldots, \hat{f}_l) | a^{\text{out}}_{\text{in}\,\nu_i}{}^*(\mathbf{q}) a^{\text{out}}_{\text{in}\,\mu_k}(\mathbf{q}') \times \varphi^{\text{out}}_{\text{in}\,\mu_1 \cdots \mu_m}(\hat{g}_1 \cdots \hat{g}_m))$$
$$\times (a^{\text{out}}_{\text{in}\,\nu_i}{}^*(\mathbf{q})\Omega | Q^f a^{\text{out}}_{\text{in}\,\mu_k}{}^*(\mathbf{q}')\Omega) d^3\mathbf{q} d^3\mathbf{q}' .$$

Here, $a^{\text{in}}_{\nu}(\mathbf{q}), a^{\text{in}}_{\nu}{}^{*}(\mathbf{q})$ denote the usual annihilation and creation operators with commutation relations

$$[a^{\text{in}}_{\mu}(\mathbf{q}), a^{\text{in}}_{\nu_1}{}^{*}(\mathbf{q}')] = 2(\mathbf{q}^2+m_\nu^2)^{1/2}\delta_{\mu\nu}\delta^3(\mathbf{q}-\mathbf{q}') .$$

Proof: On the left-hand side we have

$$\lim_{r\to\infty}\int (B^{f_1,t}_{\nu_1}\cdots B^{f_l,t}_{\nu_l}\Omega\,|\,e^{ipy}\tilde{j}^0(y,p)\cdots B^{g_m,t}_{\mu_m}\Omega)$$
$$\times\bar{f}(p)\vartheta_r(\mathbf{y})\eta(y^0)d^4p d^4y$$
$$=\int (B^{f_1,t}_{\nu_1}\cdots B^{f_l,t}_{\nu_l}\Omega\,|\,e^{ipy}\tilde{j}^0(y,p)\cdots B^{g_m,t}_{\mu_m}\Omega)$$
$$\times\bar{f}(p)\eta(y^0)d^4p d^4y$$

since the last integral is absolutely convergent by, e.g., Lemma 7.1.

We now follow Araki and Haag. By rewriting the integrand in terms of truncated functions, we get four types of expressions:

(i) $\langle\tilde{j}^0\rangle^T\{\Sigma\Pi\langle\cdots\rangle^T\}$,

(ii) $\langle\tilde{j}^0 B^{f_i,t}_{\nu_i}\rangle^T\cdot\{\Sigma\Pi\langle\cdots\rangle^T\}$,

(iii) $\langle(B^{f_i,t}_{\nu_i})^*\tilde{j}^0 B^{g_k,t}_{\mu_k}\rangle^T\{\cdots\}$,

(iv) $\langle(B^{f_i,t}_{\nu_i})^*\cdots(B^{f_k,t}_{\nu_k})^*\tilde{j}^0 B^{g_j,t}_{\mu_j}\cdots B^{g_l,t}_{\mu_l}\rangle\{\cdots\}$,

with at least two B's either before or behind $\tilde{j}^0$. (i) and (ii) do not contribute by the normalization of j^μ and Lemma 4.1. The contribution of (iv) vanishes in the limit $t\to\pm\infty$ by Lemma 7.1. Consider the contribution of (iii). The first factor is time independent and gives

$$\langle(B^{f_i,t}_{\nu_i})^*Q^f B^{g_k,t}_{\mu_k}\rangle^T = (a^{\text{in}}_{\nu_i}{}^{*}(\hat{f}_i)\Omega\,|\,Q^f a^{\text{in}}_{\mu_k}{}^{*}(\hat{g}_k)\Omega) .$$

Its coefficient $\{\cdots\}$ converges to

$$(\varphi^{\text{in}}_{\nu_1\cdots\nu_{i-1}\nu_{i+1}\cdots\nu_l}(\hat{f}_1,\ldots,\hat{f}_{i-1},\hat{f}_{i+1},\ldots,\hat{f}_l)$$
$$\times\varphi^{\text{in}}_{\mu_1\cdots\mu_{k-1}\mu_{k+1}\cdots\mu_m}(\hat{g}_1,\ldots,\hat{g}_{k-1},\hat{g}_{k+1}\cdots\hat{g}_m))$$
$$=(a^{\text{in}}_{\nu_i}(f_i)\varphi^{\text{in}}_{\nu_1\cdots\nu_l}(\hat{f}_1\cdots\hat{f}_l)\,|$$
$$\times a^{\text{in}}_{\mu_k}(g_k)\varphi^{\text{in}}_{\mu_1\cdots\mu_m}(\hat{g}_1,\ldots,\hat{g}_m)) .$$

(For the states considered the test functions have non-overlapping support.) □

Assume now asymptotic completeness.

Corollary 7.3: Q^f maps one-particle states into one-particle states.

Proof: Consider

$$(B^{f_1,t}_{\nu_1}\cdots B^{f_l,t}_{\nu_l}\Omega\,|\,Q^f B^{g_\mu,t}_{\mu}\Omega) .$$

The one-particle state $B^{g,t}_{\mu}\Omega$ is independent of t and lies in the domain of Q^f (Lemma 3.1). By Theorem 7.2, the matrix element converges to zero for $t\to\pm\infty$ if $l>1$. Since the states on the left are dense, the statement follows. □

Theorem 7.2 leads to a definition of Q^f as quadratic forms on the Hilbert spaces of asymptotic n-particle states which extend the action of Q^f on one-particle states. On their domains they coincide with the forms of the formal operators

$$Q^f_{\substack{\text{out}\\\text{in}}} = \int\sum_{\nu,\mu} Q_{\nu\mu}(\mathbf{q},\mathbf{q}')(a^{\text{in}}_{\nu}(\mathbf{q}))^* a^{\text{in}}_{\mu}(\mathbf{q}')$$
$$\times\frac{d^3\mathbf{q}}{2(\mathbf{q}^2+m_\nu^2)^{1/2}}\frac{d^3\mathbf{q}'}{2(\mathbf{q}'^2+m_\mu^2)^{1/2}} , \tag{7.2}$$

where the kernel

$$Q_{\nu\mu}(\mathbf{q},\mathbf{q}') := (a^{\text{in}}_{\nu}{}^{*}(\mathbf{q})\Omega\,|\,Q^f a^{\text{in}}_{\mu}{}^{*}(\mathbf{q}')\Omega) \tag{7.3}$$

is the same and is fixed by the action of Q on one-particle states. In the following Q^f_{ex} stands for either Q^f_{out} or Q^f_{in}. The form domain of Q^f_{ex} includes

$$\hat{D}_{\text{ex}} := \{a^{\text{ex}}_{\nu_1}{}^{*}(\hat{f}_1)\cdots a^{\text{ex}}_{\nu_n}{}^{*}(\hat{f}_n)\Omega\}_{f_i\in\mathcal{S}(\mathbb{R}^3),\,n\in\mathbb{N}} .$$

Obviously, Q^f_{ex} commutes with the corresponding particle number operator. For $Q^f_{\text{ex}}(y) = U(y)Q^f_{\text{ex}}U^{-1}(y)$ this leads to

Lemma 7.4: Let $A\in\mathcal{P}_{\text{SL}}$ and denote by $A^{\text{ex}}(x)$ the incoming or outgoing field belonging to $A(x)$. Then, in the sense of matrix elements between scattering states from $\hat{D}_{\text{ex}}$

$$[Q^f(y),A]^{\text{ex}}(x) = [Q^f_{\text{ex}}(y+x),A^{\text{ex}}(x)] .$$

Proof: $A^{\text{ex}}(x)$ is defined for Hermitian A as an operator on $\hat{D}_{\text{ex}}$ by

$$A^{\text{ex}}(x) = \sum_\rho\int\frac{d^3\mathbf{p}}{2(\mathbf{p}^2+m_\rho^2)^{1/2}}\{(a^{\text{ex}}_{\rho}{}^{*}(\mathbf{p})\Omega\,|\,A(x)\Omega)$$
$$\times a^{\text{ex}}_{\rho}{}^{*}(\mathbf{p}) + (A(x)\Omega\,|\,a^{\text{ex}}_{\rho}{}^{*}(\mathbf{p})\Omega)a^{\text{ex}}_{\rho}(\mathbf{p})\} .$$

$A^{\text{ex}}(x)$ leaves $\hat{D}_{\text{ex}}$ invariant. Therefore

$$i[Q^f_{\text{ex}}(y+x),A^{\text{ex}}(x)]$$
$$= i\sum_{\mu,\nu,\rho}\int\frac{d^3\mathbf{p}}{2(\mathbf{p}^2+m_\rho^2)^{1/2}}\frac{d^3\mathbf{q}}{2(\mathbf{q}^2+m_\nu^2)^{1/2}}\frac{d^3\mathbf{q}'}{2(\mathbf{q}'^2+m_\mu^2)^{1/2}}$$
$$\times(a^*_\nu(\mathbf{q})\Omega|\,Q^f(y+x)a^*_\mu(\mathbf{q}')\Omega)\{[a^*_\nu(\mathbf{q})a_\mu(\mathbf{q}'),a^*_\rho(\mathbf{p})]$$
$$\times(a^*_\rho(\mathbf{p})\Omega\,|\,A(x)\Omega) + [a^*_\nu(\mathbf{q})a_\mu(\mathbf{q}'),a_\rho(\mathbf{p})]$$
$$\times A(x)\Omega\,|\,a^*_\rho(\mathbf{p})\Omega)\}$$
$$= i\sum_{\nu,\mu}\int\frac{d^3\mathbf{q}}{2(\,)^{1/2}}\frac{d^3\mathbf{q}'}{2(\,)^{1/2}}(a^*_\nu(\mathbf{q})\Omega\,|\,Q^f(y+x)\,a^*_\mu(\mathbf{q}')\Omega)$$
$$\times\{a^*_\nu(\mathbf{q})(a^*_\mu(\mathbf{q}')\Omega\,|\,A(x)\Omega)$$
$$- a_\mu(\mathbf{q}')(A(x)\Omega\,|\,a^*_\nu(\mathbf{q})\Omega)\} .$$

By Corollary 7.3 this is equal to

$$i\sum_\nu\int\frac{d^3\mathbf{q}}{2(\,)^{1/2}}(a^*_\nu(\mathbf{q})\Omega\,|\,Q^f(y+x)A(x)\Omega)a^*_\nu(\mathbf{q})$$
$$-\sum_\mu\frac{d^3\mathbf{q}'}{2(\,)^{1/2}}(A(x)\Omega\,|\,Q^f(y+x)a^*_\mu(\mathbf{q}')\Omega)a_\mu(\mathbf{q}')$$
$$=\sum_\nu\int\frac{d^3\mathbf{q}}{2(\,)^{1/2}}(a^*_\nu(q)\Omega\,|\,i[Q^f(y),A](x)\Omega)a^*_\nu(\mathbf{q}) + \text{h.c.}$$
$$= i[Q^f(y),A]^{\text{ex}}(x) .$$ □

8. INVARIANCE OF THE PARTICLE MASS UNDER THE SYMMETRY

Up to now, we only used the fact that asymptotic one-particle states can be reached by polynomials in the smeared fields. For the following, we assume the existence of an interpolating Wightman field $\phi_\nu(x)$ for every asymptotic particle with quantum number ν. We may assume that the ϕ_ν are Hermitian fields.

Lemma 8.1:

$$[Q^f(y), \phi_\nu(z)]^{ex}(x) = \sum_\mu P^{y,z}_{\nu\mu}(\partial_x)\phi^{ex}_\mu(z+x)$$

$$P^{y,z}_{\nu\mu}(\partial_x) := \Xi^{y,z}_{\nu\mu}(\partial_x) + \Lambda^{y,z}_{\nu\mu}(\partial_x)\partial^0_x .$$

$\Xi^{y,z}_{\nu\mu}, \Lambda^{y,z}_{\nu\mu}$ are polynomials in the spatial derivatives ∂_x. Both sides are meant as form-valued tempered distributions in z, x, y on the domain $\hat{D}_{ex} \times \hat{D}_{ex}$.

(The additional y dependence will later be convenient in Jacobi identities.)

Proof: (i) Assume first that all asymptotic fields are scalar. $[Q^f(y), \phi_\nu(z)]^{ex}(x)$ is localized at $z+x$.

$$F_{\nu\mu} := [[Q^f(y), \phi_\nu(z)]^{ex}(x), \phi^{ex}_\mu(u)]$$

depends for fixed y as a c-number only on $x-u, z$, fulfills a Klein–Gordon equation of mass m_μ in u, and vanishes for $z+x-u$ spacelike by locality. (For the latter we have used the fact that transition to asymptotic fields preserves the local structure.[22,23])

F considered as a generalized function of $\xi = z+x-u, z$ fulfills a Klein–Gordon equation in ξ. Hence it has an initial value representation

$$F_{\nu\mu}(\xi, z) = \int \{F_{\nu\mu}(\xi', z)\partial_0 \Delta_{m_\mu}(\xi - \xi') + \partial_0 F_{\nu\mu}(\xi', z)\Delta_{m_\mu}(\xi - \xi')\} d^3\xi' ,$$

where the initial values have support in $\xi' = 0$ and are therefore finite linear combinations of derivatives of δ functions,

$$F_{\nu\mu}(\xi, z) = i(\Xi_{\nu\mu}(\partial_\xi) + \Lambda_{\nu\mu}(\partial_\xi)\partial_{\xi 0})\Delta_{m_\mu}(\xi) = \sum_{\mu'} P_{\nu\mu'}(\partial_x)[\phi^{ex}_{\mu'}(z+x), \phi^{ex}_\mu(u)]$$

(for fixed parameters y and z). Irreducibility of the $\{\phi^{ex}_\mu(x)\}$ now implies the statement.

(ii) The same method of proof can be used for the general case too: One may replace the system of free fields $\{\phi^{ex}_\mu\}$ by an "equivalent" system of free fields $\{\psi^{ex}_\alpha\}$ with commutation relations

$$[\psi_\alpha(x), \psi_\beta(y)]_\pm = i\eta_{\alpha\beta}\delta_{m_\alpha m_\beta}\Delta_{m_\alpha}(x-y)$$

so that no derivatives on Δ_{m_α} occur. Here $\eta_{\alpha\beta}$ is an invertible Hermitian matrix, real for integer spin, and imaginary otherwise. "Equivalent" means that the ϕ^{ex}_μ can be obtained from the $\{\psi^{ex}_\alpha\}$ by differentiation and linear combination.[24]

Diagonalizing $\eta_{\alpha\beta}$ by a unitary matrix $w_{\alpha\alpha'}$, one gets another set of irreducible fields

$$\chi^{ex}_\alpha(x) = w_{\alpha\alpha'}\psi^{ex}_{\alpha'}(x) ,$$

and associated fields

$$\tilde{\chi}^{ex}_\alpha(x) = \lambda^{-1}_\alpha \overline{w}_{\alpha\alpha'}\psi^{ex}_{\alpha'}(x)$$

(λ_α denotes the eigenvalues of η) with

$$[\tilde{\chi}^{ex}_\alpha(x), \chi^{ex}_\beta(y)]_\pm = i\delta_{\alpha\beta}\Delta_{m_\alpha}(x-y) .$$

Now one proceeds as in (i) and exploits the fact that $\tilde{\chi}^{ex}_\alpha(x)$ is a linear combination of derivatives of the ϕ^{ex}_μ. □

We note that the assumption of the existence of local interpolating fields for all particles has been used in the locality argument.

Lemma 8.2:

$$[Q^f_{ex}(y+x), \phi^{ex}_\nu(z+x)] = \sum_\mu P^{y-z}_{\nu\mu}(\partial_x)\phi^{ex}_\mu(z+x) ,$$

where the polynomials Ξ, Λ in the derivatives ∂_x depend only on the parameter $y-z$. Both sides are meant as form-valued tempered distributions in z, x, y on the domain $\hat{D}_{ex} \times \hat{D}_{ex}$.

Proof: Combining Lemmas 7.4 and 8.1 we get

$$[Q^f_{ex}(y+x), \phi^{ex}_\nu(z+x)] = \sum_\mu P^{y,z}_{\nu\mu}(\partial_x)\phi^{ex}_\mu(z+x) .$$

Translation by a yields

$$\sum_\mu (\Xi^{y,z}_{\nu\mu}(\partial_x) + \Lambda^{y,z}_{\nu\mu}(\partial_x)\partial^0_x)\phi^{ex}_\mu(z+x+a) = \sum_\mu (\Xi^{z+a,y+a}_{\nu\mu}(\partial_x) + \Lambda^{z+a,y+a}_{\nu\mu}(\partial_x)\partial^0_x)\phi^{ex}_\mu(z+x+a) .$$

By commuting with $\phi^{ex}_{\mu'}(z+x'+a)$, the invariance of Ξ and Λ follows. □

To show that the matrix elements of Q between one-particle states of different mass vanish, we need a further restriction on Ξ and Λ:

Lemma 8.3: $P^\xi_{\nu\mu}(\partial_\eta)\Delta_{m_\mu}(\eta) = P^{\eta+\xi}_{\mu\nu}(\partial_{-\eta})\Delta_{m_\nu}(-\eta)$

without summation over μ or ν.

Proof: Commute both sides of Lemma 8.2 with $\phi_{\mu'}(z'+y)$ and use the Jacobi identity. □

Lemma 8.4: (i) $P^\xi_{\mu\nu}(\partial_\eta) = 0$ for $m_\mu \neq m_\nu$, (ii) $P^\xi_{\mu\nu}(\partial_\eta)$ is a polynomial also with respect to ξ.

Proof: Consider the polynomial

$$P^\xi_{\mu\nu}(\partial_\eta) = \sum_{k_1,k_2,k_3} (c^{k_1k_2k_3}_{\mu\nu}(\xi) + d^{k_1k_2k_3}_{\mu\nu}(\xi)\partial^0_\eta)(\partial^1_\eta)^{k_1}(\partial^2_\eta)^{k_2}(\partial^3_\eta)^{k_3}$$

$$(\partial^\lambda_\eta = \partial/\partial\eta_\lambda,\ \lambda = 0,1,2,3),$$

with its Fourier transform in ξ and η,

$$\tilde{P}^q_{\mu\nu}(p) = \sum(\tilde{c}^{k_1k_2k_3}_{\mu\nu}(q) - i\tilde{d}^{k_1k_2k_3}_{\mu\nu}(q)\cdot p^0)(ip^1)^{k_1}(ip^2)^{k_2}(ip^3)^{k_3} .$$

Lemma 8.3 reads

$$\sum_{k_1k_2k_3} (\tilde{c}^{k_1k_2k_3}_{\nu\mu}(q) - ip^0\tilde{d}^{k_1k_2k_3}_{\nu\mu}(q)) \times (ip^1)^{k_1}(ip^2)^{k_2}(ip^3)^{k_3}\delta(p^2 - m^2_\mu)\epsilon(p^0) \quad (8.1)$$

$$= \sum_{k_1k_2k_3} (\tilde{c}^{k_1k_2k_3}_{\mu\nu}(q) + i(p^0 - q^0)\tilde{d}^{k_1k_2k_3}_{\mu\nu}(q))(-ip^1 + iq^1)^{k_1} \times (-ip^2 + iq^2)^{k_2}(-ip^3 + iq^3)^{k_3}\delta((p-q)^2 - m^2_\nu)\epsilon(-p^0 + q^0) .$$

We next show that the support in q of this expression is at most $\{q = 0\}$: Consider for a fixed $q \neq 0$ (more correctly, one should smear in q with test functions with small support not containing the origin) Eq. (8.1) as a function of p. The intersection of the supports of both sides of this equation is contained in $\{p : p^2 = m^2_\mu\} \cap \{p : (p-q)^2 = m^2_\nu\} =: S$. For $q \neq 0$ the complement of S in $\{p : p^2 = m^2_\mu\}$ contains both for $p^0 > 0$ and $p^0 < 0$ open three-

dimensional sets, symmetric with respect to $p^0 = 0$, which we call U_+ and U_-. On these, the right-hand side of (8.1) and hence the left-hand side vanish, and therefore

$$\sum \tilde{c}^{k_1k_2k_3}_{\nu\mu}(q)(ip^1)^{k_1}(ip^2)^{k_2}(ip^3)^{k_3} = 0\,,$$
$$\sum \tilde{d}^{k_1k_2k_3}_{\nu\mu}(q)(ip^1)^{k_1}(ip^2)^{k_2}(ip^3)^{k_3} = 0$$

for all $\mathbf{p}: (p^0, \mathbf{p}) \in U_+ \cup U_-$. Since the sum is finite,

$$\tilde{c}^{k_1k_2k_3}_{\nu\mu}(q) = \tilde{d}^{k_1k_2k_3}_{\nu\mu} = 0 \tag{8.2}$$

for $q \neq 0$. For $q = 0$, (8.1) leads again to (8.2) unless $m_\mu = m_\nu$. This proves (i).

For $m_\mu = m_\nu$, we have shown that $\tilde{c}^{k_1k_2k_3}_{\mu\nu}(q)$ and $\tilde{d}^{k_1k_2k_3}_{\mu\nu}(q)$ have support at most at $q = 0$, hence (ii) follows. □

We collect our main results on the symmetry generator:

Theorem 8.5:

(i) $$[Q^f_{\mathrm{ex}}, \phi^{\mathrm{ex}}_\nu(x)] = \sum_{m_\mu = m_\nu} P^x_{\nu\mu}(\partial_x)\phi^{\mathrm{ex}}_\mu(x),$$

where

$$P^x_{\nu\mu}(\partial_x) = \Xi^x_{\nu\mu}(\partial_x) + \Lambda^x_{\nu\mu}(\partial_x)\partial^0_x\,.$$

$\Xi_{\nu\mu}, \Lambda_{\nu\mu}$ are polynomials both in x and the spatial derivatives ∂_x, and the sum is restricted to asymptotic fields with mass m_ν.

(ii) Q^f_{ex} does not depend on $\tilde{f}(p)$ for $p \neq 0$.

(iii) There exists an integer N such that the N-fold commutator of Q^f_{ex} with the momentum operator vanishes,

$$[\cdots[Q^f_{\mathrm{ex}}, P^{\mu_1}], \ldots, P^{\mu_N}] = 0\,.$$

(iv) The commutator of Q^f_{ex} with the mass operator P^2 vanishes on one-particle states

$$[Q^f_{\mathrm{ex}}, P^2]\phi^{\mathrm{ex}}_\nu(x)\Omega = 0\,, \quad [Q^f, P^2]\phi^{\mathrm{ex}}_\nu(x)\Omega = 0\,.$$

All equations containing Q^f_{ex} are meant as matrix elements between states from $\hat{D}_{\mathrm{ex}}$.

Proof: (i) is a restatement of Lemma 8.4.

(ii) Consider Lemma 8.2 for which we now may write

$$[Q^f_{\mathrm{ex}}(y + x), \phi^{\mathrm{ex}}_\nu(x)] = \sum_\mu P^y_{\nu\mu}(\partial_x)\phi^{\mathrm{ex}}_\mu(x)\,.$$

Therefore,

$$(\phi^{\mathrm{ex}}_\rho(z)\Omega \,|\, Q^f_{\mathrm{ex}}(y + x)\phi^{\mathrm{ex}}_\nu(x)\Omega) = P^y_{\nu\rho}(\partial_x)(\phi^{\mathrm{ex}}_\rho(z)\Omega \,|\, \phi^{\mathrm{ex}}_\rho(x)\Omega)$$
$$= (\phi^{\mathrm{ex}}_\rho(z)\Omega \,|\, Q^f(y + x)\phi^{\mathrm{ex}}_\nu(x)\Omega)\,.$$

Integration with $g(y)$ gives for the last expression

$$(\phi^{\mathrm{ex}}_\rho(z)\Omega \,|\, Q^h(x)\phi^{\mathrm{ex}}_\nu(x)\Omega)$$

with $\tilde{h}(p) = (2\pi)^4\tilde{f}(p)\tilde{g}(-p)$. This has to be independent of $\tilde{g}$ if $\mathrm{supp}\tilde{g} \ni 0$ by Lemma 8.4 (ii).

(iii) From (i) one gets

$$[\cdots[Q^f_{\mathrm{ex}}, P^{\mu_1}], \ldots, P^{\mu_N}]\phi^{\mathrm{ex}}_\nu(x)\Omega$$
$$= \sum_\mu \{(i\partial^{\mu_1}_y)\cdots(i\partial^{\mu_N}_y)P^y_{\nu\mu}(\partial_x)\}_{y=-x}\phi^{\mathrm{ex}}_\mu(x)\Omega$$

which vanishes for N large enough by Lemma 8.4 (ii). Hence the local operator

$$[[\cdots[Q^f_{\mathrm{ex}}, P^{\mu_1}], \ldots, P^{\mu_N}], \phi^{\mathrm{ex}}_\nu(x)]$$

annihilates Ω and thus vanishes. The irreducibility of $\{\phi_\nu\}_\nu$ implies (iii).

(iv) From (i)

$$[Q^f_{\mathrm{ex}}, P^2]\phi^{\mathrm{ex}}_\nu(x)\Omega = 0$$

and Q^f_{ex} coincides with Q^f on one-particle states (Corollary 7.3). □

Remark: (iii) of Theorem 8.5 implies [without using (iv)] that the N-fold commutator of Q^f_{ex} with P^2 vanishes too, see the corollary to Lemma I in Ref. 1.

ACKNOWLEDGMENTS

We thank D. Buchholz and R. Haag for helpful criticism, in particular concerning a former version of the proof of Lemma 8.1.

[1] L. O'Raifeartaigh, Phys. Rev. Lett. **14**, 575–7 (1965).
[2] S. Coleman and J. Mandula, Phys. Rev. **159**, 1251–6 (1967).
[3] R. Haag, J. T. Lopuszanski, and M. Sohnius, Nucl. Phys. B **88**, 257–74 (1975).
[4] J. T. Lopuszanski, Commun. Math. Phys. **14**, 158–64 (1969).
[5] L. J. Landau, Commun. Math. Phys. **17**, 156–76 (1970).
[6] K. Kraus and L. J. Landau, Commun. Math. Phys. **24**, 243–52 (1972).
[7] H. Reeh, Nuovo Cimento A **4**, 566–74 (1971); "Symmetries, Currents, and Infinitesimal Generators," in *Statistical Mechanics and Field Theory*, edited by R. N. Sen and C. Weil (Wiley, New York, 1972).
[8] J. T. Lopuszanski, J. Math. Phys. **12**, 2401–12 (1971).
[9] A. H. Völkel, Rep. Math. Phys. **6**, 313–25 (1974).
[10] F. P.-W. Lam, Diplomarbeit, unpublished, University of Göttingen (1976).
[11] R. F. Streater and A. S. Wightman, *PCT, Spin, and Statistics and All That* (Benjamin, New York, 1964).
[12] Y. Dothan and E. Gal-Ezer, Nuovo Cimento A **12**, 465–79 (1972). Let ϕ be a free field of mass m and $h(x)$ a positive frequency solution of the Klein–Gordon equation for mass m. The conserved local current

$$j_\mu(x) = (\partial_\mu\bar{h}(x))\phi(x) - \bar{h}(x)\partial_\mu\phi(x) + \mathrm{h.c.}$$

leads to a symmetric combination of particle creation and annihilation operators.
[13] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[14] D. Kastler, D. W. Robinson, and A. Swieca, Commun. Math. Phys. **2**, 108–20 (1966). Their Lemma III can be immediately adapted to the present situation of unbounded operators.
[15] R. Jost, *The General Theory of Quantized Fields* (American Math. Soc., Providence, 1965).
[16] K. Hepp, Commun. Math. Phys. **1**, 95–111 (1965).
[17] H. Araki and R. Haag, Commun. Math. Phys. **4**, 77–91 (1967).
[18] A. H. Völkel, J. Math. Phys. **13**, 257–68 (1972).
[19] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966).
[20] D. Ruelle, Helv. Phys. Acta **35**, 147–63 (1962).
[21] We could have enlarged $\mathcal{P}_{\mathrm{SL}}$ and $\mathcal{P}_1$ from the outset such that $\mathcal{D}_Q$ contains arbitrary local and quasilocal operators, respectively, without any principal difficulty. Then the B_ν could be chosen such that $\sigma_\nu(\mathbf{q}) > 0$ for all $\mathbf{q}$.[17]
[22] For the case considered here, this is not immediately obvious. A proof can be found in Ref. 5; see also Lemma 4.4.2 in Ref. 23.
[23] H. Reeh, Fortschr. Phys. **16**, 687–706 (1968).
[24] J. Schwartz, J. Math. Phys. **2**, 271–90 (1961). See in particular the last paragraph of Sec. 8.

On the wave-mechanical representation of a Bose-like oscillator

Y. Ohnuki

Department of Physics, Nagoya University, Nagoya 464, Japan

S. Kamefuchi

Institute of Physics, University of Tsukuba, Ibaraki 300-31, Japan

(Received 29 March 1977)

A detailed study is made of the wave-mechanical representation of a one-dimensional Bose-like oscillator whose canonical variables satisfy the general commutation relations first proposed by Wigner. The eigenvalue problems of the momentum and Hamiltonian operators are completely solved, and this is made possible only when wavefunctions in general are allowed to be hyperfunctions. The equivalence between the wave- and matrix-mechanical representations is thereby established for any value of c (a characteristic parameter of the theory), contrary to the conclusion reached previously by Yang. It is also found that for the case $-1/2 < c < 0$ or $0 < c < 1/2$ there exist two classes of eigenfunctions that are mutually separated by a superselection rule.

I. INTRODUCTION

The system of a one-dimensional, Bose-like oscillator is described by the Hamiltonian of the classical form $H = (p^2 + q^2)/2$ given in suitable units. Here the pair of canonical variables q and $p \equiv \dot{q}$ are assumed to satisfy the commutation relations (1.1) below, which follow from the Heisenberg equation of motion $i\dot{A} = [A, H]$ with $A = q, p$, combined with the requirement that the variable q satisfy the equation of motion again of the classical form $\ddot{q} + q = 0$,

$$ip = [q, \tfrac{1}{2}(p^2 + q^2)], \quad -iq = [p, \tfrac{1}{2}(p^2 + q^2)]. \tag{1.1}$$

The most general matrix-mechanical representation of the operators q and p that satisfy (1.1) were obtained by Wigner.[1] In terms of the annihilation and creation operators

$$a \equiv 2^{-1/2}(q + ip), \quad a^\dagger \equiv 2^{-1/2}(q - ip), \tag{1.2}$$

they are given as

$$a_{n,n+1} = a^\dagger_{n+1,n} = \begin{cases} (2E_0 + n)^{1/2} & \text{for } n = \text{even}, \\ (n+1)^{1/2} & \text{for } n = \text{odd}, \end{cases} \tag{1.3}$$

where the quantum number $n = 0, 1, 2, \cdots$ is related to energy eigenvalues through $E_n = E_0 + n$ with $E_0 > 0$. The case of the usual canonical commutation relation $[q, p] = i$, describing an ordinary Bose oscillator, corresponds to $E_0 = \frac{1}{2}$. In this connection it is of interest to note that a set of operators such as $J_1 \equiv (aa + a^\dagger a^\dagger)/4$, $J_2 \equiv i(aa - a^\dagger a^\dagger)/4$, and $J_3 \equiv (a^\dagger a + aa^\dagger)/4$ with a and $a^\dagger$ given by (1.3) satisfy the commutation relations of the Lie algebra SO(2, 1) [or equivalently Sp(2, R)]: $[J_1, J_2] = -iJ_3$, $[J_2, J_3] = iJ_1$, $[J_3, J_1] = iJ_2$. In terms of unitary, irreducible representations of the group SO(2, 1) first obtained by Bargmann,[2] Wigner's representation specified by the single parameter $E_0 > 0$ corresponds to a direct sum $D^{(+)}(-E_0/2) + D^{(+)}(-(E_0+1)/2)$. We also note that the commutation relations (1.1), when expressed in a and $a^\dagger$, take the same form as the so-called parabose commutation relations for a system of one degree of freedom.[3]

Now the purpose of the present paper is to investigate the corresponding wave-mechanical representation for the operators q and p. A preliminary investigation on this problem was made by Yang[4] some time ago, whose results, however, seem to be unsatisfactory for the following reason. On the basis of a very restrictive presumption that wavefunctions for the oscillator should be analytic at the origin $q = 0$, Yang concluded that only for the case of some discrete values of c, a characteristic parameter which is related linearly to E_0 [see (4.37) below], can the eigenvalue problem of H be solved, and further that the expansion theorem of quantum mechanics does not hold unless $c = 0$. Since the matrix-mechanical result (1.3) is valid for any value of c or $E_0 > 0$, Yang's conclusion implies that the equivalence between the wave- and matrix-mechanical representations does not hold in general. However, in view of the fact that the formalism of quantum mechanics does not usually depend upon the choice of representations we would expect that the above conclusion might only be specific to his too restrictive presumption. In fact, according to the general principle of quantum mechanics anything that is susceptible of the probabilistic interpretation should be entitled to a wavefunction. Taking such a standpoint we shall show in what follows that when wavefunctions are allowed to be hyper- as well as ordinary functions,[5] eigenvalue problems can be solved for any value of c, so that the equivalence between the two representations can be recovered in its complete form.

In Sec. II the expression for the momentum operator p in the wave-mechanical representation is given, and its eigenvalue problem is solved in Sec. III. It is shown that corresponding to the two cases (i) $c > -\frac{1}{2}$ and (ii) $c < \frac{1}{2}$, $c \neq 0$ there exist the respective classes of eigenfunctions, and that the eigenfunctions are given in general by hyperfunctions. The eigenvalue problem of the Hamiltonian H can also be dealt with in essentially the same manner, and is solved in Sec. IV. Again there exist two classes of eigenfunctions corresponding to the above-mentioned ranges of c. It is found in particular that the ground states of the cases (i) and (ii) are given

 0022-2488/78/1901-0067$1.00

by even and odd (hyper-) functions, respectively. For any value of c the resulting matrix elements of a and $a^\dagger$ completely agree with (1.3).

One of the most characteristic features of our results is that for the case when c falls in the range $-\frac{1}{2} < c < 0$ or $0 < c < \frac{1}{2}$ both classes of eigenfunctions (i) and (ii) become possible. It is shown in Sec. V, however, that superposition of eigenfunctions belonging to different classes is absolutely forbidden, so that the Hilbert spaces spanned by the respective classes of eigenfunctions are separated by a superselection rule. The paper is concluded, in Sec. VI, with some supplementary remarks concerning the use of hyperfunctions in quantum mechanics.

II. THE EXPRESSION FOR THE MOMENTUM OPERATOR

When the operators q and p satisfy the canonical commutation relations $[q, p] = i$ we may put $p = -id/dq$ in the wave-mechanical representation. The purpose of the present section is to derive the corresponding expression for p when the operators concerned satisfy the general commutation relations (1.1). For completeness of arguments we describe the derivation in some detail which is originally due to Yang.[4]

Introducing the operator

$$S \equiv [q, p] - i, \tag{2.1}$$

we can rewrite (1.1) in the form

$$\{p, S\} = 0, \quad \{q, S\} = 0. \tag{2.2}$$

In discussing the wave-mechanical representation it is necessary, first of all, to make the following assumption:

Eigenvalues q' of the operator q have a continuous spectrum such as $-\infty < q' < \infty$. Further, the Hilbert space H can be expanded in the eigenfunctions $|q'\rangle$ of q, or H is a subspace of the space spanned by $|q'\rangle$'s. The validity of this assumption will be seen to be justified *a posteriori*.

Sandwiching the second relation in (2.2) with $|q'\rangle$ and $|q''\rangle$ we obtain

$$(q' + q'')\langle q'|S|q''\rangle = 0, \tag{2.3}$$

whence

$$\langle q'|S|q''\rangle = 2ic(q')\delta(q' + q''). \tag{2.4}$$

The relation $S = -S^\dagger$ imposes a restriction on the function $c(q')$ such as

$$c^*(q') = c(-q'). \tag{2.5}$$

Using (2.1) we write the left-hand side of (2.4) explicitly

$$\begin{aligned}\langle q'|S|q''\rangle &= \langle q'|([q, p] - i)|q''\rangle \\ &= (q' - q'')\langle q'|p|q''\rangle - i\delta(q' - q'').\end{aligned} \tag{2.6}$$

This together with (2.4) then gives

$$\begin{aligned}\langle q'|p|q''\rangle = &-i\delta'(q' - q'') + \frac{ic(q')}{q'}\delta(q' + q'') \\ &+ B(q')\delta(q' - q''),\end{aligned} \tag{2.7}$$

where the undetermined $B(q')$ should be a real function owing to the hermiticity of p, and $1/x = f(x)$ is an odd function satisfying $xf(x) = 1$. Writing as usual the wave-function for a state $|\ \rangle$ as $\psi(q') = \langle q'|\ \rangle$, and applying to (2.7) the completeness condition $\int dq'\,|q'\rangle\langle q'| = 1$, we find

$$\begin{aligned}\langle q'|p|\ \rangle = &\frac{1}{i}\frac{d}{dq'}\psi(q') + \frac{ic(q')}{q'}\psi(-q') \\ &+ B(q')\psi(q'),\end{aligned} \tag{2.8}$$

and hence

$$p = \frac{1}{i}\frac{d}{dq} + \frac{ic(q)}{q}R + B(q), \tag{2.9}$$

where R is a unitary operator such that

$$R\psi(q) = \psi(-q), \quad R = R^{-1} = R^\dagger. \tag{2.10}$$

The undetermined function $B(q)$ can be eliminated by means of a unitary transformation $\exp[iF(q)]$ with $F(q) = -\int^q dq'\,B(q')$:

$$\begin{aligned}&p \to \exp[-iF(q)]\,p\,\exp[iF(q)] \\ &= \frac{1}{i}\frac{d}{dq} + \frac{i\exp[-i(F(q) - F(-q))]\,c(q)}{q}R.\end{aligned} \tag{2.11}$$

Since $\exp[-i(F(q) - F(-q))]c(q)$ satisfies (2.5), we may write the whole expression as $c(q)$. Thus, as an expression which is unitarily equivalent to (2.9) we can adopt the following,

$$p = \frac{1}{i}\frac{d}{dq} + \frac{ic(q)}{q}R. \tag{2.12}$$

So far use has been made only of the second equation in (2.2). We now make use of the first equation thereof. From (2.4) we obtain

$$S = 2ic(q)R. \tag{2.13}$$

Substituting this and (2.12) in the first equation of (2.2) we obtain

$$\left\{\frac{d}{dq}, c(q)R\right\} = \left\{\frac{c(q)}{q}R, c(q)R\right\}. \tag{2.14}$$

When use is made of the relations

$$\begin{aligned}&R^2 = 1, \quad R\left(\frac{1}{q}\right)R = -\frac{1}{q}, \\ &R\left(\frac{d}{dq}\right)R = -\frac{d}{dq},\end{aligned} \tag{2.15}$$

which are all obtainable from (2.10), (2.14) leads to $dc(q)/dq = 0$ so that

$$c(q) = c = \text{real const.} \tag{2.16}$$

on account of (2.5). We thus arrive at the final expression for p,

$$p = \frac{1}{i}\frac{d}{dq} + i\frac{c}{q}R. \tag{2.17}$$

The theory does not impose any restriction on the constant c, except that it is a real constant. In the following sections we shall see, by actually solving the eigenvalue problems of p and H, how the value of c affects the structure of the theory.

III. THE EIGENVALUE PROBLEM OF THE MOMENTUM OPERATOR

Let $\psi_{p'}(q)$ be the eigenfunction of the momentum operator p with eigenvalue p'. The equation to be solved is then given by

$$\left(\frac{d}{dq} - ip'\right)\psi_{p'}(q) = \frac{c}{q}\psi_{p'}(-q), \tag{3.1}$$

where use is made of (2.17). Since $\psi_{p'}(q) \sim \exp(ip'q)$ for $|q|$ being sufficiently large, the condition for the probabilistic interpretation of $\psi_{p'}(q)$ to be possible may be written as

$$\int_\mu^\lambda dq\,|\,\psi_{p'}(q)|^2 = \text{finite} \quad (\lambda > \mu \text{ finite}). \tag{3.2}$$

To begin with let us consider the case $p' \neq 0$. Putting

$$x = p'q, \quad \psi_{p'}(q) = \psi(x), \tag{3.3}$$

we obtain from (3.1):

$$\begin{aligned}\left(\frac{d}{dx} - i\right)\psi(x) &= \frac{c}{x}\psi(-x),\\ \left(\frac{d}{dx} + i\right)\psi(-x) &= \frac{c}{x}\psi(x).\end{aligned} \tag{3.4}$$

For the functions $\psi^{(\pm)}(x)$ defined by

$$\psi^{(\pm)}(x) \equiv \psi(x) \pm \psi(-x), \tag{3.5}$$

we have

$$\begin{aligned}\left(\frac{d}{dx} - \frac{c}{x}\right)\psi^{(+)}(x) &= i\psi^{(-)}(x),\\ \left(\frac{d}{dx} + \frac{c}{x}\right)\psi^{(-)}(x) &= i\psi^{(+)}(x).\end{aligned} \tag{3.6}$$

Special caution must be exercised at $x = 0$ where Eqs. (3.6) have a singularity. Let us therefore solve the equations first in the region $x \neq 0$ and then pass to the region including $x = 0$.

Applying $(d/dx + c/x)$ to both sides of the first equation in (3.6) and then using the second equation therein, we find

$$\left(\frac{d^2}{dx^2} + 1 - \frac{c(c-1)}{x^2}\right)\psi^{(+)}(x) = 0 \quad (x \neq 0). \tag{3.7}$$

For $f(x)$ defined by

$$\psi^{(+)}(x) \equiv |x|^{1/2} f(x), \tag{3.8}$$

we have

$$\left(\frac{d^2}{dx^2} + \frac{1}{x}\frac{d}{dx} + 1 - \frac{(c-1/2)^2}{x^2}\right) f(x) = 0 \quad (x \neq 0), \tag{3.9}$$

which is Bessel's differential equation. According to (3.2), on the other hand, $\psi^{(+)}(x)$ and $f(x)$ are required to satisfy the conditions

$$\begin{aligned}\lim_{\mu\to 0}\int_\mu^\lambda dx\ |\psi^{(+)}(x)|^2 &= \lim_{\mu\to 0}\int_\mu^\lambda dx|x|\ |f(x)|^2\\ &= \text{finite},\end{aligned} \tag{3.10}$$

for $\lambda > \mu > 0$. Thus, by taking account of the fact that $f(x)$ is an even function, we obtain the following three candidates for $\psi^{(+)}(x)$:

$$\begin{aligned}&\text{(i)}\ |x|^{1/2}J_{c-1/2}(|x|) \quad (c > -\tfrac{1}{2}),\\ &\text{(ii)}\ |x|^{1/2}J_{1/2-c}(|x|) \quad (c < \tfrac{3}{2}),\\ &\text{(iii)}\ |x|^{1/2}Y_0(|x|) \quad (c = \tfrac{1}{2}).\end{aligned} \tag{3.11}$$

When such $\psi^{(+)}(x)$'s with $x \neq 0$ are substituted in the first equation in (3.6), the corresponding $\psi^{(-)}(x)$'s with $x \neq 0$ are obtained. Like (3.10) such $\psi^{(-)}(x)$'s have to satisfy

$$\lim_{\mu\to 0}\int_\mu^\lambda dx\,|\psi^{(-)}(x)|^2 = \text{finite}. \tag{3.12}$$

Let us first consider the case (iii). Here $\psi^{(-)}(x)$ with $x \neq 0$ is found to be

$$\begin{aligned}&\frac{1}{i}\left(\frac{d}{dx} - \frac{1}{2x}\right)\left[|x|^{1/2}Y_0(|x|)\right]\\ &\quad = \frac{1}{i}\epsilon(x)|x|^{1/2}Y_1(|x|),\end{aligned} \tag{3.13}$$

with

$$\epsilon(x) = \begin{cases} 1 & x > 0,\\ -1 & x < 0.\end{cases} \tag{3.14}$$

We know that in the neighborhood of $x = 0$, $Y_1(|x|)$, a Bessel function of the second kind, behaves as $\sim -\pi^{-1}[|x|^{-1} - |x|\log|x| + O(|x|)]$, so that the $\psi^{(-)}(x)$ given by (3.13) does not satisfy (3.12). The case (iii) is thereby excluded. We are thus left with the cases (i) and (ii). The $\psi^{(-)}(x)$'s with $x \neq 0$ obtained by substituting (3.11) in the first equation of (3.6) are $i\epsilon(x)|x|^{1/2}J_{c+1/2}(|x|)$ and $-i\epsilon(x)|x|^{1/2}J_{-1/2-c}(|x|)$ for the cases (i) and (ii), respectively. Now the conditions under which the $\psi^{(-)}(x)$'s thus obtained satisfy (3.12) are given as follows: In the case (i) the condition agrees with the one obtained above, i.e., $c > -\frac{1}{2}$, whereas in case (ii) a stronger condition has to be imposed, that is, $c < \frac{1}{2}$. Summarizing, the $\psi^{(\pm)}(x)$'s with $x \neq 0$ which satisfy (3.6), (3.10), and (3.12) consist of the following two classes:

$$\begin{aligned}&\text{(i)}\ c > -\tfrac{1}{2},\\ &\quad \psi^{(+)}(x) = |x|^{1/2}J_{c-1/2}(|x|),\\ &\quad \psi^{(-)}(x) = i\epsilon(x)|x|^{1/2}J_{c+1/2}(|x|),\end{aligned} \tag{3.15}$$

$$\begin{aligned}&\text{(ii)}\ c < \tfrac{1}{2},\\ &\quad \psi^{(+)}(x) = |x|^{1/2}J_{1/2-c}(|x|),\\ &\quad \psi^{(-)}(x) = -i\epsilon(x)|x|^{1/2}J_{-1/2-c}(|x|).\end{aligned} \tag{3.16}$$

Let us now try to extend the above results to the region including $x = 0$. In such an extended region we may interpret (3.15) and (3.16) as

$$|x|^{1/2}J_\nu(|x|) = \sum_{n=0}^\infty \frac{(-1)^n}{2^{2n+\nu}n!\,\Gamma(n+\nu+1)}|x|^{2n+\nu+1/2}, \tag{3.17}$$

or

$$\begin{aligned}\epsilon(x)|x|^{1/2}J_\nu(|x|) &= \sum_{n=0}^\infty \frac{(-1)^n}{2^{2n+\nu}n!\,\Gamma(n+\nu+1)}\\ &\quad\times\epsilon(x)|x|^{2n+\nu+1/2}.\end{aligned} \tag{3.18}$$

Formally these are none other than the power series expansions of the Bessel function, but our intention here is to try to interpret each term in the

expansions, i.e., $|x|^\alpha$ or $\epsilon(x)|x|^\alpha$ $(\alpha > -\frac{1}{2})$ as a hyperfunction. Thus $\epsilon(x)|x|^\alpha$ as a whole is to be regarded as a single hyperfunction,[6] and should not, in general, be identified with the product[7] of two hyperfunctions $\epsilon(x)$ and $|x|^\alpha$: For distinction such a product will hereafter be denoted by $\epsilon(x)\cdot|x|^\alpha$.

Now, in the present formalism the basic operation consists of q and p, and it is necessary, first of all, to make clear how the operators x, d/dx, and $1/x$ act on the hyperfunctions $|x|^\alpha$ and $\epsilon(x)|x|^\alpha$. According to the theory of hyperfunctions[6] we have the following rules[8]:

$$\begin{aligned} &x\cdot|x|^\alpha=\epsilon(x)|x|^{\alpha+1} \quad (\alpha>-\tfrac{1}{2}),\\ &x\cdot\epsilon(x)|x|^\alpha=|x|^{\alpha+1} \quad (\alpha>-\tfrac{1}{2}), \end{aligned} \tag{3.19}$$

and

$$\begin{aligned} &\frac{d}{dx}|x|^\alpha=\alpha\epsilon(x)|x|^{\alpha-1} \quad (\alpha>-\tfrac{1}{2}),\\ &\frac{d}{dx}[\epsilon(x)|x|^\alpha]=\alpha|x|^{\alpha-1} \quad (\alpha>-\tfrac{1}{2},\ \alpha\neq 0),\\ &\frac{d}{dx}\epsilon(x)=2\delta(x), \end{aligned} \tag{3.20}$$

where the region of α can further be extended, but this will not be needed for our present purpose. Although in (3.20) the third formula follows from the second by taking the limit $\alpha\to 0$, we have written them separately for the sake of later convenience. As for the product of $1/x$ and the hyperfunctions let us adopt the following rules:

$$\begin{aligned} &\frac{1}{x}\cdot|x|^\alpha=\epsilon(x)|x|^{\alpha-1} \quad (\alpha>-\tfrac{1}{2}),\\ &\frac{1}{x}\cdot\epsilon(x)|x|^\alpha=|x|^{\alpha-1} \quad (\alpha>-\tfrac{1}{2},\ \alpha\neq 0). \end{aligned} \tag{3.21}$$

The case $\alpha=0$ is excluded in the second equation of (3.21) for the following reason: The right-hand side of the equation $(1/x)\cdot\epsilon(x)=|x|^{-1}$ contains, as a hyperfunction, an ambiguity of the form $c\delta(x)$ with c being a constant, and it is not possible in general to fix the value of c in such a way that any calculation becomes consistent at every step.[9] As will be seen below, however, there exists a fortunate situation for the present case such that we can develop the discussions without directly touching upon this point.

In this connection let us insert a supplementary remark concerning products of the type (3.21), since as is well known the product of two hyperfunctions is not always well definable. In the present case $|x|^\alpha$ and $\epsilon(x)|x|^\alpha$ are regarded, respectively, as the even and odd hyperfunctions among the solutions $f(x)$ of the equation $xdf(x)/dx=\alpha f(x)$.[10] We therefore have the relation $(1/x)\cdot f=(1/\alpha)(df/dx)+c\delta(x)$ which involves an undetermined term. The products (3.21) correspond to $c=0$, and are, in fact, the extensions to $x=0$ of the relations valid for $x\neq 0$. The relations (3.19) and (3.20) correspond also to the same kind of extensions. When the eigenfunctions which have been found for the region $x\neq 0$ are made to satisfy (3.4) with $x=0$, it is quite a reasonable procedure to adopt the relations (3.21) extended from the region $x\neq 0$. In fact, it can easily be confirmed that when the term $c\delta(x)$ $(c\neq 0)$ is retained in the above relation it is not possible to satisfy (3.4). Stated otherwise, the ambiguity inherent in the product of the above-mentioned two hyperfunctions can be uniquely fixed when considered within the framework of the eigenvalue problem under consideration.[11] Accordingly, in dealing with hyperfunctions in what follows we shall consistently use the relations (3.19)~(3.21).

In the case (i) the relations (3.17)~(3.21) enable us to show easily that the functions (3.15) satisfy (3.6). In the case (ii), on the other hand, we find that (3.16) fails to satisfy the same equation only when $c=0$. In fact, for the case (ii) with $c=0$ we have $\psi^{(+)}(x)=(2/\pi)^{1/2}\sin|x|$ and $\psi^{(-)}(x)=i(2/\pi)^{1/2}\epsilon(x)\cos|x|$, whereas according to (3.20) $(d/dx)[\epsilon(x)\cos|x|]=2\delta(x)-\sin|x|$, thereby implying that the second equation in (3.6) with $c=0$ is not satisfied. In this respect let us make the following remark. For $c\neq 0$ the formulas (3.21) give $(c/x)\cdot\epsilon(x)|x|^{-\alpha}=c|x|^{-\alpha-1}$, whereas $\lim_{c\to 0}c|x|^{-\alpha-1}=-2\delta(x)$ as known in the theory of hyperfunctions. Thus the second equation in (3.6) is satisfied in the sense that $\lim_{c\to 0}(d/dx+c/x)[-i\epsilon(x)|x|^{1/2}J_{-1/2-c}(|x|)]=i|x|^{1/2}J_{1/2}(x)$. It should be noticed, however, that this expression does not coincide with the corresponding result $(d/dx)[-i\epsilon(x)|x|^{1/2}\times J_{-1/2}(|x|)]$ for the case in which c is set equal to zero from the outset, where $p=-id/dx$.

For $p'\neq 0$ we have thus found the following, two classes of eigenfunctions:

(i) $c>-\frac{1}{2}$,

$$\psi_{p'}(q)=\tfrac{1}{2}|p'q|^{1/2}[J_{-1/2+c}(|p'q|)+i\epsilon(p'q)J_{1/2+c}(|p'q|)], \tag{3.22}$$

(ii) $c<\frac{1}{2}$, $c\neq 0$,

$$\psi_{p'}(q)=\tfrac{1}{2}|p'q|^{1/2}[J_{1/2-c}(|p'q|)-i\epsilon(p'q)J_{-1/2-c}(|p'q|)]. \tag{3.23}$$

Let us now turn to the case $p'=0$. Since both (3.22) and (3.23) have $|p'q|$ as arguments, they may be interpreted either as hyperfunctions of q with $-\infty<q<\infty$, $p'\neq 0$, or as the same kind of functions of $x\equiv p'q$ with $-\infty<x<\infty$. In the latter case $\psi_{p'}(q)$'s may be regarded as hyperfunctions of p' with $-\infty<p'<\infty$. In other words, (3.22) and (3.23), when regarded as such functions of p', can be extended to the point $p'=0$, thereby providing the required eigenfunctions. Throughout the following arguments $\psi_{p'}(q)$'s are to be understood in this sense. That q and p' can be treated in such a symmetrical manner is due to the reciprocity, or invariance under the transformation $q\to -p$, $p\to q$ of the basic relations (1.1).

Lastly we show that in any of the two classes (i) and (ii), the eigenfunctions $\psi_{p'}(q)$ form a complete, orthonormalized system. Since $\psi_{p'}(q)\not\to 0$ as $|q|\to\infty$, and since they are not square integrable, let us define their inner products by introducing a damping factor as follows:

$$(\psi_{p'},\psi_{p''})\equiv\lim_{\xi\to 0}\int_{-\infty}^{\infty}dq\,\exp(-\xi q^2)\psi_{p'}^*(q)\psi_{p''}(q) \quad (\xi>0), \tag{3.24}$$

where both $\psi_{p'}$ and $\psi_{p''}$ are taken to belong to one and the same class (i) or (ii). (The case when two eigenfunctions belong to different classes will be discussed in Sec. V.) The expression (3.24) contains products of two hyperfunctions. For such products let us assume, as usual, the following relations:

$$\begin{aligned}
&|p'q|^{\alpha}\cdot|p''q|^{\beta}=|p'|^{\alpha}|p''|^{\beta}|q|^{\alpha+\beta},\\
&\epsilon(p'q)|p'q|^{\alpha}\cdot\epsilon(p''q)|p''q|^{\beta}\\
&\quad=\epsilon(p')|p'|^{\alpha}\epsilon(p'')|p''|^{\beta}|q|^{\alpha+\beta},\\
&|p'q|^{\alpha}\cdot\epsilon(p''q)|p''q|^{\beta}\\
&\quad=|p'|^{\alpha}\epsilon(p'')|p''|^{\beta}\epsilon(q)|q|^{\alpha+\beta}\quad(\alpha,\beta>-\tfrac{1}{2}).
\end{aligned}\tag{3.25}$$

Further, the integral (3.24) for the product of an even hyperfunction and an odd hyperfunction is taken to be vanishing. The assumption (3.25) is a reasonable one in the sense that each element on either side of (3.25) is a function integrable for any finite interval. In fact it will be confirmed in later discussions that the above assumption does not cause any difficulties at all.

In evaluating the integral in (3.24) we make use of the formula[12]:

$$\begin{aligned}
&\int_0^{\infty}dq\exp(-\xi q^2)[|p'q|^{1/2}J_{\alpha}(|p'q|)][|p''q|^{1/2}J_{\alpha}(|p''q|)]\\
&\quad=\frac{|p'p''|^{1/2}}{2\xi}\exp\left(-\frac{|p'|^2+|p''|^2}{4\xi}\right)I_{\alpha}\left(\frac{|p'p''|}{2\xi}\right)\\
&\qquad(\alpha>-1),
\end{aligned}\tag{3.26}$$

where $I_{\alpha}(z)$ stands for a modified Bessel function with the asymptotic behavior: $I_{\alpha}(z)\sim e^{z}/(2\pi z)^{1/2}$ as $z\to\infty$. We thus obtain

$$(3.26)=\frac{1}{2(\pi\xi)^{1/2}}\exp\left(-\frac{(|p'|-|p''|)^2}{4\xi}\right)\qquad\text{for small }\xi,\tag{3.27}$$

and hence

$$\lim_{\xi\to 0}(3.26)=\delta(|p'|-|p''|).\tag{3.28}$$

This together with (3.24) and (3.25) then gives

$$\begin{aligned}
(\psi_{p'},\psi_{p''})&=\lim_{\xi\to 0}\int_{-\infty}^{\infty}dq\exp(-\xi q^2)\psi_{p'}^{*}(q)\psi_{p''}(q)\\
&=\delta(p'-p'').
\end{aligned}\tag{3.29}$$

Interchanging the roles of p' and q we also obtain

$$\lim_{\xi\to 0}\int_{-\infty}^{\infty}dp'\exp(-\xi p'^2)\psi_{p'}^{*}(q)\psi_{p'}(q')=\delta(q-q'),\tag{3.30}$$

which is our completeness condition. [At this point the following remark may be in order. As can be seen from (3.26), (3.29) has been derived under the condition $p',p''\neq 0$. However, the points of either or both of p' and p'' being zero yield merely vanishing contributions with measure zero. Let $f_i(p')$ $(i=1,2)$ be good functions and let $\tilde{f}_i(q)$ be defined by

$$\tilde{f}_i(q)\equiv\int_{-\infty}^{\infty}dp'\psi_{p'}(q)f_i(p')\quad(i=1,2).\tag{3.31}$$

Taking account of the fact that $\psi_{p'}(q)$'s are hyperfunctions, we then find from (3.26) and (3.27) that

$$\lim_{\xi\to 0}\int_{-\infty}^{\infty}dq\exp(-\xi q^2)\tilde{f}_1^{*}(q)\tilde{f}_2(q)=\int_{-\infty}^{\infty}dp'\,f_1^{*}(p')f_2(p').\tag{3.32}$$

As a matter of course, the delta function on the right-hand side of (3.29) is to be understood in this sense. The same remark applies to (3.30) and (5.18) given below.]

We have thus solved the eigenvalue problem completely for any given value of c. Particularly noteworthy is the fact that for the case when c happens to be in the range $(-\frac{1}{2},0)$ or $(0,\frac{1}{2})$ both classes of eigenfunctions, (i) and (ii), become available. The interrelation between the two classes will be discussed in Sec. V.

IV. THE EIGENVALUE PROBLEM OF THE HAMILTONIAN

In discussing the eigenvalue problem of the Hamiltonian H the expression for H that is obtained by substituting (2.17) in $H=(p^2+q^2)/2$ is not convenient. Instead we shall use the one that is obtained by substituting in $H=\{a^{\dagger},a\}/2$ the following relations [cf. (1.2)]:

$$\begin{aligned}
a&=2^{-1/2}\left(\frac{d}{dq}+q-\frac{c}{q}R\right),\\
a^{\dagger}&=2^{-1/2}\left(-\frac{d}{dq}+q+\frac{c}{q}R\right).
\end{aligned}\tag{4.1}$$

Since H and R are commutable, these operators possess simultaneous eigenfunctions $\psi^{(\pm)}(q)$ such that

$$R\psi^{(\pm)}(q)=\pm\psi^{(\pm)}(q),\tag{4.2}$$

$$\begin{aligned}
&\left[\left(\frac{d}{dq}+q\pm\frac{c}{q}\right)\left(\frac{d}{dq}-q\mp\frac{c}{q}\right)\right.\\
&\quad\left.+\left(\frac{d}{dq}-q\pm\frac{c}{q}\right)\left(\frac{d}{dq}+q\mp\frac{c}{q}\right)\right]\psi^{(\pm)}(q)=-4E\psi^{(\pm)}(q).
\end{aligned}\tag{4.3}$$

As in the case of p we first try to solve (4.3) for the case $q\neq 0$. Breaking up the bracketed factors in (4.3), we obtain

$$\begin{aligned}
&\left(\frac{d^2}{dq^2}-2q\frac{d}{dq}-\frac{c(c\mp 1)}{q^2}\right)f^{(\pm)}(q)\\
&\quad=-(2E-1)f^{(\pm)}(q)\quad(q\neq 0),
\end{aligned}\tag{4.4}$$

where we have put

$$\psi^{(\pm)}(q)\equiv\exp(-q^2/2)f^{(\pm)}(q).\tag{4.5}$$

It is now sufficient to consider (4.4) with $q>0$. In fact, once a solution $f(q)$ of (4.4) for $q>0$ is known, the corresponding solutions for $q\neq 0$ will be obtained as $f^{(+)}(q)=f(|q|)$ and $f^{(-)}(q)=\epsilon(q)f(|q|)$. By putting

$$x^{1/2}\equiv q,\quad f^{(\pm)}(q)\equiv x^{\pm c/2}g^{(\pm)}(x),\tag{4.6}$$

we can further rewrite (4.4) in the form

$$\left[x\frac{d^2}{dx^2}+(\beta^{(\pm)}-x)\frac{d}{dx}-\alpha^{(\pm)}\right]g^{(\pm)}(x)=0,\tag{4.7}$$

where

$$\alpha^{(\pm)}\equiv-\tfrac{1}{2}(E\mp c-\tfrac{1}{2}),\quad\beta^{(\pm)}\equiv\pm c+\tfrac{1}{2}.\tag{4.8}$$

Thus, our equation has been reduced to a confluent hypergeometric equation.

We now solve (4.7) by dividing the possibilities into the following three cases: (1) $\beta^{(\pm)} \neq$ integer, (2) $\beta^{(\pm)} = n+1$, and (3) $\beta^{(\pm)} = -n$, where $n = 0, 1, 2, \cdots$.

(1) *The case* $\beta^{(\pm)} \neq$ *integer*

In this case the two independent solutions of (4.7) are given by confluent hypergeometric functions such as

$$\Phi(\alpha^{(\pm)}, \beta^{(\pm)}; x) = 1 + \frac{\alpha^{(\pm)}}{\beta^{(\pm)}} x + \frac{\alpha^{(\pm)}(\alpha^{(\pm)}+1)}{\beta^{(\pm)}(\beta^{(\pm)}+1)} \frac{x^2}{2!} + \cdots, \tag{4.9}$$

which is regular at $x = 0$, and

$$x^{1-\beta^{(\pm)}} \Phi(\alpha^{(\pm)} - \beta^{(\pm)} + 1, 2 - \beta^{(\pm)}; x), \tag{4.9'}$$

which is singular at $x = 0$. In order to see the behavior of $|\psi^{(\pm)}(q)|$ for $|q| \to \infty$ we make use of the asymptotic form of $\Phi(a, b; x)$ for $x \to \infty$,

$$\Phi(a, b; x) \underset{x \to \infty}{\sim} \frac{\Gamma(b)}{\Gamma(a)} e^x x^{a-b} \qquad (a \neq 0, -1, -2, \cdots). \tag{4.10}$$

Thus, the asymptotic forms of $|\psi^{(\pm)}(q)|$ obtained from (4.9) and (4.9′) are given, respectively, by

$$|\psi^{(\pm)}(q)| \underset{|q| \to \infty}{\sim} \left| \frac{\Gamma(\beta^{(\pm)})}{\Gamma(\alpha^{(\pm)})} \right| \exp(q^2/2)(q^2)^{-(\beta+1/2)/2} \tag{4.11}$$

if $\alpha^{(\pm)} \neq 0, -1, -2, \cdots$,

and

$$|\psi^{(\pm)}(q)| \underset{|q| \to \infty}{\sim} \left| \frac{\Gamma(2-\beta^{(\pm)})}{\Gamma(\alpha^{(\pm)} - \beta^{(\pm)} + 1)} \right| \times \exp(q^2/2)(q^2)^{-(\beta+1/2)/2} \tag{4.11'}$$

if $\alpha^{(\pm)} - \beta^{(\pm)} + 1 \neq 0, -1, -2, \cdots$.

As a consequence, $|\psi^{(\pm)}(q)|$ would have the wrong behavior for $|q| \to \infty$ unless

$$\begin{aligned} &\alpha^{(\pm)} = -k \quad \text{for (4.11)}, \\ &\alpha^{(\pm)} - \beta^{(\pm)} + 1 = -k \quad \text{for (4.11')}, \end{aligned} \tag{4.12}$$

where

$$k = 0, 1, 2, \cdots.$$

Conversely, if (4.12) holds, then the Φ's in (4.9) and (4.9′) will become polynomials in x of the kth degree, and hence the corresponding $|\psi^{(\pm)}(q)|$'s will rapidly tend to 0 as $|q| \to \infty$. The eigenvalues of E can then be obtained from (4.8) and (4.12). In this way we are led to the following results:

$$E = 2k \pm c + \tfrac{1}{2},$$

$$\psi^{(\pm)}(q) = \begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} |q|^{\pm c} \exp(-q^2/2) \Phi(-k, \pm c + \tfrac{1}{2}; q^2), \tag{4.13}$$

$$E = (2k+1) \mp c + \tfrac{1}{2},$$

$$\psi^{(\pm)}(q) = \begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} |q|^{1 \mp c} \exp(-q^2/2) \Phi(-k, \mp c + \tfrac{3}{2}; q^2), \tag{4.13'}$$

where $\begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix}$ implies that 1 and $\epsilon(q)$ are to be chosen for $\psi^{(+)}(q)$ and $\psi^{(-)}(q)$, respectively.

Here, c cannot be an arbitrary real constant, but is subject to the restriction $\beta^{(\pm)} \neq$ integer. Further restrictions will come out when we examine the behavior of $\psi^{(\pm)}(q)$ in the neighborhood of $q = 0$. It is required that $\psi^{(\pm)}(q)$ satisfy the conditions corresponding to (3.10) and (3.12),

$$\lim_{\mu \to 0} \int_{\mu}^{\lambda} dq \, |\psi^{(\pm)}(q)|^2 = \text{finite}, \tag{4.14}$$

and further

$$\begin{aligned} &\lim_{\mu \to 0} \int_{\mu}^{\lambda} dq \, |a\psi^{(\pm)}(q)|^2 = \text{finite}, \\ &\lim_{\mu \to 0} \int_{\mu}^{\lambda} dq \, |a^{\dagger}\psi^{(\pm)}(q)|^2 = \text{finite}, \end{aligned} \tag{4.15}$$

for $\lambda > \mu > 0$. To evaluate the contributions to (4.14) or (4.15) from the region $q \sim \mu$ we may put $\psi^{(\pm)}(q) \sim \begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} |q|^{\pm c}$ for (4.13) and $\psi^{(\pm)}(q) \sim \begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} |q|^{1 \mp c}$ for (4.13′). When this is done, (4.14) and (4.15) provide the following conditions:

$$\pm c > -\tfrac{1}{2} \quad \text{for (4.13)}, \tag{4.16}$$

$$\mp c > -\tfrac{1}{2} \quad \text{for (4.13')}. \tag{4.16'}$$

[Under the condition (4.15) we may also think of (4.13′) with $1 \mp 2c = 0$, which contradicts our premise $\beta^{(\pm)} \neq$ integer, however.]

Let us now extend the region of q into the one including $q = 0$, and regard (4.13) and (4.13′) as hyperfunctions of q defined for $-\infty < q < \infty$. We then have to examine whether these functions satisfy (4.3) in the extended region of q. Since we have already seen that this is the case with $q \neq 0$, we have only to examine the problem in the neighborhood of $q = 0$. Here we may make use of approximate expressions for $\psi^{(\pm)}(q)$. Since $\psi^{(\pm)}(q)$ are hyperfunctions, we rewrite (4.13), for example, as a sum of hyperfunctions such as

$$\begin{aligned} \psi^{(\pm)}(q) = &\begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} \exp(-q^2/2) |q|^{\pm c} + \frac{-k}{(\pm c + 1/2)} \begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} \\ &\times \exp(-q^2/2) |q|^{\pm c + 2} + \frac{(-k)(-k+1)}{(\pm c + 1/2)(\pm c + 3/2)} \\ &\times \begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} \exp(-q^2/2) |q|^{\pm c + 4} + \cdots, \end{aligned} \tag{4.17}$$

by use of (4.9), where it is to be understood that

$$\begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} \exp(-q^2/2) |q|^{\pm c + 2n} = \begin{cases} \sum_{r=0}^{\infty} \frac{(-1/2)^r}{r!} |q|^{c + 2n + 2r}, \\ \sum_{r=0}^{\infty} \frac{(-1/2)^r}{r!} \epsilon(q) |q|^{-c + 2n - 2r}. \end{cases} \tag{4.18}$$

In what follows, all hyperfunctions including (4.13′) are to be understood in a similar manner. Clearly, we do not have to invoke such expansions for $q \neq 0$. Thus, by applying the formulas (3.20) and (3.21), we obtain the following result: For $\psi^{(+)}(q)$ of (4.13) and $\psi^{(-)}(q)$ of (4.13′), equation (4.3) is satisfied under the same conditions as (4.16) and (4.16′), respectively, whereas for $\psi^{(-)}(q)$ of (4.13) and $\psi^{(+)}(q)$ of (4.13′) it becomes neces-

sary to impose a new restriction $c \neq 0$ in addition to (4.16) and (4.16′), respectively. Needless to say, in both cases we have, of course, the restriction $\beta^{(\pm)} = \pm c + \frac{1}{2} \neq$ integer.

Summarizing, we have the following two classes of eigenfunctions and eigenvalues:

(i) $c > -\frac{1}{2}$, $c \neq \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$,

$$\psi^{(+)}(q) = |q|^c \exp(-q^2/2)\Phi(-k, c+\tfrac{1}{2}; q^2) \quad (4.19)$$

with $E = \frac{1}{2} + c + 2k$,

$$\psi^{(-)}(q) = \epsilon(q)|q|^{1+c} \exp(-q^2/2)\Phi(-k, c+\tfrac{3}{2}; q^2) \quad (4.20)$$

with $E = \frac{1}{2} + c + (2k+1)$;

(ii) $c < \frac{1}{2}$, $c \neq 0, -\frac{1}{2}, -\frac{3}{2}, -\frac{5}{2}, \cdots$,

$$\psi^{(-)}(q) = \epsilon(q)|q|^{-c} \exp(-q^2/2)\Phi(-k, -c+\tfrac{1}{2}; q^2) \quad (4.21)$$

with $E = \frac{1}{2} - c + 2k$,

$$\psi^{(+)}(q) = |q|^{1-c} \exp(-q^2/2)\Phi(-k, -c+\tfrac{3}{2}; q^2) \quad (4.22)$$

with $E = \frac{1}{2} - c + (2k+1)$.

(2) *The case* $\beta^{(\pm)} = n+1$ $(n = 0, 1, 2, \cdots)$

The solution of (4.7), which is regular at the origin, is given by (4.9) but with $\beta^{(\pm)} = n+1$, i.e.,

$$\Phi(\alpha^{(\pm)}, n+1; x), \quad (4.23)$$

whereas the solution corresponding to the so-called logarithmic case is obtained by the Frobenius method as

$$\Phi(\alpha^{(\pm)}, n+1; x)\log x + x^{-n}\sum_{r=0}^{\infty} h_r^{(\pm)} x^r, \quad (4.23')$$

with $h_r^{(\pm)}$ being appropriately chosen constants. The $\psi^{(\pm)}(q)$ constructed from (4.23′) behaves for small $|q|$ like

$$\psi^{(\pm)}(q) \sim \begin{pmatrix} 1 \\ \epsilon(q) \end{pmatrix} |q|^{\pm c} (\log q^2 + h_0 q^{-2n}). \quad (4.24)$$

On the other hand, $\psi^{(\pm)}(q)$ must satisfy, for any positive integer m, the following conditions corresponding to a generalization of (4.15):

$$\lim_{\mu \to 0} \int_{\mu}^{\lambda} dq\, |a^m \psi^{(\pm)}(q)|^2 = \text{finite},$$
$$\lim_{\mu \to 0} \int_{\mu}^{\lambda} dq\, |a^{\dagger m} \psi^{(\pm)}(q)|^2 = \text{finite}, \quad (4.25)$$

for $\lambda > \mu > 0$. However, when (4.24) is substituted in the left-hand sides of (4.25) we find that the integrals do not become finite for large m. That is to say, from (4.23′) we cannot construct $\psi^{(\pm)}(q)$ which satisfies (4.25). Hence (4.23′) should be excluded.

Let us next examine (4.23). In this case we can argue in the same way as in (4.9) of (1), and find that $\alpha^{(\pm)} = -k$ with $k = 0, 1, 2, \cdots$. Further, $\beta^{(\pm)} = n+1$ gives $\pm c = n + \frac{1}{2}$. For any of such values of c it is easy to show that (4.14), (4.15), and (4.25) are all satisfied. In the same way as in (1) we can also show that $\psi^{(\pm)}(q)$, as a hyperfunction defined for $-\infty < q < \infty$, satisfies (4.3). The permissible eigenfunctions are thus classified into the following:

(i) $c = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$,

$$\psi^{(+)}(q) = |q|^c \exp(-q^2/2)\Phi(-k, c+\tfrac{1}{2}; q^2) \quad (4.26)$$

with $E = \frac{1}{2} + c + 2k$,

(ii) $c = -\frac{1}{2}, -\frac{3}{2}, -\frac{5}{2}, \cdots$,

$$\psi^{(-)}(q) = \epsilon(q)|q|^{-c} \exp(-q^2/2)\Phi(-k, -c+\tfrac{1}{2}; q^2) \quad (4.27)$$

with $E = \frac{1}{2} - c + 2k$.

(3) *The case* $\beta^{(\pm)} = -n$ $(n = 0, 1, 2, \cdots)$

One of the solutions of (4.7) is given by (4.9′) but with $\beta^{(\pm)} = -n$, i.e.,

$$x^{1+n}\Phi(\alpha^{(\pm)} + n + 1, n+2; x), \quad (4.28)$$

and the other, corresponding to the logarithmic case, is given by

$$x^{1+n}\Phi(\alpha^{(\pm)} + n + 1, n+2; x)\log x + \sum_{r=0}^{\infty} h_r^{(\pm)\prime} x^r. \quad (4.28')$$

However, the $\psi^{(\pm)}(q)$ constructed from (4.28′) does not satisfy (4.25), and therefore is excluded. Arguing in a way similar to the above we obtain from (4.28) the following eigenfunctions:

(i) $c = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$,

$$\psi^{(-)}(q) = \epsilon(q)|q|^{1+c} \exp(-q^2/2)\Phi(-k, c+\tfrac{3}{2}; q^2) \quad (4.29)$$

with $E = \frac{1}{2} + c + (2k+1)$,

(ii) $c = -\frac{1}{2}, -\frac{3}{2}, -\frac{5}{2}, \cdots$,

$$\psi^{(+)}(q) = |q|^{1-c} \exp(-q^2/2)\Phi(-k, -c+\tfrac{3}{2}; q^2) \quad (4.30)$$

with $E = \frac{1}{2} - c + (2k+1)$.

We have thus solved the eigenvalue problem completely. Writing the eigenfunction with eigenvalue E_n as $\psi_n(q)$ $(n = 0, 1, 2, \cdots)$ we can summarize the results of (1), (2), and (3) as follows:

(i) $c > -\frac{1}{2}$, $E_n = \frac{1}{2} + c + n$,

$$\psi_n(q) = (-1)^k \left(\frac{k!}{\Gamma(k+c+1/2)}\right)^{1/2} |q|^c \times \exp(-q^2/2) L_k^{c-1/2}(q^2) \quad \text{for } n = 2k,$$
$$\psi_n(q) = (-1)^k \left(\frac{k!}{\Gamma(k+c+3/2)}\right)^{1/2} \epsilon(q)|q|^{c+1} \times \exp(-q^2/2) L_k^{c+1/2}(q^2) \quad \text{for } n = 2k+1; \quad (4.31)$$

(ii) $c < \frac{1}{2}$, $c \neq 0$, $E_n = \frac{1}{2} - c + n$,

$$\psi_n(q) = (-1)^k \left(\frac{k!}{\Gamma(k-c+1/2)}\right)^{1/2} \epsilon(q)|q|^{-c} \times \exp(-q^2/2) L_k^{-c-1/2}(q^2) \quad \text{for } n = 2k,$$
$$\psi_n(q) = (-1)^k \left(\frac{k!}{\Gamma(k-c+3/2)}\right)^{1/2} |q|^{-c+1} \times \exp(-q^2/2) L_k^{-c+1/2}(q^2) \quad \text{for } n = 2k+1, \quad (4.32)$$

where $L_k^{\alpha}(x)$ denotes the generalized Laguerre polynomial defined by

$$L_k^\alpha(x) \equiv \binom{k+\alpha}{k} \Phi(-k, \alpha+1; x)$$
$$= \sum_{r=0}^{k} \binom{k+\alpha}{k-r} \frac{(-x)^r}{r!}. \tag{4.33}$$

By use of the orthogonality relation for generalized Laguerre polynomials

$$\int_0^\infty dx\, e^{-x} x^\alpha L_k^\alpha(x) L_{k'}^\alpha(x) = \delta_{kk'} \Gamma(\alpha+k+1)/k! \qquad (\alpha > -1), \tag{4.34}$$

we can prove that the $\psi_n(q)$'s form an orthonormalized system for either of the classes (i) and (ii). Using (4.1), (3.20), (3.21), and the formulas[13] for $L_k^\alpha(x)$,

$$\begin{aligned} &\frac{d}{dx} L_k^\alpha(x) = -L_{k-1}^{\alpha+1}(x), \\ &L_k^\alpha(x) - \frac{d}{dx} L_k^\alpha(x) = L_k^{\alpha+1}(x), \\ &x \frac{d}{dx} L_k^\alpha(x) + \alpha L_k^\alpha(x) = (k+\alpha) L_k^{\alpha-1}(x), \\ &x \frac{d}{dx} L_k^\alpha(x) + (\alpha - x) L_k^\alpha(x) = (k+1) L_{k+1}^{\alpha-1}(x), \end{aligned} \tag{4.35}$$

we can also prove[14] that the following relations hold in either of the classes (i) and (ii):

$$\begin{aligned} a\psi_n(q) &= \begin{cases} n^{1/2}\psi_{n-1}(q) & \text{for } n = \text{even}, \\ (2E_0 + n - 1)^{1/2}\psi_{n-1}(q) & \text{for } n = \text{odd}, \end{cases} \\ a^\dagger \psi_n(q) &= \begin{cases} (2E_0 + n)^{1/2}\psi_{n+1}(q) & \text{for } n = \text{even}, \\ (n+1)^{1/2}\psi_{n+1}(q) & \text{for } n = \text{odd}, \end{cases} \end{aligned} \tag{4.36}$$

in agreement with (1.3). The equivalence between the wave- and matrix-mechanical representations is now complete. Here the parameter E_0, the lowest eigenvalue of H, is given by

$$E_0 = \begin{cases} \frac{1}{2} + c & \text{for the class (i)}, \\ \frac{1}{2} - c & \text{for the class (ii)}. \end{cases} \tag{4.37}$$

Thus, we find $0 < E_0 < \infty$ for the class (i) and $0 < E_0 < \frac{1}{2}$ or $\frac{1}{2} < E_0 < \infty$ for the class (ii). It is to be noted that while the case of canonical quantization $E_0 = \frac{1}{2}$ is included only in the class (i), the cases of all other values of E_0 are included in either of the two calsses.[15] In this connection we recall that the situation was similar for the eigenvalue problem of p: The eigenstate of $p = -id/dq$ satisfying the canonical commutation relation $[q, p] = i$ is included only in the class (i).

The results obtained above are summarized as follows. For any given value of c we have solved the eigenvalue problems of p and H completely. For $c = 0$ or $c \geq \frac{1}{2}$ the eigenfunctions belong to the class (i), and for $c \leq -\frac{1}{2}$ they belong to the class (ii). For $-\frac{1}{2} < c < 0$ or $0 < c < \frac{1}{2}$ both classes of eigenfunctions become available. The last case is a very unusual situation and is discussed in detail under a separate heading.

We conclude the present section by making the following remark. In the foregoing we have been able to solve the eigenvalue problems only by invoking hyperfunctions If, on the contrary, we had restricted ourselves to ordinary analytic functions, the wave-mechanical representation would have been possible only for discrete values of c such that for the calss (i) c equals zero or a positive even integer, and for the class (ii) c equals a negative odd integer. In fact this is precisely what was done by Yang.[4] Further he was also led to conclude that the expansion theorem does not hold as far as $c \neq 0$: For example, wavefunctions that are nonvanishing at the origin, such as $\delta(q - q')$ with $q' = 0$, cannot be expanded in the eigenfunctions of H because the latter functions all vanish at the point.[16] In our opinion, however, the first principle of quantum mechanics does not seem to require that wavefunctions be analytic functions, but only that they be susceptible of the probabilistic interpretation. Accordingly, in solving eigenvalue problems we should not impose any other unnecessarily restrictive conditions on eigenfunctions. What should be done here instead is only to examine whether the wavefunctions thereby obtained are consistent with their inner products as well as with hermiticity of the basic operators q and p. This will be discussed in the following section.

V. THE RELATION BETWEEN THE TWO CLASSES OF EIGENFUNCTIONS AND HERMITICITY OF THE MOMENTUM OPERATOR

We now examine, for the case $-\frac{1}{2} < c < 0$ or $0 < c < \frac{1}{2}$, the relation between the two classes of eigenfunctions (i) and (ii), placing emphasis upon hermiticity of the operator p. Let us begin by noting some general features common to the eigenfunctions of p and of H obtained in the preceding sections: First the ranges of c are the same, and second the singular parts are of the same form. Here by singular parts we mean the following. Any wavefunction can be written as a sum of even and odd functions, and each of them in turn can be written as a product of analytic and nonanalytic functions. Such nonanalytic functions are called the singular parts. For example, as can be seen from (3.22), the singular parts of even functions in the space spanned by $\psi_{p'}(x)$'s of the class (i) are given by $|q|^c$. From (4.31) we also see that these are precisely the singular parts of even functions in the space spanned by $\psi_n(q)$'s of the class (i). Similarly, we see from (3.22), (3.23), (4.31), and (4.32) that for either of the classes (i) and (ii) $\psi_{p'}(q)$'s and $\psi_n(q)$'s have the singular parts of the same form as given in Table I.

From now on let us denote any wavefunction of the class (i) and that of the class (ii) by $\phi(q)$ and $\tilde{\phi}(q)$, re-

TABLE I. The singular parts of even and odd functions in the space of eigenfunctions. Any even or odd function belonging to the space can be written as a product of the analytic function $F(q^2)$ and the singular part given in this table.

	singular part	
region of c	even function	odd function
(i) $(-\frac{1}{2}, \infty)$	$\|q\|^c$	$\epsilon(q)\|q\|^{c+1}$
(ii) $(-\infty, 0)$ $(0, \frac{1}{2})$	$\|q\|^{-c+1}$	$\epsilon(q)\|q\|^{-c}$

sepctively. In general, these are, of course, given as hyperfunctions. For each of the classes we have so far been defining inner products as

$$(\phi_1, \phi_2) \equiv \int_{-\infty}^{\infty} dq\, \phi_1^*(q)\phi_2(q), \qquad (\tilde{\phi}_1, \tilde{\phi}_2) \equiv \int_{-\infty}^{\infty} dq\, \tilde{\phi}_1^*(q)\tilde{\phi}_2(q). \tag{5.1}$$[17]

For the case when c is in the range $-\frac{1}{2} < c < 0$ or $0 < c < \frac{1}{2}$, it may appear reasonable to define inner products between ϕ and $\tilde{\phi}$ by $(\phi, \tilde{\phi}) = \int_{-\infty}^{\infty} dq\, \phi^*(q)\tilde{\phi}(q)$. If this were possible, we could then define inner products for arbitrary wavefunctions $f(q)$ and $g(q)$ by

$$(f, g) \equiv \int_{-\infty}^{\infty} dq\, f^*(q)g(q), \tag{5.2}$$

irrespective of which class they belong to. The situation, however, is not so simple as this. As an example, let us take $f(q) = \psi_0(q)$ given by (4.31) and $g(q) = \psi_1(q)$ given by (4.32). The corresponding eigenvalues are $(\frac{1}{2} + c)$ and $(\frac{3}{2} - c)$ and are unequal to each other as far as $-\frac{1}{2} < c < 0$ or $0 < c < \frac{1}{2}$. We might therefore expect that the inner product (f, g) would be vanishing. A calculation shows the contrary, however:

$$\begin{aligned}(f, g) &= [\Gamma(c + \tfrac{1}{2})\Gamma(\tfrac{3}{2} - c)]^{-1/2} \int_{-\infty}^{\infty} dq\, |q| \exp(-q^2) \\ &\quad \times L_0^{c-1/2}(q^2) L_0^{c+1/2}(q^2) \\ &= 2\left(\frac{2\cos(\pi c)}{\pi(1-2c)}\right)^{1/2} \neq 0.\end{aligned} \tag{5.3}$$

Where does such a paradox originate from? To answer this question is then our last task in the present paper. At any rate this example at least indicates that to define inner products between ϕ and $\tilde{\phi}$ by (5.2) is not a legitimate procedure.

In the theory concerned the basic operators are chosen to be q and p, and these operators have been considered Hermitian. Generally speaking, however, hermiticity should be guaranteed, in connection with inner products, by $(f, qg) = (qf, g)$ and $(f, pg) = (pf, g)$. Stated otherwise, inner products should be so defined as to be consistent with hermiticity of q and p. When definition (5.2) is employed, hermiticity of q is obvious. The above result (5.3) thus implies that the inner product (5.2) taken between two wavefunctions f and g belonging to different classes does not make p Hermitian: For, if p is Hermitian, the two wavefunctions corresponding to different eigenvalues of $(p^2 + q^2)/2$ must be orthogonal to each other. In order to clarify the nature of what is happening here, let us therefore examine the question of under what conditions hermiticity of p is guaranteed.

From the results of Table I we see that the general forms of even and odd wavefunctions belonging to the class (i) or (ii) are such as given in the second row of Table II. The way in which the basic operators q and p act on such wavefunctions can then be easily found when (3.19) ~ (3.21) are taken into account, the results being given in the third and fourth rows of Table II, respectively. Now, in checking the relation $(f, pg) = (pf, g)$ we have only to consider the case in which one of the functions f and g is an even function and the other an odd function, because both sides of the above relation vanish when both f and g are even or odd functions.

(1) *The case* $f, g \in (i)$

Putting $f(q) = |q|^c F(q^2)$ and $g(q) = \epsilon(q)|q|^{c+1}G(q^2)$ we find from Table II, (3.25), and (3.20) that

$$\begin{aligned}D(f, g) &\equiv f^*(q)\cdot\{pg(q)\} - \{pf(q)\}^* \cdot g(q) \\ &= \frac{1}{i}\{(2c+1)|q|^{2c}F^*(q^2)G(q^2) + 2|q|^{2c+2} \\ &\quad \times [F^*(q^2)G'(q^2) + F^{*\prime}(q^2)G(q^2)]\} \\ &= \frac{1}{i}\frac{d}{dq}[\epsilon(q)|q|^{2c+1}F^*(q^2)G(q^2)],\end{aligned} \tag{5.4}$$

which guarantees hermiticity of p.

(2) *The case* $f, g \in (ii)$

Putting $f(q) = |q|^{-c+1}F(q^2)$ and $g(q) = \epsilon(q)|q|^{-c}G(q^2)$ we similarly obtain the following relation,

$$D(f, g) = \frac{1}{i}\frac{d}{dq}[\epsilon(q)|q|^{-2c+1}F^*(q^2)G(q^2)], \tag{5.5}$$

which guarantees hermiticity of p.

Next we turn to those cases in which one of the functions of f and g belongs to the class (i) and the other to the class (ii). Here the following two cases must be examined, where of course $|c| < \frac{1}{2}$.

(1′) *The case* $f \in (ii)$, $g \in (i)$

Putting $f(q) = |q|^{-c+1}F(q^2)$ and $g(q) = \epsilon(q)|q|^{c+1}G(q^2)$ we find that

$$D(f, g) = \frac{1}{i}\frac{d}{dq}[\epsilon(q)|q|^2 F^*(q^2)G(q^2)], \tag{5.6}$$

which is consistent with hermiticity of p.

(2′) *The case* $f \in (i)$, $g \in (ii)$

TABLE II. The general forms of even and odd wavefunctions belonging to the class (i) or (ii) and the effects of the basic operators q and p on such functions. Here $F(z)$ denotes a function analytic on the real axis.

	the class (i): $-\frac{1}{2} < c < \infty$		the class (ii): $-\infty < c < \frac{1}{2}$, $c \neq 0$	
$f(q)$	$\|q\|^c F(q^2)$	$\epsilon(q)\|q\|^{c+1}F(q^2)$	$\|q\|^{-c+1}F(q^2)$	$\epsilon(q)\|q\|^{-c}F(q^2)$
$qf(q)$	$\epsilon(q)\|q\|^{c+1}F(q^2)$	$\|q\|^c\{q^2F(q^2)\}$	$\epsilon(q)\|q\|^{-c}\{q^2F(q^2)\}$	$\|q\|^{-c+1}F(q^2)$
$pf(q)$	$\frac{2}{i}\epsilon(q)\|q\|^{c+1}F'(q^2)$	$\frac{1}{i}\|q\|^c\{2c+1)F(q^2) + 2q^2F'(q^2)\}$	$\frac{1}{i}\epsilon(q)\|q\|^{-c}\{(-2c+1)F(q^2) + 2q^2F'(q^2)\}$	$\frac{2}{i}\|q\|^{-c+1}F'(q^2)$

Putting $f(q)=|q|^c F(q^2)$ and $g(q)=\epsilon(q)|q|^{-c}G(q^2)$ we find this time that

$$D(f,g)=\frac{2}{i}|q|[F^{*\prime}(q^2)G(q^2)+F^*(q^2)G'(q^2)]$$

$$=\frac{1}{i}\epsilon(q)\frac{d}{dq}[F^*(q^2)G(q^2)], \tag{5.7}$$

which results in

$$(f,pg)-(pf,g)=\frac{1}{i}\int_{-\infty}^{\infty}dq\,\epsilon(q)\frac{d}{dq}[F^*(q^2)G(q^2)]$$

$$=2iF^*(0)G(0). \tag{5.8}$$

This implies that hermiticity of p is violated because $F^*(0)G(0)$ does not always vanish.

We have thus found that inner products (5.1) within either of the classes (i) and (ii) guarantee hermiticity of q and p, and hence that of H, whereas this is not the case with (5.2) for which f and g belong to different classes. The origin of the paradox encountered above can thus be attributed to the latter situation.

We are led in this way to the consistent definition of inner products:

$$(f,g)=\begin{cases}\int_{-\infty}^{\infty}dq\,f^*(q)g(q) & \text{for both } f \text{ and } g\in \text{(i)} \\ & \quad \text{or } \in \text{(ii)}, \\ 0 & \text{for } f\in \text{(i) and } g\in \text{(ii)} \\ & \quad \text{or } f\in \text{(ii) and } g\in \text{(i)}.\end{cases} \tag{5.9}$$

The latter part of this definition does not give rise to any difficulty: As can be seen from Table II, the singular parts are transformed only within the same class under the action of q and p, or in other words, the two classes are separated by a superselection rule. For this reason it would be nonsensical to consider a superposition of wavefunctions belonging to the two different classes. We may therefore say that even when the parameter c takes a common value the eigenfunctions belonging to the class (i) and those belonging to the class (ii) span the different Hilbert spaces. In fact, it has already been shown in (3.30) that the eigenfunctions of p form a complete set of basis functions within either of the classes (i) and (ii). The situation is the same for the eigenfunctions of H.

In connection with this let us make a few remarks. In either of the classes (i) and (ii), the L_2 space spanned by the eigenfunctions $\psi_n(q)$ of H is contained in the one spanned by the eigenfunctions $\psi_{p'}(q)$ of p, and any $\psi_n(q)$ can be expanded in the $\psi_{p'}(q)$'s within the same class. In fact, by using the formula[18]

$$\tfrac{1}{2}\int_{-\infty}^{\infty}dq[|p'q|^{1/2}J_\alpha(|p'q|)][|q|^{\alpha+1/2}\exp(-q^2/2)L_k^\alpha(q^2)]$$

$$=(-1)^k\exp(-p'^2/2)|p'|^{\alpha+1/2}L_k^\alpha(p'^2)\quad(\alpha>-1), \tag{5.10}$$

we can easily compute $(\psi_{p'},\psi_n)$,

$$(\psi_{p'},\psi_n)=A\psi_n(p'), \tag{5.11}$$

where $A=(-1)^k$ and $-i(-1)^k$ for $\psi_n(q)$ given by the first and the second equations in (4.31), respectively, and $A=(-1)^k$ and $i(-1)^k$ for $\psi_n(q)$ given by the first and second equations in (4.32), respectively. Performing an interchange $q\rightleftarrows p'$ in (5.10) and using (5.11) we immediately find

$$\psi_n(q)=\int_{-\infty}^{\infty}dp'(\psi_{p'},\psi_n)\psi_{p'}(q). \tag{5.12}$$

When the completeness relation (3.30) for $\psi_{p'}(q)$'s is considered, it is realized that (5.12) is the equation which holds true only when the space spanned by $\psi_n(q)$'s is contained in the one spanned by $\psi_{p'}(q)$'s.

There still remains a question as to whether we can conversely expand $\psi_{p'}(q)$ in $\psi_n(q)$'s. Since $\psi_{p'}(q)$ does not tend to zero as $|q|\to\infty$ whereas $\psi_n(q)$ does so, it may appear at first sight that such an expansion is not possible. Let us now examine the situation a little further. We start with a formula which holds in the sense of uniform and absolute convergence[19]:

$$\sum_{k=0}^{\infty}\frac{k!(-t)^k}{\Gamma(k+\alpha+1)}[|q|^{\alpha+1/2}\exp(-q^2/2)L_k^\alpha(q^2)]$$

$$\times[|p'|^{\alpha+1/2}\exp(-p'^2/2)L_k^\alpha(p'^2)]$$

$$=\frac{1}{(1+t)t^{\alpha/2}}\exp\left(-\frac{1-t}{2(1+t)}(p'^2+q^2)\right)$$

$$\times|p'q|^{1/2}J_\alpha\left(\frac{2t^{1/2}}{1+t}|p'q|\right) \tag{5.13}$$

$$(-1<t<1,\ \alpha>-1),$$

and substitute (5.10) in the second bracket on the left-hand side of (5.13), thereby obtaining

$$\tfrac{1}{2}\sum_{k=0}^{\infty}\left(\frac{k!t^k}{\Gamma(k+\alpha+1)}[|q|^{\alpha+1/2}\exp(-q^2/2)L_k^\alpha(q^2)]\right.$$

$$\times\left\{\int_{-\infty}^{\infty}dq[|q|^{\alpha+1/2}\exp(-q^2/2)L_k^\alpha(q^2)]\right.$$

$$\left.\left.\times[|p'q|^{1/2}J_\alpha(|p'q|)]\right\}\right)$$

$$=\frac{1}{(1+t)t^{\alpha/2}}\exp\left(-\frac{1-t}{2(1+t)}(p'^2+q^2)\right)$$

$$\times|p'q|^{1/2}J_\alpha\left(\frac{2t^{1/2}}{1+t}|p'q|\right). \tag{5.14}$$

By using (3.22), (3.23), (4.31), and (4.32) we thus find that for either of the classes (i) and (ii)

$$\sum_{n=0}^{\infty}(\psi_n,\psi_{p'})\psi_n(q)t^{[n/2]}$$

$$=\frac{2}{(1+t)t^{\alpha/2}}\exp\left(-\frac{1-t}{2(1+t)}(p'^2+q^2)\right)$$

$$\times\psi_{p'}\left(\frac{2t^{1/2}}{1+t}q\right)\quad(0<t<1), \tag{5.15}$$

that is,

$$\lim_{t\to1-0}\left(\sum_{n=0}^{\infty}(\psi_n,\psi_{p'})\psi_n(q)t^{[n/2]}\right)=\psi_{p'}(q). \tag{5.16}$$

If we carry out the limit for each term separately and put $t=1$, then the uniform convergence of the series for $-\infty<q<\infty$, namely

$$|\psi_{p'}(q)-\sum_{n=0}^{N}(\psi_n,\psi_{p'})\psi_n(q)|<\mu$$

for any q will be spoilt. This may easily be understood by observing the difference between the behavior of $\psi_n(q)$ and that of $\psi_{p'}(q)$ for $|q|\to\infty$. Thus, all we can say about our problem is as follows: The formal expansion

of $\psi_{p'}(q)$ in the $\psi_n(q)$'s, i.e., $\sum_{n=0}^{\infty}(\psi_n, \psi_{p'})\psi_n(q)$, does not uniformly converge, and the expansion theorem holds only in the sense of (5.15) or Abel summability (5.16).

It can be seen further that the replacement $t \to -t$ and $p' - q'$ in (5.13) gives

$$\sum_{k=0}^{\infty} \frac{k!\, t^k}{\Gamma(k+\alpha+1)} [|q|^{\alpha+1/2} \exp(-q^2/2) L_k^{\alpha}(q^2)]$$

$$\times [|q'|^{\alpha+1/2} \exp(-q'^2/2) L_k^{\alpha}(q'^2)]$$

$$= \frac{1}{(1-t)t^{\alpha/2}} \exp\left(-\frac{1+t}{2(1-t)}(q^2+q'^2)\right)$$

$$\times |qq'|^{1/2} I_{\alpha}\left(\frac{2t^{1/2}}{1-t}|qq'|\right) \quad (-1 < t < 1). \qquad (5.17)$$

Using the asymptotic form of $I_{\alpha}(z)$ for $z \to \infty$ which is given immediately after (3.26) we obtain the following relation for either of the classes (i) and (ii),

$$\lim_{t \to 1-0}\left(\sum_{n=0}^{\infty} \psi_n(q)\psi_n^*(q')t^n\right) = \delta(q-q'). \qquad (5.18)$$

The relation (3.30) implies that any eigenfunction $\delta(q-q')$ of q can be expressed as a superposition of the eigenfunction of p. In the beginning of Sec. II we have started by assuming that the spectrum of q ranges continuously from $-\infty$ to ∞, i.e., $-\infty < q' < \infty$. The above results thus makes it clear that this assumption in fact has been a consistent one. As repeatedly stated, for the case $-\frac{1}{2} < c < 0$ or $0 < c < \frac{1}{2}$ the expansion in the energy eigenfunctions of one and the same eigenfunction $\delta(q-q')$ can be made in two different ways.

The existence of the two classes of eigenfunctions, (i) and (ii), has been revealed by working in the wave-mechanical representation for q and p. The operator calculus of matrix mechanics, on the other hand, is too formal and straightforward to exhibit such an intricate structure of the theory, although of course nothing is wrong with this method. The existence of the two classes of wavefunctions thus implies that ket vectors also have a corresponding structure, that is, there exist two kinds of ket vectors to be denoted by $|\ \rangle_{(\mathrm{i})}$ and $|\ \rangle_{(\mathrm{ii})}$. In view of (5.9) their inner products must, of course, be assumed to be ${}_{(\mathrm{i})}\langle\ |\ \rangle_{(\mathrm{ii})} = 0$, and the two spaces spanned by $|\ \rangle_{(\mathrm{i})}$'s and by $|\ \rangle_{(\mathrm{ii})}$'s form the respective superselection sectors. This explains the reason why in the matrix-mechanical treatment assuming the existence of the unique vacuum the two spaces make no appearance simultaneously.

VI. CONCLUDING REMARKS

We have seen above that in the wave-mechanical representation of a Bose-like oscillator, wavefunctions can be obtained completely by invoking the use of hyperfunctions and that only in this way can the equivalence with the matrix-mechanical representation be restored. The use of hyperfunctions, of course, is nothing new in quantum mechanics: We usually employ the δ functions as eigenfunctions of operators such as q. However, we shall argue in what follows that the present case possesses a characteristically new feature.

Generally speaking, a hyperfunction encountered in quantum mechanics is of the following two-sided aspect: It may be interpreted either as a hyperfunction of the eigenvalue concerned when the variable q is kept fixed, or as that of q when the eigenvalue is kept fixed. Of course, the former applies only for the case of continuous spectra. As for the eigenfunction $\delta(q-q')$ either interpretation is applicable. In this case, however, it is possible to form a square-integrable wavefunction by superposing the eigenfunctions $\delta(q-q')$ with q' ranging over a certain finite interval, the wavefunction being thereby reduced to an ordinary function. In this way the explicit use of hyperfunctions may be avoided if so desired.

In contrast with this we should say that the wavefunctions obtained in the preceding sections are, essentially, hyperfunctions of the variable q. Indeed, our Hamiltonian H has discrete eigenvalues, and the corresponding eigenfunctions are square-integrable by themselves. Hence it is not possible to remove the singular parts, with respect to q, of the eigenfunctions by the method applicable to the case of $\delta(q-q')$. The situation is essentially the same for the eigenfunctions of the momentum operator p: Even when a square-integrable wavefunction is formed by a superposition of the eigenfunctions (3.22) and (3.23), the singular parts still survive in general cases. This is the point where the hyperfunctions encountered in the present case are basically different from those of the usual case including $\delta(q-q')$.

In ordinary quantum mechanics there exist very few cases in which hyperfunctions must be employed for square-integrable wavefunctions (an example will be given below). We should say, however, that such a situation is merely an accidental one, and it is impossible, in fact, to deduce on some physical grounds that wavefunctions shall be restricted to those functions which are nonsingular with respect to q. Thus, the necessary and sufficient conditions for $\psi_i(q)$'s to be wavefunctions may be stated as follows: (I) The product of $\psi_i^*(q)$ and $\psi_j(q)$ can be defined, and the inner product $\int_{-\infty}^{\infty} dq \times \psi_i^*(q)\psi_j(q)$ exists; (II) the action of the basic operators q and p on wavefunctions is well defined, and the resulting functions can also be regarded as wavefunctions, where hermiticity of q and p must of course be guaranteed in connection with the inner product considered in (I). It is to be noticed here that (I) implies $\int_{\mu}^{\lambda} dq\, |\psi(q)|^2 < \infty$ for arbitrary μ and λ with $\mu < \lambda$, thereby enabling us to adopt the probabilistic interpretation.

Contrary to the case of ordinary functions a hyperfunction does not always have the functional value for a given value of q. Such a property, however, does not contradict the above conditions (I) and (II). What is necessary to do here is to define products of the hyperfunctions concerned in a proper way: They must be consistent with all operations that arise within the framework of a given theory. In the preceding sections we have faithfully carried out such a program in accordance with the conditions (I) and (II) for the specific model.

Before concluding the present paper let us give an example in ordinary quantum mechanics that necessitates the use of hyperfunctions, i.e., the eigenvalue problem of the dilatation operator,

$$D \equiv \tfrac{1}{2}\{q, p\} = \frac{1}{i}\left(q\frac{d}{dq} + \frac{1}{2}\right). \tag{6.1}$$

The problem can be solved completely only in terms of hyperfunctions, and the even and odd eigenfunctions with eigenvalue λ are given, respectively, by

$$\psi_\lambda^{(+)}(q) = (4\pi)^{-1/2}|q|^{-1/2+\lambda i},$$
$$\psi_\lambda^{(-)}(q) = (4\pi)^{-1/2}\epsilon(q)|q|^{-1/2+\lambda i} \tag{6.2}$$
$$(-\infty < \lambda < \infty).$$

Here, the action of the operators q and d/dq on these functions is defined by the same formulas as (3.19) and (3.20) but with α being replaced by a complex number $-\frac{1}{2} + \lambda i$. As for products of these functions we also use the previous formulas (3.25) with α and β being replaced by complex numbers. It is then easy to check that the above eigenfunctions $\psi_\lambda^{(\pm)}(q)$ form a complete, orthonormal system:

$$\int_{-\infty}^{\infty} dq\, \psi_\lambda^{(\pm)*}(q)\psi_{\lambda'}^{(\pm)}(q) = \delta(\lambda - \lambda'),$$
$$\int_{-\infty}^{\infty} dq\, \psi_\lambda^{(\pm)*}(q)\psi_{\lambda'}^{(\mp)}(q) = 0, \tag{6.3}$$
$$\int_{-\infty}^{\infty} d\lambda\{\psi_\lambda^{(+)}(q)\psi_\lambda^{(+)*}(q') + \psi_\lambda^{(-)}(q)\psi_\lambda^{(-)*}(q')\} = \delta(q - q').$$

It is to be noticed that even a square-integrable wavefunction formed by superposing the eigenfunctions with λ ranging over a finite interval does not always become a nonsingular function. In other words, such square-integrable wavefunctions must also be treated as hyperfunctions of q.

ACKNOWLEDGMENTS

The authors would like to thank Professor N. Nakanishi of Kyoto University for valuable comments on the original manuscript.

[1]E. P. Wigner, Phys. Rev. **77**, 711 (1950).
[2]V. Bargmann, Ann. Math. **48**, 568 (1947).
[3]H. S. Green, Phys. Rev. **90**, 270 (1953).
[4]L. M. Yang, Phys. Rev. **84**, 788 (1951).
[5]Throughout the present paper we shall use the terminology "hyperfunctions" instead of generalized functions or distributions.
[6]See, for example, L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1950, 1951), Vols. 1 and 2; I. M. Gel'fand and I. M. Silov, *Generalized Functions* (Moscow, 1958), Vols. 1, 2, and 3; M. Sato, Proc. Jpn Acad. **34**, 126, 604 (1958); M. J. Lighthill, *An Introduction to Fourier Analysis and Generalized Functions* (Cambridge U. P., Cambridge, 1958).
[7]For the case $-\frac{1}{2} < \alpha < 0$, for example, the application of the usual rule of differentiation $d(\epsilon(x)|x|^\alpha)/dx = d\epsilon(x)/dx \cdot |x|^\alpha + \epsilon(x)\cdot d|x|^\alpha/dx$ results in a contradiction with (3.20) below.
[8]$|x|^0 = 1$, $\epsilon(x)|x|^0 = \epsilon(x)$.
[9]See, for example, M. J. Lighthill, Ref. 6.
[10]See, for example, M. Sato, Ref. 6.
[11]Restated in terms of the theory of hyperfunctions, the products prescribed by (3.21) are the so-called formal products of $1/x$ and the hyperfunctions $|x|^\alpha$ and $\epsilon(x)|x|^\alpha$, whose generating functions are given, respectively, by $\{z^\alpha - (-z)^\alpha\}/(2i\sin\pi\alpha)$ and $i\{z^\alpha + (-z)^\alpha\}/(2\sin\pi\alpha)$ (or by the limit $\alpha \to n$ of these functions for the cases $\alpha = n$ with $n = 1, 2, \cdots$). Here z^α and $(-z)^\alpha$ are to be understood, respectively, as $\exp(\alpha\,\mathrm{Log}z)$ and $\exp[\alpha\,\mathrm{Log}(-z)]$ with $\mathrm{Log}z$ denoting the principal value of logarithm with a cut on the negative real axis. Incidentally, the relation between a hyperfunction $f(x)$ and its generating function $F(z)$ is given by $f(x) = \mathrm{H.F.}\,[F(z)] \equiv F(x + i0) - F(x - i0)$, and the formal product of $1/x$ and $f(x)$ is defined as $(1/x)\cdot f(x) \equiv \mathrm{H.F.}\,[(1/z)\cdot F(z)]$.
[12]A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vols. 1, 2, and 3, use Eq. (25), p. 50, Vol. 2.
[13]Derived from Ref. 12: Sec. 6.1, Vol. 1 and Sec. 10.12, Vol. 2.
[14]For $q \neq 0$ we use (4.35), whereas for $q \simeq 0$ we express $\psi_n(q)$ as a sum of hyperfunctions in a way similar to (4.17) and then use (3.20) and (3.21).
[15]In the case $E_0 = 0$, $a^\dagger\psi = a\psi = 0$ or $q\psi = p\psi = 0$ and hence $q = p = 0$. This case has been excluded from our discussions, because we have started from the assumption $-\infty < q' < \infty$.
[16]Such an argument seems to be too naive since $\delta(q - q')$ is a hyperfunction. The expansion theorem can in fact be shown to be true: cf. (3.30), the statement made thereafter, and (5.18) below.
[17]For the eigenfunctions of p it is to be understood that a damping factor $e^{-\xi q^2}$ is included in the integrals on the right-hand sides; cf. (3.24).
[18]A. Erdélyi, Ed., *Tables of Integral Transforms* (McGraw-Hill, New York, 1953), Vols. 1 and 2, use Eq. (4), p. 43, Vol. 2.
[19]A modification of Eq. (20), p. 189, Vol. 2, Ref. 18.

The three-dimensional convolution of reduced Bessel functions and other functions of physical interest

E. Filter and E. O. Steinborn

Institut für Chemie, Universität Regensburg, D—8400 Regensburg, West Germany
(Received 31 May 1977)

A method for evaluating convolution integrals over rather general functions is suggested, based on the analytical evaluation of convolution integrals over functions

$$B^M_{\nu,L}(\mathbf{r}) = (2/\pi)^{1/2} r^{L+\nu} K_\nu(r)\, Y^M_L(\theta,\phi),$$

which are products of modified Bessel functions of the second kind $K_\nu(r)$, regular solid spherical harmonics $r^L Y^M_L(\theta,\phi)$, and powers r^ν.

1. INTRODUCTION

In many branches of mathematical physics one encounters the problem of how to evaluate three-dimensional convolution integrals

$$\Phi(\mathrm{R}) = \int d\mathbf{r}\, g(\mathbf{r}) f(\mathrm{R}-\mathbf{r}) \tag{1.1}$$

with given functions g and f.

A common method for evaluating these integrals is based upon the Fourier transform convolution theorem[1–4] which states that the Fourier transforms $\bar{\Phi}$, $\bar{g}$, and $\bar{f}$ are related by

$$\bar{\Phi}(\mathbf{k}) = \bar{g}(\mathbf{k})\bar{f}(\mathbf{k})(2\pi)^{3/2}. \tag{1.2}$$

This relationship allows the transformation of the *two-centric* convolution integral $\Phi(\mathrm{R})$ of Eq. (1.1) into the *one-centric* Fourier integral

$$\Phi(\mathrm{R}) = (2\pi)^{-3/2} \int d\mathbf{k} \exp(i\mathbf{k}\cdot\mathrm{R})\bar{\Phi}(\mathbf{k}) = \int d\mathbf{k} \exp(i\mathbf{k}\cdot\mathrm{R})\bar{g}(\mathbf{k})\bar{f}(\mathbf{k}) \tag{1.3}$$

which usually is easier to calculate. This method, however, requires the evaluation of three integrals, namely $\bar{f}$, $\bar{g}$, and the integral of Eq. (1.3). Often this is rather cumbersome or even impossible.

In the present article we shall suggest another method of evaluating convolution integrals $\Phi(\mathrm{R})$. This method is easy to handle and leads to simple results in many cases. It is based on some properties of the so-called reduced Bessel functions (RBF's) which so far do not seem to have been recognized properly. It will be shown in this paper: First, the convolution product of two RBF's is simple, in fact it is given by a combination of some other RBF's. Second, many functions can be expanded in terms of RBF's. Therefore, also for these rather general functions the convolution integrals can be obtained analytically in terms of RBF's, making any further integration superfluous.

In addition to the mathematical aspects discussed so far, there are strong physical reasons for investigating the integral properties of RBF's. The RBF's are closely connected with Slater-type orbitals (STO's) which play an important role in molecular theory.

The RBF's are of physical interest not only because of their peculiar integral properties, but also because of their local behavior, which allows the construction of physically meaningful wavefunctions. Shavitt suggested that the RBF's may be used as the radial part of a generalized type of orbital in atomic and molecular calculations.[5] Recent calculations of the ground state energies of small molecules have shown that in fact the RBF's represent a good basis set for LCAO-MO (linear combination of atomic orbitals-molecular orbitals) calculations.[6–8] This is another reason for investigating molecular integrals over RBF's, which are always connected with convolution integrals.

2. DEFINITIONS AND BASIC FORMULAS

The following definitions and formulas will be used.

We define a function $E^M_{\nu,L}(\alpha\mathbf{r})$ by

$$B^M_{\nu,L}(\alpha\mathbf{r}) = (\alpha r)^L \hat{k}_\nu(\alpha r) Y^M_L(\Omega_\mathbf{r}) = \hat{k}_\nu(\alpha r)\mathcal{Y}^M_L(\alpha\mathbf{r}). \tag{2.1}$$

The order ν is arbitrary. For half-integer order $\nu = N - \frac{1}{2}$, it is advantageous to define

$$\mathcal{E}^M_{N-1/2,L}(\alpha\mathbf{r}) = [2^{N+L}(N+L)!\,]^{-1} \times (\alpha r)^L \hat{k}_{N-1/2}(\alpha r) Y^M_L(\Omega_\mathbf{r}). \tag{2.2}$$

The anisotropy is expressed by the surface spherical harmonic $Y^M_L(\Omega_\mathbf{r})$ using Condon—Shortley phases.[9] The function $\mathcal{B}^M_{\nu,L}(\alpha\mathbf{r})$ is a product of a regular solid spherical harmonic $\mathcal{Y}^M_L(\alpha\mathbf{r}) = (\alpha r)^L Y^M_L(\theta,\phi)$ and a reduced Bessel function $\hat{k}_\nu(\alpha r)$ of arbitrary order which is defined by

$$\hat{k}_\nu(\alpha r) = (2/\pi)^{1/2}(\alpha r)^\nu K_\nu(\alpha r), \tag{2.3}$$

where $K_\nu(\alpha r)$ is a modified Bessel function of the second kind.[10] Here, α denotes a scaling parameter. $\hat{k}_\nu$ should not be confused with the spherical Bessel function k_l, which for integer l is defined by

$$k_l(\alpha r) = (2/\pi)^{1/2}(\alpha r)^{-1/2} K_{l+1/2}(\alpha r). \tag{2.4}$$

Certain properties of the reduced Bessel function are of special interest. The function $\hat{k}_\nu(x)$ is finite at the origin if $\nu \geq 0$ and increases like $x^{-2\nu}$ for $x \to 0$ if $\nu < 0$, while for $x \to \infty$ it decreases exponentially. For half-integer order ν, one has

$$\hat{k}_{-1/2}(x) = x^{-1}e^{-x}, \tag{2.5}$$

$$\hat{k}_{N-1/2}(x) = x^{-1}e^{-x} \sum_{p=1}^{N} \frac{(2N-p-1)!}{(p-1)!\,(N-p)!}\, 2^{p-N} x^p, \tag{2.6}$$

 0022-2488/78/1901-0079$1.00

for natural numbers $N=1,2,\cdots$. The reduced Bessel function obeys the relationship[11]

$$D_x^N \hat{k}_\nu(x) = (-1)^N \hat{k}_{\nu-N}(x) \tag{2.7}$$

where the "Bessel operator" is defined by

$$D_x^N = (x^{-1}\, d/dx)^N. \tag{2.8}$$

The following recursion formula[12] holds for all ν

$$x^2 \hat{k}_{\nu-1/2}(x) = \hat{k}_{\nu+3/2}(x) - (2\nu+1)\hat{k}_{\nu+1/2}(x). \tag{2.9}$$

If $\Gamma(x)$ denotes the gamma function,[13] the Pochhammer symbol[14] $(a)_N$ stands for

$$(a)_N = a(a+1)\cdots(a+N-1) = \Gamma(a+N)/\Gamma(a), \quad (a)_0 = 1. \tag{2.10}$$

Use will be made of the Jacobi polynomials, which can be represented by the following series expansion[15]

$$P_N^{(\mu,\nu)}(x) = \frac{\Gamma(\mu+N+1)}{\Gamma(\mu+\nu+N+1)} \cdot \sum_{p=0}^{N} \frac{\Gamma(\mu+\nu+N+p+1)}{p!\,(N-p)!\,\Gamma(\mu+p+1)} \times \left(\frac{x-1}{2}\right)^p. \tag{2.11}$$

Because[16]

$$P_N^{(\mu,\nu)}(-1) = (-1)^N \frac{\Gamma(N+\nu+1)}{N!\,\Gamma(\nu+1)} = (-1)^N \binom{N+\nu}{N}, \tag{2.12}$$

Eq. (2.11) becomes Vandermonde's theorem[17] for $x=-1$.

For the integral over the product of three spherical harmonics, which is the Gaunt coefficient[18] and may be expressed by Clebsch–Gordan coefficients[19] as indicated, the Dirac notation is used,

$$\begin{aligned}\langle L_1 M_1 | L_2 M_2 | lm\rangle &= \int d\Omega\, Y_{L_1}^{M_1*}(\Omega) Y_{L_2}^{M_2}(\Omega) Y_l^m(\Omega) \\ &= [(2l+1)(2L_2+1)]^{1/2}[4\pi(2L_1+1)]^{-1/2} \\ &\quad \times C(l, L_2, L_1; 0, 0) C(l, L_2, L_1; m, M_2).\end{aligned} \tag{2.13}$$

The angular momentum quantum numbers L_1, L_2, and l often appear in certain combinations for which we introduce the following abbreviations, where Δ should remind us of the triangular condition:

$$\Delta l = (L_1 + L_2 - l)/2, \quad \Delta L_1 = (L_2 + l - L_1)/2,$$
$$\Delta L_2 = (l + L_1 - L_2)/2, \tag{2.14}$$

$$\sigma(L_1) = \sigma(L_2) = \sigma(l) = (L_1 + L_2 + l)/2 = \sigma. \tag{2.15}$$

The double factorial function is defined by

$$(2n)!! = 2\cdot 4\cdot 6\cdots(2n) = 2^n n!, \tag{2.16}$$

$$(2n+1)!! = 1\cdot 3\cdot 5\cdots(2n+1) = (2^n n!)^{-1}(2n+1)!, \tag{2.17}$$

$$0!! = (-1)!! = 1!! = 1. \tag{2.18}$$

3. CONVOLUTION OF REDUCED BESSEL FUNCTIONS WITH UNEQUAL SCALING PARAMETERS

In this section it will be shown that the convolution product of the functions $\beta_{N_1-1/2,L_1}^{M_1}(\alpha\mathbf{r})$ and $\beta_{N_2-1/2,L_2}^{M_2}(\beta\mathbf{r})$ with $N_1, N_2 = 0, 1, 2, \cdots$ and with different scaling parameters α and β leads to the following formula:

$$\begin{aligned}\int d\mathbf{r}\beta_{N_1-1/2,L_1}^{M_1*}(\alpha\mathbf{r})\beta_{N_2-1/2,L_2}^{M_2}[\beta(\mathbf{R}-\mathbf{r})] &= \sum_l \langle L_2 M_2 | L_1 M_1 | lm\rangle \\ &\times \left\{ \sum_{t=0}^{N_1+L_1} C_{L_2 L_1 l}^{N_2 N_1 t}(\beta,\alpha)\beta_{t-l-1/2,t}^{m}(\alpha\mathbf{R}) \right. \\ &\left. + \sum_{t=0}^{N_2+L_2} C_{L_1 L_2 l}^{N_1 N_2 t}(\alpha,\beta)\beta_{t-l-1/2,t}^{m}(\beta\mathbf{R}) \right\}.\end{aligned} \tag{3.1}$$

Hence, the resulting linear combination contains only functions $\beta_{m+1/2,t}^{m}$, i.e., functions of the same type as those contained in the integrand of the convolution integral. The coefficients

$$C_{\lambda_1\lambda_2 l}^{\nu_1\nu_2 t}(\alpha,\beta) = (-1)^{\nu_2+\lambda_2+t} 4\pi\alpha^{-3}\left(\frac{\beta}{\alpha}\right)^{\lambda_1}\left[1-\left(\frac{\beta}{\alpha}\right)^2\right]^{-\nu_1-\lambda_1-1} \times P_{\nu_2+\lambda_2-t}^{(-\nu_2-\Delta\lambda_1+t,\,\nu_1+\Delta\lambda_2)}\left(\frac{(\alpha/\beta)^2+1}{(\alpha/\beta)^2-1}\right) \tag{3.2}$$

are related to the Jacobi polynomials defined by Eq. (2.11). The expansion Eq. (3.1) is finite: The summation index l is restricted to the values allowed by the triangular condition

$$|L_1 - L_2| \le l \le L_1 + L_2. \tag{3.3}$$

Due to the Gaunt coefficient one has

$$m = M_2 - M_1. \tag{3.4}$$

In Eq. (3.1) it is not possible to set $\alpha=\beta$ due to the occurrence of singularities in this case. Although these singularities cancel each other, the compact formula for the case $\alpha=\beta$ will be derived in a different manner in the next section.

For the derivation of Eq. (3.1), we consider the following integral:

$$S_{N_1 N_2}^{L_1 L_2}(\alpha,\beta,\mathbf{R}) = \int d\mathbf{r}\beta_{N_1-1/2,L_1}^{M_1*}(\alpha\mathbf{r})\beta_{N_2-1/2,L_2}^{M_2}[\beta(\mathbf{R}-\mathbf{r})]. \tag{3.5}$$

With the help of Eq. (2.7) we obtain for $N_1=0$

$$\begin{aligned}S_{0,N_2}^{L_1 L_2}(\alpha,\beta,\mathbf{R}) &= (-2)^{-N_2-L_2}[(N_2+L_2)!]^{-1}\beta^{2N_2+L_2-1} \\ &\times \int d\mathbf{r}\beta_{-1/2,L_1}^{M_1*}(\alpha\mathbf{r}) D_\beta^{N_2+L_2}\beta^{L_2+1} \\ &\times k_{L_2}(\beta|\mathbf{R}-\mathbf{r}|)\, Y_{L_2}^{M_2}(\Omega_{\mathbf{R}-\mathbf{r}}).\end{aligned} \tag{3.6}$$

The product of the spherical Bessel function k_{L_2} [as defined by Eq. (2.4)] and the spherical harmonic $Y_{L_2}^{M_2}$ is called a modified Helmholtz harmonic.[20,21] The addition theorem for this function is given by[20,21]

$$\begin{aligned}k_{L_2}[\beta|\mathbf{R}-\mathbf{r}|]\cdot Y_{L_2}^{M_2}(\Omega_{\mathbf{R}-\mathbf{r}}) &= 4\pi\beta^{-1}\cdot\sum_{l_1 l_2}\sum_{m_1 m_2}\langle L_2 M_2 | l_1 m_1 | l_2 m_2\rangle \\ &\times (\gamma R)^{-1/2}\cdot(-1)^{l_<+l_2-L_2} \\ &\times I_{l_<+1/2}(\beta\gamma_<)\cdot K_{l_>+1/2}(\beta\gamma_>) Y_{l_1}^{m_1}(\Omega_{\mathbf{r}}) Y_{l_2}^{m_2}(\Omega_{\mathbf{R}})\end{aligned} \tag{3.7}$$

with

$$r_< = \min(r, R), \quad r_> = \max(r, R), \tag{3.8}$$

and

$$l_< = l_1, \quad l_> = l_2 \text{ if } r < R,$$

$$l_< = l_2, \quad l_> = l_1 \text{ if } R > r. \tag{3.9}$$

We now represent the product $I \cdot K$ in the addition theorem by the following "discontinuous integral[22]"

$$(-1)^{l_<+l_2} I_{l_<+1/2}(\beta r_<) K_{l_>+1/2}(\beta r_>)$$
$$= (-1)^{(l_1-l_2+n)/2} \beta^n \cdot \int_0^\infty dt\, t^{1-n}(t^2+\beta^2)^{-1} J_{l_1+1/2}(rt)$$
$$\times J_{l_2+1/2}(Rt) \tag{3.10}$$

with n being an arbitrary natural number such that $(l_1 - l_2 + n)$ is even. If we introduce Eq. (3.7) and Eq. (3.10) into Eq. (3.6), the integration over the angular variables can be performed making use of the orthogonality relations of the spherical harmonics. The application of the Bessel operator $D_\beta^{N_2+L_2}$ on $\beta^{L_2+n}(t^2+\beta^2)^{-1}$ in the integrand of the remaining integral produces an expression independent of n, yielding

$$S_{0,N_2}^{L_1L_2}(\alpha, \beta, \mathrm{R}) = 4\pi[2^{L_1} L_1!]^{-1} R^{-1/2} \alpha^{L_1-1} \beta^{2N_2+L_2+3/2}$$
$$\times \sum_{l_1 l_2} \sum_{m_1 m_2} (-1)^{\Delta l_2} \langle L_2 M_2 | l_1 m_1 | l_2 m_2 \rangle Y_{l_2}^{m_2}(\Omega_\mathrm{R})$$
$$\times \delta_{l_1, L_1} \delta_{m_1, M_1} \int_0^\infty dr \int_0^\infty dt\, e^{-\alpha r}$$
$$\times r^{L_1+1/2} (t^2+\beta^2)^{-N_2-L_2-1} t^{L_2+1}$$
$$\times J_{l_1+1/2}(rt) J_{l_2+1/2}(Rt). \tag{3.11}$$

The integration over r can be performed according to[23]

$$\int_0^\infty dr\, r^{L_1+1/2} J_{L_1+1/2}(tr)\, e^{-\alpha r}$$
$$= (2/\pi)^{1/2} \cdot 2^{L_1} L_1!\, t^{L_1+1/2} (t^2+\alpha^2)^{-L_1-1}. \tag{3.12}$$

If the Bessel operator $D_\alpha^{N_1}$ is applied to both sides of Eq. (3.11), because of the definition Eq. (2.8) together with Eqs. (2.7) and (2.6), one obtains

$$S_{N_1 N_2}^{L_1 L_2}(\alpha, \beta, \mathrm{R})$$
$$= \sum_{l_2 m_2} 2^{5/2} \pi^{1/2} (-1)^{\Delta l_2} \alpha^{2N_1+L_1-1} \beta^{2N_2+L_2-1} R^{-1/2} Y_{l_2}^{m_2}(\Omega_\mathrm{R})$$
$$\times \langle L_2 M_2 | L_1 M_1 | l_2 m_2 \rangle\, V_{l_2}^{\alpha,\beta} \tag{3.13}$$

and the remaining integral V_{l_2} becomes

$$V_{l_2}^{\alpha,\beta} = V_{l_2}^{\alpha,\beta}(N_1, N_2, L_1, L_2; R)$$
$$= \int_0^\infty dt\, \frac{t^{L_1+L_2+3/2} J_{l_2+1/2}(tR)}{(t^2+\alpha^2)^{N_1+L_1+1} (t^2+\beta^2)^{N_2+L_2+1}}. \tag{3.14}$$

Because for $\alpha \neq \beta$ it is true that

$$[(t^2+\alpha^2)(t^2+\beta^2)]^{-1}$$
$$= (\alpha^2-\beta^2)^{-1}[(t^2+\beta^2)^{-1} - (t^2+\alpha^2)^{-1}], \tag{3.15}$$

one obtains

$$V_{l_2}^{\alpha,\beta} = \frac{(-1)^{N_1+N_2+L_1+L_2}}{(N_1+L_1)!\,(N_2+L_2)!} \left(\frac{\partial}{\partial \alpha^2}\right)^{N_1+L_1} \left(\frac{\partial}{\partial \beta^2}\right)^{N_2+L_2}$$
$$\times (\alpha^2-\beta^2)^{-1} \cdot \{W_0^\beta - W_0^\alpha\}, \tag{3.16}$$

where W_0^α—and equivalently W_0^β—are special cases ($q = 0$) of the integral

$$W_q^\alpha = \int_0^\infty dt\, t^{L_1+L_2+3/2} (t^2+\alpha^2)^{-q-1} J_{l_2+1/2}(Rt),$$
$$q = 0, 1, 2, 3, \cdots. \tag{3.17}$$

It is advantageous, however, to define the more general integral Eq. (3.17), because we obtain integrals of this type after having carried out the derivatives with respect to α and β in Eq. (3.16) with the help of Leibniz' rule. In the next step, the integral Eq. (3.17) can be reduced to a definite integral, which is known,[24] if the properties of the Bessel operator Eq. (2.7) are utilized. One obtains

$$W_q^\alpha = (\pi/2)^{1/2} (2^q q!)^{-1} R^{-l_2-1/2} D_R^{\Delta l_2}$$
$$\times \{R^{2q} \hat{k}_{\sigma(l_2)-q+1/2}(\alpha R)\}. \tag{3.18}$$

Collecting the results of the various calculations as indicated, we arrive at the following relationship:

$$S_{N_1,N_2}^{L_1,L_2}(\alpha, \beta, \mathrm{R})$$
$$= \alpha^{2N_1+L_1-1} \beta^{2N_2+L_2-1} (\alpha^2-\beta^2)^{-N_1-N_2-L_1-L_2-1} \cdot R^{-1/2}$$
$$\times \sum_{l_2} \langle L_2 M_2 | L_1 M_1 | l_2 m_2 \rangle Y_{l_2}^{m_2}(\Omega_\mathrm{R})$$
$$\times \sum_q q!\,(\alpha^2-\beta^2)^q \{a_q(N_1, N_2, L_1, L_2) \cdot W_q^\beta$$
$$+ (-1)^{N_1+N_2+L_1+L_2-q+1} a_q(N_2, N_1, L_2, L_1) \cdot W_q^\alpha q!\,(\alpha^2-\beta^2)^q\} \tag{3.19}$$

with the coefficient

$$a_q(N_1, N_2, L_1, L_2) = \pi^{1/2} 2^{5/2} (q!)^{-1} \binom{-N_1-L_1-1}{N_2+L_2-q}. \tag{3.20}$$

In order to obtain an analytical closed form expression for the integrals of the type W_q^α, in Eq. (3.18) the derivatives with respect to R are to be taken. Before doing so, it is advantageous to expand the expression of Eq. (3.18), on which the Bessel operator is to be applied, in terms of pure $\hat{k}$ functions, because then Eq. (2.7) can be utilized and cumbersome expressions are avoided. A method for expanding certain given functions in terms of reduced Bessel functions is given in Sec. 5. Here we use the following expansion:

$$R^{2q} \hat{k}_{\sigma(l_2)-q+1/2}(\alpha R)$$
$$= \alpha^{-2q} \sum_{p \geq 0} 2^p \binom{q}{p} \left(-\frac{2\sigma+1}{2}\right)_p \hat{k}_{\sigma+q-p+1/2}(\alpha R) \tag{3.21}$$

which is a special case of Eq. (6.1). The expansion is finite, because the summation limits are determined by the binomial coefficient and/or the Pochhammer symbol, which only for certain p values do not vanish.

If the Bessel operator is applied to each term of the series given by Eq. (3.21), one obtains, due to Eq. (2.7), a shift of the indices of the $\hat{k}$ functions. In Eq.

(3.19) there remains a double summation over q and p. Because the indices of the $\hat{k}$ functions contain, among others, $p+q$, a rearrangement of the terms of the double sum by collecting $\hat{k}$ functions with the same index gives a simple sum of $\hat{k}$ functions whose coefficients turn out to be Jacobi polynomials. With the help of the transformation[16]

$$\binom{n}{l}P_n^{(-l,\beta)}(x)=\binom{n+\beta}{l}\left(\frac{x-1}{2}\right)^l P_{n-l}^{(l,\beta)}(x)$$

we obtain the formula given in Eq. (3.1).

4. CONVOLUTION OF REDUCED BESSEL FUNCTIONS WITH EQUAL SCALING PARAMETERS

For equal scaling parameters $\alpha=\beta$ the relationship Eq. (3.1) does not hold, because then it contains indeterminate expressions like $\infty-\infty$ or $0/0$ although the limit $\alpha\to\beta$ should exist. It seems easier, however, to consider an independent evaluation of the integral $S_{N_1^1 N_2^2}^{L_1^1 L_2^2}(\alpha,\alpha,\mathrm{R})$ which will be given in this section. The result is the following compact formula:

$$\begin{aligned}&S_{N_1^1,N_2^2}^{L_1^1,L_2^2}(\alpha,\alpha,\mathrm{R})\\&=\int d\mathbf{r}\,\beta_{N_1^1-1/2,L_1}^{M_1^1*}(\alpha\mathbf{r})\beta_{N_2^2-1/2,L_2}^{M_2}[\alpha(\mathbf{R}-\mathbf{r})]\\&=4\pi\alpha^{-3}\sum_l\langle L_2M_2|L_1M_1|lm\rangle\sum_t(-1)^t\binom{\Delta l}{t}\\&\quad\times\beta_{N_1^1+N_2^2+L_1+L_2-l-t+1/2,\,l}^{m}(\alpha\mathrm{R}).\end{aligned}\tag{4.1}$$

The summation indices l and t run over all values for which the Gaunt coefficients and the binomial coefficients do not vanish. This is the case for

$$|L_2-L_1|\le l\le L_1+L_2,\quad m=M_2-M_1,\quad 0\le t\le\Delta l.\tag{4.2}$$

The derivation follows the lines explained in the preceding section up to Eq. (3.14). Setting $\alpha=\beta$ in Eq. (3.14), one sees that the integral $V_{l_2}^{\alpha,\alpha}$ defined by Eq. (3.14) is equal to W_q^α defined by Eq. (3.17) with $q=N_1+N_2+L_2+1$. Obviously q is fixed now; there is no summation over q anymore. This is due to the fact that one does not need the partial fraction expansion Eq. (3.15) as was necessary for the case $\alpha\neq\beta$. The remaining integral W_q^α is again given by Eq. (3.18). Carrying out the derivatives with respect to R with the help of Eqs. (2.7) and (3.21) we obtain the final result Eq. (4.1) if we use Eq. (6.1) and Eqs. (2.11), (2.12).

If one "angular momentum quantum number," for instance L_2, vanishes, the sums over l and t in Eq. (4.1) reduce to one term, because $\Delta l=0$ for $L_2=M_2=0$,

$$\begin{aligned}&\int d\mathbf{r}\,\beta_{N_1^1-1/2,L_1}^{M_1^1*}(\alpha\mathbf{r})\beta_{N_2-1/2,0}^{0}[\alpha(\mathrm{R}-\mathbf{r})]\\&=2\pi^{1/2}\alpha^{-3}\beta_{N_1^1+N_2+1/2,L_1}^{M_1^1*}(\alpha\mathrm{R}).\end{aligned}\tag{4.3}$$

5. CONVOLUTION OF FUNCTIONS BY EXPANSIONS IN TERMS OF REDUCED BESSEL FUNCTIONS

As has been shown in the preceding section, the convolution integral of two β functions,

$$\int d\mathbf{r}\,\beta_{\nu,L}^{M}[\alpha\mathbf{r}]\beta_{\nu',L'}^{M'}[\beta(\mathbf{R}-\mathbf{r})]\tag{5.1}$$

for half-integer ν,ν', leads to simple results. The result is a simple sum of β functions, which for the case of unequal scaling parameters α and β is given by Eq. (3.1), for the case of equal scaling parameters by Eq. (4.1). The simplicity of these relationships also makes it possible to evaluate the convolution integrals of functions of the type $g(r)\mathcal{Y}_L^M(\mathbf{r})$, if the radial part of these functions can be expanded in terms of reduced Bessel functions $\hat{k}_\nu(r)$. In this case, the corresponding convolution integral can be expressed by a series of integrals of the type given by Eq. (4.1), because the functions $\beta_{\nu,L}^{M}(\mathbf{r})$ are given—apart from a constant—by $\hat{k}_\nu(r)\mathcal{Y}_L^M(\mathbf{r})$ as defined by Eq. (2.2).

In this section we give a method which, under certain conditions, enables us to expand a given function $f(x)$ in a series of reduced Bessel functions according to

$$f(x)=\sum_p a_p\hat{k}_{p+\mu-1/2}(x).\tag{5.2}$$

The functions $f(x)$ considered here must fulfill the following rather general conditions:

(a) $f(x)$ can be represented by a Laplace integral

$$f(x)=\int_0^\infty ds\,\exp(-sx^2)\tilde{f}(s),\quad x>0.\tag{5.3a}$$

(b) The Laplace transform $\tilde{f}(s)$ has the property that the following power series expansion holds:

$$\begin{aligned}&\exp[(4s)^{-1}]\cdot\tilde{f}(s)\\&=(2/\pi)^{1/2}\sum_p a_p(2s)^{-p-\mu-1/2},\quad s>0.\end{aligned}\tag{5.3b}$$

(c) The following series, which represents the integrand of the Laplace integral Eq. (5.3a),

$$\sum_p a_p(2s)^{-p-\mu-1/2}\exp[-(4s)^{-1}-sx^2]\tag{5.3c}$$

converges uniformly with respect to s for $0\le s<\infty$ and for all $x>0$.

Under the conditions (a), (b), and (c), $f(x)$ can be expanded in terms of reduced Bessel functions according to Eq. (5.2) with μ and a_p being determined by Eq. (5.3b). The proof is easy: Insert Eq. (5.3b) into Eq. (5.3a) and interchange the order of integration and summation, which is allowed because of (c), then use the well-known Laplace transformation for reduced Bessel functions given by,[25]

$$\hat{k}_\lambda(x)=(2/\pi)^{1/2}2^{-\lambda-1}\int_0^\infty ds\cdot s^{-\lambda-1}\cdot\exp[-(4s)^{-1}-sx^2].\tag{5.4}$$

This yields Eq. (5.2). Q.E.D.

Obviously, the recipe is rather simple: If the Laplace transform $\tilde{f}(s)$ is known, one has to expand the left-hand side of Eq. (5.3b) as a power series of s. A comparison of the coefficients of the left-hand side with those of the right-hand side of Eq. (5.3b) furnishes μ and the coefficients a_p, which can be inserted into Eq. (5.2). This yields the expansion of $f(x)$ as a series of reduced Bessel functions. For functions whose expansions contain only reduced Bessel functions of half-integer order (i.e., if μ is integer), the convolution integral can be evaluated by using the formulas given in the preceding sections.

6. EXAMPLES

As examples we give some expansions which were found by application of the method described above.

The following formula allows the product of a power of x times a reduced Bessel function to be represented by a finite sum of pure reduced Bessel functions:

$$x^t \hat{k}_{n-1/2}(x) = \sum_{p\geq 0} 2^p \binom{t/2}{p} \binom{\frac{1-2n-t}{2}}{p} \hat{k}_{t-p+n-1/2}(x) \qquad (6.1)$$

for $n = 1, 2, 3, \cdots$, $t = -1, 0, 1, 2, \cdots$.

The summation limits are determined by the binomial coefficient and the Pochhammer symbol. This formula Eq. (6.1) was obtained with the help of the Laplace integral[26]

$$x^{2\mu-2} I_{2\nu}(x) = \int_0^\infty ds \exp[-sx^2]\{s^{-\mu+1/2} \times \exp[-(4s)^{-1}] W_{\mu-1/2,\nu}[(4s)^{-1}]\}, \qquad (6.2)$$

where $W_{\lambda,\kappa}$ is Whittaker's function. Multiplying the whole equation (6.1) by $Y_L^M(\Omega_R)$, one obtains for $n=1$ and $t+L=N-1$, because $\hat{k}_{1/2}(x) = \exp(-x)$,

$$\chi_{N,L}^M(\alpha \mathbf{r}) = (\alpha r)^{N-1} e^{-\alpha r} Y_L^M(\Omega_r) = \sum_p \frac{(-2)^{p-N+L}(N-L)!}{(2p-N+L)!\,(N-L-P)!} B_{p-1/2,L}^M(\alpha \mathbf{r}). \qquad (6.3)$$

The summation index p runs from $\min p$ to $\max p = N\text{-}L$ with

$$\min p = \begin{cases} (N-L)/2 & \text{for } N-L \text{ even}, \\ (N-L+1)/2 & \text{for } N-L \text{ odd}. \end{cases} \qquad (6.4)$$

This formula gives the expansion of a Slater-type atomic orbital in terms of reduced Bessel functions. It is helpful for the evaluation of molecular integrals which occur in LCAO-MO calculations.

The Coulomb potential may be represented by the series

$$r^{-1} = \sum_{p=0}^{\infty} 2^{-p}(p!)^{-1} \hat{k}_{p-1/2}(r) \qquad (6.5)$$

which can be derived from the Laplace integral[27]

$$r^{-1} = \pi^{-1/2} \int_0^\infty ds\, s^{-1/2} \exp(-sr^2). \qquad (6.6)$$

A generalization of Eq. (6.5) is the expansion of an irregular solid spherical harmonic $\mathcal{Z}_L^M(\mathbf{r}) = r^{-L-1} Y_L^M(\Omega_r)$ in terms of β functions,

$$\mathcal{Z}_L^M(\alpha \mathbf{r}) = [(2L-1)!!]^{-1} (\beta/\alpha)^{L+1} \sum_{p=0}^{\infty} \beta_{p-L-1/2,L}^{M}(\beta \mathbf{r}). \qquad (6.7)$$

Another useful relationship is the following expansion of the product of two RBF's as a finite sum of RBF's:

$$\hat{k}_{N_1+1/2}(\alpha r) \hat{k}_{N_2+1/2}(\beta r) = \sum_{s=0}^{N_1+N_2} \sum_{q=s}^{2s+1} \frac{(q+1)!\, 2^{s-N_1-N_2}}{(q-s)!\,(2s-q+1)!\,(\alpha+\beta)^q} (-1)^{s+q} \alpha^q \times \Pi_q^{N_1 N_2}\left(\frac{\beta}{\alpha}\right) \hat{k}_{s+1/2}[(\alpha+\beta)r]. \qquad (6.8)$$

Here the polynomials $\Pi_q^{N_1 N_2}$ are given by

$$\Pi_q^{N_1 N_2}(\beta/\alpha) = \sum_{p=\max(0,q-N_1)}^{\min(q,N_2)} \frac{(2N_1-q+p)!\,(2N_2-p)!}{(q-p)!\,p!\,(N_1-q+p)!\,(N_2-p)!} \left(\frac{\beta}{\alpha}\right)^p. \qquad (6.9)$$

As an illustration of the advantages of the method suggested here we will discusss the evaluation of the two-centric nuclear attraction integral over Slater-type atomic orbitals (STO's) as defined by Eq. (6.3),

$$\int d\mathbf{r}\, \chi_{\nu,\lambda}^{\mu}(\alpha \mathbf{r}) \frac{1}{|\mathbf{r}-\mathbf{R}|} \chi_{\nu',\lambda'}^{\mu'}(\beta \mathbf{r}). \qquad (6.10)$$

This, of course, is also the potential at the point R which is generated by a charge distribution $\chi_{\nu,\lambda}^{\mu}(\alpha \mathbf{r}) \chi_{\nu',\lambda'}^{\mu'}(\beta \mathbf{r})$. Because of Eqs. (6.3) and (6.8), this molecular integral can be expressed by a sum of integrals of the form

$$A_{N,L}^M(\alpha, \mathbf{R}) = \int d\mathbf{r}\, (|\mathbf{r}-\mathbf{R}|)^{-1} \mathcal{B}_{N-1/2,L}^M(\alpha \mathbf{r}). \qquad (6.11)$$

If the operator $(|\mathbf{r}-\mathbf{R}|)^{-1}$ is expressed by the series Eq. (6.5), one obtains *with the help of the convolution theorem* Eq. (4.3) the formula

$$A_{N,L}^M(\alpha, \mathbf{R}) = 4\pi \alpha^{-2} (-\alpha R)^L Y_L^M(\Omega_R) D_{\alpha R}^L\{Q_{N,L}(\alpha R)\}, \qquad (6.12)$$

where $Q_{N,L}$ stands for the infinite series

$$Q_{N,L}(\alpha R) = \sum_{p=0}^{\infty} 2^{-p-N-1}[(p+N+L+1)!]^{-1} \hat{k}_{p+N+L+1/2}(\alpha R). \qquad (6.13)$$

Putting $p+N+L+1 = q$, a comparison with Eq. (6.5) shows that $Q_{N,L}$ can be represented as

$$Q_{N,L}(\alpha R) = (\alpha R)^{-1} - \sum_{q=0}^{N+L} 2^{-q}(q!)^{-1} \hat{k}_{q-1/2}(\alpha R). \qquad (6.14)$$

The application of the Bessel differential operator $D_{\alpha R}^L$ is now straightforward and leads to the remarkably simple result for the crucial integrals $A_{N,L}^M$ of Eq. (6.11),

$$A_{N,L}^M(\alpha, \mathbf{R}) = 4\pi \alpha^{-2}\{(2L-1)!!\, \mathcal{Z}_L^M(\alpha \mathbf{R}) - \sum_{q=0}^{N+L} \mathcal{E}_{q-L-1/2,L}^M(\alpha \mathbf{R})\}. \qquad (6.15)$$

This is a simple linear combination of irregular solid spherical harmonics $\mathcal{Z}_L^M$ and $\mathcal{E}$ functions.

The analytical evaluation of other types of molecular integrals based on this method will be discussed somewhere else.[28]

7. RELATION TO DIFFERENTIAL EQUATIONS

As is well known, a solution f of the inhomogeneous differential equation

$$Lf(\mathbf{r}) = \rho(\mathbf{r}) \qquad (7.1)$$

is given by the convolution integral

$$f(\mathbf{r}) = \int d\mathbf{r}'\, G(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}'), \qquad (7.2)$$

where L is a linear differential operator and G is the corresponding Green's function with

$$LG(\mathbf{r}-\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}'). \qquad (7.3)$$

A solution of the differential equation (7.1) can be obtained as an expansion in terms of β functions if both the Green function G as well as the function ρ, which represents the inhomogenity, can be expanded in terms of β functions: In this case the convolution integral given by Eq. (3.1) of Sec. 3 or Eq. (4.1) of Sec. 4 can be utilized. Therefore, the method described in Sec. 5 is helpful also for the solution of certain differential equations.

For the following differential equation

$$\left[\left(\frac{\partial}{\partial \mathbf{r}}\right)^2 - \alpha^2\right] f(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \tag{7.4}$$

which is an inhomogeneous "modified" Helmholtz equation, the Green function itself is a special RBF, namely $\hat{k}_{-1/2}$, because[29]

$$\left[\left(\frac{\partial}{\partial \mathbf{r}}\right)^2 - \alpha^2\right] \hat{k}_{-1/2}(\alpha\,|\,\mathbf{r}-\mathbf{r}'\,|) = -4\pi\alpha\,\delta(\mathbf{r}-\mathbf{r}'). \tag{7.5}$$

Therefore, with the help of the Green function $\hat{k}_{-1/2}$, a solution f of the differential equation (7.4) can be written as the convolution integral

$$f(\mathbf{r}) = \int d\mathbf{r}'\,\rho(\mathbf{r}')\hat{k}_{-1/2}(\alpha\,|\,\mathbf{r}-\mathbf{r}'\,|). \tag{7.6}$$

In this case, only ρ needs to be expanded in terms of RBF's in order to utilize the convolution integrals Eqs. (3.1) and (4.1).

If ρ itself is a B function, a comparison of the convolution integrals Eqs. (7.6) and (4.1) shows that a solution f is also a B function, fulfilling the following differential equation (7.7), which of course can be verified directly,

$$\left(\frac{\partial}{\partial \mathbf{r}}\right)^2 B^M_{N-1/2,L}(\alpha\mathbf{r}) = \alpha^2[B^M_{N-1/2,L}(\alpha\mathbf{r}) - 2(N+L+1)B^M_{N-3/2,L}(\alpha\mathbf{r})]. \tag{7.7}$$

This relationship is very helpful for the evaluation of matrix elements of the kinetic energy operator, because it allows the expression of these matrix elements in terms of convolution integrals.

These considerations show that not only the properties of the β functions with respect to integration, but also their properties with respect to differentiation are of physical interest.

8. CONCLUSION

In this article, some properties of reduced Bessel functions (RBF's) have been investigated. For greater applicability, the so-called β functions were introduced, which are defined by Eq. (2.2) as products of a RBF times a regular solid spherical harmonic. We have shown that the convolution products of two β functions can be represented by a rather simple sum of—again—β functions. The result for the convolution integral of β functions with equal scaling parameters is given by Eq. (4.1), whereas the corresponding formula for the case of unequal scaling parameters is given by Eq. (3.1).

In Sec. 5, a rather general method is suggested which allows us to expand arbitrary functions in terms of β functions. By this method it is possible to use the simple results of Secs. 3 and 4 also for the evaluation of three-dimensional convolution integrals over other functions than β functions.

Convolution integrals are of special interest in quantum chemical problems which require the application of LCAO-MO methods. In this case, the multicentric molecular integrals, which occur, may be considered as convolution products. For instance, Gaussian-type orbitals (GTO's) are frequently used in molecular calculations because the convolution product of two primitive Gaussians is a simple one-centric integral over another Gaussian.[30] This advantage, however, is partly compensated by the drawbacks of the GTO's due to their intrinsic shape. Therefore, it is desirable to use other basis functions, e.g., STO's, which may lead to a better convergence. So far, this was not possible because most of the multicentric molecular integrals, which occur for these basis functions, are extremely difficult to evaluate. Because the β functions considered in this article are closely related to STO's, the results obtained here are of great advantage for the investigation of this problem.[28]

ACKNOWLEDGMENTS

We thank the Fonds der Chemischen Industrie and the Verein der Freunde der Universität Regensburg for financial support.

[1]F. P. Prosser and C. H. Blanchard, J. Chem. Phys. **36**, 1112 (1962).
[2]H. J. Silverstone, J. Chem. Phys. **45**, 4337 (1966).
[3]K. Ruedenberg, Theor. Chim. Acta (Berlin) **7**, 359 (1967).
[4]H. J. Silverstone, J. Math. Phys. **47**, 537 (1967).
[5]I. Shavitt, in *Methods in Computational Physics*, edited by B. Alder, S. Fernbach, M. Rotenberg (Academic, New York, 1963), p. 16.
[6]C. P. Yue and R. L. Somorjai, Chem. Phys. Lett. **5**, 537 (1970).
[7]D. M. Bishop and B. E. Schneider, J. Math. Phys. **11**, 2711 (1970).
[8]E. O. Steinborn and E. J. Weniger, Int. J. Quantum Chem. Symp. **11** (1977) (to be published).
[9]E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge U. P., Cambridge, 1953), p. 52.
[10]G. N. Watson, *Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1966), p. 78.
[11]See Ref. 10, p. 79.
[12]E. O. Steinborn and E. Filter, Theor. Chim. Acta (Berlin) **38**, 273 (1975). Equation (4.3) of this reference contains a misprint. Equation (2.9) of the present article is correct.
[13]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), p. 1.
[14]See Ref. 13, p. 3.
[15]See Ref. 13, p. 209.
[16]G. Szegö, *Orthogonal Polynomials*, (American Mathematical Society Colloquium, Providence, R. I., 1975), Vol. 13, p. 64, Eq. (4.22.2). See also Ref. 13, p. 210.
[17]See Ref. 10, p. 143.
[18]M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1967), p. 62.
[19]See Ref. 18, p. 32.
[20]R. Nozawa, J. Math. Phys. **7**, 1841 (1966).
[21]E. O. Steinborn and E. Filter, Int. J. Quantum Chem. Symp. **9**, 435 (1975).
[22]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), p. 694, Eq. (6.577.1,2).
[23]See Ref. 22, p. 694, Eq. (6.576.7).
[24]See Ref. 22, p. 686, Eq. (6.565.4).
[25]See Ref. 5, p. 15, Eq. (55).
[26]W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. 1, p. 217, Eq. (4.22.22).
[27]See Ref. 26, p. 137, Eq. (4.3.1).
[28]E. Filter and E. O. Steinborn, to be published.
[29]G. F. D. Duff and D. Naylor, *Differential Equations of Applied Mathematics* (Wiley, New York, 1966), p. 332.
[30]See. Ref. 5., p. 30.

A generalized prolongation structure and the Bäcklund transformation of the anticommuting massive Thirring model

H. C. Morris

School of Mathematics, Trinity College, Dublin 2, Ireland
(Received 15 February 1977)

The prolongation structure method of Wahlquist and Estabrook is generalized to Grassmann algebra valued differential forms and used to determine a Bäcklund transformation for the equations of the anticommuting massive Thirring model.

1. INTRODUCTION

The massive Thirring[1] model is defined by the equations

$$i\frac{\partial}{\partial X}\phi_1 = m\phi_2 + g\bar{\phi}_2\phi_2\phi_1, \tag{1.1}$$

$$i\frac{\partial}{\partial T}\phi_2 = m\phi_1 + g\bar{\phi}_1\phi_1\phi_2. \tag{1.2}$$

It has been shown by Kulish and Nissimov[2] that in the case where the fields ϕ_1 and ϕ_2 anticommute, the equations possess an infinite number of conserved quantities. Such a situation has always signaled the existence of soliton type behavior and a related Bäcklund transform.

In Sec. 2 we will determine a prolongation structure for the Thirring equations by generalizing the Wahlquist and Estabrook[3] method to Grassmann algebra-valued differential forms. Following this in Sec. 3 we determine the Bäcklund transformation as a diffeomorphism from the original ideal to the prolonged ideal.

Finally, in Sec. 4 we prolong the ideal further by introducing a potential ω. It is then shown that this leads to the Kulish and Nissimov result of an infinite set of conservation laws.

2. GRASSMAN ALGEBRA VALUED PROLONGATION FORMS

Equations (1.1) and (1.2) and their complex conjugates can be represented by the closed ideal of Grassmann algebra-valued forms spanned by

$$\alpha_1 = d\phi_1 \wedge dT + i[m\phi_2 + g\bar{\phi}_2\phi_2\phi_1]dX\wedge dT, \tag{2.1}$$

$$\alpha_2 = d\phi_2 \wedge dX - i[m\phi_1 + g\bar{\phi}_1\phi_1\phi_2]dX\wedge dT, \tag{2.2}$$

$$\bar{\alpha}_1 = d\bar{\phi}_1\wedge dT - i[m\bar{\phi}_2 + g\bar{\phi}_2\phi_2\bar{\phi}_1]dX\wedge dT, \tag{2.3}$$

$$\bar{\alpha}_2 = d\bar{\phi}_2 \wedge dX + i[m\bar{\phi}_1 + g\bar{\phi}_1\phi_1\bar{\phi}_2]dX\wedge dT. \tag{2.4}$$

We seek to determine a 1-form Ω having the property that the ideal spanned by α_1, α_2, $\bar{\alpha}_1$, $\bar{\alpha}_2$, Ω, $\bar{\Omega}$ is also closed.

The prolongation form Ω is taken in the form

$$\Omega = dY + F(\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2, Y, \bar{Y})dX + G(\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2, Y, \bar{Y})dT, \tag{2.5}$$

where Y, $\bar{Y}$ are Grassmann algebra-valued pseudopotentials.

If we denote the Grassmann algebra which is assumed to be over the complex field by A then F and G both belong to $M(A, A)$, the space of maps of A into itself.

The derivative of a function $H \in M(A, A)$ can be defined as follows. As every element Y of A has the property $Y^2 = 0$ there is a unique element H_Y of A such that

$$H(Y) = H_Y Y. \tag{2.6}$$

The element H_Y is defined to be the right partial derivative of $H(Y)$ with respect to Y. Further details of differential geometry on Grassmann manifolds can be found in the work of Fronsdal[4] and Mansouri.[5] Using the normal prolongation structure technique we find, taking care to maintain the order of terms, that

$$G_{\phi_1}\alpha_1 + F_{\phi_2}\alpha_2 + G_{\bar{\phi}_1}\bar{\alpha}_1 + F_{\bar{\phi}_2}\bar{\alpha}_2 + [F, G] = 0 \tag{2.7}$$

and

$$G_{\phi_2} = G_{\bar{\phi}_2} = F_{\phi_1} = F_{\bar{\phi}_1} = 0, \tag{2.8}$$

where

$$[F, G] \overset{\text{def}}{=} G_Y F - F_Y G + G_{\bar{Y}}\bar{F} - F_{\bar{Y}}G. \tag{2.9}$$

If we take F and G in the forms

$$F = X_0(Y, \bar{Y}) + X_1(Y, \bar{Y})\phi_2 + X_2(Y, \bar{Y})\bar{\phi}_2\phi_2 \tag{2.10}$$

and

$$G = X_3(Y, \bar{Y}) + X_4(Y, \bar{Y})\phi_1 + X_5(Y, \bar{Y})\bar{\phi}_1\phi_1, \tag{2.11}$$

then substitution of these forms into (2.7) gives the relations

$$[X_0, X_3] = 0, \quad [X_0, X_5] = 0, \quad [X_2, X_5] = 0, \quad [X_2, X_3] = 0, \tag{2.12}$$

$$[X_0, X_4\phi_1] = imX_1\phi_1, \tag{2.13}$$

$$[X_1\phi_2, X_4\phi_1] = -im(X_2 + X_5)(\bar{\phi}_1\phi_2 + \phi_1\bar{\phi}_2), \tag{2.14}$$

$$[X_1\phi_2, X_3] = -imX_4\phi_2, \tag{2.15}$$

$$[X_1\phi_2, X_5] = igX_1\phi_2, \tag{2.16}$$

$$[X_2, X_4\phi_1] = -igX_4\phi_1. \tag{2.17}$$

In the case of a normal manifold these relationships would be the basis of the prolongation structure. However in this case we cannot remove the fields ϕ_1, ϕ_2 to the outside of the Lie brackets without assuming something about the $X_i(Y, \bar{Y})$. The situation is, on the other hand, simpler as we know that each X_i must have the

 0022-2488/78/1901-0085$1.00

form

$$X_i(Y, \bar{Y}) = x_i + y_i Y + z_i \bar{Y} + w_i \bar{Y} Y. \quad (2.18)$$

We will restrict ourselves to the case when x_i, y_i, z_i, and w_i are complex numbers. Substitution of such general forms will lead to a number of quadratic relations between the x_i, y_i, z_i, and w_i.

For the Grassmann algebra case we will refer to that set of quadratic constraints as the prolongation structure of the original equations.

We will not attempt to determine the most general prolongation structure but choose special forms from the start to reduce the algebra. Take the generators X_i in the forms

$$X_0 = x_0 Y, \quad (2.19)$$

$$X_1 = x_1 + y_1 \bar{Y} Y, \quad (2.20)$$

$$X_2 = x_2 Y, \quad (2.21)$$

$$X_3 = x_3 Y, \quad (2.22)$$

$$X_4 = x_4 + y_4 \bar{Y} Y, \quad (2.23)$$

$$X_5 = x_5 Y. \quad (2.24)$$

The commutator relations (2.12) are automatically satisfied by this choice and so we have only to consider Eqs. (2.13)–(2.17).

$$[X_0, X_4\phi_1] = -(x_0 x_4 - y_4 \bar{x}_0 \bar{Y} Y)\phi_1 = im(x_1 + y_1 \bar{Y} Y) \quad (2.25)$$

or

$$x_0 x_4 = -imx_1, \quad y_4 \bar{x}_0 = imy_1, \quad (2.26)$$

$$[X_1\phi_2, X_4\phi_1] = \bar{Y}(y_1 x_4 + y_4 x_1)\phi_2\phi_1 + Y(y_1 \bar{x}_4 \bar{\phi}_1 \phi_2 + y_4 \bar{x}_1 \phi_1 \bar{\phi}_2)$$
$$= -im(x_5 + x_2) Y(\bar{\phi}_1 \phi_2 + \phi_1 \bar{\phi}_2), \quad (2.27)$$

and so

$$y_1 x_4 + y_4 x_1 = 0, \quad y_1 \bar{x}_4 = -im(x_5 + x_2) = y_4 \bar{x}_1, \quad (2.28)$$

$$[X_1\phi_2, X_3] = (x_3 x_1 - y_1 \bar{x}_3 \bar{Y} Y)\phi_2 = -im(x_4 + y_4 \bar{Y} Y) \quad (2.29)$$

which yields the constraints

$$x_3 y_1 = -imx_4, \quad y_1 \bar{x}_3 = imy_4, \quad (2.30)$$

$$[X_1\phi_2, X_5] = (x_5 x_1 - y_1 \bar{x}_5 \bar{Y} Y)\phi_2 = ig(x_1 + y_1 \bar{Y} Y)\phi_2, \quad (2.31)$$

giving

$$x_5 x_1 = igx_1, \quad y_1 \bar{x}_5 = -igy_1, \quad (2.32)$$

$$[X_2, X_4\phi_1] = -(x_2 x_4 - y_4 \bar{x}_2 \bar{Y} Y)\phi_1 = -ig(x_4 + y_4 \bar{Y} Y)\phi_1, \quad (2.33)$$

which leads to the final pair of constraints

$$x_2 x_4 = igx_4, \quad y_4 \bar{x}_2 = -igy_4. \quad (2.34)$$

Equations (2.26), (2.28), (2.30), (2.32), and (2.34) form a prolongation structure for the massive Thirring equations with anticommuting fields. This prolongation structure has the two-parameter solution

$$x_0 = -\lambda m, \quad (2.35)$$

$$x_1 = 2gm\bar{\rho}^{-1}, \quad y_1 = i\lambda\rho, \quad (2.36)$$

$$x_2 = ig, \quad (2.37)$$

$$x_3 = m\lambda^{-1}, \quad (2.38)$$

$$x_4 = 2img(\lambda\bar{\rho})^{-1}, \quad y_4 = \rho, \quad (2.39)$$

$$x_5 = ig, \quad (2.40)$$

where λ is a real constant. By rescaling Y we can make ρ take any value. For convenience we make $y_4 = ig$ to match with x_2 and x_5. This gives the prolongation form

$$\Omega = dY + [-\lambda m Y + (2mi - \lambda g \bar{Y} Y)\phi_2 + ig\bar{\phi}_2\phi_2 Y]dx + [m\lambda^{-1} Y + (-2m\lambda^{-1} + ig\bar{Y} Y)\phi_1 + ig\bar{\phi}_1\phi_1 Y]dT. \quad (2.41)$$

3. THE BÄCKLUND TRANSFORMATION

In terms of prolonged ideals a Bäcklund transformation can be considered[6] to be a diffeomorphism from the original ideal spanned by α_1, α_2, $\bar{\alpha}_1$, $\bar{\alpha}_2$ to the prolonged ideal spanned by α_1, α_2, $\bar{\alpha}_1$, $\bar{\alpha}_2$, Ω, $\bar{\Omega}$. Thus we seek new functions $\phi_1' = \phi_1'(\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2, Y, \bar{Y})$ and $\phi_2' = \phi_2'(\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2, Y, \bar{Y})$ having the property that

$$\alpha_1' = d\phi_1' \wedge dT + i[m\phi_2' + g\bar{\phi}_2'\phi_2'\phi_1']dX \wedge dT = f_1\alpha_1 + f_2\alpha_2 + f_3\bar{\alpha}_1 + f_4\bar{\alpha}_2 + \eta_1 \wedge \Omega + \eta_2 \wedge \bar{\Omega}, \quad (3.1)$$

$$\alpha_2' = d\phi_2' \wedge dX - i[m\phi_1' + g\bar{\phi}_1'\phi_1'\phi_2']dX \wedge dT = g_1\alpha_1 + g_2\alpha_2 + g_3\bar{\alpha}_1 + g_4\bar{\alpha}_2 + \eta_3 \wedge \Omega + \eta_4 \wedge \bar{\Omega}. \quad (3.2)$$

As we are dealing with a Grassmann algebra we know that

$$\phi_i' = \sigma_i(\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2) + \rho_i(\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2)Y + \omega_i(\phi_1, \bar{\phi}_2, \phi_2, \bar{\phi}_2)\bar{Y} + K_i(\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2)\bar{Y}Y. \quad (3.3)$$

By substituting these forms into (3.1) and (3.2) it is possible to determine the functions σ_i, ρ_i, ω_i, and K_i. The calculations are very tedious so we present only the final result. The mappings B_λ defined by

$$B_\lambda : (\phi_1, \phi_2, Y) \to \left(-\phi_1 + Y + \frac{i\lambda g}{2m}\bar{Y}Y\phi_1, \phi_2 + i\lambda Y + \frac{i\lambda g}{2m}\bar{Y}Y\phi_2, Y\right) \quad (3.4)$$

has the property that

$$\alpha_1' = -\alpha_1 + \left(\frac{i\lambda g}{2m}\bar{Y}\phi_1 - 1\right)dT \quad \Omega - \left(\frac{i\lambda g}{2m}dT\right) \wedge \bar{\Omega}, \quad (3.5)$$

$$\alpha_2' = \alpha_2 + \left(\frac{i\lambda g}{2m}\bar{Y}\phi_2 - 1\right)dX \wedge \Omega - \left(\frac{i\lambda g}{2m}dX\right) \wedge \bar{\Omega}. \quad (3.6)$$

and is therefore a Bäcklund transformation.

To obtain a more conventional form of the transformation we must section the prolongation form (2.41) onto a solution manifold of the ideal (2.1)–(2.4). This gives

$$Y_X = \lambda m Y - (2mi - \lambda g \bar{Y} Y)\phi_2 - ig\bar{\phi}_2\phi_2 Y, \quad (3.7)$$

$$Y_T = -m\lambda^{-1}Y - (-2m\lambda^{-1} + ig\bar{Y}Y)\phi_1 - ig\bar{\phi}_1\phi_1 Y, \quad (3.8)$$

which together with Eq. (3.4),

$$\phi_1' = -\phi_1 + Y + \frac{i\lambda g}{2m}\bar{Y}Y\phi_1, \quad (3.9)$$

$$\phi_2' = \phi_2 + i\lambda Y + \frac{i\lambda g}{2m}\bar{Y}Y\phi_2, \quad (3.10)$$

give the complete form of the Bäcklund transformation.

We note that this transformation has also been determined by Izergin and Stehr[7] using different methods.

From the trivial solution $\phi = 0$ one easily generates the single soliton solution

$$\phi_1' = a\ \exp[m(\lambda X - \lambda^{-1}T)], \quad (3.11)$$

$$\phi_2' = i\lambda\, a\ \exp[m(\lambda X - \lambda^{-1}T)], \quad (3.12)$$

where $a \in A$ and $\lambda \in R$.

4. FURTHER PROLONGATION AND CONSERVATION LAWS

As we have managed to prolong our original ideal α_1, α_2, $\bar{\alpha}_1$, $\bar{\alpha}_2$, to α_1, α_2, $\bar{\alpha}_1$, $\bar{\alpha}_2$, Ω, $\bar{\Omega}$ it is natural to enquire if we can prolong this ideal also. It is not difficult to show that the form ω defined by

$$\omega = \lambda(\bar{\phi}_2 Y + \bar{Y}\phi_2)\,dX + i(\bar{\phi}_1 Y - \bar{Y}\phi_1)\,dT \quad (4.1)$$

is a potential for the prolonged ideal. In fact

$$d\omega = \lambda(\bar{\alpha}_2 Y + \bar{Y}\alpha_2) + i(\bar{\alpha}_1 Y - \bar{Y}\alpha_1) - (\lambda\bar{\phi}_2\,dX + i\bar{\phi}_1\,dT)\wedge\Omega + (\lambda\phi_2\,dX - i\phi_1\,dT)\wedge\bar{\Omega}. \quad (4.2)$$

Following the arguments for prolongation structures on a normal manifold,[3] we deduce from the Stokes theorem,

$$\int_{\partial M}\omega = \int_M d\omega, \quad (4.3)$$

that if we choose a curve C lying in a solution manifold of the prolonged ideal which is closed, then

$$\int_C \tilde{\omega} = 0. \quad (4.4)$$

Expanding ω as a series in λ,

$$\omega = \sum_{i=0}^{\infty} \lambda^i \omega_i, \quad (4.5)$$

we see that (4.4) yields the infinite set of conservation laws

$$\int_C \tilde{\omega}_i = 0. \quad (4.6)$$

These are the Kulish and Nissimov[2] conservation laws.

[1] S. Coleman, Phys. Rev. D 11, 2088 (1975).
[2] D. Kulish and E. Nissimov, JETP Lett. 24, 247 (1976).
[3] H.D. Wahlquist and F.B. Estabrook, J. Math. Phys. 16, 1 (1975).
[4] C. Fronsdal, U.C.L.A. preprint UCLA/76/TED/1 (1976).
[5] F. Mansouri, Yale preprint C00-3075-152 (1976).
[6] R. Hermann, *Interdisciplinary Mathematics, Vol. XII: The Geometry of Non-Linear Differential Equations, Bäcklund Transformations and Solitons, Part A* (Math. Sci. Press, Brookline, Mass., 1976).
[7] A. Izergin and J. Stehr, DESY preprint 76/70 (1976).

Eigenvalues of $S\cdot\pi$ for spins 1/2, 1, and 3/2

D. L. Weaver

Department of Physics, Tufts University, Medford, Massachusetts 02155
(Received 28 February 1977)

The eigenvalues of the matrix operator $\mathbf{S}\cdot\boldsymbol{\pi}$ for a constant magnetic field are derived in a parallel way for spins 1/2, 1, and 3/2 using only the algebra of the spin matrices and the commutation relations of the components of π.

INTRODUCTION

The matrix operator $\mathbf{S}\cdot\boldsymbol{\pi}$, where S_i $(i=1,2,3)$ are the spin matrices for integer or half-integer spin S and $\boldsymbol{\pi}=\mathbf{p}-q\mathbf{A}$ with $\mathbf{p}$ the momentum operator $-i\nabla$, q the charge, and $\mathbf{A}$ the vector potential for a constant magnetic field $\mathbf{B}$ in the z direction so that $\mathbf{A}=\frac{1}{2}B(-y,x,0)$, appears in quantum physics when the interactions of charged particles with spin are considered.

The purpose of this paper is to derive the eigenvalues of $\mathbf{S}\cdot\boldsymbol{\pi}$ and indicate some of the applications for the three lowest nonzero spins, that is, $\frac{1}{2}$, 1, and $\frac{3}{2}$. The method applied in this paper is to utilize only the algebraic relations of the spin matrices found from their Lorentz transformation properties and their status as angular momentum operators, and the commutation relations of the components of π which follow from the fundamental commutator of X_i and P_j. These relations, along with the known eigenvalues of some operators such as S_3, as well as the standard methods for finding the roots of quadratic, cubic, and quartic equations, are sufficient to determine the eigenvalues.

The plan of the paper is to deal in order with the three cases starting with the well-known spin-$\frac{1}{2}$ case, which is included primarily for completeness and as the simplest example of the method.

For all spins there are two relations among the spin matrices that will be utilized. They are the commutator

$$[S_i,S_j]_-=i\epsilon_{ijk}S_k, \tag{1}$$

and the related relation

$$\epsilon_{ijk}S_iS_j=iS_k. \tag{2}$$

The components of π have the commutator

$$[\pi_i,\pi_j]_-=iq\epsilon_{ijk}B_k, \tag{3}$$

and the related relation

$$\epsilon_{ijk}\pi_i\pi_j=iqB_k. \tag{4}$$

1. SPIN $\frac{1}{2}$

The algebra of the spin-$\frac{1}{2}$ matrices is very well known, usually stated in the form

$$\sigma_i\sigma_j=\delta_{ij}+i\epsilon_{ijk}\sigma_k, \tag{5}$$

where σ_i are the Pauli spin matrices, $\boldsymbol{\sigma}=2\mathbf{s}$. Equation (5) can be found using the 2×2 matrix representation of $\boldsymbol{\sigma}$ very readily. The equivalent algebraic relations for higher spins are not so easily derived using a particular matrix representation. However, a general, representation independent method exists[1] that employs the Lorentz transformation properties of the symmetric, traceless covariantly defined spin tensor[2] $\tilde{S}_{\mu\nu\rho}\ldots$ with $2s$ indices, each Greek index ranging from 1 to 4. All the elements of the spin tensor may be derived, starting from $\tilde{S}_{44\ldots4}$, by separately considering pure rotations and pure special Lorentz transformations. The results are[2]

$$\tilde{S}_{44\ldots4}=(i)^{2s}, \tag{6}$$

$$\tilde{S}_{4\ldots4j}=(i)^{2s-1}S_j/S, \tag{7}$$

$$\tilde{S}_{4\ldots4jk}=\frac{(i)^{2s-2}}{s(2s-1)}\{S_jS_k+S_kS_j-s\delta_{jk}\}. \tag{8}$$

By combining the transformation properties for rotations with those for special Lorentz transformations, the general, algebraic relations for any spin are found to be

$$\begin{aligned}\tilde{S}_{4\cdots4i_1\cdots i_N}S_j&=\tfrac{i}{2}(2s-N)\tilde{S}_{4\cdots4ji_1\cdots i_N}-\tfrac{i}{2}\delta_{ji_1}\tilde{S}_{4\cdots4i_2\cdots i_N}\\&\quad-\tfrac{i}{2}\delta_{ji_2}\tilde{S}_{4\cdots4i_1i_3\cdots i_N}-\cdots-\tfrac{i}{2}\epsilon_{ji_1k}\\&\quad\times\tilde{S}_{4\cdots4ki_2\cdots i_N}-\tfrac{i}{2}\epsilon_{ji_2k}\tilde{S}_{4\cdots4ki_3\cdots i_N}-\cdots,\end{aligned} \tag{9}$$

where N can be any integer from 1 to $2s$. Equation (5) follows for $s=\frac{1}{2}$ and $N=1$.

The eigenvalues of $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$ follow from Eq. (5) because

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2=\sigma_i\sigma_j\pi_i\pi_j=\pi^2-q\boldsymbol{\sigma}\cdot\mathbf{B}$$

using Eq. (4). The result for $\mathbf{S}$ is

$$\begin{aligned}(\mathbf{S}\cdot\boldsymbol{\pi})^2&=\tfrac{1}{4}\{\pi^2-2q\mathbf{S}\cdot\mathbf{B}\}\\&\equiv\tfrac{1}{4}\alpha,\end{aligned} \tag{10}$$

where α is defined by

$$\alpha\equiv\pi^2-2q\mathbf{S}\cdot\mathbf{B}, \tag{11}$$

for any spin. The operator α is particularly important in magnetic field interaction problems because it commutes with $\mathbf{S}\cdot\boldsymbol{\pi}$ as well as π^2 and $\mathbf{S}\cdot\mathbf{B}$. The eigenvalues of α are $qB(2n+1-2m_s)+P_3^2$ with m_s the eigenvalues of S_3 for spin S, P_3 the eigenvalue of $-i(\partial/\partial z)$ and $n=0,1,2,\cdots$. The dependence on n follows because $\pi_x^2+\pi_y^2$ has the form of the simple harmonic oscillator Hamiltonian operator.

The spin-$\frac{1}{2}$ equation

$$\mathbf{S}\cdot\boldsymbol{\pi}\psi_\lambda=\lambda_{1/2}\psi_\lambda, \tag{12}$$

 0022-2488/78/1901-0088$1.00

with ψ the eigenfunctions and $\lambda_{1/2}$ the eigenvalues has the eigenvalues

$$\lambda_{1/2} = \pm\tfrac{1}{2}[\alpha(n, m_s, P_s)]^{1/2} = m_s[\alpha]^{1/2}, \tag{13}$$

utilizing Eq. (10), and taking ψ_λ to be the simultaneous eigenfunctions of the mutually commuting. Hermitian operators $\mathbf{S}\cdot\pi$, α, and P_3.

2. SPIN 1

The spin-1 algebraic relations are not as well known as those for spin $\frac{1}{2}$. They may be found from Eq. (9) for $s=1$ and $N=2$, the result being

$$S_iS_jS_k + S_kS_jS_i = \delta_{ij}S_k + \delta_{kj}S_i. \tag{14}$$

Contraction of Eq. (14) with $\pi_i\pi_j\pi_k$ produces the result

$$(\mathbf{S}\cdot\pi)^3 = \alpha\mathbf{S}\cdot\pi + qBP_3. \tag{15}$$

The eigenvalues then follow from the equation

$$\mathbf{S}\cdot\pi\psi_\lambda = \lambda_1\psi_\lambda, \tag{16}$$

by applying $\mathbf{S}\cdot\pi$ to ψ_λ two additional times to obtain

$$\{\alpha\mathbf{S}\cdot\pi + qBP_3\}\psi_\lambda = \lambda_1^3\psi_\lambda, \tag{17}$$

resulting in the cubic equation

$$\lambda_1^3 - \alpha\lambda_1 - qBP_3 = 0, \tag{18}$$

for the spin-1 eigenvalues of $\mathbf{S}\cdot\pi$, where the operators α and P_3 have been replaced by their eigenvalues. Using the standard methods,[3] the eigenvalues of $\mathbf{S}\cdot\pi$ for spin 1 are

$$\lambda_1 = t_1 + t_2, \quad -\tfrac{1}{2}(t_1+t_2) \pm i(\sqrt{3}/2)(t_1 - t_2), \tag{19}$$

where

$$t_1 = \left[\frac{qBP_3}{2} + \left(\frac{q^2B^2P_3^2}{4} - \frac{\alpha^3}{27}\right)^{1/2}\right]^{1/3} \tag{20}$$

and

$$t_2 = \left[\frac{qBP_3}{2} - \left(\frac{q^2B^2P_3^2}{4} - \frac{\alpha^3}{27}\right)^{1/2}\right]^{1/3}. \tag{21}$$

Only the real eigenvalues λ_1 are of physical interest, and their reality depends on the expression $-q^2B^2P_3^2/4 + \alpha^3/27$ being positive or zero. Using the specific expression for α, it is easy to see that for $n \geq 1$ and any m_s,

$$\alpha \geq qB + P_3^2, \tag{22}$$

so that

$$\frac{\alpha^3}{27} - \frac{q^2B^2P_3^2}{4} \geq \frac{(qB)^3}{27}\left\{\left(\frac{P_3^2}{qB}\right)^3 + 3\left(\frac{P_3^2}{qB}\right)^2 - 3.75\frac{P_3^2}{qB} + 1\right\}. \tag{23}$$

Reality of the eigenvalues λ_1, thus depends, for $n>0$, on the expression $b^3 + 3b^2 - 3.75b + 1$ being greater than or equal to zero, with $b \equiv P_3^2/qB$. The expression is always greater than zero except for $P_3^2 = qB/2$ when it is zero. For $n=0$, however, and $m_s = +1$, $\alpha = P_3^2 - qB$ and the eigenvalues of $\mathbf{S}\cdot\pi$ can be complex. This will occur when $P_3^2 < qB$.

These spin-1 eigenvalues of $\mathbf{S}\cdot\pi$ may be utilized[4] in deriving the enery levels of the Sakata—Taketani equation[5] for a spin-1 charged particle.

It is worth noting that when $P_3 = 0$, Eq. (15) simplifies to

$$(\mathbf{S}\cdot\pi_\perp)^3 = \alpha_\perp \mathbf{S}\cdot\pi_\perp, \tag{24}$$

for the transverse components only, resulting in the eigenvalues

$$\lambda_{1_\perp} = m_s[\alpha_\perp(n, m_s, 0)]^{1/2}, \tag{25}$$

in analogy to Eq. (13) for spin $\frac{1}{2}$.

3. SPIN $\frac{3}{2}$

With $s=\frac{3}{2}$ and $N=3$, Eq. (9) yields the algebraic equation for spin-$\frac{3}{2}$ spin matrices corresponding to Eq. (5) for spin $\frac{1}{2}$. After considerable simplification the resulting equation is

$$\begin{aligned} S_j\langle S_{i_1}S_{i_2}S_{i_3}\rangle + \langle S_{i_1}S_{i_2}S_{i_3}\rangle S_j &= -\tfrac{3}{8}\{\delta_{ji_1}\delta_{i_2i_3} + \delta_{ji_2}\delta_{i_1i_3} + \delta_{ji_2}\delta_{i_1i_3}\} \\ &\quad + \tfrac{1}{4}\{\delta_{ji_1}[S_{i_2}, S_{i_3}]_+ + \delta_{ji_2}[S_{i_1}, S_{i_3}]_+ + \delta_{ji_3}[S_{i_1}, S_{i_2}]_+\} \\ &\quad + \tfrac{7}{12}\{\delta_{i_1i_2}[S_j, S_{i_3}]_+ + \delta_{i_1i_3}[S_j, S_{i_2}]_+ + \delta_{i_2i_3}[S_j, S_{i_1}]_+\}, \end{aligned} \tag{26}$$

where $\langle S_{i_1}S_{i_2}S_{i_3}\rangle$ indicates 1/3! times a sum over all permutations of the three indices. The simplest consequence of this equation is the characteristic equation for spin $\frac{3}{2}$,

$$\{\mathbf{S}\cdot\hat{\mathbf{e}} + \tfrac{3}{2}\}\{\mathbf{S}\cdot\mathbf{e} - \tfrac{3}{2}\}\{\mathbf{S}\cdot\hat{\mathbf{e}} + \tfrac{1}{2}\}\{\mathbf{S}\cdot\hat{\mathbf{e}} - \tfrac{1}{2}\} = 0$$

which reduces to

$$(\mathbf{S}\cdot\hat{\mathbf{e}})^4 = \tfrac{5}{2}(\mathbf{S}\cdot\mathbf{e})^2 - \tfrac{9}{16}. \tag{27}$$

Other useful relations are, for example,

$$[S_j, S_k^3]_+ = \tfrac{7}{4}[S_j, S_k]_+, \quad j \neq k, \tag{28}$$

$$S_j\langle S_{i_1}S_{i_2}S_{i_3}\rangle + \langle S_{i_1}S_{i_2}S_{i_3}\rangle S_j = \tfrac{1}{4}[S_{i_2}S_{i_3}]_+,$$

$$i_1 \neq i_2 \neq i_3, \quad j = i, \tag{29}$$

and cyclic permutations.

Both sides of Eq. (26) are contracted with $\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3}$ to derive the relation between $(\mathbf{S}\cdot\pi)^4$ and lower powers. Some intermediate results are

$$\{\delta_{ji_1}\delta_{i_2i_3} + \delta_{ji_2}\delta_{i_1i_3} + \delta_{ji_3}\delta_{i_1i_2}\}\,\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3} = 3\{(\pi^2)^2 + q^2B^2\}, \tag{30}$$

$$\begin{aligned} &\{\delta_{ji_1}[S_{i_2}, S_{i_3}]_+ + \delta_{ji_2}[S_{i_1}, S_{i_3}]_+ + \delta_{ji_3}[S_{i_1}, S_{i_2}]_+\}\,\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3} \\ &\quad = 6\pi^2(\mathbf{S}\cdot\pi)^2 + 6q\pi^2\mathbf{S}\cdot\mathbf{B} \\ &\qquad + 6iq\mathbf{S}\cdot(\pi\times\mathbf{B})\mathbf{S}\cdot\pi - 3q\mathbf{S}\cdot\pi\mathbf{B}\cdot\pi, \end{aligned} \tag{31}$$

and

$$\begin{aligned} &\{\delta_{i_1i_2}[S_j, S_{i_3}]_+ + \delta_{i_1i_3}[S_j, S_{i_2}]_+ + \delta_{i_2i_3}[S_j, S_{i_1}]_+\}\,\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3} \\ &\quad = 6\pi^2(\mathbf{S}\cdot\pi)^2 + 12q\pi^2\mathbf{S}\cdot\mathbf{B} + 18iq\mathbf{S}\cdot(\pi\times\mathbf{B})\mathbf{S}\cdot\pi \\ &\qquad - 9q\mathbf{S}\cdot\pi\mathbf{B}\cdot\pi - \tfrac{45}{2}q^2B^2 + 6q^2(\mathbf{S}\cdot\mathbf{B})^2. \end{aligned} \tag{32}$$

Denoting the right-hand side of Eq. (26) contracted with $\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3}$ by RHS, the result is

$$\begin{aligned}\mathrm{RHS} = &-\tfrac{9}{8}\alpha^2 + 5\alpha(\mathbf{S}\cdot\boldsymbol{\pi})^2 + 4q\alpha\mathbf{S}\cdot\mathbf{B}\\ &+16q^2(\mathbf{S}\cdot\mathbf{B})^2 + 12iq\mathbf{S}\cdot(\boldsymbol{\pi}\times\mathbf{B})\mathbf{S}\cdot\boldsymbol{\pi}\\ &+10q\mathbf{S}\cdot\mathbf{B}(\mathbf{S}\cdot\boldsymbol{\pi})^2 - 6q\mathbf{S}\cdot\boldsymbol{\pi}\mathbf{B}\cdot\boldsymbol{\pi} - \tfrac{57}{4}q^2B^2. \qquad (33)\end{aligned}$$

Similarly, contraction of the left-hand side of Eq. (26) with $\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3}$ and denoted by LHS yields the intermediate results

$$\begin{aligned}&S_j\langle S_{i_1}S_{i_2}S_{i_3}\rangle\,\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3}\\ &\quad=(\mathbf{S}\cdot\boldsymbol{\pi})^4 + 2q\mathbf{S}\cdot\mathbf{B}(\mathbf{S}\cdot\boldsymbol{\pi})^2 + q\pi^2\mathbf{S}\cdot\mathbf{B}\\ &\quad+3iq\mathbf{S}\cdot(\boldsymbol{\pi}\times\mathbf{B})\mathbf{S}\cdot\boldsymbol{\pi} - \tfrac{15}{4}q^2B^2 + q^2(\mathbf{S}\cdot\mathbf{B})^2\\ &\quad-\tfrac{28}{8}q\mathbf{S}\cdot\boldsymbol{\pi}\mathbf{B}\cdot\boldsymbol{\pi}, \qquad (34)\end{aligned}$$

and

$$\begin{aligned}&\langle S_{i_1}S_{i_2}S_{i_3}\rangle S_j\pi_j\pi_{i_1}\pi_{i_2}\pi_{i_3}\\ &\quad=(\mathbf{S}\cdot\boldsymbol{\pi})^4 + 8q\mathbf{S}\cdot\mathbf{B}(\mathbf{S}\cdot\boldsymbol{\pi})^2 + 3q\pi^2\mathbf{S}\cdot\mathbf{B}\\ &\quad+9iq\mathbf{S}\cdot(\boldsymbol{\pi}\times\mathbf{B})\mathbf{S}\cdot\boldsymbol{\pi} - 15q^2B^2\\ &\quad+7q^2(\mathbf{S}\cdot\mathbf{B})^2 - \tfrac{121}{8}q\mathbf{S}\cdot\boldsymbol{\pi}\mathbf{B}\cdot\boldsymbol{\pi}, \qquad (35)\end{aligned}$$

with the final result for LHS being

$$\begin{aligned}\mathrm{LHS} = &\,2(\mathbf{S}\cdot\boldsymbol{\pi})^4 + 4q\alpha\mathbf{S}\cdot\mathbf{B} + 16q^2(\mathbf{S}\cdot\mathbf{B})^2\\ &+12iq\mathbf{S}\cdot(\boldsymbol{\pi}\times\mathbf{B})\mathbf{S}\cdot\boldsymbol{\pi} + 10q\mathbf{S}\cdot\mathbf{B}(\mathbf{S}\cdot\boldsymbol{\pi})^2\\ &-18q\mathbf{S}\cdot\boldsymbol{\pi}\mathbf{B}\cdot\boldsymbol{\pi} - \tfrac{75}{4}q^2B^2 \qquad (36)\end{aligned}$$

Equating LHS and RHS, one sees that remarkable cancellations occur so that only mutually commuting terms remain. The result, the spin-$\frac{3}{2}$ equivalent of Eq. (15) is

$$(\mathbf{S}\cdot\boldsymbol{\pi})^4 = \tfrac{5}{2}\alpha(\mathbf{S}\cdot\boldsymbol{\pi})^2 - \tfrac{9}{16}\{\alpha^2 - 4q^2B^2\} + 6qBP_3\mathbf{S}\cdot\boldsymbol{\pi}. \qquad (37)$$

The spin-$\frac{3}{2}$ eigenvalues of $\mathbf{S}\cdot\boldsymbol{\pi}$ then follow from the equation

$$\mathbf{S}\cdot\boldsymbol{\pi}\psi_\lambda = \lambda_{3/2}\psi_\lambda, \qquad (38)$$

which by repeated applications of $\mathbf{S}\cdot\boldsymbol{\pi}$ and the use of Eq. (37) leads to the quartic equation

$$\lambda_{3/2}^4 - \tfrac{5}{2}\alpha\lambda_{3/2}^2 - 6qBP_3\lambda_{3/2} + \tfrac{3}{16}\{\alpha^2 - 4q^2B^2\} = 0. \qquad (39)$$

Application of the standard methods[3] leads to the four eigenvalues.

A great simplification occurs when the coefficient of the term linear in $\lambda_{3/2}$ is negligible, say $P_3 \cong 0$. Then the quartic equation is readily solved to give

$$\lambda_{3/2}^2 = \tfrac{5}{4}\alpha_\perp \pm \alpha_\perp\left(1 + \frac{9}{4}\frac{q^2B^2}{\alpha_\perp^2}\right)^{1/2}. \qquad (40)$$

Expanding in powers of $q^2B^2/\alpha_\perp^2$ yields the approximate eigenvalues

$$\begin{aligned}\lambda_{3/2} = &\pm\tfrac{3}{2}[\alpha_\perp]^{1/2}\left\{1 + \frac{1}{2}\frac{q^2B^2}{\alpha_\perp^2} + \cdots\right\}^{1/2},\\ &\pm\tfrac{1}{2}[\alpha_\perp]^{1/2}\left\{1 - \frac{9}{2}\frac{q^2B^2}{\alpha_\perp^2} + \cdots\right\}^{1/2} \qquad (41)\end{aligned}$$

similar to Eqs. (13) and (25), but with corrections depending on the field strength.

4. UNITARY OPERATORS

For spin $\frac{1}{2}$, the operator

$$U_{1/2} = \exp\left[\tfrac{1}{2}\beta\alpha\cdot\frac{\boldsymbol{\pi}_\perp}{|\boldsymbol{\pi}_\perp|}\tan^{-1}\left(\frac{|\boldsymbol{\pi}_\perp|}{m}\right)\right] \qquad (42)$$

with

$$|\boldsymbol{\pi}_\perp| = [\alpha_\perp]^{1/2}$$

and

$$\boldsymbol{\alpha} \equiv \frac{1}{s}\begin{pmatrix}\mathbf{S} & 0\\ 0 & -\mathbf{S}\end{pmatrix},\quad \beta \equiv \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

is used to reduce the Dirac equation with a constant, external magnetic field and anomalous magnetic moment interaction to a form whose square is diagonal and gives the energy eigenvalues directly.[5] In the absence of the magnetic field, Eq. (42) is the Melosh transformation operator[6] proposed to relate the constituent and current quark pictures of elementary particles. Using Eqs. (15) and (37) with $P_3 = 0$, the equivalent operators for spin 1 and spin $\frac{3}{2}$ are derived below:

Spin 1: The appropriate operator is

$$U_1 = \exp\left[\beta\boldsymbol{\alpha}\cdot\frac{\boldsymbol{\pi}_\perp}{|\boldsymbol{\pi}_\perp|}\tan\left(\frac{|\boldsymbol{\pi}_\perp|}{m}\right)\right], \qquad (43)$$

with the same definitions for the matrices as in the spin-$\frac{1}{2}$ case. The result is

$$\begin{aligned}U_1 = &\,1 - \frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^2}{\alpha_\perp} + \cos\left[\tan^{-1}\left(\frac{|\boldsymbol{\pi}_\perp|}{m}\right)\right]\frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^2}{\alpha_\perp}\\ &+\sin\left[\tan^{-1}\left(\frac{|\boldsymbol{\pi}_\perp|}{m}\right)\right]\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp\beta}{[\alpha_\perp]^{1/2}}\\ = &\,1 - \frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^2}{\alpha_\perp} + \frac{m}{(\alpha_\perp + m^2)^{1/2}}\frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^2}{\alpha_\perp}\\ &+\frac{1}{(\alpha_\perp + m^2)^{1/2}}\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp\beta. \qquad (44)\end{aligned}$$

Replacement of $\tan^{-1}(|\boldsymbol{\pi}_\perp|/m)$ with other values leads to an alternate form for U. For example, $\tan^{-1}(|\boldsymbol{\pi}_\perp|/(\alpha_\perp + m^2)^{1/2})$ is appropriate for the spin-1 Foldy—Wouthuysen operator.[7,8] These operators may be used to transform Hamiltonians and other useful operators to forms appropriate for particular applications.[9]

Spin $\frac{3}{2}$: Consider the spin-$\frac{3}{2}$ operator

$$U_{3/2} = \exp\left(\tfrac{3}{2}\beta\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp}{|\boldsymbol{\pi}_\perp|}\theta\right). \qquad (45)$$

It has the expanded form

$$\begin{aligned}\exp\left(\tfrac{3}{2}\beta\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp}{|\boldsymbol{\pi}_\perp|}\theta\right) = &\,(f+1)\cos\tfrac{3}{2}u\theta - f\cos\tfrac{3}{2}v\theta\\ &+\frac{1}{u^2 - v^2}\left\{\left[\frac{u^2}{v}\sin\tfrac{3}{2}v\theta - \frac{v^2}{u}\sin\tfrac{3}{2}u\theta\right]\right.\\ &\times\beta\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp}{[\alpha_\perp]^{1/2}} + [\cos\tfrac{3}{2}u\theta - \cos\tfrac{3}{2}v\theta]\\ &\times\frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^2}{\alpha_\perp} + \left[\frac{1}{u}\sin\tfrac{3}{2}u\theta - \frac{1}{v}\sin\tfrac{3}{2}v\theta\right]\\ &\left.\times\beta\frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^3}{[\alpha_\perp]^{3/2}}\right\}, \qquad (46)\end{aligned}$$

where

$$f=\frac{\mp\lambda^{\pm 2}_{(3/2)\perp}}{2\alpha_\perp[1+\frac{9}{4}q^2B^2/\alpha_1^2]^{1/2}} \tag{47}$$

and

$$u^2=\frac{5}{2}\frac{f}{(2f+1)},\quad v^2=\frac{5}{2}\frac{(f+1)}{(2f+1)}, \tag{48}$$

choosing one of the two possible values for $\lambda^{\pm 2}_{(3/2)\perp}$.

When B is set to zero, $U_{3/2}$ reduces to the expected form for such an operator, i.e., the same form as the spin-$\frac{3}{2}$ rotation operator with iS in the rotation operator replaced by $\frac{3}{2}\beta\boldsymbol{\alpha}$.

Another, simpler form of a spin-$\frac{3}{2}$ unitary operator may be derived by replacing $|\pi_\perp|$ in Eq. (45) with $N\equiv[(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^4]^{1/4}=[\alpha_\perp(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^2-\frac{9}{16}(\alpha_1^2 4qB^2)]^{1/4}$. Then the expanded operator form is

$$\exp\ \tfrac{3}{2}\beta\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp}{N}\theta$$

$$=\tfrac{1}{2}\{\cos\tfrac{3}{2}\theta+\cosh\tfrac{3}{2}\theta\}+\tfrac{1}{2}\beta\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp}{N}\{\sin\tfrac{3}{2}\theta+\sinh\tfrac{3}{2}\theta\}$$

$$+\frac{1}{2}\frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^2}{N^2}\{\cos\tfrac{3}{2}\theta-\cosh\tfrac{3}{2}\theta\}$$

$$+\tfrac{1}{2}\beta\ \frac{(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}_\perp)^3}{N^3}\{\sin\tfrac{3}{2}\theta-\sinh\tfrac{3}{2}\theta\}. \tag{49}$$

Other forms are also readily found using the algebra of $\mathbf{S}\cdot\boldsymbol{\pi}$.

[1]D. L. Weaver, J. Math. Phys. **18**, 306 (1977).
[2]T. J. Nelson and R. H. Good, Jr., J. Math. Phys. **11**, 1355 (1970).
[3]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Segun (Dover, New York, 1965), p. 17.
[4]D. L. Weaver, Phys. Rev. D **14**, 2824 (1976).
[5]See, for example, D. L. Weaver, Phys. Rev. D **12**, 4001 (1975).
[6]H. J. Melosh, Phys. Rev. D **9**, 1095 (1974).
[7]L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
[8]D. L. Weaver, Nuovo Cimento A **53**, 665 (1968).
[9]See, for example, D. L. Weaver, Phys. Rev. D **12**, 2325 (1975).

Spaces of positive and negative frequency solutions of field equations in curved space–times. II. The massive vector field equations in static space–times

Carlos Moreno

Physique Mathématique, Collège de France, Paris
(Received 18 January 1977)

The space–times considered in this article are static, $V_n \times \mathbb{R}$, with compact space-section manifolds without boundary, V_n, and such that the trajectories of the Killing vector field are geodesics. For the physical field of spin 1 and mass $m > 0$ in these space–times, field equations are solved in any adapted atlas, by the one-parameter groups of unitary operators generated by scalar and vector Hamiltonians, $i^{-1}T_j^{-1}$, $j = 0,1$, in Sobolev spaces $H_j^l(V_n) \times H_j^{l-1}(V_n)$, $l \in \mathbb{R}$. Hilbert spaces of positive energy solutions of field equations, as well as those of reduced solutions and their canonical symplectic and complex structures, are determined. The existence and the uniqueness of Lichnerowicz's $(1-1)$ current on space–time are established, and the corresponding frequency-solution Hilbert spaces are constructed. Within the framework of Segal, a definition of quantum field operators is given, leading to the postulated commutator for the physical field concerned.

INTRODUCTION

Let V_{n+1} be a globally hyperbolic, $(n+1)$-dimensional Lorentzian manifold of class C^∞ with metric tensor g of the same class.

The physical neutral *free* field of spin 1 and mass $m > 0$ in V_{n+1} is by definition described by the real solutions of the system of equations

$$(\Delta_{n+1} - m^2)\varphi = 0, \tag{1}$$

$$\delta\varphi = 0, \tag{2}$$

where $\Delta_{n+1} = d\delta + \delta d$ is the Laplacian operator on 1-form fields on V_{n+1}, d the exterior differential operator on V_{n+1}, and δ the codifferential operator on V_{n+1}.

In the neighborhood Ω of each point of V_{n+1}, the propagator ${}_1G$ of Eq. (1) is a $(1-1)$-current[1] [or a $(1-1)$-distribution form], defined[2] as the difference of the two fundamental solutions of this equation.[3,4]

In the domain of a local chart of V_{n+1}, containing Ω, the Cauchy problem for Eq. (1) and for the spacelike hypersurface Σ is solved[2,5] locally by the expression

$$\varphi^\alpha(x) = \int_\Sigma \{\varphi^{\lambda'}(x')\nabla_{\mu'} {}_1G^\alpha_{\lambda'}(x;x') - {}_1G^\alpha_{\lambda'}(x;x') \times \nabla_{\mu'}\varphi^{\lambda'}(x')\}\, d\Sigma^{\mu'}(x'), \tag{3}$$

where α, λ', $\mu' = 0, 1, \cdots, n$; $\nabla_{\mu'}$ denotes the covariant derivative at the point x' of V_{n+1}; $\{G^\alpha_{\lambda'}(x;x')\}$ are the components of ${}_1G$ in natural frames, and $\{d\Sigma^{\mu'}(x')\}$ are the components of the surface vector of Σ.

The following expression,

$$b(\varphi;\psi) = \int_\Sigma \{\varphi^\rho(x) \cdot \nabla_\lambda \psi_\rho(x) - \nabla_\lambda \varphi^\rho(x) \cdot \psi_\rho(x)\}\, d\Sigma^\lambda(x), \tag{4}$$

where φ and ψ are two solutions of Eq. (1), is the local expression of a bilinear skew-symmetric form on the vector space of solutions of Eq. (1). Expressions (3) and (4). are independent of the choice of Σ.

The development, for the considered field, of the standard quantization program,[6–9] needs at first a suitable definition of frequency parts $\varphi^\oplus$, $\varphi^\ominus$ of any real solution φ of the system of equations (1), (2) as well as a suitable Hilbert structure on the vector space of these solutions. In other words,[8,10] the equation system (1), (2) must provide some real Hilbert space of real solutions and some suitable complex structure operator, orthogonal in this space.

In the Minkowski space–time M_{n+1}, the foregoing definitions are usually given by means of the Fourier transform.[6,11,12] Nevertheless, it is possible to give these definitions, without the help of this transform.[13]

As in the scalar case,[14] the central role in the new definition is played by the following properties of the usual ${}_1G^1$ current on M_{n+1}, determined by the distribution kernel G^1 or Δ^1, D^1 etc.: (i) It is real, symmetric, and a solution of Eq. (1); (ii) the fundamental Lichnerowicz convolution relation

$${}_1G^\alpha_{\lambda''}(x;x'') = \int_\Sigma \{{}_1G^{1\alpha}_{\beta'}(x;x')\nabla_{\mu'} {}_1G^{1\beta'}_{\lambda''}(x'';x') - {}_1G^{1\beta'}_{\lambda''}(x'';x')\nabla_{\mu'} {}_1G^{1\alpha}_{\beta'}(x;x')\}\, d\Sigma^{\mu'}(x') \tag{5}$$

is satisfied; (iii) if we define a transformation J between real solutions of the system of equations (1), (2), by the local expression

$$(J\varphi)_{\alpha'}(x') = \int_\Sigma \{\varphi_\beta(x)\nabla_\lambda {}_1G^{1\beta}_{\alpha'}(x';x) - {}_1G^{1\beta}_{\alpha'}(x';x)\nabla_\lambda\varphi_\beta(x)\}\, d\Sigma^\lambda(x), \tag{6}$$

then the expression

$$\{\varphi;\psi\} = b(\varphi;J\psi) \tag{7}$$

is a scalar product on some real space of real solutions of the system of equations (1), (2).

The usual frequency parts of the real solution φ of the system of equations (1), (2), are now the expressions

$$\varphi^\oplus = \tfrac{1}{2}(I + iJ)\varphi, \quad \varphi^\ominus = \tfrac{1}{2}(I - iJ)\varphi. \tag{8}$$

Obviously, this process of definition of $\varphi^\oplus$ and $\varphi^\ominus$ in Minkowski space–time can be formulated in any curved space–time V_{n+1}.

Because it is not possible to define a Fourier transform in a general manifold V_{n+1}, the foregoing process then becomes a natural generalization to V_{n+1}, of the

 0022-2488/78/1901-0092$1.00

notion of frequency-solution Hilbert spaces. The problem is then to find a ${}_1G^1$ current on V_{n+1} satisfying the foregoing properties.

For the commutator of quantum field operators, the expression

$$[\varphi(x);\varphi(x')] = -(\hbar/i)\{{}_1G(x;x') - (1/m^2)\,d_x d_{x'}\,{}_0G(x;x')\}, \quad (9)$$

also becomes a natural generalization[13] to any curved space–time. Here $\hbar$ is the reduced Planck constant; ${}_0G$ and ${}_1G$ are the scalar and vector propagator for the Klein–Gordon equation in V_{n+1}.

The foregoing process to introduce on any curved space–time, the notion of frequency-part solutions, has been proposed by Lichnerwicz.[13]

In the Minkowski space–time it can be proved[11,12,15,16] that there is a bijective mapping between the solutions of field equations (1), (2) (the so-called *nonreduced* solutions) and the solutions of Eq. (1) on 1-form fields on V_{n+1}, with temporal component zero in the canonical coordinates system (the so-called *reduced solutions*).

In the space–time we shall consider, definitions of reduced and nonreduced solutions of field equations can still be given, and a bijective mapping between the sets of these solutions can still be established.

The space–time considered in this work are static, $V_n \times \mathbb{R}$ with compact space-section manifolds without boundary V_n and such that the Killing vector field trajectories are geodesics. On these space–times our main goals are (a) to give an existence and uniqueness theorem for the ${}_1G^1$ current, (b) to construct the corresponding Hilbert spaces of positive and negative frequency solutions, (c) in the context of Segal's work[10,17] to give a definition of quantum field operators for the physical field concerned, leading to the postulated commutator (9). To achieve these goals, we need to determine the functional spaces of solutions of the system of equations (1), (2) and those of the reduced solutions, as well as the symplectic and complex canonical structures of the associated dynamical system (see Chernoff and Marsden[18]).

1. FIELD EQUATIONS

Let $V_n \times \mathbb{R}$ be a static space–time with compact space-section manifolds without boundary V_n and such that the trajectories of the Killing vector field are geodesics. There is then an adapted[19] atlas of the differentiable structure of $V_n \times \mathbb{R}$ such that the metric element takes the form

$$ds^2 = dt^2 - d\sigma^2, \quad (10)$$

where t is the canonical coordinate of $\mathbb{R}$—the time—and $d\sigma^2$ the Riemannian metric element on V_n, defined by the metric tensor with components g_{ij}, $i,j = 1, \ldots, n$.

A simple calculation proves that the connection symbols $\Gamma^{\gamma}_{\alpha\beta}$, the components of the Ricci tensor $R_{\alpha\beta}$, and the components of the Riemann tensor $R_{\alpha\beta\gamma\delta}$, defined by the metric element (10), are zero, if at least one of the α, β, γ, δ is zero.

Let φ be a 1-form field on $V_n \times \mathbb{R}$ and a solution of Eqs. (1), (2). Let φ^0, φ^1 be the temporal and spatial components of φ in the adapted atlas considered. φ^0 and φ^1 then satisfy the following equations:

$$\partial_0 \varphi^0 = -\delta\varphi^1, \quad (11)$$

$$\partial_0^2 \varphi^0 + M_0 \varphi^0 = 0, \quad (12)$$

$$\partial_0^2 \varphi^1 + M_1 \varphi^1 = 0, \quad (13)$$

where $\partial_0 = \partial/\partial t$, δ is the codifferential operator of V_n, $M_0 = \Delta_0 + m^2$, $M_1 = \Delta_1 + m^2$, and Δ_l is the Laplacian operator on l-form fields on V_n.

We shall write Eqs. (12) and (13) in the differential system form

$$\frac{d}{dt}\begin{pmatrix}\varphi^0\\(d/dt)\varphi^0\end{pmatrix} = T_0^{-1}\begin{pmatrix}\varphi^0\\(d/dt)\varphi^0\end{pmatrix}, \quad (14)$$

$$\frac{d}{dt}\begin{pmatrix}\varphi^1\\(d/dt)\varphi^1\end{pmatrix} = T_1^{-1}\begin{pmatrix}\varphi^1\\(d/dt)\varphi^1\end{pmatrix}, \quad (15)$$

where

$$T_0^{-1} = \begin{pmatrix}0 & I\\ -M_0 & 0\end{pmatrix}, \quad T_1^{-1} = \begin{pmatrix}0 & I\\ -M_1 & 0\end{pmatrix}.$$

2. FUNCTIONAL SPACES AND OPERATORS T_0^{-1}, T_1^{-1}

(a) Let $\Lambda_r(V_n)$, $r = 1, \cdots, n$, be the vector bundle of r-forms on V_n and H_r^s the usual[20] Sobolev spaces of sections of $\Lambda_r(V_n)$.

We recall[20] that the space of distribution sections of $\Lambda_r(V_n)$, $\mathcal{D}_r'(V_n)$, is the inductive limit of spaces H_r^s and the space of smooth sections of $\Lambda_r(V_n)$, $C_r^\infty(V_n)$, their projective limit.

Clearly the space H_r^0 is identical to the completion of $C_r^\infty(V_n)$ space with respect to the scalar product[21]

$$(\alpha;\beta)_{r,0} = \int_{V_n} \alpha^{i_1\cdots i_r}\bar{\beta}_{i_1\cdots i_r}\eta, \quad \alpha,\beta \in C_r^\infty(V_n), \quad (16)$$

where η is the volume element of V_n defined by the metric tensor $\{g_{ij}\}$, and $\alpha^{i_1\cdots i_n}$ are the components of α. H_r^0 becomes, in this way, a Hilbert space with scalar product (16).

The exterior differential operator d on V_n, is a continuous operator from H_r^s to H_{r+1}^{s-1} and the codifferential operator δ is a continuous operator from H_r^s to H_{r-1}^{s-1}.

We recall some properties of d, δ, and Δ operators.

Proposition 1: If $\alpha \in H_r^1$, $\beta \in H_{r+1}^1$, then

$$(d\alpha;\beta)_{r+1,0} = (\alpha;\delta\beta)_{r,0}. \blacksquare$$

Corollary 2: If α, $\beta \in H_r^2$ then

$$(\Delta\alpha;\beta)_{r,0} = (\alpha,\Delta\beta)_{r,0} \quad \text{and} \quad (\Delta\alpha;\alpha)_{r,0} \geq 0. \blacksquare$$

Proposition 3: If $\alpha \in \mathcal{D}_r'(V_n)$, then

$$d\Delta\alpha = \Delta d\alpha, \quad \delta\Delta\alpha = \Delta\delta\alpha. \blacksquare$$

(b) $M = (\Delta + m^2)$ is a second-order differential operator on the vector bundles $\Lambda_r(V_n)$. By Corollary 2, M is invertible, strictly positive in H_r^0, and for $\lambda < 0$ we have

$$\|(M - \lambda I)^{-1}\|_{r,0} \leq |\lambda|^{-1}.$$

From the results of Seeley[22,23] and Balakrishnan,[24] complex powers operators M^l, $l \in \mathbb{C}$ of M can be defined. In particular,[23] M^l is a top linear isomorphism from

H_r^s to $H_r^{s-2\mathrm{Re}(l)}$, $r=1,\ldots,n$. Hence, the expression

$$(\alpha;\beta)_{r,k}=(M^{k/2}\alpha;M^{k/2}\beta)_{r,0},\quad \alpha,\beta\in H_r^k, \tag{17}$$

is a scalar product on H_r^k compatible with the topology; H_r^k becomes, in this way, a Hilbert space in which operators M^p, $p\in\mathbb{R}$, are self-adjoints and strictly positive. Consequently, space $H_r^k\times H_r^{k-1}$ becomes a Hilbert space with respect to the scalar product

$$\{\alpha;\beta\}_{n,k}=(\alpha_1;\beta_1)_{r,k}+(\alpha_2;\beta_2)_{r,k-1},\quad \alpha_1,\beta_1\in H_r^k,\quad \alpha_2,\beta_2\in H_r^{k-1}. \tag{18}$$

We also have $dM^p=M^pd$ and $\delta M^p=M^p\delta$.

The following theorem can now be easily proved:

Theorem 4: For $j=0,1$, we have

(i) T_j^{-1} is a toplinear isomorphism from $H_j^{l+1}\times H_j^l$ to $H_j^l\times H_j^{l-1}$.

(ii) $i^{-1}T_j^{-1}$ is self-adjoint in $H_j^l\times H_j^{l-1}$ with domain $H_j^{l+1}\times H_j^l$; and for $\alpha\in H_j^{l+1}\times H_j^l$ and $\lambda\in\mathbb{R}-\{0\}$ we have

$$\|(T_j^{-1}-\lambda)^{-1}\|_{j,l}\leq|\lambda|^{-1},\quad \|T_j^{-1}\alpha\|_{j,l}=\|\alpha\|_{j,l+1}.$$

(iii) iT_j with

$$T_j=\begin{pmatrix}0 & -M_j^{-1}\\ I & 0\end{pmatrix}$$

is a compact (Rellich theorem) self-adjoint operator on $H_j^l\times H_j^{l-1}$; $\lambda=0$ is an accumulation point of the point spectrum; the eigenfunctions are in $C_j^\infty(V_n)\times C_j^\infty(V_n)$ and $T_jT_j^{-1}=T_j^{-1}T_j=I$. ∎

Remark 1: By straightforward calculation on the Balakrishnan[24,25] expression defining T_j^{-l}, we obtain

$$T_r^{-l}=\begin{pmatrix}\cos\pi(-l/2)M_r^{l/2} & -\sin\pi(-l/2)M_r^{l-1/2}\\ \sin\pi(-l/2)M^{l+1/2} & \cos\pi(-l/2)M_r^{l/2}\end{pmatrix}.$$

(See Refs. 14, 26.) We can now verify the relation

$$\{\alpha;\beta\}_{r,l}=\{T_r^{-l+1}\alpha;\ T_r^{-l+1}\beta\}_{r,s},\quad \alpha,\beta\in H_r^l\times H_r^{l-1}.$$

3. SOLUTION SPACES OF THE KLEIN-GORDON VECTOR EQUATION. THE VECTOR PROPAGATOR

(a) Let $\{{}_jU(t)\}$ (or $\{{}_jU_t\}$), $j=0,1$, be the one-parameter group of unitary operators generated by T_j^{-1} in the space $H_j^l\times H_j^{l-1}$. We have $d^k/dt^k\,({}_jU_t\varphi)=T_j^{-k}({}_jU_t\varphi)$ with $\varphi\in H_j^{l+k}\times H_j^{l+k-1}$. The restriction of this group to the space $H_j^s\times H_j^{s-1}$, $s>l$, is identical with the group of unitary operators generated by T_j^{-1} in $H_j^s\times H_j^{s-1}$ still denoted ${}_jU(t)$. Cauchy problem for Eqs. (14) or (15) is solved by the expressions

$$\varphi^j(t')={}_jU(t'-t)\varphi^j(t),\qquad \varphi^j(t)\in H_j^l\times H_j^{l-1},$$

with $\varphi^j(\circ)$; $\mathbb{R}\to H_j^{l+1}\times H_j^l$ a continuously differentiable mapping. $\varphi^j(\circ)$ will be called a solution of Eqs. (14) or (15).

For a fixed time to each element of $H_j^l\times H_j^{l-1}$ is the Cauchy data at time t_0 of a *unique* solution of Eqs. (14) or (15). We can then to identify each solution with its Cauchy data at t_0. It has then a unambigous meaning to say that $H_j^l\times H_j^{l-1}$ is a space of solutions.

With obvious changes of notation, the proof of the next propostion is the same as that given in detail, for a similar proposition in the scalar case.[14]

Proposition 5: In the space $X_j=C_j^\infty(V_n)\times C_j^\infty(V_n)$ the operator T_j^{-1} generates a uniform equicontinuous one-parameter group of operators. The operators of this group are the restrictions to X_j of the operators ${}_jU(t)$. In $X_j'\equiv\mathcal{D}_j'(V_n)\times\mathcal{D}_j'(V_n)$ the operator T_j^{-1} generates a uniform equicontinuous one-parameter group of operators. The restriction to $H_j^l\times H_j^{l-1}$ of the operators of this group are the operators ${}_jU(t)$. The groups engendered by T_j^{-1} in X_j and X_j' will still be denoted by ${}_jU(t)$ and we have $d^k/dt^k\,{}_jU_t=T_j^{-k}\,{}_jU_t$ in the sense of the uniform topology of operators in X_j or X_j'. ∎

Notation: Hereafter, $\mathcal{D}_{(r,s)}(V_{n+1}^2)$ denotes the space of smooth (r,s)-currents with compact support in V_{n+1}^2. $C_r^\infty(V_{n+1})\hat{\otimes}\mathcal{D}_s'(V_{n+1})$ denotes the space of regular-on-the-left (r,s)-currents in V_{n+1}, and $\mathcal{D}_r'(V_{n+1})\hat{\otimes}C_s^\infty(V_{n+1})$ the space of regular on the right (r,s)-currents. The same meaning is supposed for similar notations with V_n instead of V_{n+1}.

Let ${}_jU_t^{ab}$, $a,\ b=1,2$ be the matrix-element operators of the operator ${}_jU_t$. By Proposition 5, ${}_jU_t^{ab}$ determines a regular (j,j)-current, ${}_jG_t^{ab}$ on V_n defined by linearity and continuity from the expression

$$\langle {}_jG^{ab}(t);\varphi\otimes\psi\rangle_{V_n^2}=-({}_jU_t^{ab}\varphi;\bar{\psi})_{j,0},\quad \varphi,\psi\in C_j^\infty(V_n). \tag{19}$$

The following proposition is an immediate consequence of Proposition 5.

Proposition 6: The mapping

$$\mathbb{R}\to C_j^\infty\hat{\otimes}\mathcal{D}_j'(V_n),$$
$$t\to{}_jG^{ab}(t)$$

is of class C^∞. See Also Ref. 14. ∎

Let φ_j be in $C^\infty(\mathbb{R}^2;C_{(j,j)}^\infty(V_n^2))$. Then

Corollary 7: The function from $\mathbb{R}^2$ to $\mathbb{C}$ defined by

$$(t';t)\to\langle G_j^{ab}(t'-t);\varphi_j(t';t)\rangle_{V_n^2},$$

is of class C^∞, and the Leibniz formula for derivatives with respect to t or t' holds. ∎

Corollary 8: The mapping from $\mathbb{R}^2$ to $C_j^\infty(V_n)$ defined by

$$(t';t)\to\langle {}_jG^{ab}(t'-t);\psi_j(t)\rangle_{V_n^2},$$

is of class C^∞, and the Leibniz formula for derivatives with respect to t or t' holds. ∎

Let $\hat{\Phi}_{l,k}$, $l,k=0,1$ be mappings of class C^∞ from $\mathbb{R}^2$ to $C_{(l,k)}^\infty(V_n^2)$, canonically determined by the element Φ of $\mathcal{D}_{(1,1)}(V_{n+1}^2)$. Then

Corollary 9: The expression

$$\langle {}_1G;\Phi\rangle=\sum_{j=0,1}\int_{\mathbb{R}^2}\langle {}_jG^{12}(t'-t);\hat{\Phi}_{j,j}(t',t)\rangle_{V_n^2}\,dtdt' \tag{20}$$

defines a regular element of $\mathcal{D}'_{(1,1)}(V_{n+1}^2)$, which is the propagator of the Klein–Gordon vector equation (1). ∎

(c) Let $\hat{M}_j$, $j=0,1$ be the symemtric $(j-j)$-current on $V_n\times V_n$ defined by operator M_j. Then

Proposition 10:

(i) $({}_jG^{ab}(t))'=(-1)^{a+b}\,{}_jG^{ab}(-t)$; (ii) $({}_jG^{ab}(t))'={}_jG^{ab}(t)$;

(iii) $\hat{M}_j * {}_jG^{ab}(t) = {}_jG^{ab}(t) * \hat{M}_j$, (iv) ${}_jG$ is skew-symmetric.

Here * denotes the Volterra–Schwartz[27] convolution of currents, and the apostrophe the transposition of currents.

Proof: (i), (ii), and (iii) follow immediately from the relations

$$ {}_jU^*(t) = {}_jU(-t), \quad {}_jU(t) \cdot T_j^{-1} = T_j^{-1} \cdot {}_jU(t) \tag{21} $$

and the representation formula[25] for one-parameter groups of operators. (See Proposition 5.) (iv) is a consequence of (i) and definition (20). ∎

(d) We now wish to establish a relation between scalar and vector propagators for the Klein–Gordon equation. See Lichnerowicz.[2,5] This relation will be used in Sec. 8.

Let us consider the operator

$$ (d \oplus d)\colon H^l \times H^{l-1} \to H_1^{l-1} \times H_1^{l-2} $$
$$ (u; v) \to (du; dv) $$

Proposition 11:

$$ d \oplus d)\, {}_0U_t = {}_1U_t\, (d \oplus d). \tag{22} $$

Proof: This relation follows, by continuity, from relation (21), the relation $dM = Md$, and Proposition 5. ∎

Each matrix element operator on the left side of expression (22) is continuous from $C^\infty(V_n)$ to $C_1^\infty(V_n)$; the corresponding adjoint operators are continuous from $C_1^\infty(V_n)$ to $C^\infty(V_n)$; the former operators then determine regular $(1-0)$-currents,[1,27] H_t^{ab}, defined by linearity and continuity from the expression

$$ \langle H_t^{ab}; \varphi \otimes \psi \rangle = -(d({}_0U_t^{ab}\varphi); \bar{\psi})_{1,0}, $$
$$ a, b = 1, 2, \quad \varphi \in C^\infty(V_n), \quad \psi \in C_1^\infty(V_n). $$

From expression (19) we now obtain

$$ \langle H_t^{ab}; \varphi \otimes \psi \rangle_{\mathcal{V}_n^2} = \langle (d \hat{\otimes} I)_1 G_t^{ab}; \varphi \otimes \psi \rangle_{\mathcal{V}_n^2}. \tag{23} $$

From the right side of expression (22), we similarly find

$$ \langle H_t^{ab}; \varphi \otimes \psi \rangle_{\mathcal{V}_n^2} = \langle (I \hat{\otimes} \delta), {}_1G_t^{ab}; \varphi \otimes \psi \rangle_{\mathcal{V}_n^2}. \tag{24} $$

Then

$$ (d \hat{\otimes} I)\, {}_0G_t^{ab} = (I \hat{\otimes} \delta)\, {}_1G_t^{ab}. $$

This relation, for $(ab) = (12)$, and relation $(d/dt')\, {}_0G^{12}_{t'-t} = -(d/dt)\, {}_0G^{12}_{t'-t}$, implies in view of defintion (20), the desired relation

$$ (\tilde{d} \hat{\otimes} I)\, {}_0G = (I \hat{\otimes} \tilde{\delta})\, {}_1G, \tag{25} $$

where $\tilde{d}$ is the exterior differential operator and $\tilde{\delta}$ the codifferential operator on $V_n \times \mathbb{R}$. Relation (25) holds locally[5] in any space–time V_{n+1}. See also Ref. 21, pp. 170–77, for similar relation in Riemannian geometry.

4. ENERGY SCALAR PRODUCT AND SPACE OF REDUCED SOLUTIONS

(a) The following proposition is immediate:

Proposition 12: Let $\varphi^j \equiv (\varphi_1^j; \varphi_2^j)$ be an element of $\mathcal{D}_j'(V_n) \times \mathcal{D}_j'(V_n)$, $j = 0, 1$. A solution of Eqs. (14) and (15), with Cauchy data φ^0 and φ^1 respectively (at some fixed time), is also a solution of Eq. (11) if and only if $\varphi^0 = L\varphi^1$ where

$$ L = \begin{pmatrix} 0 & M_0^{-1}\delta \\ -\delta & 0 \end{pmatrix}. \blacksquare $$

(b) From the Lagrangian density[5]

$$ L(\varphi) = \tfrac{1}{4} g^{\nu\rho} g^{\mu\sigma} (\nabla_\mu \varphi_\nu - \nabla_\nu \varphi_\mu)(\nabla_\sigma \varphi_\rho - \nabla_\rho \varphi_\sigma) + \tfrac{1}{2} m^2 g^{\mu\nu} \varphi_\mu \varphi_\nu, $$

we obtain (1) and (2), as Lagrange equations, and

$$ T^{\alpha\gamma}(\varphi) = \nabla^\gamma \varphi^\rho \circ \nabla_\rho \varphi^\alpha - \nabla^\gamma \varphi^\sigma \circ \nabla^\alpha \varphi_\sigma + \nabla^\sigma \varphi^\gamma \circ \nabla^\alpha \varphi_\sigma $$
$$ - \nabla^\rho \varphi^\gamma \cdot \nabla_\rho \varphi^\alpha + \tfrac{1}{2} g^{\alpha\gamma} \nabla^\rho \varphi^\sigma \circ \nabla_\rho \varphi_\sigma - \tfrac{1}{2} g^{\alpha\gamma} \nabla^\sigma \varphi^\rho \circ \nabla_\rho \varphi_\sigma $$
$$ + m^2 \varphi^\alpha \varphi^\gamma - \tfrac{1}{2} m^2 g^{\alpha\gamma} \varphi^\mu \circ \varphi_\mu, $$

as components of the energy–momentum tensor.[28] This tensor is not symmetric, and it satisfies $\nabla_\alpha T^{\alpha\beta}(\varphi) = \nabla_\beta T^{\alpha\beta}(\varphi) = 0$, φ being a solution of field equations (1) and (2).

In the space–time considered, the notion of total energy of a smooth solution φ, of the field equations has a meaning.[29] Its expression, in the chosen atlas on $V_n \times \mathbb{R}$, is

$$ K(\varphi) = \int_{V_n} 2T^{00}(\varphi)\eta. $$

Now let $\varphi^j \in C_j^\infty(V_n) \times C_j^\infty(V_n)$, $j = 0, 1$, be Cauchy data (at some fixed time t_0) for a solution φ of Eqs. (14) and (15). A straightforward calculation gives

$$ K(\varphi) = -(\varphi_2^0; \varphi_2^0)_{0,0} - (\varphi_1^0; \varphi_1^0)_{0,1} + (\varphi_2^1; \varphi_2^1)_{1,0} + (\varphi_1^1; \varphi_1^1)_1 $$
$$ = -\{\varphi^0; \varphi^0\}_{0,1} + \{\varphi^1; \varphi^1\}_{1,1}. $$

This expression is not definite-positive, and is independent of t_0 because ${}_jU_t$ is unitary on $H_j^1 \times H_j^0$. If, moreover, φ is also a solution of Eq. (11), we must have, from Propsoition 12, $\varphi^0 = L\varphi^1$, and then

$$ K(\varphi) = ((\delta d + m^2)\varphi_1^1; \varphi_1^1)_{1,0} + (M_1^{-1}(\delta d + m^2)\varphi_2^1; \varphi_2^1)_{1,0}. \tag{26} $$

The energy of any smooth solution of the system of equations (11), (12), (13) is then definite-positive.

Theorem 13: There is a unique linear operator P in $C_1^\infty(V_n)$, symmetric with respect to the scalar product of H_1^0, commuting with d and δ, and such that if

$$ PA_1 = \varphi_1^1, \quad PA_2 = \varphi_2^1 $$

for some A_1, A_2 in $C_1^\infty(V_n)$, expression (26) becomes

$$ K(\varphi) = (A_2; A_2)_{1,0} + (M_1 A_1; A_1)_{1,0}. \tag{27} $$

This operator is

$$ P = I + (1/m)(M_1^{1/2} + m)^{-1} d\delta \tag{28} $$

with inverse

$$ P^{-1} = I - (M_1^{1/2} + m)^{-1} M_1^{-1/2} d\delta. \tag{29} $$

Thus P and P^{-1} are self-adjoint, strictly positive in Hilbert spaces H_1^l, and P^{-1} is bounded.

Proof: From expression (26), we have

$$ K(\varphi) = ((\delta d + m^2)P^2 A_1; A_1)_{1,0} + (M_1^{-1}(\delta d + m^2)P^2 A_2; A_2)_{1,0}. $$

This expression is identical to (27) if and only if

$$M_1^{-1}(\delta d + m^2)P^2 = I, \quad P^2 M_1^{-1}(\delta d + m^2) = I.$$

Therefore,

$$(P^2)^{-1} = M_1^{-1}(\delta d + m^2).$$

In H_1^1, $(P^2)^{-1}$ is a bounded self-adjoint positive operator; it then possesses a unique positive square root $((P^2)^{-1})^{1/2}$, which we can verify to be given by the right side of expression (29). We now define $P^{-1} \equiv ((P^2)^{-1})^{1/2}$ and we find that $P \circ P^{-1} = P^{-1} \circ P = I$. ■

Corollary 14: Let $A \equiv (A_1; A_2)$ be a solution of Eq. (15) in $H_1^1 \times H_1^{1-1}$. Set $\varphi^1 = (P \times P)A$; $\varphi^0 = L\varphi^1$. Then, $\varphi^0 \in H^{1-1} \times H^{1-2}$, $\varphi^1 \in H_1^{1-1} \times H_1^{1-2}$, and $(\varphi^0; \varphi^1)$ is a solution of the system of equations (11), (14), (15). Conversely, let $(\varphi^0; \varphi^1)$, $\varphi^0 = L\varphi^1$ be a solution of the system of Eqs. (11), (14), (15) with $\varphi^1 \in H_1^1 \times H_1^{1-1}$. Then $A = (P^{-1} \times P^{-1})\varphi^1$ is a solution of Eq. (15) and $A \in H_1^1 \times H_1^{1-1}$. ■

Definition: Solutions A of Eq. (15) are called *reduced* solutions of the field equations. Solutions $(\varphi^0; \varphi^1)$ with $\varphi^0 = L\varphi^1$ of the system of equations (11), (14), (15) are called *nonreduced* solutions of the field equations. ■

From Corollary 14, it follows the there is a bijective mapping between the set of reduced solutions and the set of nonreduced solutions of the field equations and that $H_1^1 \times H_1^{1-1}$ is a Hilbert space of reduced solutions.

5. CANONICAL SYMPLECTIC AND COMPLEX STRUCTURES ON SOLUTION SPACES

(a) Following Segal[10,17] (see also P. Chernoff and J. Marsden[18]) let $J_j = T_j(-T_j^2)^{-1/2}$ be the unitary part of the polar[30] decomposition of T_j in the real space $H_j^1 \times H_j^{1-1}$, $j = 0, 1$. The proof of the following theorem is immediate

Theorem 15:

$$J_j = \begin{pmatrix} 0 & -M_j^{-1/2} \\ M_j^{1/2} & 0 \end{pmatrix},$$

this operator defines a complex structure on the real space $H_j^1 \times H_j^{1-1}$ of real solutions of Eqs. (14) or (15), i.e., (i) J_j is a top linear isomorphism of $H_j^1 \times H_j^{1-1}$, (ii) $J_j^2 = -I$, (iii) ${}_jU_t \cdot T_j^{-1} = T_j^{-1} \cdot {}_jU_t$. ■

Because the range of J_j in $H_j^1 \times H_j^{1-1}$ is *all* of $H_j^1 \times H_j^{1-1}$ and that of T_j is *only* $H_j^{1+1} \times H_j^1$, we obtain (i) and (ii) in the following theorem directly from the definition of weakly and strongly symplectic forms on Hilbert spaces.[18] (iii) follows by simple calculation.

Theorem 16: For $\alpha, \beta \in H_j^1 \times H_j^{1-1}$:

(i) The skew-symmetric form $\omega_t(\alpha, \beta) = \{J_j\alpha; \beta\}_{j,1}$ is strongly symplectic on $H_j^1 \times H_j^{1-1}$.

(ii) The skew-symmetric form $b_t(\alpha; \beta) = \{T_j\alpha; \beta\}_{j,1}$ is weakly symplectic on $H_j^1 \times H_j^{1-1}$.

(iii) The restriction of $\omega_{1-1/2}$ to the subspace $H_j^1 \times H_j^{1-1}$ is b_1. ■

Following Chernoff and Marsden,[18] p. 28, Eqs. (14) and (15) define Hamiltonian dynamical systems on $H_j^1 \times H_j^{1-1}$; $\frac{1}{2}\{\alpha; \alpha\}_{j,1}$ is the energy of $\alpha \in H_j^{1+1} \times H_j^1$, $i^{-1}T_j^{-1}$ is the Hamitonian, and b_t the weak symplectic form, of the dynamical system.

Theorem 17: On the real space $H_j^1 \times H_j^{1-1}$, there is an unique complex structure operator J_j' satisfying (i) $J_j'T_j = T_jJ_j'$ and (ii) $\omega_t(\alpha; J_j'\alpha) > 0$ for $\alpha \neq 0$. Thus we have $J_j' = J_j$. ■

Remarks 2: (a) Let φ, ψ be two smooth solutions of Eqs. (1), (2). Let A, B be the corresponding reduced solutions. Then $b(\varphi; \psi) = -b_1(A; B)$, where the left-side is the expression (4) and the right side is defined in Theorem 16, (ii). Also, $b_1(A; B) = (A_1; B_2)_{1,0} - (A_2; B_1)_{1,0}$.

(b) In the space–time considered, transformation J in (iii) of the Introduction *will be defined* by $(J\varphi)^0 = J_0\varphi^0$ and $(J\varphi)^1 = -J_1\varphi^1$.

(c) From (a) and (b) we obtain $b(\varphi; J\psi) = \{A; B\}_{1,1/2}$. See expression (7).

(d) For $j = 1$, Theorem 17 is the uniqueness theorem for the complex structure of spaces of reduced solutions. ■

6. THE LICHNEROWICZ (1 - 1)-CURRENT

(a) From operators J_0, J_1 we now obtain a $(1-1)$-current on $V_n \times \mathbb{R}$, ${}_1G^1$ satisfying the properties required in the Introduction.

The matrix element operators of ${}_jU_t^1 = {}_jU_t \cdot J_j$ determine regular $(j-j)$-currents on $V_n \times V_n$ [see Section 3(b)]. Their definitions are given by relations similar to (20). A simple calculation assigns to the operator ${}_jU_t^1$ the following matrix:

$$\begin{pmatrix} -\frac{d}{dt'}\,{}_jG^1_{t'-t} & -{}_jG^1_{t'-t} \\ -\frac{d^2}{dt'^2}\,{}_jG^1_{t'-t} & -\frac{d}{dt'}\,{}_jG^1_{t'-t} \end{pmatrix}.$$

Here the derivatives are taken in the sense of Proposition 6.

(b) Let $\hat{M}_j^{-1/2}$ be the regular, symmetric $(j-j)$-current on $V_n \times V_n$ determined by the operator $M_j^{-1/2}$.

Writing the relations $(d/dt)\,{}_jU_t^1 = T_j^{-1}\,{}_jU_t^1$, $({}_jU_t^1)^* = -{}_jU_{-t}^1$, and ${}_jU^1_{t''-t'} \cdot {}_jU^1_{t'-t} = -{}_jU_{t''-t}$ in terms of currents, we obtain in particular

Proposition 18:

(i) $(M_j \hat{\otimes} I)\,{}_jG^1_{t'-t} + \frac{d^2}{dt'^2}\,{}_jG^1_{t'-t} = 0$,

(ii) $(I \hat{\otimes} M_j)\,{}_jG^1_{t'-t} + \frac{d^2}{dt^2}\,{}_jG^1_{t'-t} = 0$,

(iii) $${}_jG^1_{t'-t} = -\hat{M}_j^{-1/2} * \frac{d}{dt'}\,{}_jG_{t'-t} = -\frac{d}{dt'}\,{}_jG_{t'-t} * \hat{M}_j^{-1/2},$$

(iv) $${}_jG_{t''-t} = {}_jG^1_{t''-t'} * \frac{d}{dt'}\,{}_jG^1_{t'-t} - \frac{d}{dt'}\,{}_jG^1_{t''-t'} * {}_jG^1_{t'-t},$$

(v) $\hat{M}_j^{1/2} * {}_jG_t^1 = {}_jG_t^1 * \hat{M}_j^{1/2}$,

(vi) ${}_jG_t^1 = {}_jG_{-t}^1$, $({}_jG_t^1)' = {}_jG_{-t}^1$,

where * denotes convolution of currents in the sense of Volterra–Schwartz[27] and the apostrophe, transposition of currents. See also Combet.[31,32]■

Theorem 19: With Φ and $\Phi_{j,j}$ being the functions in Corollary 9, the expression

$$\langle {}_1G^1; \Phi \rangle = \sum_{j=0,1} \int_{\mathbb{R}^2} \langle {}_jG^1_{t'-t}; \Phi_{j,j}(t',t) \rangle_{V_n^2} dt \circ dt' \tag{30}$$

defines a regular element of $\mathcal{D}'_{(1,1)}(V^2_{n+1})$, and satisfies properties (i), (ii), and (iii) of the Introduction. Theorem 15 is also a uniqueness theorem for ${}_1G^1$. Relation (25) is satisfied with ${}_0G^1$ and ${}_1G^1$ in place of ${}_0G$ and ${}_1G$ respectively.

Proof: Corollaries 8 and 9 hold with ${}_jG^1_{t'-t}$ in place of ${}_jG^{ab}_{t'-t}$; ${}_1G^1$ is then regular. T_j^{-1} is a real operator and therefore ${}_jU^1_t$ as well. In Proposition 18, (i) and (ii) prove that ${}_1G^1$ is a solution of field equations (12), (13); (iv) is the Lichnerowicz fundamental convolution relation (3); (vi) proves that ${}_1G^1$ is symmetric (and separately symmetric in time and space). Finally from Remarks 2(c) expression (7) is a scalar product.■

7. POSITIVE AND NEGATIVE FREQUENCY SPACES

On the complex space $H^l_j \times H^{l-1}_j$ of reduced solutions, let us consider the operators[13]

$$\oplus = \tfrac{1}{2}(I + iJ_1), \quad \ominus = \tfrac{1}{2}(I - iJ_1).$$

Then

Theorem 20: (i) $\oplus$ and $\ominus$ are complementary projectors in $H^l_1 \times H^{l-1}_1$, and $\oplus\, {}_1U_t = {}_1U_t \oplus$, $\ominus\, {}_1U_t = {}_1U_t \ominus$.

(ii) If $A \in H^l_1 \times H^{l-1}_1$ is real,

$$2\{\oplus A; \oplus A\}_{1,l} = 2\{\ominus A; \ominus A\}_{1,l} = \{A; A\}_{1,l}.\blacksquare$$

By definition $\oplus A$ and $\ominus A$ are the positive and negative frequency parts of the real reduced solution A, and the closed subspaces

$$E^{\oplus}_l = \oplus(H^l_1 \times H^{l-1}_1), \quad E^{\ominus}_l = \ominus(H^l_1 \times H^{l-1}_1),$$

are the Hilbert spaces of frequency parts of reduced solutions.

The frequency parts of non-reduced solutions can be defined in a similar way.[14]

8. QUANTUM FIELD OPERATORS. CREATION AND ANNIHILATION OPERATORS

For classical fields described by some kind[10,17] of nonlinear system of equations, the corresponding quantum field operators (see Segal[10,17]) are a special type of vector fields on the solutions manifold of the system of equations. Hence, commutators for these operators can be calculated.

For the linear field of spin 1 and mass $m > 0$, described by reduced solutions, the quantum field operators are those deduced from the general definition of Segal, here called *reduced quantum field operators*.

The aim of this section is to justify, in this context, the postulated commutator (9). For this, we need to give a definition of quantum field operators $\varphi(x)$ on the left side of expression (9), which are here called *nonreduced quantum field operators*.

For simplicity's sake, we limit ourself to the reduced solutions (manifold) space $H^{1/2}_1 \times H^{-1/2}_1$.

(a) The complex structure operator J_1 provides the space $H^{1/2}_1 \times H^{-1/2}_1$ of real reduced solutions, with a complex Hilbert structure, as follows[32,33]: multiplication by complex numbers is defined by $iA = J_1 A$, $A \in H^{1/2}_1 \times H^{-1/2}_1$, and the scalar product is defined by

$$\langle A; B \rangle_{1/2} = \{A; B\}_{1,1/2} - i\{J_1 A; B\}_{1,1/2}.$$

(b) Let $\mathcal{J}$ be the algebra of holomorphic functions on $H^{1/2}_1 \times H^{-1/2}_1$ with values in a complete, locally convexe, Haussdorf space[34] H.

Following Segal,[10] let us consider in $\mathcal{J}$ the linear operator $\psi(A)$ defined by

$$(\psi(A)F)C = i(\hbar/2)\, d_C F(A) + \omega_{1/2}(A; C)\, F(C), \tag{31}$$

where $F \in \mathcal{J}$, A, $C \in H^{1/2}_1 \times H^{-1/2}_1$, $\omega_{1/2}$ is defined in (i) of Theorem 18, $\hbar$ is the reduced Planck constant, and $d_C F(A)$ is the value at the point A, of the differential mapping of F at the point C.

Let us introduce some notations

$$\psi^1(A) = \psi(J_1 A),$$
$$\psi^{\oplus}(A) = \tfrac{1}{2}(\psi + i\psi^1)A, \quad \psi^{\ominus}(A) = \tfrac{1}{2}(\psi - i\psi^1)A.$$

A simple calculation gives the commutators

$$[\psi(A); \psi(B)] = i\hbar\omega_{1/2}(A; B)I,$$
$$[\psi^{\ominus}(A); \psi^{\oplus}(B)] = -(\hbar/2)\langle A; B \rangle_{1/2} I, \tag{32}$$

where I is the identity mapping in $\mathcal{J}$. See also Ref. 31.

(c) Let $A(\circ)$ be a continuous mapping from a compact K in $\mathbb{R}$ to space $H^{1/2}_1 \times H^{-1/2}_1$. For t, $t' \in K$, $A(t)$ and $A(t')$ are Cauchy data at times t and t' respectively, for two (generally differents) reduced solutions

Let us consider[32] the following linear operator in $\mathcal{J}$:

$$(\overline{\psi}(A(\circ))F)C = i(\hbar/2)\int_K d_C F(A(t))dt + \int_K \omega_{1/2}(A(t); C)F(C)dt. \tag{33}$$

Let $B(\cdot)$ be another continuous mapping from a compact K' in $\mathbb{R}$ to $H^{1/2}_1 \times H^{-1/2}_1$. We find the following expression for commutators:

$$[\overline{\psi}(A(\circ)); \overline{\psi}(B(\circ))] = i\hbar \int_{K\times K'} \omega_{1/2}(A(t); B(t'))dtdt' \cdot I. \tag{34}$$

Now let us suppose $A(\cdot)$ and $B(\cdot)$ to be of the following particular[17] type

$$K \to H^{1/2}_1 \times H^{-1/2}_1, \quad K' \to H^{1/2}_1 \times H^{-1/2}_1,$$
$$t \to (0; a(t)), \quad t' \to (0; b(t')) \tag{35}$$

and $a(t); b(t') \in C^\infty_1(V_n)$. Then set

$$\Phi(a) = \overline{\psi}(A(\circ)), \quad \Phi(b) = \overline{\psi}(B(\cdot)),$$
$$\Phi^{\oplus}(a) = \overline{\psi}^{\oplus}(A(\cdot)), \quad \Phi^{\ominus}(b) = \overline{\psi}^{\ominus}(B(\cdot)).$$

By definition (19), a straightforward calculation on the expression (34) gives

$$[\Phi(a); \Phi(b)] = i\hbar \int_{K\times K'} \langle {}_1G_{t-t'}; b(t') \otimes a(t) \rangle_{V_n^2} dt \cdot dt' I,$$
$$[\Phi^{\ominus}(a); \Phi^{\oplus}(b)] = i\hbar \int_{K\times K'} \langle {}_1G^{\ominus}_{t-t'}; b(t') \otimes a(t) \rangle_{V_n^2} dt \cdot dt' I, \tag{36}$$
$$[\Phi(a); \Phi^1(b)] = i\hbar \int_{K\times K'} \langle {}_1G^1_{t-t'}; b(t') \otimes a(t) \rangle_{V_n^2} dt \cdot dt' I,$$

where ${}_1G^{\ominus}_{t-t'} = \tfrac{1}{2}({}_1G_{t-t'} - i\,{}_1G^1_{t-t'})$.

We call the operators $\Phi(a)$, with a in $\mathcal{D}_1(V_n\times\mathbb{R})$, the *reduced quantum field operators*.

(d) Our purpose is now to interpret the postulated commutator (9) in the aforementioned framework of Segal.[10] For simplicity's sake, we shall only deal with smooth solutions of field equations.

Let $A\in C_1^\infty(V_n)\times C_1^\infty(V_n)\subset H_1^{1/2}\times H_1^{-1/2}$ be a reduced solu-solution, and let $\varphi\equiv(\varphi^0;\varphi^1)$, $\varphi^0\in L\varphi^1$, $\varphi^1=(P\times P)A$ be the corresponding nonreduced solutions. (see Corollary 14.) We give the following:

Definition: Let us define $\psi(\varphi)$ as being the operator (31), determined by the *reduced solution* $(d\oplus d)\varphi^0+\varphi^1$; i.e.,

$$\psi(\varphi)=\psi((d\oplus d)\varphi^0+\varphi^1).\blacksquare \tag{37}$$

Let $(\varphi^0(\cdot);\varphi^1(\cdot))$ be the nonreduced solutions corresponding to the reduced solutions $A(\cdot)$ of type (35). Set

$$\Phi(\varphi_a)=\overline{\psi}((d\oplus d)\varphi^0(\cdot)+\varphi^1(\cdot)), \tag{38}$$

where the right side is given by expression (33). By relation (25) and the equalities

$$I+(1/m)\Delta_0(M_0^{1/2}+m)^{-1}=(1/m)M_0^{1/2},$$

$$(2/m)(M_0^{1/2}+m)^{-1}+(1/m)\Delta_0(M_0^{1/2}+m)^{-2}=(1/m^2)I,$$

a straightforward calculation on expression (34), with operator (38), gives

$$[\Phi(\varphi_a);\Phi(\varphi_b)]$$
$$=i\hbar\langle{}_1G-(1/m^2)(\tilde{d}\otimes\tilde{d})_0G;(\delta b(\cdot);b(\cdot))$$
$$\otimes(\delta a(\cdot);a(\cdot))\rangle I, \tag{39}$$

where the right side is defined by an expression similar to (20). Here $\tilde{d}$ is the exterior differential operator on $V_n\times\mathbb{R}$.

If $a(\cdot)$ and $b(\cdot)$ are canonically determined by two elements of $\mathcal{D}_1(V_n\times\mathbb{R})$ with zero temporal components in the adapted atlas considered, $(\delta a(\cdot);a(\cdot))$ and $(\delta b(\cdot);b(\cdot))$ are in $\mathcal{D}_1(V_n\times\mathbb{R})$.

The sense of commutator (9) is that of commutator (39).

Commutators for the creation and annihilation operators $\Phi^{\oplus}(\varphi_a)$ and $\Phi^{\ominus}(\varphi_b)$ can be similarly obtained. (See Ref. 26.)

We call the operators $\Phi(\varphi_a)$ the *nonreduced quantum field operators*.

9. CONCLUSIONS

The space–times considered in this paper are static, $V_n\times\mathbb{R}$, with compact space-section manifolds without boundary, V_n, and such that the trajectories of the Killing vector field are geodesics.

The Klein–Gordon vector equation (1) becomes the couple of Eqs. (14), (15) in an adapted atlas where the metric element is of the form (10). These equations have been solved by the one-parameter groups of unitary operators $\{{}_jU_t;j=0,1\}$ on spaces $H_j^1\times H_j^{-1}$ generated by the scalar $i^{-1}T_0^{-1}$ and vector $i^{-1}T_1^{-1}$ Hamiltonians. The restrictions of these groups on spaces $C_j^\infty(V_n)\times C_j^\infty(V_n)$ determine one-parameter uniform equicontinuous groups of operators. These groups induce one-parameter uniform equicontinuous groups of operators on spaces $\mathcal{D}_j'(V_n)\times\mathcal{D}_j'(V_n)$. The propagator of the Klein–Gordon vector equation (1) and its symmetry properties have been determined.

For field equations of the physical field of spin 1 and mass $m>0$ solution spaces and reduced solution spaces have been determined. Canonical complex structures on these spaces have been calculated, and we have proved that a unique Lichnerowicz current, ${}_1G^1$, on $V_n\times\mathbb{R}$ simultaneously determines all these structures. By means of the ${}_1G^1$ current, frequency Hilbert spaces of reduced and nonreduced solutions have been constructed. Finally, in the context of Segal's work,[10,17] we have given a definition of nonreduced quantum field operators leading to an expression for the commutator of two such operators which interprets the expression of the postulated commutator[5] for the physical field concerned.

[1] J. Dieudonné, *Elements d'analyse* (Gauthier-Villars, Paris, 1974).
[2] A. Lichnerowicz, "Propagateurs et commutateurs en relativité générale," Publ. Math. I.H.E.S. **10** (1961).
[3] Y. Choquet-Bruhat, in *Battelle Rencontres* 1967. *Lectures in Mathematics and Physics* (Benjamin, New York, 1968).
[4] J. Leray, *Hyperbolic differential equations* (Institut for Advanced Study, Princeton, N.J., 1954).
[5] A. Lichnerowicz, Bull. Soc. Math. France 92, 11 (1964).
[6] D. Kastler, *Introduction à l'électrodynamique quantique* (Dunod, Paris, 1961).
[7] S.S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, New York, 1962).
[8] I.E. Segal, J. Math. Phys. **1**, 468 (1960).
[9] R.F. Streater and A.S. Wightman, *PCT Spin, Statistics and All That* (Benjamin, New York, 1964).
[10] I.E. Segal, "La variété des solutions d'une equation hyperbolique non linéaire d'ordre deux" (Séminaire Physique Mathématique, Collège de France, Paris, 1965).
[11] N.N. Bogolioubov and D.V. Chirkov, *Introduction à la théorie quantique des champs* (Dunod, Paris, 1960).
[12] P. Roman, *Theory of Elementary Particles* (North-Holland, Amsterdam, 1960).
[13] A. Lichnerowicz, in *Relativity Groups and Topology*, edited by DeWitt and DeWitt (Gordon & Breach, New York, London, Paris, 1965).
[14] C. Moreno, J. Math. Phys. **18**, 2153 (1977).
[15] W. Pauli, Rev. Mod. Phys. **13**, 203 (1941).
[16] A. Visconti, *Théorie quantique des Champs, Tome I* (Gauthier Villars, Paris, 1961).
[17] I.E. Segal, J. Math. Pures Appl. 44, 71 (1965).
[18] P. Chernoff and J. Marsden, *Properties of Infinite Dimensional Hamiltonian Systems*, Lectures Notes in Mathematiques No. 425 (Springer-Verlag, Berlin, Heidelberg, 1974).
[19] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnetisme* (Masson, Paris, 1965).
[20] R.S. Palais, *Seminar on the Atiyah-Singer Index Theorem*, Ann. of Math. Stud, No. 57 (Princeton U.P., Princeton, N.J., 1965).
[21] G. de Rham, *Variétés différentiables* (Hermann, Paris, 1960).
[22] R.T. Seeley, Amer. Math.Soc. Proc. Symp. Pure Math. **10**, 288 (1966).
[23] R.T. Seeley, *Topics in Pseudo-Differential Operators* (C.I.M.E. Edizioni Cremonese, Roma, 1969).
[24] A.V. Balakrishnan, Pac. J. Math. **10**, 419 (1960).
[25] K. Yosida, *Functional Analysis* (Springer-Verlag, Berlin, Heidelberg, 1968).
[26] C. Moreno, "Sur les noyaux de la théorie quantique des

champs sur un espace-temps courbe," Thèse (Université de Lyon I, 1975).
[27]L. Schwartz, "Théorie des distributions à valeurs vectorielles" (Ann. Inst. Fourier, extraits des tomes VII et VIII, 1959).
[28]C. Roche, "Les équations de Hamilton covariantes en Relativité Générale," Thèse (Faculté des Sciences, Paris, 1968).
[29]A. Trautman, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[30]N. Dunford and J.T. Schwartz, *Linear Operators, Vol. II* (Interscience, New York, 1963).
[31]E. Combet, C.R. Acad. Sci. A 262, 1077 (1966).
[32]E. Combet, "Sur une formule de M. Lichnerowicz: Fonction Generatrice des opérateurs de creation et d'annihilation" (Séminaire Physique Mathematique, Collège de France, 1964).
[33]R. Hermann, *Vector Bundles in Mathematical Physics* (Benjamin, New York, 1970), Vol. I.
[34]N. Bourbaki, *Variétés différentielles et analytiques. Fascicule de résultats* (Hermann, Paris, 1967).

A variational derivation of the Bach–Lanczos identity[a)]

John R. Ray

Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29631
(Received 18 July 1977)

A discussion of a modified Hilbert variational principle is presented. The Bach–Lanczos identity is then derived from this variational principle.

I. INTRODUCTION

Recently we presented a modified Hilbert variational principle in general relativity.[1] The modified principle varies the metric and affine connection independently but adds the constraint, via Lagrange multipliers, that the affine connection equals the Christoffel symbols constructed from the metric. In Ref. 1 the modified variational principle was used to explain the origin of the, so-called, Palatini variational principle in general relativity. We shall review this result in Sec. III of this paper since the results are needed in deriving the Bach–Lanczos identity which is contained in Sec. IV.

Several other workers[2–4] have used the modified variational principle to study alternate theories of gravity. Reference 2 used the variational principle to study nonlinear Lagrangians, Reference 3 used it to study scalar-tensor theories while Ref. 4 used it to study theories with torsion. Our applications will deal only with general relativity.

II. MODIFIED HILBERT PRINCIPLE

Consider a Lagrangian L which is a function of the metric g_{ij} and its first two derivatives

$$L = L(g_{ij}, g_{ij,k}, g_{ij,kn}). \tag{2.1}$$

The Hilbert variational principle extremizes the action

$$I = \int L\, dx, \tag{2.2}$$

with respect to variations of the metric which yields

$$\delta I = \int \frac{\delta L}{\delta g_{su}}\, \delta g_{su} dx, \tag{2.3}$$

where the variational derivative of L is defined by

$$\frac{\delta L}{\delta g_{su}} = \frac{\partial L}{\partial g_{su}} - \left(\frac{\partial L}{\partial g_{su,r}}\right)_{,r} + \left(\frac{\partial L}{\partial g_{su,rp}}\right)_{,pr}. \tag{2.4}$$

The field equations for g_{su} are then

$$\frac{\delta L}{\delta g_{su}} = -\sqrt{-g}\, T^{su}, \tag{2.5}$$

where T^{su} is the energy–momentum tensor which is obtained by varying the matter Lagrangian with respect to the metric. We assume the matter Lagrangian does not depend on derivatives of the metric.

For example, if one chooses $L = \sqrt{-g}R$ then there obtains

$$\frac{\delta L}{\delta g_{su}} = -\sqrt{-g}\,(R^{su} - 1/2 g^{su}R) = -\sqrt{-g}\, T^{su}, \tag{2.6}$$

the Einstein equations.

The modification of the foregoing variational principle presented in Ref. 1 is to rewrite the Lagrangian L in terms of the metric and the symmetric affine connection Γ^i_{jk} and its first derivative

$$L = L(g_{ij}, \Gamma^i_{jk}, \Gamma^i_{jk,n}). \tag{2.7}$$

We now carry out independent variations of the metric and affine connection but impose the constraint, via Lagrange multipliers, that the affine connection equals the Christoffel symbols $\{^i_{jk}\}$. Hence, the action

$$I = \int [L + \sqrt{-g}\, \lambda_i{}^{jk}(\Gamma^i_{jk} - \{^i_{jk}\})]\, dx, \tag{2.8}$$

is obtained.

The field equations follow by varying g_{ij}, Γ^i_{jk}, $\lambda_i{}^{jk}$ and are

$$\frac{\delta L}{\delta g_{su}} + \sqrt{-g}\,\Lambda^{su} = -\sqrt{-g}\, T^{su} \tag{2.9a}$$

$$\frac{\delta L}{\delta \Gamma^p_{su}} + \sqrt{-g}\,\lambda_p{}^{su} = 0, \tag{2.9b}$$

$$\Gamma^i_{jk} - \{^i_{jk}\} = 0, \tag{2.9c}$$

where

$$\Lambda^{su} = \tfrac{1}{2}(g^{up}\lambda_i{}^{sr}{}_{;r} + g^{si}\lambda_i{}^{ur}{}_{;r} - g^{ri}\lambda_i{}^{su}{}_{;r}). \tag{2.10}$$

Note the variations in (2.9a) are carried out holding Γ^i_{jk} constant. The field equations for g_{ij} are obtained by solving (2.9b) and (2.9c) for the Lagrange multipliers in terms of the metric, calculating Λ^{su} from (2.10) and then substituting Λ^{su} and (2.9c) into (2.9a) which yields the desired equations. It is important to realize that we obtain the same field equations from this modified Hilbert principle as from the original Hilbert principle. The function of the Lagrange multipliers is to decouple the metric and affine connection variations. A similar procedure is used in classical mechanics if we wish to find the constraint forces which are related to the Lagrange multipliers. Even though the modified Hilbert variational principle gives the same field equations it contains additional information in the Lagrange multipliers[2] and also gives a shorter calculation.

III. PALATINI VARIATIONAL PRINCIPLE

The, so-called, Palatini variational principle is different than the Hilbert principle discussed in the previous section. It uses the action (2.7) but varies with respect to independent variations of the metric and affine connection without imposing any constraint. The field equations are

$$\frac{\delta L}{\delta g_{su}} = -\sqrt{-g}\, T^{su}, \tag{3.1a}$$

[a)] Dedicated to the memory of my father.

 0022-2488/78/1901-0100$1.00

$$\frac{\delta \mathcal{L}}{\delta \Gamma^{p}_{su}} = 0. \tag{3.1b}$$

For the Lagrangian $\sqrt{-g}R$ if we substitute (2.9c) into (2.9b) we obtain

$$\frac{\delta \mathcal{L}}{\delta \Gamma^{p}_{su}} = 0 \quad \text{(identically)} \tag{3.2}$$

and hence

$$\lambda^{su}_{p} = 0. \tag{3.3}$$

Thus, the field equations (2.9a) and (3.1a) are the same in terms of the metric and affine connection. Now (2.9c) is the solution to (3.1b) in this case. Therefore, identical field equations for the metric result from either the Hilbert (modified or ordinary) or Palatini variational principle for the Lagrangian $\sqrt{-g}R$. This is the mathematics behind the Palatini variational principle. Finally, note that these results hold only for the Lagrangian $\sqrt{-g}R$ and therefore there is not, in general, a Palatini variational principle for other Lagrangians.[2]

IV. BACH-LANCZOS IDENTITY

It has been proven using the ordinary Hilbert variational principle that in four dimensions the most general Lagrangian of the form (2.1) whose variational derivative contains no higher than second order derivatives of the metric[5] is

$$\mathcal{L} = \alpha\sqrt{-g}R - 2\lambda\sqrt{-g} + \beta\epsilon^{ijkn}R^{ab}{}_{ij}R_{abkn} + \gamma\sqrt{-g}(R^2 - 4R_{ij}R^{ij} + R_{ijkn}R^{ijkn}), \tag{4.1}$$

where α, β, γ, and λ are arbitrary constants and ϵ^{ijkn} is the permutation symbol.

It is known from Noether's theory that the variational derivative of any $\mathcal{L}$ satisfies the Bianchi identities

$$\sqrt{-g}E^{su} = \frac{\delta \mathcal{L}}{\delta g_{su}}, \tag{4.2}$$

$$E^{su}{}_{;u} = 0. \tag{4.3}$$

Also it has been proven that the only tensor of the form

$$A^{su} = A^{su}(g_{ij}, g_{ij,k}, g_{ij,kn}), \tag{4.4}$$

which satisfies (4.3) in four dimensions is[5]

$$A^{su} = G^{su} + \lambda g^{su}, \tag{4.5}$$

where G^{su} is the Einstein tensor and λ is a constant. It follows that the equations of motion resulting from (4.1) must be the Einstein equations with cosmological constant

$$G^{su} + \lambda g^{su} = T^{su}. \tag{4.6}$$

Because of parity conservation we shall set $\beta = 0$ and for simplicity $\lambda = 0$ and study the resulting Lagrangian

$$\mathcal{L} = \alpha\sqrt{-g}R + \gamma\sqrt{-g}\,(R^2 - 4R_{ij}R^{ij} + R_{ijkn}R^{ijkn}), \qquad \mathcal{L} = \alpha\sqrt{-g}R + \mathcal{L}_1. \tag{4.7}$$

We now wish to obtain the field equations from (4.7) using the modified Hilbert principle. The field equations are (2.9). (2.9a) yields

$$\frac{\delta \mathcal{L}}{\delta g_{su}} + \sqrt{-g}\,\Lambda^{su} = -\sqrt{-g}\,T^{su}, \tag{4.8}$$

or

$$-\alpha\sqrt{-g}\,G^{su} + \frac{\delta \mathcal{L}_1}{\delta g_{su}} + \sqrt{-g}\,\Lambda^{su} = -\sqrt{-g}\,T^{su}. \tag{4.9}$$

Since the modified Hilbert principle gives the same equations of motion as the Hilbert principle, we know the field equations in this case are just the Einstein equations. Hence from (4.9) we have

$$\frac{\delta \mathcal{L}_1}{\delta g_{su}} + \sqrt{-g}\,\Lambda^{su} = AG^{su}\sqrt{-g}, \tag{4.10}$$

where A is a constant. For the first term in (4.10) we obtain

$$\frac{1}{\gamma\sqrt{-g}}\frac{\delta \mathcal{L}_1}{\delta g_{su}} = \tfrac{1}{2}g^{su}(R^2 - 4R_{ij}R^{ij} + R_{ijkn}R^{ijkn}) + (-2RR^{su} + 8R^{s}{}_{n}R^{un} - 2R^{s}{}_{mnk}R^{umnk}). \tag{4.11}$$

In order to calculate Λ^{su} we must calculate $\lambda_p{}^{su}$ using (2.9b) and (2.9c). Combining (2.9b) and (2.9c) gives

$$\frac{\delta \mathcal{L}_1}{\delta \Gamma^{p}_{su}} + \sqrt{-g}\,\lambda^{su}_{p} = 0, \tag{4.12}$$

where we have used the fact that $\sqrt{-g}\,R$ gives no contribution to the Lagrange multiplier. We then construct Λ^{su} using (2.10). The final result is

$$\frac{1}{\gamma}\Lambda^{su} = 4R^{snur}R_{nr} - 4R^{s}{}_{n}R^{un}, \tag{4.13}$$

where to obtain this form for Λ^{su} we must use the identities given in the Appendix. Adding (4.11) and (4.13), Eq. (4.10) yields

$$\tfrac{1}{4}g^{su}(4R_{ij}R^{ij} - R^2 + R_{ijkn}R^{ijkn}) - (2R^{s}{}_{n}R^{un} + 2R^{snur}R_{nr} - RR^{su} - R^{s}_{mnk}R^{umnk}) = (A/\gamma)\,G^{su}, \tag{4.14}$$

In order to evaluate A, we note that under the transformation $g_{ij} \to -g_{ij}$ the left-hand side of (4.14) changes sign whereas the right-hand side does not. Therefore, $A = 0$ and (4.14) is precisely the Bach–Lanczos identity.[6] As far as we are aware, this is the first proof of this identity from a variational point of view. All other derivations of this identity use an algebraic method.[7] Note that since the Hilbert and modified Hilbert principles give the same equations of motion we could derive the Bach–Lanczos identity from the ordinary Hilbert principle. The necessary calculations are given in Ref. 8. It has, however, been our purpose to give an application of the modified Hilbert principle.

V. CONCLUSIONS

We have given a detailed discussion of a modified Hilbert variational principle and shown in two examples how it may be used to arrive at important results. In Sec. II we used it to explain the Palatini variational principle for the Lagrangian $\sqrt{-g}R$. In Sec. IV we used it to study the general Lagrangian for the Einstein equations. This general Lagrangian differs from $\sqrt{-g}R$ by terms quadratic in the curvature tensor $\mathcal{L}_1$. When we apply the modified Hilbert variational principle to this Lagrangian and use the uniqueness of the Einstein equa-

tions, we obtain a variational derivation of the Bach–Lanczos identity. It also follows that the quadratic invariant $\mathcal{L}_1$ satisfies the Hilbert field equations identically and is therefore a divergence. In what follows below we shall mean by the Bach–Lanczos identity either the statement that $\mathcal{L}_1$ is a divergence or (4.14) with $A=0$ since they are equivalent. *One may now say the existence of the Bach–Lanczos identity is due to the fact that there is an invariant which satisfies the Hilbert field equations identically together with the uniqueness of the Einstein equations.* Also it should be mentioned that the Bach–Lanczos is a generalization of the Gauss–Bonnet formula to four-dimensions.

Our proof of the Bach–Lanczos identity was dependent upon two remarkable results concerning the form of the general Lagrangian (4.1) and the fact that the Einstein equations are unique (4.5).[5] Both of these results hold only in four dimensions. Thus, it is not clear if this method of arriving at identities can be used in other theories or generalized to other dimensions.

As a final point we note that the Bach–Lanczos identity has recently attracted attention in discussions associated with the renormalization problems in general relativity. The Bach–Lanczos identity implies the one-loop renormalizability of pure gravitation.[9] Here we have shown that the Bach–Lanczos identity is implied by the uniqueness of classical general relativity together with the existence of $\mathcal{L}_1$. Hence, the uniqueness of classical general relativity implies the Bach–Lanczos identity which in turn implies the one-loop renormalizability of quantum gravity. It is interesting (accidental?) that classical general relativity implies something about quantum gravity.

APPENDIX

In this Appendix we shall give the Lagrange multipliers for the three invariants R^2, $R_{ij}R^{ij}$, and $R_{ijkl}R^{ijkl}$ which make up $\mathcal{L}_1$:

$$R^2:\lambda_p{}^{su}=2g^{su}R_{,p}-\delta^u{}_pR^{,s}-\delta^s{}_pR^{,u}, \tag{A1}$$

$$R^{ij}R_{ij}:\lambda_p{}^{su}=2R^{su}{}_{;p}-\delta^u{}_pR^{sr}{}_{;r}-\delta^s{}_pR^{ur}{}_{;r}, \tag{A2}$$

$$R_{ijkn}R^{ijkn}:\lambda_p{}^{su}=-2R_p{}^{sur}{}_{;r}-2R_p{}^{usr}{}_{;r}. \tag{A3}$$

The Λ^{su} for each invariant can be constructed for each invariant by using (2.10). The Λ^{su} so calculated give the same field equations given for these invariants in Ref. 8. The Λ^{su} in (4.13) can then be obtained from these these by taking the appropriate linear combination $(\Lambda-4\Lambda+\Lambda)$. To reduce the Λ^{su} so obtained to the exact form (4.13) one must use the results

$$R^{snur}{}_{;n;r}=R^{su;r}{}_{;r}-R^{sr;u}{}_{;r}, \tag{A4}$$

$$R^{sn;u}{}_{;n}=\tfrac{1}{2}R^{;su}-R^{snur}R_{nr}+R^s{}_nR^{un}, \tag{A5}$$

$$R^{su}{}_{;su}=\tfrac{1}{2}R^{;n}{}_{;n}, \tag{A6}$$

which can be derived from the Bianchi identities.

[1] J. R. Ray, Nuovo Cimento B **25**, 706 (1975).
[2] J. L. Safko and F. Elston, J. Math. Phys. **17**, 1531 (1976).
[3] U. Lindström, Nuovo Cimento B **35**, 130 (1976).
[4] J. L. Safko, M. Tsamparlis, and F. Elston, Phys. Lett. A **60**, 1 (1977); W. K. Atkins, W. M. Baker, and W. R. Davis, Phys. Lett. A **61**, 363 (1977).
[5] D. Lovelock and H. Rund, *Tensors, Differential Forms and Variational Principles* (Wiley, New York, 1975).
[6] R. Bach, Math. Z. **9**, 110 (1921); C. Lanczos, Ann. Math. (N.Y.) **39**, 842 (1938).
[7] D. Lovelock, Lincei Rend. Sci. Fis. Mat. Not. **42**, 187 (1967).
[8] B. S. DeWitt, in *Relativity Groups and Topology*, edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1964), p. 719.
[9] G. 't Hooft and M. Veltman, Ann. Inst. H. Poincaré **20**, 69 (1974).

Uniqueness connection between charge conjugation and statistics

C. L. Hammer

Ames Laboratory—ERDA, Iowa State University, Ames, Iowa 50011

B. DeFacio

Department of Physics, University of Missiouri—Columbia, Columbia, Missouri 65201
(Received 4 April 1977)

The charge conjugation properties of bilinear quantum field theories are examined in considerable detail. It is shown that the connection between charge conjugation and statistics is unique. The relation between spin and statistics for a large class of these theories and the statistics of unusual fields such as Faddeev–Popov ghost fields and Gupta's regularizing fields with negative norm are discussed.

1. INTRODUCTION

Recently DeFacio and Hammer[1] extended the quantization formalism of Takahashi and Umezawa[2] and Hammer and Tucker[3] to the interacting case. This formalism is interesting because it does not depend upon a Lagrangian or canonical formulation and even applies when such formulations do not exist.[1] Instead it postulates the existence of field equations and a conserved current, which is derivable from them, of the bilinear form,

$$\partial_\mu J_\mu(\psi_1,\psi_2)=0,$$
$$a^* J_\mu(\psi_1,\psi_2)=J_\mu(a\psi_1,\psi_2)=J_\mu(\psi_1,a^*\psi_2), \tag{1}$$

where * indicates complex conjugation; a^* is a complex number. The fields ψ_i are any two solutions to the differential equations

$$D(\partial)\psi=j(x) \quad \text{and} \quad \bar{\psi}\overleftarrow{D}(-\partial)=\bar{j}(x), \tag{2}$$

where $j(x)$ is a source term, $D(\partial)$ is a configuration space differential operator, and

$$\bar{\psi}(x)=[\gamma_4\psi(x)]^\dagger, \quad \bar{j}(x)=[\gamma_4 j(x)]^\dagger, \tag{3}$$

$$D^\dagger(\partial)\gamma_4=-\gamma_4^\dagger D(-\partial), \quad \gamma_4^{-1}=\gamma_4^\dagger. \tag{4}$$

The quantity γ_4 is the $[2(2s+1)\times 2(2s+1)]$-dimensional generalization of the Dirac matrix γ_4 when ψ is a field constructed from symmetric spinors of like indices but otherwise must be chosen to insure the Lorentz covarience of the conserved current. A small, but important, point to keep in mind is that γ_4 and $D(\partial)$ must be put in a form such that Eq. (4) applies since otherwise J_μ will not have the convenient Hermitian property

$$J_i^\dagger(\psi_1,\psi_2)=J_i(\psi_2,\psi_1),$$
$$J_0^\dagger(\psi_1,\psi_2)=J_0(\psi_2,\psi_1). \tag{5}$$

Because J_μ is conserved and has the bilinear form expressed by Eq. (1), all physical operators are also bilinear. Hence the name bilinear quantum field theory.

The existence of a free (in, out) limit is also assumed with a *free* conserved current J_μ^f which satisfies

$$\partial_\mu J_\mu^f(\psi_1^f,\psi_2^f)=\{\bar{\psi}_1^f\vec{D}(\partial)\psi_2^f-\psi_1^f\overleftarrow{D}(-\partial)\psi_2^f\}$$
$$=0, \tag{6}$$

$$a^* J_\mu^f(\psi_1^f,\psi_2^f)=J_\mu^f(a\psi_1^f,\psi_2^f)=J_\mu^f(\psi_1^f,a^*\psi_2^f). \tag{7}$$

Thus the *free* theory is also bilinear. It is worth noting at this point that any quantity a, having the property shown by the right-hand equality of Eq. (7), is a c-number constant since, as seen by Eq. (6), it must have commuted with $(\psi^f)^\dagger$, γ_4, and $D(\partial)$.

Most field theories are bilinear quantum field theories. All cannonical field theories,[1] including quantum electrodynamics and the recent gauge theories,[4] as well as noncannonical theories such as Klauder's ultralocal models[1,5] and theories which contain nonnormalizable c-number transformations, such as Umezawa's boson transformations,[6] are special cases. A study of the charge conjugation properties of bilinear quantum field theories, a subject largely ignored in our previous work,[1,3] is therefore of general significance.

We find that the basic postulates of the theory relating to antiparticle operators can be replaced by the usual q-number charge conjugation relationships

$$\mathcal{C}a_k^\dagger(\mathrm{p})\mathcal{C}^{-1}=\eta_C^* b_k^\dagger(\mathrm{p}), \tag{8}$$

$$\mathcal{C}\psi(x)\mathcal{C}^{-1}=C\psi^\dagger(x)\equiv\psi^c, \tag{9}$$

where a_k, b_k are the usual *free* particle and antiparticle destruction operators; η_C is a complex constant (a phase factor); $\mathcal{C}, C$ are the q-number, c-number charge conjugation operators. The basic postulates along with the usual requirements $\mathcal{C}^\dagger=\mathcal{C}=\mathcal{C}^{-1}$, $C\gamma_4^* C^{-1}=-\gamma_4$ are sufficient to show that $C^*=C^\dagger=C^{-1}$.

We also show that the assumption of charge conjugation invariance of the basic field equations gives information on the connection between spin and statistics not given in the usual derivations.[2,7–11] Specifically the postulates force the fields, which are the solutions to *one class* of field equations, to have *anticommutation* relationships whereas the fields which are solutions to a *second class* of field equations are forced to have *commutation relationships*. The connection between spin and statistics is made by assuming *local* commutation rules in the usual manner. The discussion is in some respects the same as that due to Takahashi.[12]

As specific examples, the field equations of several authors are discussed.[11,13–15] In addition it is shown that this unique connection between charge conjugation and statistics has special implications for the commuta-

 0022-2488/78/1901-0103$1.00

tion properties of unusual fields such as Fadde'ev–Popov[16] ghosts and Gupta's[17] regularizing fields.

The notation used throughout this work will be that a general 4-vector A_μ is written, in terms of real quantities A_i $(i=1,2,3)$ and $A_0=-iA_4$, as

$$A_\mu=\begin{pmatrix}A_i\\ iA_0\end{pmatrix} \text{ and } \partial_\mu=\begin{pmatrix}\partial_i\\ -i\partial_0\end{pmatrix}\equiv\begin{pmatrix}\nabla\\ -i\partial/\partial t\end{pmatrix}. \tag{10}$$

2. BASIC POSTULATES AND ASSUMPTIONS

We assume the existence of a unitary charge conjugation, q-number, operator with the following properties:

1. $C|\Omega>=|\Omega>$, where $|\Omega>$ is the vacuum state; (11)

2. $C^{-1}=C=C^\dagger$; (12)

3. $C a_k^\dagger(\mathbf{p})C^{-1}=\eta_C^* b_k^\dagger(\mathbf{p})$,
$C a_k(\mathbf{p})C^{-1}=\eta_C b_k(\mathbf{p})$, (13)

where a_k and b_k are free particle and antiparticle destruction operators;

4. $C\psi(x)C^{-1}=C\psi^\dagger(x)\equiv\psi^c$, (14)

where C is the c-number charge conjugation matrix and $\psi(x)$ is the Heisenberg configuration field operator, or, in the weak limit sense, the corresponding free field operator.

Because of postulates 1–3, it follows that the n-particle state $|n,+\rangle$ is related to the n-antiparticle state $|n,-\rangle$ by:

$$|n,+\rangle=(\eta_C^*)^n|n,-\rangle, \tag{15}$$

$$|n,-\rangle=(\eta_C^*)^{-n}|n,+\rangle; \tag{16}$$

2. $\langle m,+|n,+\rangle=(\eta_C^*)^n\eta_C^m\langle m,-|n,-\rangle$ (17)

which requires

$$|\eta_C|^2=1 \tag{18}$$

for normalization.

Postulates 2 and 4 establish

$$C^*=C^{-1}. \tag{19}$$

The basic postulates needed from the bilinear quantum field theory, in addition to those mentioned in the introduction, are:

1. $a_k(\mathbf{p})|\Omega\rangle=0$; (20)

2. $z_3^{1/2}a_k(\mathbf{p})\equiv-\text{w-lim}\int d\sigma_\mu J_\mu^f(u_k(\mathbf{p},x),\psi(x))$, (21)

or

$$a_k(\mathbf{p})=-\int d\sigma_\mu J_\mu^f(u_k(\mathbf{p},x),\psi^f(x)),$$
$$=\int d\mathbf{x} J_0^f(u_k(\mathbf{p},x),\psi^f(x)), \tag{22}$$

where $d\sigma_\mu(x)$ is an integration variable over a spacelike surface $\sigma_\mu(x)$ which contains the point x, $z_3^{1/2}$ is a renormalization constant, and $u_k(\mathbf{p},x)$ is any positive energy, c-number solution to the equations

$$D(\partial)u_k(\mathbf{p},x)=0; \tag{23}$$

3. $[a_k(\mathbf{p}),a_l^\dagger(\mathbf{q})]_\pm=-\int d\sigma_\mu J_\mu^f(u_k(\mathbf{p},x),u_l(\mathbf{q},x))$. (24)

For physical fields the right-hand side of Eq. (24) is positive definite, whereas it is negative definite or zero for ghost fields.[16–18]

It will sometimes be convenient to use positive energy, plane wave solutions rather than wavepacket solutions $u_k(\mathbf{p},x)$. We designate those functions as $f_k(\mathbf{p},x)$, and assume the normalization

$$-\int d\sigma_\mu J_\mu^f(f_k(\mathbf{p},x),f_l(\mathbf{q},x))=\rho(E_p)\delta_{k,l}\delta(\mathbf{p}-\mathbf{q}), \tag{25}$$

where $\rho(E_p)$ is an energy density of states factor. Because $u_k(\mathbf{p},x)$ is expandable in terms of $f_k(\mathbf{p},x)$, it follows from the right-hand side of Eq. (25) that

$$\int d\sigma_\mu J_\mu^f(u_k(\mathbf{p},x),u_l(\mathbf{q},x))=\int d\sigma_\mu J_\mu^f(u_l(\mathbf{q},x),u_k(\mathbf{p},x)). \tag{26}$$

As shown in the previous studies,[1,3] these postulates lead to the useful results:

1. $[\psi^f(x),\bar\psi^f(y)]_\pm=G(x-y)$, (27)

where $G(x-y)$ is the homogeneous Green's function defined from the advanced and retarded Green's functions as

$$G(x)=G_a(x)-G_r(x),$$

$$D(\partial_x)G_{a,r}(x-y)=-\delta(x-y); \tag{28}$$

2. Any solution, ψ^f to Eq. (23) satisfies

$$\bar\psi^f(x)=-\int d\sigma_\mu(y)J_\mu^f(\bar\psi^f(y),G(y-x)),$$

or

$$\psi^f(x)=-\int d\sigma_\mu(y)J_\mu^f(G(y-x),\psi^f(y)), \tag{29}$$

where

$$\bar G(y-x)=\gamma_4 G^\dagger(y-x)\gamma_4^\dagger=G(x-y). \tag{30}$$

Central to the discussion is the assumed invariance of the field equations (2) to the charge conjugation operation. If we define

$$C j(x)C^{-1}\equiv j^c(x), \tag{31}$$

charge conjugation of the fields in Eq. (2) implies

$$D(\partial)\psi^c=j^c(x), \tag{32}$$

which by Eq. (14) becomes

$$CD^*(\partial)C^{-1}\psi(x)=C[j^c(x)]^\dagger. \tag{33}$$

Charge conjugation invariance requires the right-hand side to be proportional to $j(x)$. Thus,

$$CD^*(\partial)C^{-1}\psi(x)=-\eta_A^{-1}j(x), \tag{34}$$

where η_A^{-1} is a c-number proportionality constant. The required invariance is assured if

$$CD^*(\partial)C^{-1}=-\eta_A^{-1}D(\partial) \tag{35}$$

and

$$j^c(x)=-\eta_A^{*-1}Cj^\dagger(x). \tag{36}$$

Clearly in this analysis, the factor η_A arises from the structure of the source $j(x)$. Although perhaps more unphysical, it would be possible to postulate Eq. (35) without Eq. (36), that is, without requiring charge conjugation invariance of the interacting theory. Intuitively

one expects η_A to be at most a phase, and as will be seen later, it is limited to be ± 1 by the field theory postulates. But at this stage, without Eq. (36), this phase is undetermined. To see this, let $\eta_A = \exp(i\delta)$. Then Eq. (35) becomes

$$CD^*(\partial)C^{-1} = -e^{-i\delta}D(\partial).$$

If we make the change

$$D(\partial) \Rightarrow D(\partial)\exp[(i/2)(\delta-\theta)],$$

which in no way affects the free field equations, Eq. (35) becomes

$$CD^*(\partial)C^{-1} = -e^{-i\theta}D(\partial),$$

for arbitrary θ. The constraints imposed by Eq. (4) are also satisfied if γ_4 is changed according to

$$\gamma_4 \Rightarrow \gamma_4 \exp[i\tfrac{1}{2}(\delta-\theta)].$$

The choice for η_A is therefore arbitrary for the free field case. A similar arbitrary shift in phase may be for Takahashi's analysis.[12]

However, this ambiguity can be avoided by imposing the additional constraint

$$C\gamma_4^* C^{-1} = -\gamma_4. \tag{37}$$

This fixes the phase of γ_4 and forces $e^{i\delta} = e^{i\theta}$ thereby making η_A unique, and independent of the source term $j(x)$. If $j(x)$ satisfies Eq. (36), then the interacting theory is charge configuration invariant. If $j(x)$ does not satisfy Eq. (36), only the free field limit has the appropriate symmetry. Either way there is only one choice for η_A.

In the following it will be convenient to define the charge conjugate counterparts to the c-number solutions $u_k(\mathrm{p},x)$, $f_k(\mathrm{p},x)$ as

$$u_k^c(\mathrm{p},x) = Cu_k^*(\mathrm{p},x) \equiv v_k(\mathrm{p},x), \tag{38}$$

and

$$f_k^c(\mathrm{p},x) = Cf_k^*(\mathrm{p},x) \equiv g_k(\mathrm{p},x). \tag{39}$$

Because of Eq. (19), these equations can be used to show

$$u_k(\mathrm{p},x) = Cv_k^*(\mathrm{p},x),$$
$$f_k(\mathrm{p},x) = Cg_k^*(\mathrm{p},x). \tag{40}$$

3. ANALYSIS

We begin by taking the charge conjugation transformation of $u_k(\mathrm{p})$ as defined in Eqs. (21) and (22). The charge conjugation postulates, Eqs. (13) and (14), require

$$z_3^{1/2}\eta_C b_k(\mathrm{p}) = -\,\mathrm{w\text{-}lim}\int d\sigma_\mu\, J_\mu^f(u_k(\mathrm{p},x), C\psi^\dagger(x)),$$
$$\eta_C b_k(\mathrm{p}) = -\int d\sigma_\mu\, J_\mu^f(u_k(\mathrm{p},x), C(\psi^f(x))^\dagger),$$
$$= \int d\mathbf{x}\, J_0^f(u_k(\mathrm{p},x), C(\psi^f(x))^\dagger). \tag{41}$$

Similarily, from Eq. (24) we have

$$[b_k(\mathrm{p}), b_l^\dagger(\mathrm{q})]_\pm = -\int d\sigma_\mu J_\mu^f(u_k(\mathrm{p},x), u_l(\mathrm{q},x)). \tag{42}$$

Clearly Eqs. (41) and (42) must be made consistant with each other. It is this requirement that establishes η_A.

To show this relationship, we need a knowledge of the transposition properties of J_μ^f. These can be determined from Eq. (6), which, for the integrand of Eq. (41), becomes

$$\partial_\mu J_\mu^f(u_k(\mathrm{p},x), C(\psi^f(x))^\dagger)$$
$$= \bar{u}_k[\vec{D}(\partial)C(\psi^f)^\dagger] - [\bar{u}_k\overleftarrow{D}(-\partial)]C(\psi^f)^\dagger,$$
$$= \bar{\psi}^f\gamma_4 C^T\vec{D}^T(\partial)\gamma_4^* u_k^* - \bar{\psi}^f\gamma_4 C^T(\vec{D}(-\partial)\gamma_4^* u_k^*), \tag{43}$$

where T indicates the transpose matrix. After some manipulation, Eqs. (4), (19), (35), and (37) give

$$\gamma_4 C^T D^T(\partial)\gamma_4^* = -\eta_A D(-\partial)C^T. \tag{44}$$

This result can be used to rewrite Eq. (43) as

$$\partial_\mu J_\mu^f(u_k, C(\psi^f)^\dagger) = \eta_A \partial_\mu J_\mu^f(\psi^f, C^T C^* v_k), \tag{45}$$

where Eq. (40) has been used to express u_k in terms of the conjugate function v_k.

At this point we establish that $C^T C^* = 1$. To see this, note that in the above proof that if $(\psi^f)^\dagger$ is replaced by $v_l^*(\mathrm{q},x)$,

$$\partial_\mu J_\mu^f(u_k(\mathrm{p},x), u_l(\mathrm{q},x)) = \partial_\mu J_\mu^f(v_l(\mathrm{q},x), \eta_A C^T C^* v_k(\mathrm{p},x)). \tag{46}$$

If the Hermitian conjugate of both sides is taken and the k, l; p, q indices interchanged, we have

$$\partial_\mu J_\mu^f(u_k(\mathrm{p},x), u_l(\mathrm{q},x)) = \partial_\mu J_\mu^f(\eta_A C^T C^* v_l(\mathrm{q},x), v_k(\mathrm{p},x)). \tag{47}$$

Comparison of Eqs. (46) and (47) shows that $\eta_A C^T C^*$ obeys the linearity requirement, Eq. (7). Thus $C^T C^*$ is a c-number constant,

$$C^T C^* = k',$$

whose transpose is

$$C^\dagger C = k'.$$

Thus k' is real and positive. Furthermore

$$\det(C^\dagger C) = \det(C^* C) = \det k' = 1,$$

which for k' real and positive requires $k' = 1$ and

$$C^\dagger = C^{-1} = C^*. \tag{48}$$

Because the matrix transposition can introduce no new functions into Eq. (6), we may conclude

$$J_\mu^f(u_k, C\psi^\dagger) = \eta_A J_\mu^f(\psi, v_k),$$
$$J_\mu^f(u_k(\mathrm{p},x), u_l(\mathrm{q},x)) = \eta_A J_\mu^f(v_l(\mathrm{q},x), v_k(\mathrm{p},x)). \tag{49}$$

This last result, when applied to the plane wave solutions, gives

$$-\eta_A \int d\sigma_\mu\, J_\mu^f(g_k(\mathrm{p},x), g_l(\mathrm{q},x)) = \rho(E_\mathrm{p})\delta(\mathrm{p}-\mathrm{q})\delta_{k,l} \tag{50}$$

and, when applied to Eq. (41), gives

$$\eta_C b_k(\mathrm{p}) = -\eta_A \int d\sigma_\mu\, J_\mu^f(\psi^f(x), v_k(\mathrm{p},x)). \tag{51}$$

This, along with Eqs. (27) and (29), can be used to

show

$$[b_k^\dagger(\mathbf{p}), b_l(\mathbf{q})]_\pm = -|\eta_A|^2 \int d\sigma_\mu J_\mu^f(v_k(\mathbf{p},x), v_l(\mathbf{q},x)),$$
$$= -\eta_A^* \int d\sigma_\mu J_\mu^f(u_l(\mathbf{q},x), u_k(\mathbf{p},x)). \quad (52)$$

These equations are obtained directly from the definition of $a_k(\mathbf{p})$ and the charge conjugation postulates. They must be consistent with the basic quantization postulate, Eq. (24), which with the charge conjugation postulates give the commutation rules expressed by Eq. (42). Comparison shows

$$[b_k^\dagger(\mathbf{p}), b_l(\mathbf{q})]_\pm = \eta_A^* [b_l(\mathbf{q}), b_k^\dagger(\mathbf{p})]_\pm. \quad (53)$$

Therefore $\eta_A^* = \eta_A = \pm 1$, unless the commutators vanish, and

$$[b_l(\mathbf{q}), b_k^\dagger(\mathbf{p})]_{\eta_A} = -\eta_A \int d\sigma_\mu J_\mu^f(v_l(\mathbf{q},x), v_k(\mathbf{p},x)), \quad (54)$$

where the indices have been exchanged as allowed by the symmetry of Eq. (50).

This is our primary result. If $\eta_A = \eta_F = 1$, the free fields obey Fermi statistics. Then, in this theory, the free fermion fields must satisfy a differential equation with the property

$$CD^*(\partial)C^{-1} = -D(\partial).$$

The connection is unique, and by the arguments following Eq. (36), independent of the charge conjugation properties of the source term.

Similarily, if $\eta_A = \eta_B = -1$, the free fields obey Bose statistics. The free boson fields must then satisfy a differential equation with the property

$$CD^*(\partial)C^{-1} = D(\partial).$$

Again the connection is unique and independent of the source term. We may conclude that for physical fields, or fields with strictly negative norms, there are two classes of differential equations, one for fermions and the other for bosons, *providing* the appropriate field equations satisfy Eq. (35).

The connection between spin and statistics occurs in the usual way.[7-11] Because of Eq. (29), we have

$$\bar{f}_k(\mathbf{p},x) = -\int d\sigma_\mu(y) J_\mu^f(f_k(\mathbf{p},y), G(y-x)),$$

and

$$\bar{g}_k(p,x) = -\int d\sigma_\mu(y) J_\mu^f(g_k(\mathbf{p},y), G(y-x)).$$

For these equations to be consistent with the normalization, Eqs. (25) and (50),

$$G(y-x) = \sum_k \int d\mathbf{p}\, \rho^{-1}(E_p)[f_k(\mathbf{p},y)\bar{f}_k(\mathbf{p},x$$
$$+ \eta_A g_k(\mathbf{p},y)\bar{g}_k(\mathbf{p},x)],$$
$$= [\psi(y), \bar{\psi}(x)]_{\eta_A}. \quad (55)$$

This commutator is *local* only if it vanishes for spacelike separations. If this requirement is also imposed, then only a subclass of the solutions to each of the two classes of free field equation is acceptable. If $\eta_A = 1$, this subclass hopefully is the half-integral spin solutions, whereas for $\eta_A = -1$, the subclass hopefully is the integral spin solutions. This will be discussed more completely in the next section.

Some further results, which are straightforward to prove are:

1. If $\psi^f(x)$ is expanded in terms of the plane wave solutions,

$$\psi^f(x) = \sum_k \int d\mathbf{p}\, \rho^{-1}(E_p)[a_k(\mathbf{p})f_k(\mathbf{p},x) + \eta_C^* b_k^\dagger(\mathbf{p})g_k(\mathbf{p},x)], \quad (56)$$

where $a_k(\mathbf{p}), b_k^\dagger(\mathbf{p})$ have been redefined to be

$$a_k(\mathbf{p}) = -\int d\sigma_\mu J_\mu^f(f_k(\mathbf{p},x), \psi^f(x)), \quad (57)$$

$$\eta_C^* b_k^\dagger(\mathbf{p}) = -\eta_A \int d\sigma_\mu J_\mu^f(g_k(\mathbf{p},x), \psi^f(x)), \quad (59)$$

$$[a_k(\mathbf{p}), a_l^\dagger(\mathbf{q})]_{\eta_A} = [b_k(\mathbf{p}), b_l^\dagger(\mathbf{q})]_{\eta_A} = \rho(E_p)\delta_{k,l}\delta(\mathbf{p}-\mathbf{q}). \quad (59)$$

2. The charge conjugation of any operator Q,[1]

$$Q \equiv -\int d\sigma_\mu J_\mu(\psi, q\psi), \quad (60)$$

where q is the corresponding c-number operator, gives

$$Q^c = \int d\sigma_\mu J_\mu(q^c\psi, \psi),$$
$$\equiv \int d\sigma_\mu J_\mu(\psi, q^{c\dagger}\psi), \quad (61)$$

where

$$CQC^{-1} \equiv Q^c,$$
$$Cq^*C^{-1} \equiv q^c. \quad (62)$$

In deriving Eq. (61), the transposition of the Heisenberg fields introduces an additional factor of $-\eta_A$ not present in the derivation leading to Eq. (45).

For the Hamiltonian and momentum operators, $q = i\partial_0$ and $q = -\nabla$, respectively. Equation (61) then yields

$$H^c = -\int d\sigma_\mu J_\mu(\psi, i\partial_0\psi)$$
$$= H$$

and

$$\mathbf{p}^c = -\int d\sigma_\mu J_\mu(\psi, (-i\nabla\psi))$$
$$= \mathbf{p}$$

as required.

4. RELATIVISTIC EQUATIONS

The two classes of wave equation for arbitrary spin, developed by Hammer, McDonald, and Pursey[14] nicely illustrate the above discussion. The "linear" equation is generated from the Foldy type equation for the $2(2s+1)$ component field,

$$E\beta\phi_F^f = i\partial_0\phi_F^f, \quad (63)$$

where β is the $2(2s+1)$ generalization of the Dirac β matrix and $E = (-\nabla^2 + m^2)^{1/2}$. The transformation

$$\psi_F^f(x) = m^s E^{-1/2} S \phi_F^f(x), \quad (64)$$

where S is the Lorentz transformation, is used to generate the field equation for ψ_F^f. If one uses for the Lorentz transformation

$$S = \exp[s\alpha\cdot(\mathbf{q}/q)\operatorname{arctanh}(p/E)], \quad (65)$$

where α is the $2(2s+1)$ generalization of the Dirac α matrix, $\mathbf{q}=\hat{\epsilon}\mathbf{p}$ with $\hat{\epsilon}=1$ for positive energy solutions and $\hat{\epsilon}=-1$ for negative energy solutions, $p=-i\nabla$, $q=p=(-\nabla^2)^{1/2}$, one generates the wave equations of Weaver, Hammer, and Good[15] for particles of arbitrary spin s.

If one uses the hyperplane formalism of Flemming[19] to describe S, then one obtains[14]

$$mp_{[\mu]}\gamma_{[\mu]}\psi_F^f=(i)^{2s}\hat{\epsilon}^{2s+1}(-p_\mu p_\mu)^{s+1/2}\psi_F^f, \tag{66}$$

where

$$\gamma_{[\mu]}=\gamma_{[\mu]}^\dagger,$$
$$p_{[\mu]}\gamma_{[\mu]}\equiv p_{\mu_1}p_{\mu_2}\cdots p_{\mu_{2s}}\gamma_{\mu_1\mu_2\cdots\mu_{2s}}, \tag{67}$$

and $\gamma_{\mu_1\mu_2\cdots\mu_{2s}}$ are the $2(2s+1)$ Dirac-like matrices studied by Barut, Muzinich, and Williams[20] and by Weinberg.[13] These equations are the manifestly covariant form of the Weaver, Hammer, and Good equations.[15] They are particularly simple for half integral spins since $\hat{\epsilon}^{2s+1}=1$ and minimal coupling to other fields can be used[21] since there are no auxiliary conditions to be satisfied. Given the equality

$$\gamma_{[\mu]}p_{[\mu]}\gamma_{[\nu]}p_{[\nu]}=(p_\mu p_\mu)^{2s}, \tag{68}$$

it is straightforward to show

$$(p_\mu p_\mu)^{2s}(p_\mu p_\mu+m^2)\psi_F^f=0. \tag{69}$$

Then ψ_F^f is seen to be a solution of either

$$[p_\mu p_\mu+m^2]\psi_F^f=0 \tag{70}$$

or

$$p_\mu p_\mu\psi_F^f=0, \tag{71}$$

so that ψ_F^f describes both a massive and massless particle. For the massive particle part Eq. (66) reduces to the Weinberg equations,[13]

$$[\gamma_{[\mu]}p_{[\mu]}-(i)^{2s}\hat{\epsilon}^{2s+1}m^{2s}]\psi_F^f=0, \tag{72}$$

which has Eq. (70) as an auxiliary condition. These equations have been shown[11,13] to have nonlocal commutation rules for integral spin because of the $\hat{\epsilon}$ factor. Thus, Eqs. (66) or (72) can be thought of as equations which admit a subclass of solutions that correspond to half-integral spins. The $\gamma_{[\mu]}$ matrices have the properties $\beta=\gamma_{[4]}$ and

$$\gamma_{[\mu]}p_{[\mu]}^*\gamma_{[4]}=\gamma_{[4]}\gamma_{[\mu]}p_{[\mu]},$$

and, for Eq. (66) to be charge conjugation invariant (s half-integral),

$$C_F\gamma_{[\mu]}^*p_{[\mu]}^*C_F^{-1}=(-1)^{2s}\gamma_{[\mu]}p_{[\mu]}. \tag{73}$$

These properties can be used to show that

$$\gamma_4\equiv\gamma_{[4]},$$
$$D(\partial)=m\gamma_{[\mu]}p_{[\mu]}-(i)^{2s}(-p_\mu p_\mu)^{s+1/2}],$$
$$D^\dagger(\partial)\gamma_{[4]}=-\gamma_{[4]}D(-\partial),$$
$$C_F\gamma_{[4]}^*C_F^{-1}=-\gamma_{[4]},$$
$$C_FD^*(\partial)C_F^{-1}=-D(\partial), \tag{74}$$

for s half-integral. Thus $\eta_A=\eta_F=1$ and the fields are Fermion fields.

The second class of equations is generated by the transformation, Eq. (64), on a "second order" Foldy type equation[14]

$$2E^2\phi_B^f=-(1+\beta)\partial_0^2\phi_B^f, \tag{75}$$

which, for $\beta=1$, is the Klein–Gordon equation. If the hyperplane formalism is used to describe S, one obtains[14]

$$[(-i)^{2s}\hat{\epsilon}^{2s}\gamma_{[\mu]}p_{[\mu]}-(-p_\mu p_\mu)^{s-1}(2m^2+p_\mu p_\mu)]\psi_B^f=0. \tag{76}$$

This equation is particularly simple for the integral spin case since $\hat{\epsilon}^{2s}=1$. These equations also have no auxiliary conditions and may be used with minimal coupling.[17] As before, Eq. (68) can be used to show that, for spin greater than 1, ψ_B^f contains both massive and massless solutions. The massive case reduces to Weinberg's equation[13]

$$[(-i)^{2s}\hat{\epsilon}^{2s}\gamma_{[\mu]}p_{[\mu]}-m^{2s}]\psi_B^f=0,$$
$$[p_\mu p_\mu+m^2]\psi_B^f=0. \tag{77}$$

These equations, Eqs. (76) and (77), have also been shown[11,13] to exhibit nonlocal commutation rules when s is half-integral because of the $\hat{\epsilon}$ factor. Thus the acceptable solutions for this class are the subset of integral s solutions.

The charge conjugation properties for this case are (s integral)

$$C_B\gamma_{[\mu]}^*p_{[\mu]}^*C_B^{-1}=\gamma_{[\mu]}p_{[\mu]},$$
$$C_B(i\gamma_{[4]})^*C_B^{-1}=-i\gamma_{[4]}. \tag{78}$$

These properties can be used to show

$$\gamma_4\equiv i\gamma_{[4]},$$
$$D(\partial)=[(-i)^{2s}\gamma_{[\mu]}p_{[\mu]}-(-p_\mu p_\mu)^{s-1}(2m^2+p_\mu p_\mu)]$$
$$=D(-\partial),$$
$$D^\dagger(\partial)\gamma_4=-\gamma_4^\dagger D(-\partial),$$
$$C_B\gamma_4^*C_B^{-1}=-\gamma_4,$$
$$C_BD^*(\partial)C_B^{-1}=D(\partial), \tag{79}$$

for s integral. The acceptable solutions are therefore boson fields, $\eta_A=\eta_B=-1$. The choice for C_B used by Nelson and Good is

$$C_B=C_F\gamma_5, \tag{80}$$

where γ_5 is the $2(2s+1)$ generalization of the Dirac γ_5 matrix.

The fields ψ_F^f and ψ_B^f are closely related to symmetric spinor fields of dotted indices and as such are solutions of the Dirac–Pauli–Fierz equations.[22] The proofs will therefore apply also to these equations.

5. UNUSUAL FIELDS

We have shown for physical bilinear quantum field theories that there is a unique relationship between the free field equations and the statistics of the quantized fields. This is interesting because it emphasizes the caution that must be exercised when setting up a field theory. It would, for example, be totally inappropriate

to postulate that a Fermion field satisfy the Klein–Gordon equation in the interaction free limit. It is clear that interactions should be added only to those field equations that satisfy Eqs. (4), (35), and (37), if the statistics are to be included in the appropriate manner. There are, however, some exceptions, namely those cases for which Eq. (35) is not satisfied or when the field has a zero norm.

Schrödinger fields are one of the exceptions to the present study because nonrelativistic wave-equations are not charge conjugation invariant. The statistics of these fields are only fixed in the sense that they satisfy appropriate relativistic, charge conjugation invariant wave equations whose nonrelativistic limit destroys the charge conjugation invariance. However, the phonon field, the quanta of elastic excitations of a lattice, which is a nonrelativistic phenomenon, is very different. The equation satisfied by the phonon field is a Klein–Gordon-like equation.[23] Since they have a nonzero norm, they must be bosons.

The other exceptions are those fields with vanishing norms. It has recently been shown that Goldstone bosons[18] and the gauge fixing field $B(x)$ of Nakanishi[23] fall into this category. It is possible[24,25] to describe these fields with negative norms without altering the accompanying physical fields. Thus these fields must be bosons.

The Fadde'ev–Popov ghost fields[16] satisfy a massless Klein–Gordon equation, but are interpreted as fermions because of their signature, $\det^{-1}(\eta)$, in the path-integral formulation. They are considered as anticommuting c-numbers

$$[\eta,\bar{\eta}]_+ = [\eta,\eta)_+ = [\bar{\eta},\bar{\eta}]_+ = 0,$$

and therefore they present no contradiction to the results of this paper. However, our work does show that charge conjugation invariance requires that any such q-number ghost field must have a zero norm. Such a field theory will be the subject of a future investigation.

Lastly, the Gupta regularizing fields[17] must have a negative norm to be useful. Since they satisfy the same free field equations as their physical field counterparts, they must have the same statistics. Thus, Gupta's method is consistent with the present study.

ACKNOWLEDGMENT

This work was supported by the U.S. Energy Research and Development Administration, Division of Physical Research.

[1]B. DeFacio and C. L. Hammer, J. Math. Phys. **17**, 267 (1976).
[2]Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, London, 1969); Y. Takahashi and H. Umezawa, Prog. Theor. Physics **9**, 14 (1953); Nucl. Phys. **51**, 193 (1964); and H. Umezawa, University of Alberta, Edmonton, lecture notes on quantum field theory, Chap. II: "Free Field Equations" (1977).
[3]C. L. Hammer and R. H. Tucker, J. Math. Phys. **12**, 1327 (1971).
[4]E. S. Abers and B. W. Lee, Phys. Rep. C **9**, 1 (1973); S. Weinberg, Rev. Mod. Phys. **46**, 255 (1974).
[5]J. R. Klauder, in *Lectures in Theoretical Physics, Mathematical Methods in Theoretical Physics, Vol. XIV.B*, edited by W. Britten (Colorado Associated U. P., Boulder, 1973), pp. 329–421.
[6]H. Umezawa, in *Renormalization and Invariance in Quantum Field Theory*, edited by E. R. Cainiello (Plenum, New York, 1974), pp. 275–328.
[7]W. Pauli, Phys. Rev. **58**, 716 (1940).
[8]G. Luders and B. Zumino, Phys. Rev. **110**, 1450 (1958).
[9]C. L. Hammer and R. H. Good, Jr., Phys. Rev. **111**, 342 (1958).
[10]N. Burgoyne, Nuovo Cimento **8**, 604 (1958).
[11]T. J. Nelson and R. A. Good, Jr., Rev. Mod. Phys. **40**, 508 (1968).
[12]See Ref. 2, pp. 91, 165.
[13]S. Weinberg, Phys. Rev. **133**, B1318 (1964).
[14]C. L. Hammer, S. C. McDonald, and D. L. Pursey, Phys. Rev. **171**, 1349 (1968).
[15]D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135**, B241 (1964).
[16]L. D. Fadde'ev and V. N. Popov, Phys. Lett. B **25**, 29 (1967).
[17]S. N. Gupta, Proc. Roy. Soc. A **64**, 695 (1951); D. Barua and S. N. Gupta, Phys. Rev. D **10**, 3303, 3306 (1974).
[18]C. L. Hammer and B. DeFacio, Phys. Rev. D **10**, 1225 (1974).
[19]G. N. Fleming, J. Math. Phys. **7**, 1459 (1966).
[20]A. O. Barut, I. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963).
[21]R. H. Tucker and C. L. Hammer, Phys. Rev. D **3**, 2448 (1971).
[22]See, for example, H. Umezawa, *Quantum Field Theory* (Interscience, New York, 1956), Chap. IV, Sec. 3.
[23]See esp. pp. 35–44 of Takahashi in Ref. 2.
[24]N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **50**, 1388 (1973).
[25]N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **49**, 640 (1973).

Theory of vibrations of coated, thermopiezoelectric laminae

M. Cengiz Dökmeci[a)]

Massachusetts Institute of Technology, R.5—207, Cambridge, Massachusetts 02139
(Received 10 September 1976)

This study presents a theory for dynamic problems of coated laminae in which there is coupling between mechanical and electrical as well as thermal fields. The laminae is coated completely with perfectly conducting electrodes on both its faces, and it may comprise any number of bonded layers, each with a distinct but uniform thickness, curvature and electromechanical properties. First, a generalized variational theorem is derived so as to describe the complete set of the fundamental equations of thermopiezoelectricity. Next, by the use of this theorem, a system of two-dimensional, approximate governing equations of the coated laminae is constructed for the case when the mechanical displacement, electric potential, and temperature fields vary linearly across the laminae thickness. The effects of elastic stiffnesses of, and the interactions between, layers of the laminae and its electrodes are all taken into account. Also, the uniqueness of the governing equations is examined, and a theorem which includes the conditions sufficient for the uniqueness is given.

1. INTRODUCTION

Observations of elastic strain caused by applying an electrical field as well as mechanical stress, and of electrical polarization produced by impressing mechanical stress instead of an electrical field have been long attributed to certain classes of crystalline solids.[1–4] The latter is called the direct piezoelectric effect discovered by the Curie brothers,[5,6] and the former the converse (also reciprocal or inverse) piezoelectric effect predicted theoretically by Lippmann[7] as a thermodynamic consequence of the direct effect and verified by the Curies.[6,8] The piezoelectric crystalline solids are always anisotropic; of the 32 crystal classes, 21 lack a center of symmetry, and with the exception of one class (the cubic class 432), all of these are piezoelectric.[9] Recently, the observation has also been established of the piezoelectric effects in certain synthetic materials[10] (for instance, piezoelectric ceramics[11]) and a number of biological substances[12,13] (the so-called piezoelectric textures such as wood[14] and bone[15]). The discovery and manufacturing of synthetic piezoelectric materials, and, in particular, of piezoelectric ceramics made it possible to produce a virtually unlimited variety of advantageous shapes for piezoelectric devices. Piezoelectric materials are so extensively used in electro- and hydro-acoustics, communications, and measurement techniques that a whole branch of industry is devoted to the development of piezoelectric devices.[16–19] As a result, it became, in fact, desirable to investigate, both theoretically and experimentally, piezoelectric structures of any geometric shape.

Remarkable interest in dynamic problems of piezoelectric structures seemed to begin with Mindlin's classical paper[20] on vibrations of elastic plates. Following Cauchy[21] and Poisson,[22] and using the integral method of Kirchhoff,[23] Mindlin[24] introduced a systematic method of approximation for deducing one- and two-dimensional equations for rods and plates from the three-dimensional theory of elastodynamics. Then, he and Tiersten[25] applied the method to analyze certain vibrations of piezoelectric crystal plates. From this onward, Mindlin[26–28] and his associates investigated extensively dispersion of waves in, and vibrational modes and frequencies of, a bounded crystal plate. These works have been elaborated by Tiersten[29] in a notable monograph and by Lee and Haines[30] in a comprehensive survey article with an up to date bibliography. Also, the works of Holland and Eer Nisse[31–33] who, in recent years, fruitfully exploited a variational technique in dealing with complicated geometric structures, and those of Bleustein,[34] Paul,[35] Kagawa and Yamabuchi,[36] Schmidt[37] and Keuning[38] are mentioned on vibrations of piezoelectric crystal plates. Besides the dynamic problems, we cite the papers by Paul[39] and Vekovishcheva[40,41] for bending of a rectangular piezoelectric plate. Along the two-dimensional problems of piezoelectricity, worth mentioning is also the work of Rizzo and Nowinski,[42] who presented, following Muskhelishvili[43] and Lekhnitskii,[44] a method of solution for plane piezoelectric problems as a counterpart of the complex variable technique of elasticity. And mention is further made of the paper of Kudriavtsev, Parton and Rakitin[45] on a crack growth problem in piezoelectric medium.

Investigations concerning the analysis of piezoelectric shells have been directed, in general, toward solutions of specific problems. Toupin[46] derived the piezoelectric relations and equations of equilibrium for a polarized, elastic spherical shell. Stephenson[47] reported in detail the radial vibrations of polarized, hollow cylinders; though his analysis can be called uncoupled, his results were in good agreement with those of experiments. Drumheller and Kalnins[48] presented a theory for vibrations of piezoelectric shells of revolution and free vibration solutions for a hollow ceramic cylinder. Also, Haskins and Walsh,[49] Paul,[50] Martin,[51] and Dianov and Kuz'menko[52] studied polarized cylinders, as did Adelman, Stavsky, and Segal.[53,54] Recently, Lazutkin[55] dealt with the vibrations of a ceramic sphere, he and his associates[56–58] with those of polarized piezoelectric ceramic rings, and Viktorov[59] with surface waves on cylindrical crystal surfaces. In addition, the paper which includes high frequency vibrations of piezoelectric shells in the current open litera-

[a)]Present address: Yildiz Posta Caddesi, No:2, Omay Apt., D.1, Besiktas, Istanbul, Turkey.

 0022-2488/78/1901-0109$1.00

ture, was due to Dökmeci[60,61] who presented a linear theory of piezoelectric crystal finite surfaces and theorems of uniqueness in this theory as well.

A layered type of piezoelectric structures is of a new design feature and demand in piezoelectric devices operating by the inverse effect and primarily devised for application in control engineering.[17,18] Greenspan and Wilmotte[62] and Newell[63] proposed one use for this type of structure as did Fry and Dunn[64] and Sittig[65] who also presented a state-of-the-art article[66] to get a better insight on the subject. Schmidt and Voltmer[67] dealt with piezoelectric elastic surface waves in a layered media. Recently, Fahmy and Adler[68,69] described a Thomson–Haskell transfer matrix formalism for studying acoustic surface wave propagation in a planar, multilayer structure. Cheng and Sun[70] considered wave propagation and thickness vibrations in two-layered piezoelectric plates. And more recently, Pauley and Dong[71] presented a finite element analysis for predicting dispersive characteristics in a laminated piezoelectric plate under plane strain conditions. However, the works tended to be limited to planar, layered structures and to disregard the effects of the elastic stiffnesses and inertia of electrodes and the mutual coupling of layers.

To describe continuum problems, variational formulations have been supplied as an alternative for differential formulations which are fundamental and at hand. The former, if possible,[72] may be useful in obtaining approximate solutions, and it may suggest fruitful analogies and generalizations.[73–75] Hence, the foregoing works mostly relied on a variational principle. Variational principles for linear piezoelectricity have been primarily derived by Mindlin,[24] Tiersten and Mindlin,[25] Eer Nisse,[76] Holland and Eer Nisse,[31,77] and then by Tiersten[29] who unified the earlier works by a modification of Hamilton's principle, and Vekovishcheva.[78] By the use of Lagrange multipliers and following Tiersten,[79] Dökmeci[80] presented variational principles which generate all the basic equations of linear piezoelectricity. Recently, Mindlin[28] discussed a variational principle for thermopiezoelectricity, which is analogous to Biot's[81] variational principle for the thermoelastic case. Yet his work is in need of further study to produce the complete set of the basic equations of linear thermopiezoelectricity.

The purpose of this paper is (i) to derive a variational theorem which describes all the basic equations of the classical, linear theory of thermopiezoelectricity, by the use of the theorem; (ii) to construct a theory for the dynamic problems of coated laminae which accounts for coupling of mechanical, electrical, and thermal fields and includes the effect of elastic stiffnesses and inertia of, and all the interactions between, layers and electrodes of the laminae; and then (iii) to present a theorem of uniqueness in the theory, which enumerates the sufficient conditions of the uniqueness.

The remainder of the section introduces the notation to be used in the subsequent development. The next section contains, for ease of quick reference, the principal results of the classical, linear three-dimensional theory of thermopiezoelectricity. In Sec. 3, a description of the laminae geometry is presented. The laminae is of a finite, regular region[82] with no singularities of any type, and it is completely coated with perfectly conducting electrodes on both its lower and upper faces. And it may have a construction comprising any number of bonded piezoelectric layers, each with a distinct but uniform thickness, curvature and electromechanical properties. Also recorded is the relations between space and surface tensors that are needed in the sequel. In Sec. 4, the result of Mindlin[28,83] is extended and a generalized variational theorem is formulated. The theorem generates the divergence, gradient and constitutive equations, and natural boundary and initial conditions for the mechanical, electrical, and thermal fields of linear thermopiezoelectricity. Section 5 deals with the kinematics and all the continuity conditions between layers and electrodes of the laminae and Sec. 6 with the definitions of mechanical stress, load, acceleration, traction, body force, heat flux, entropy resultants and gross electric displacements for the laminae. The distributions of mechanical strain, electric field, and heat flux are introduced in Sec. 7. The constitutive equations corresponding to the resultants are given in Sec. 8. In Sec. 9, by the use of the variational theorem together with a chosen field of mechanical displacement, electric potential, and temperature increment, the macroscopic field equations and the natural boundary and initial conditions of the coated, thermopiezoelectric laminae are constructed in a consistent and systematic manner within the framework of three-dimensional theory of linear thermopiezoelectricity. The field is linearized with respect to the thickness coordinates of layers, and it may accommodate all the types of vibrations of the laminae.[28,48,61,84] This approach may be considered in the same spirit as those of Mindlin[28] and Tiersten.[29] In Sec. 10, a theorem of uniqueness is established for the initial mixed-boundary value problems defined by the governing equations of the laminae. The proof of the theorem is given by means of an argument similar to that used by Weiner.[85] Also, the sufficient conditions to assure the uniqueness are enumerated. A brief discussion of the results, special cases of interest, and extensions to other types of piezoelectric structures are presented in the last section.

Notation

In the paper, standard tensor notation[86,87] is freely used in a Euclidean 3-space Ξ. The θ^i system in the space Ξ is identified with a fixed, right-handed system of geodesic normal coordinates. Latin indices (subscripts or superscripts), except for m and r, have the range 1, 2, 3, and Greek indices 1, 2. The former stands for space tensors, while the latter for surface tensors in the space Ξ. Einstein's summation convention is implied for all repeated Latin and Greek indices, unless they are enclosed with parentheses. The index m (or r) within parentheses represents the mth constituent from the lower face of the laminae, and it takes the values $1, 2, \ldots, N$. Accordingly, 1 (or a prime) and N (or a double prime) are assigned to the lower and upper electrodes of the laminae and $2, 3, \ldots, N-1$ to the layers of the laminae. A comma stands for partial differentiation

with respect to the indicated space coordinate and a superposed dot for time differentiation. A semicolon and a colon are used to designate covariant differentiation with respect to the indicated coordinate, using the space and surface metrics, respectively. Also, an asterisk denotes prescribed quantities and an overbar denotes the field quantities referred to the base vectors of constituent midsurface. Further, the symbol $B(t)$ refers to the region B at time t in the space Ξ, BXT refers to the Cartesian product of the region B and the time interval $T=[t_0, t_1)$, where $t_1 > t_0$ may be infinity, and C_{mn} refers to the functions with derivatives of order up to and including (m) and (n) with respect to space coordinates and time.

2. THREE-DIMENSIONAL LINEAR FUNDAMENTAL EQUATIONS OF THERMOPIEZOELECTRICITY

In the three-dimensional Euclidean space Ξ, let B stand for an arbitrary, simply connected, finite and bounded region of space[82] occupied by the thermopiezoelectric medium at $t=t_0$. The region is referred to by a fixed, right-handed system of geodesic normal[86] coordinates θ^i in this space. The boundary surface of B is denoted by ∂B and its closure (i.e., $B \cup \partial B$) by $\bar{B}$, and the outward unit vector normal to ∂B by $\mathbf{n}$. Also, the complementary regular subsurfaces of ∂B are indicated by (S_u, S_t), (S_σ, S_ϕ), and (S_θ, S_h), that is, $S_u \cup S_t = \partial B$, $S_u \cap S_t = \emptyset$, and so on. Further, let $\bar{B}X[t_0, t_1) = \bar{B}XT$ represent the domain of definitions for the functions of (θ^i, t).

Now, we state the three-dimensional, linear fundamental equations of thermopiezoelectricity[83]:

Divergence equations:

$$T^{ij}{}_{;i} + f^j = \rho b^j, \quad \epsilon_{ijk} T^{jk} = 0 \quad \text{in } \bar{B}XT, \tag{2.1}$$

$$D^i{}_{;i} = 0 \quad \text{in } \bar{B}XT, \tag{2.2}$$

$$h^i{}_{;i} = -\Theta_0 \dot{\eta} \quad \text{in } \bar{B}XT, \tag{2.3}$$

with the definitions

T^{ij} = symmetric, spatial (contravariant) components of the stress tensor,

$\mathbf{f}$ = body force vector per unit volume,

$\mathbf{u}$ = mechanical displacement vector,

$\mathbf{b} = \ddot{\mathbf{u}}$ = acceleration vector,

ρ = mass density,

ϵ_{ijk} = components of the alternating tensor,

$\mathbf{D}$ = electric displacement vector,

$\mathbf{h}$ = heat flux vector,

η = entropy density

Θ_0 = constant, positive, reference temperature.

Gradient equations:

$$S_{ij} = \tfrac{1}{2}(u_{i;j} + u_{j;i}) \quad \text{in } \bar{B}XT, \tag{2.4}$$

$$E_i = -\phi_{,i} \quad \text{in } \bar{B}XT, \tag{2.5}$$

$$h^i = -\kappa^{ij}\Theta_{,j} \quad \text{in } \bar{B}XT, \tag{2.6}$$

where

S_{ij} = components of the symmetric strain tensor,

$\mathbf{E}$ = quasistatic electric field vector,

ϕ = electric potential,

Θ = temperature increment from the reference temperature,

κ^{ij} = components of symmetric, positive semidefinite conductivity tensor.

Constitutive equations:

$$T^{ij} = \frac{\partial G}{\partial S_{ij}}, \quad D^i = -\frac{\partial G}{\partial E_i}, \quad \eta = -\frac{\partial G}{\partial \theta} \quad \text{in } \bar{B}XT, \tag{2.7}$$

where

$$G = G(S_{ij}, E_i, \theta) = U - E_i D^i - \eta\theta$$

= electric Gibbs function[16],

U = potential energy density

A quadratic form of the thermodynamic potential G is given as follows:

$$G = \tfrac{1}{2}C^{ijkl}S_{ij}S_{kl} - C^{ij}E_iE_j - \rho C_v \Theta_0^{-1}\Theta^2 - C^{ijk}E_iS_{jk} - \lambda^i\Theta E_i - \lambda^{ij}S_{ij}\Theta, \tag{2.8}$$

and, in view of Eqs. (2.7), it yields the linear constitutive equations of thermopiezoelectricity in the form:

$$\begin{aligned} T^{ij} &= C^{ijkl}S_{kl} - C^{kij}E_k - \lambda^{ij}\Theta \\ D^i &= C^{ijk}S_{jk} + C^{ij}E_j + \lambda^i\Theta \quad \text{in } \bar{B}XT. \\ \eta &= \lambda^{ij}S_{ij} + \lambda^i E_i + \alpha\Theta \end{aligned} \tag{2.9a}$$

Here, C^{ijkl} are the elastic constants, C^{ijk} the piezoelectric strain constants, C^{ij} the dielectric permittivity, λ^{ij} the thermal stress constants, λ^i the pyroelectric constants, and α the material constant ($\alpha = \rho C_v \Theta_0^{-1}$, where C_v is the specific heat per unit volume). Of these material constants, C^{ijkl} refer to free constants (since C^{ijkl} describe the stress–strain relations when the electric and temperature fields are absent), while the remaining refer to clamped constants.[88] Further, we note that

$$\begin{aligned} &C^{ijkl} = C^{jikl} = C^{klij}, \quad C^{ijk} = C^{ikj}, \\ &C^{ij} = C^{ji}, \quad \lambda^{ij} = \lambda^{ji}. \end{aligned} \tag{2.9b}$$

Boundary conditions:

$$T^{*j} - n_i T^{ij} = 0 \quad \text{on } S_t XT, \quad u_i^* - u_i = 0 \quad \text{on } S_u XT, \tag{2.10}$$

$$\sigma^* - n_i D^i = 0 \quad \text{on } S_\sigma XT, \quad \phi^* - \phi = 0 \quad \text{on } S_\phi XT, \tag{2.11}$$

$$\Theta^* - \Theta = 0 \quad \text{on } S_\theta XT, \quad v^* - n_i h^i = 0 \quad \text{on } S_h XT. \tag{2.12}$$

Here,

T^j = stress vector = $n_i T^{ij}$,

σ = surface charge,

v = normal component of the heat flux (across the surface),

and S_t, S_u, S_σ, S_ϕ, S_θ, and S_h are the surface portions of ∂B where the tractions, displacements, surface charge, electric potential, temperature, and heat flux are prescribed, respectively.

Initial conditions:

$$\begin{aligned}&\mathbf{u}(\theta^i, t_0) - \mathbf{v}^*(\theta^i) = 0, \quad \dot{\mathbf{u}}(\theta^i, t_0) - \mathbf{w}^*(\theta^i) = 0 \\ &\phi(\theta^i, t_0) - \Phi^*(\theta^i) = 0, \quad \Theta(\theta^i, t_0) - \Psi^*(\theta^i) = 0\end{aligned} \quad \text{in } B(t_0). \tag{2.13}$$

Equations (2.1) represent the local balance of momenta (Cauchy's first and second laws of motion), Eqs. (2.2) and (2.5) the charge equation of electrostatics and the electric field—electric potential relations (Maxwell's equations of the quasistatic[89] electric field, where the electric charge density, the conduction current and the rate of change of magnetic induction are neglected for nonconductors at frequencies far below optical frequencies, that is, polarizable but not magnetizable dielectrics), Eq. (2.3) represents the heat conduction equation, Eq. (2.4) the strain-mechanical displacement relations, Eqs. (2.6) Fourier's law of heat conduction (in which the conductivity tensor is symmetric, since a strong magnetic field or Coriolis forces are absent; this is due to Onsager's principle), and Eqs. (2.9) represent the linear constitutive equations. These equations completely describe the linear theory of thermopiezoelectricity, and the uniqueness of their solutions has been assured by the boundary conditions (2.10)—(2.12) and the initial conditions (2.13) as shown by Mindlin.[28]

Variational theorem: By the use of the principle of conservation of energy, a variational theorem due to Mindlin[28] may have the form

$$\begin{aligned}&\int_B [(T^{ij}{}_{;i} + f^j - \rho b^j)\delta u_j - D^i{}_{;i}\delta\phi - (\dot{\eta} + h^i{}_{;i}\Theta_0^{-1})\delta\theta]\, dV \\ &- \int_{\partial B} [(n_i T^{ij} - T^j_*)\delta u_j - (n_i D^i - \sigma_*)\delta\phi \\ &- (n_i h^i - v_*)\Theta_0^{-1}\delta\theta]\, dS = 0\end{aligned} \tag{2.14}$$

which evidently generates the stress equations of motion (2.1), the charge equation of electrostatics (2.2), the equation of heat conduction (2.3), and the natural boundary conditions (2.10)—(2.12).

Continuity conditions: Since multilayer structures of piezoelectricity are of interest, we shall also use the conditions on the interfaces of constituents of the laminae. The so-called continuity conditions are expressed by

$$\begin{aligned}&\mathbf{T}^{(m)} + \mathbf{T}^{(m+1)} = 0, \quad \nu^{(m)} + \nu^{(m+1)} = 0, \\ &\sigma^{(m)} + \sigma^{(m+1)} = 0, \quad \mathbf{u}^{(m)} - \mathbf{u}^{(m+1)} = 0, \\ &\Theta^{(m)} - \Theta^{(m+1)} = 0, \quad \phi^{(m)} - \phi^{(m+1)} = 0, \quad \text{on } S_b X T.\end{aligned} \tag{2.15}$$

These relations state the continuity of displacements, temperature and electric potential on, and tractions, surface charge, and heat flux across, the bonding surface $S_b = S_{m,m+1}$ between the mth and $(m+1)$th constituent.

Further, we record the kinetic energy density k_e,

$$k_e = \tfrac{1}{2}\rho \dot{u}^i \dot{u}_i, \tag{2.16}$$

the Biot generalized free energy density B,

$$B = B(S_{ij}, D_i, \eta) = U - \Theta_0\eta = G + E_i D^i + \eta\Theta \tag{2.17}$$

with

$$T^{ij} = \frac{\partial B}{\partial S_{ij}}, \quad E_i = -\frac{\partial B}{\partial D^i}, \quad \Theta = \frac{\partial B}{\partial \eta}, \tag{2.18}$$

and the dissipation function F,

$$F = \tfrac{1}{2}\kappa^{ij}\Theta_{,i}\Theta_{,j}\Theta_0^{-1}, \tag{2.19}$$

and recall that the individual k_e, B, and F are positive definite, by definition, and initially zero.

3. DESCRIPTION OF THE LAMINAE GEOMETRY

The coated, thermopiezoelectric laminae treated herein is composed of N constituents: two perfectly conducting, lower and upper face electrodes and $(N-2)$ piezoelectric layers between them. Each constituent may possess distinct but uniform thickness $2h_m$, curvature, and electromechanical properties. The laminae occupies a regular, finite region of space, $V + S$, with its entire boundary surface S and overall thickness $2H$. And it is referred to by the θ^i fixed, right-handed system of geodesic normal coordinates in the space Ξ so that the equations

$$\theta^3 = -h_1, \quad \theta^3 = 2H - h_1, \quad f(\theta^1, \theta^2) = 0 \tag{3.1}$$

define the lower and upper faces, S_{1f} and S_{uf}, and edge boundary surface, S_e, of the laminae. The edge boundary surface is taken as a right cylindrical surface whose generators lie along the normal to S_{1f} and S_{uf}, and it intersects them along closed, smooth, and nonintersecting Jordan curves C. The bonding surface between the mth and $(m+1)$th constituent is denoted by $A_{m,m+1}$, and the edge boundary surface and midsurface of the mth constituent by S_e^m and A_m. Also, $\mathbf{n}$ and $\boldsymbol{\nu}$ are used to designate the outward unit vectors normal to the faces and to the edge boundary surface, respectively.

The θ^i-normal coordinate system is located on the midsurface of the lower electrodes, A_1. θ^3 is chosen positively upward and $\theta^3 = 0$ is taken as a reference surface A which coincides with the midsurface A_1. The θ^α coordinate curves form a system of curvilinear coordinates on this surface. Further, we introduce a set of local coordinates θ_m^i situated on the midsurface A_m of the mth constituent. Thus, we have

$$\theta_m^\alpha = \theta^\alpha, \quad \theta_m^3 = \theta^3 - z_m, \quad m = 1, 2, \ldots, N. \tag{3.2}$$

Here, z_m is the distance between the reference surface and the midsurface of the mth constituent. Hence, the parametric equations

$$\theta_m^3 = 0, \quad \theta^3 - z_m = 0 \tag{3.3a}$$

and

$$\begin{aligned}&\theta_m^3 - h_m = 0, \quad \theta_{m+1}^3 + h_{m+1} = 0, \\ &\theta^3 - (z_m + h_m) = 0, \quad \theta^3 - (z_{m+1} - h_{m+1}) = 0\end{aligned} \tag{3.3b}$$

with

$$z_m = \sum_{r=1}^{m} (2 - \delta_{1r} - \delta_{mr})h_r \tag{3.3c}$$

clearly define the midsurface A_m and the bonding surface $A_{m,m+1}$.

With reference to the θ^i-system of normal coordinates,[86] the position vector $\mathbf{R}$ of a generic point P in the laminae space V takes the form

$$\mathbf{R}(\theta^i) = \mathbf{r}(\theta^\alpha) + \theta^3 \mathbf{a}_3(\theta^\alpha) \tag{3.4a}$$

with

$$\mathbf{a}_\alpha \cdot \mathbf{a}_3 = 0, \quad \mathbf{a}_3 \cdot \mathbf{a}_3 = 1. \tag{3.4b}$$

Here, $\mathbf{r}$ represents the position vector of the projection of P on, $\mathbf{a}_\alpha = \mathbf{R}_{,\alpha}(\theta^\beta, 0)$ the covariant base vectors of, and $\mathbf{a}_3$ the unit vector normal to, the reference surface $\mathcal{A}$ $(\theta^3 = 0)$. Thus, the base vectors, and metric and conjugate tensors of the space V are defined by

$$\begin{aligned} &\mathbf{g}_\alpha = \mathbf{a}_\alpha + \theta^3 \mathbf{a}_{3,\alpha} = \mu_\alpha^\beta \mathbf{a}_\beta, \\ &\mathbf{g}^\alpha = (\mu^{-1})_\beta^\alpha \mathbf{a}^\beta, \quad \mathbf{g}^3 = \mathbf{g}_3 = \mathbf{a}_3 = \mathbf{a}^3 \end{aligned} \tag{3.5a}$$

and

$$\begin{aligned} &g_{\alpha\beta} = \mu_\alpha^\sigma \mu_\beta^\nu a_{\sigma\nu}, \quad g^{\alpha\beta} = (\mu^{-1})_\sigma^\alpha (\mu^{-1})_\nu^\beta a^{\sigma\nu}, \\ &g_{\alpha 3} = 0, \quad g_{33} = 1, \end{aligned} \tag{3.5b}$$

where

$$\begin{aligned} &\mu_\beta^\alpha = \delta_\beta^\alpha - b_\beta^\alpha \theta^3, \quad \mu_\sigma^\alpha (\mu^{-1})_\beta^\sigma = \delta_\beta^\alpha, \\ &(\mu^{-1})_\beta^\alpha = [\delta_\beta^\alpha + \theta^3 (b_\beta^\alpha - b_\sigma^\alpha \delta_\beta^\alpha)] \, |\mu_\beta^\alpha|^{-1}, \end{aligned} \tag{3.6}$$

and those of the reference surface $\mathcal{A}$ are

$$\begin{aligned} &a_{\alpha\beta} = a_\alpha \cdot a_\beta = g_{\alpha\beta}(\theta^\sigma, 0), \\ &a^{\alpha\beta} = g^{\alpha\beta}(\theta^\sigma, 0), \quad a^{\alpha\sigma} a_{\sigma\beta} = \delta_\beta^\alpha \end{aligned} \tag{3.7}$$

in which $a_{\alpha\beta}$, $b_{\alpha\beta}$, and $c_{\alpha\beta} = b_{\alpha\sigma} b_\beta^\sigma$ stand for the first, second, and third fundamental forms of $\mathcal{A}$, respectively.

Consider a vector field $\mathbf{P}$ in the space Ξ. The field can be referred to either by the base vectors $(\mathbf{g}_i, \mathbf{g}^i)$ or those $(\mathbf{a}_\alpha, \mathbf{a}_3; \mathbf{a}^\alpha, \mathbf{a}^3)$, and it may be written

$$\mathbf{P} = P^i \mathbf{g}_i = P_i \mathbf{g}^i = \bar{P}_\alpha \mathbf{a}^\alpha + \bar{P}_3 \mathbf{a}^3 = \bar{P}^\alpha \mathbf{a}_\alpha + \bar{P}^3 \mathbf{a}_3. \tag{3.8}$$

The shifted components $(\bar{P}_i, \bar{P}^i)$ of $\mathbf{P}$ are associated with the components (P_i, P^i) by the relations:

$$\begin{aligned} &P_\alpha = \mu_\alpha^\nu \bar{P}_\nu, \quad P^\alpha = (\mu^{-1})_\nu^\alpha \bar{P}^\nu; \quad \bar{P}^\alpha = \mu_\nu^\alpha \bar{P}^\nu, \\ &\bar{P}_\alpha = (\mu^{-1})_\alpha^\nu P_\nu, \quad P_3 = P^3 = \bar{P}_3 = \bar{P}^3. \end{aligned} \tag{3.9}$$

In addition, we record the relations

$$\begin{aligned} &P_{\alpha;\beta} = \mu_\alpha^\nu (\bar{P}_{\nu;\beta} - b_{\nu\beta} \bar{P}^3), \quad P^\alpha{}_{;\beta} = (\mu^{-1})_\nu^\alpha (\bar{P}^\nu{}_{;\beta} - b_\beta^\nu \bar{P}^3), \\ &P_{\alpha;3} = \mu_\alpha^\nu \bar{P}_{\nu,3}, \quad P_{3;\alpha} = \bar{P}_{3,\alpha} + b_\alpha^\sigma \bar{P}_\sigma, \\ &P^\alpha{}_{;3} = (\mu^{-1})_\nu^\alpha \bar{P}^\nu{}_{,3}, \quad P^3_{;\alpha} = \bar{P}^3{}_{,\alpha} + b_{\alpha\sigma} \bar{P}^\sigma, \\ &P^3{}_{;3} = P_{3;3} = P_{3,3} = \bar{P}^3{}_{,3} = \bar{P}_{3,3} \end{aligned} \tag{3.10}$$

for later convenience. Here, a semicolon and a colon are used to designate covariant differentiation with respect to the indicated coordinate by the use of space and surface metrics, respectively.

Further, the elements of volume dV, of surface dS on $\mathcal{S}$, of area dA on $\mathcal{A}$ and of line ds along $\mathcal{C}$ are of the forms:

$$\begin{aligned} &dV = \sqrt{g}\, d\theta^1 d\theta^2 d\theta^3 = dS\, d\theta^3 = \mu\, dA\, d\theta^3 \\ &n_\alpha dS = \mu \nu_\alpha ds\, d\theta^3 \end{aligned} \tag{3.11}$$

with

$$\begin{aligned} &\mu = |\mu_\beta^\alpha| = (g/a)^{1/2} = 1 - 2\theta^3 H + (\theta^3)^2 K, \\ &a = |a_{\alpha\beta}|, \quad g = |g_{ij}|. \end{aligned} \tag{3.12}$$

Here, $H = \frac{1}{2} b_\alpha^\alpha$ and $K = |b^\alpha{}_\beta|$ are the mean and Gaussian curvature of the reference surface. In the foregoing relations, μ_β^α and its inverse $(\mu^{-1})_\beta^\alpha$ are of particular importance. They play the role of shifters in the space of normal coordinates, and they do exist when $|\theta^3| < |R_{\min}|$, where $R_{\min}$ denotes the least principal radius of curvature. This sufficient condition is evidently satisfied by the fundamental assumption

$$2H < |R_{\min}| \tag{3.13}$$

which allows the laminae to be treated as a two-dimensional continuum.

4. VARIATIONAL THEOREMS FOR LINEAR THERMOPIEZOELECTRICITY

In the variational theorem (2.14) given for the thermopiezoelectric region of space, $B + \partial B$, with its boundary surface ∂B, the admissible states $\Lambda_M = [\mu_i, \phi, \theta]$ are required to meet the mechanical displacement—strain relations (2.4), the electric field—electric potential relations (2.5), Fourier's law of heat conduction (2.6), the constitutive equations (2.9), the boundary conditions of displacements, elastic potential and temperature, (2.10)—(2.12), and the initial conditions (2.13) as constraints. However, to use variational theorems in which the admissible states satisfy either no constraints or as few constraints as possible is, in general, desirable in applications. Hence, the constraints put aside in theorem (2.14) can be removed, for instance, through Friedrichs's transformation described in the book of Courant and Hilbert.[90] De Veubeke[91,92] illustrated this technique of removing constraints in the classical theory of elasticity, as Tiersten[29,79] systematically did in the linear theory of piezoelectricity. Due to its familiarity and relative simplicity the technique was of wide use in the literature.[93] Another is the method of the "mirror equation" or the "adjoint equation method" of Morse and Feschbach,[73] used extensively in continuum physics. In removing variational constraints, we further mention the "quasi-variational method" of Biot,[81] the "restricted variational methods" of Rosen,[94] Glandsdorf and Prigogine,[95] and recently the Gurtin method of "convolution."[96,97]

To begin with, we integrate equation (2.14) from $t = t_0$ to $t = t_1$, and then define the functional J,

$$J = J_{i\alpha}^{i\alpha}. \tag{4.1a}$$

Here, the first variations of $J_{i\alpha}^{i\alpha}$ are given by

$$\begin{aligned} &\delta J_{11}^{11} = \int_T dt \int_B (T^{ij}{}_{;i} + f^j - \rho b^j) \delta u_j \, dV, \\ &\delta J_{12}^{12} = \int_T dt \Big\{ \int_{S_t} (T^{*j} - n_i T^{ij}) \delta u_j \, dS \\ &\qquad + \int_{S_u} (u_i - u_i^*) \delta T^i \, dS \Big\}, \end{aligned} \tag{4.1b}$$

$$\begin{aligned} &\delta J_{21}^{21} = \int_T dt \int_B - D^i{}_{;i} \delta\phi \, dV, \\ &\delta J_{22}^{22} = \int_T dt \Big\{ \int_{S_\sigma} (n_i D^i - \sigma^*) \delta\phi \, dS + \int_{S_\phi} (\phi - \phi^*) \delta\sigma \, dS \Big\}, \end{aligned} \tag{4.1c}$$

and

$$\begin{aligned} &\delta J_{31}^{31} = \int_T dt \int_B - (\dot{\eta} + \Theta_0^{-1} h^i{}_{;i}) \delta\Theta \, dV, \\ &\delta J_{32}^{32} = \int_T dt \Big\{ \int_{S_h} \Theta_0^{-1} (n_i h^i - v^*) \delta\Theta \, dS \\ &\qquad + \int_{S_\theta} n_i (\Theta - \Theta^*) \delta h^i \, dS \Big\} \end{aligned} \tag{4.1d}$$

for each $t \in T = [t_0, t_1)$. Since the variations of $\delta\Lambda_M = [\delta u_i, \delta\phi, \delta\theta]$ in B, and those of δu_i, $\delta\phi$, $\delta\Theta$, T^i, $\delta\sigma$, and δh_i, in this order, on S_u, S_ϕ, S_θ, S_t, S_σ, and S_h, are arbitrary and independent, we have the field equations (2.1)—(2.3) and the natural boundary conditions

(2.10)–(2.12), as appropriate Euler equations, for the case when $\delta J = 0$.

Next, following the methodology of Tiersten[79] and De Veubeke,[92] the variational theorem (4.1) is suitably modified in order to include the rest of the fundamental equations of linear thermopiezoelectricity. Thus, by the use of the dislocation potentials and the Lagrange multipliers for each subsidiary condition, we readily remove all the constraints as in a recent paper by Dökmeci.[80] Hence, we conclude a generalized variational theorem as follows.

Theorem: Let $B + \partial B$ denote a regular, finite, thermopiezoelectric region of space in Ξ, with its entire boundary surface ∂B ($S_u \cup S_t = S_\sigma \cup S_\phi = S_h \cup S_\theta = \partial B$ and $S_u \cap S_t = S_\sigma \cap S_\phi = S_\theta \cap S_h = \emptyset$) and its closure B, and define a functional $I(\Lambda_D)$ whose first variation is given by

$$\delta I = \delta J^{i\alpha}_{i\alpha} + \delta I^{ij}_{ij} \tag{4.2a}$$

with

$$\Lambda_D = [u_i \in C_{12},\ S_{ij} \in C_{00},\ T^{ij} \in C_{10};\ D^i \in C_{10},\ E_i \in C_{00},\ \phi \in C_{10};\ h_i \in C_{10},\ \Theta \in C_{10},\ \eta \in C_{01}] \text{ in } \bar{B}XT, \tag{4.2b}$$

and

$$\delta I^{11}_{11} = \int_T dt \int_B \left[T^{ij} - \frac{1}{2}\left(\frac{\partial G}{\partial S_{ij}} + \frac{\partial G}{\partial S_{ji}}\right)\right]\delta S_{ij}\, dV, \tag{4.2c}$$

$$\delta I^{12}_{12} = \int_T dt \int_B [S_{ij} - \tfrac{1}{2}(u_{i;j} + u_{j;i})]\delta T^{ij}\, dV, \tag{4.2d}$$

$$\delta I^{13}_{13} = \int_B \{\rho[\dot{u}^i(\theta, t_0) - w^{*i}(\theta)]\delta u_i(\theta, t_0) + \rho[u^i(\theta, t_0) - v^{*i}(\theta)]\delta \dot{u}_i(\theta, t_0)\} dV, \tag{4.2e}$$

and

$$\delta I^{21}_{21} = \int_T dt \int_B - \left(D^i + \frac{\partial G}{\partial E_i}\right)\delta E_i\, dV, \tag{4.2f}$$

$$\delta I^{22}_{22} = \int_T dt \int_B - (E_i + \phi_{,i})\delta D^i\, dV, \tag{4.2g}$$

$$\delta I^{23}_{23} = \int_B [\phi(\theta, t_0) - \Phi^*(\theta)]\delta\phi(\theta, t_0)\, dV, \tag{4.2h}$$

and

$$\delta I^{31}_{31} = \int_T dt \int_B - \left(\eta + \frac{\partial G}{\partial \Theta}\right)\delta\eta\, dV, \tag{4.2i}$$

$$\delta I^{32}_{32} = \int_T dt \int_B - (h^i + \kappa^{ij}\Theta_{,j})\delta h_i\, dV, \tag{4.2j}$$

$$\delta I^{33}_{33} = \int_B [\Theta(\theta, t_0) - \Psi^*(\theta)]\delta\Theta(\theta, t_0)\, dV. \tag{4.2k}$$

Then, of all the admissible states Λ_D, only those which admit the functional I have zero first variation, if and only if, they satisfy the divergence equations [the stress equations of motion, (2.1), the charge equation of electrostatics, (2.2), and the heat conduction equation (2.3)], the gradient equations [the mechanical displacement–strain relations, (2.4), the electric field–electric potential relations, (2.5), and Fourier's law of heat conduction, (2.6)], the constitutive equations of stress, electric displacement and entropy, (2.7), the natural boundary conditions, (2.10)–(2.12), and the initial conditions (2.13) as appropriate Euler equations.

Further, in view of Eqs. (2.9), (4.2c), (4.2f) and (4.2i), we get the linearized form of $\delta I^{11}_{(i)1}$ as

$$\delta I^{11}_{11} = \int_T dt \int_B [T^{ij} - (C^{ijkl}S_{kl} - C^{kij}E_k - \lambda^{ij}\Theta)]\delta S_{ij}\, dV,$$
$$\delta I^{21}_{21} = \int_T dt \int_B [D^i - (C^{ijk}S_{jk} + C^{ij}E_j + \lambda^i\Theta)]\delta E_i\, dV, \tag{4.3}$$
$$\delta I^{31}_{31} = \int_T dt \int_B [\eta - (\lambda^{ij}S_{ij} + \lambda^i E_i + \alpha\Theta)]\delta\eta\, dV,$$

which lead to the linear constitutive equations of thermopiezoelectricity.

The variational theorem $\delta I(\Lambda_D) = 0$ evidently generates, by the use of the fundamental lemma of the calculus of variations, the complete set of the basic equations of thermopiezoelectricity, and it contains certain earlier variational theorems [29,80] of piezoelectricity as well as Mindlin's theorem,[28] as special cases.

5. DISPLACEMENT, POTENTIAL, AND TEMPERATURE FIELDS. CONTINUITY CONDITIONS

Due to the foregoing suitable regularity and smoothness, and basic assumption (3.13), and absence of any kind of singularities for the laminae region in the space Ξ, the laminae is considered to be a two-dimensional continuum, and all the field quantities together with their derivatives, are assumed to exist, and be single-valued and continuous functions in VXT, where V is the closure of the laminae region. Among the fields, those of displacements, electric potential, and temperature are chosen as a basis for the systematic derivation of two-dimensional, approximate theory of the coated, thermopiezoelectric laminae. And then, in accordance with Eq. (3.13), they are represented by series expansions.

A. Displacement field

Referring to the displacement vector $\mathbf{u}^m$ of a constituent by the base vectors of the midsurface $\mathcal{A}$, we have

$$\mathbf{u} = \bar{u}_\alpha \mathbf{a}^\alpha + \bar{u}_3 \mathbf{a}^3 \tag{5.1}$$

in the notation of Eq. (3.8), with its components represented by[98,99]

$$\bar{u}_i(\theta, t) = \sum_{r=1}^{R} P_r(\lambda) u_i^{(r)}(\theta^\alpha, t) \tag{5.2a}$$

with

$$\lambda = \theta^3_m/2H, \tag{5.2b}$$

where P_r are the Legendre polynomials defined by

$$P_0(\lambda) = 1,\quad P_1(\lambda) = \lambda,$$
$$P_r(\lambda) = (1/2^r r!)d^r(\lambda^2 - 1)^r/d\lambda^r \quad \text{(no sum over } r \text{ and } \lambda). \tag{5.2c}$$

In Eq. (5.2), the integer R is called the order of approximation, and $R = 1$ is the closest to the classical theory of shells.[84] Thus, only the zeroth and first order terms are retained in the series expansion (5.2), and the displacement field linearized with respect to the thickness coordinate is used in the analysis, namely

$$\bar{u}^m_\sigma(\theta, t) = \alpha^m_\sigma(\theta^\nu, t) + \theta^3\beta^m_\sigma(\theta^\nu, t),$$
$$\bar{u}^m_3(\theta, t) = \alpha^m_3(\theta^\nu, t) + \theta^3\beta^m_3(\theta^\nu, t) \tag{5.3a}$$

with

$$\alpha_i^m \equiv u_i^{(0)m} - z_{(m)} u_i^{(1)m}, \quad \beta_i^{(m)} \equiv u_i^{(1)m}. \tag{5.3b}$$

Here, α_σ characterizes the extensional motions (or the stretching), α_3 and β_α the flexural motions (or the bending), and β_3 the thickness stretching of the laminae. By the use of this representation, we automatically abrogate the Kirchhoff–Love hypothesis of classical shells in which $\beta_3 = 0$ and $\beta_\sigma = -(\alpha_{3,\sigma} + b_\sigma^\nu \alpha_\nu)$, that is, only α_i are chosen independently, and its contradictions. Hence, the effects of transverse shear and transverse normal strains, the rotatory inertia and the coupling of adherent constituents are all taken into account. Further, we note that Eqs. (5.3) were shown to account for the coupled vibrations of piezoelectric shells (see, e.g., Refs. 61 and 48, where $\beta_3 = 0$ is assumed) and single and layered plates.[26,29,70]

B. Electric potential and temperature fields

The electric potential and temperature fields of a constituent are represented by a truncated form of the Legendre polynomials of order r, (5.2c), as

$$\phi^m(\theta, t) = \kappa^m(\theta^\alpha, t) + \theta^3 \zeta^m(\theta^\alpha, t) \tag{5.4}$$

and

$$\Theta^m(\theta, t) = \tau^m(\theta^\alpha, t) + \theta^3 \gamma^m(\theta^\alpha, t) \tag{5.5}$$

which are linearized with respect to the thickness coordinate in consistence with the displacement field (5.3).

C. Continuity conditions

The constituents of the laminae are well attached one to each other, and relative deformations at interfaces are prevented. Hence, the continuity of displacements, (2.15), is imposed at interfaces. To this end, using this equation and Eq. (5.3), the continuity is expressed by

$$\alpha^m + (z_m + h_m)\beta^{(m)} = \alpha^{m+1} + (z_{m+1} - h_{m+1})\beta^{(m+1)} \quad \text{on } \mathcal{A}_{m,m+1}XT, \quad m = 1, 2, \ldots, N-1. \tag{5.6}$$

This represents $3(N-1)$ constraints. Because of the constraints, the number of independent displacement functions, $6N$, in Eq. (5.3) is now reduced to $3(N+1)$. And they are suitably chosen as

$$\alpha^1 \in C_{12} \text{ and } \beta^m \in C_{12}, \quad m = 1, 2, \ldots, N. \tag{5.7}$$

Equation (5.6) is now solved for the rest of the displacement functions,

$$\alpha^m = \alpha^1 + \sum_{r=1}^{m} z_{rm}\beta^r, \quad m = 2, 3, \ldots, N, \tag{5.8a}$$

with

$$z_{rm} = (2 - \delta_{1r} - \delta_{mr})h_{(r)} - \delta_{mr} z_{(m)} \tag{5.8b}$$

in terms of the independent displacement functions (5.7).

In a similar manner, upon substituting Eqs. (5.4) into Eq. (2.15), we express the continuity of electric potential in the form

$$\kappa^m + (z_m + h_m)\zeta^{(m)} = \kappa^{m+1} + (z_{m+1} - h_{m+1})\zeta^{(m+1)} \quad \text{on } \mathcal{A}_{m,m+1}XT, \quad m = 2, 3, \ldots, N-2 \tag{5.9a}$$

at the interfaces of layers, and

$$\phi' = \kappa^2 + (z_2 - h_2)\zeta^2 \quad \text{on } \mathcal{A}_{12}XT, \tag{5.9b}$$

$$\phi'' = \kappa^{N-1} + (z_{N-1} + h_{N-1})\zeta^{N-1} \quad \text{on } \mathcal{A}_{N-1,N}XT \tag{5.9c}$$

at the interfaces of the lower electrode and the second constituent (i.e., first layer) and that of the upper electrode and the $(N-1)$th constituent [i.e., $(N-2)$th layer]. In Eqs. (5.9), apparently, $\zeta^1 = \zeta^N = 0$, since the electrodes are perfectly conducting, and $\phi' \equiv \phi^1$, $\phi'' \equiv \phi^N$, and $\phi^1 = -\phi^N$. We choose the independent functions of electric potential,

$$\zeta^m \in C_{10}, \quad m = 2, 3, \ldots, N-2 \tag{5.10}$$

and then solve Eqs. (5.9) for the dependent functions to obtain

$$\begin{aligned} &\kappa^2 = \phi' - (z_2 - h_2)\zeta^2, \\ &\kappa^m = \phi' + \sum_{r=1}^{m} z_{rm}\zeta^r, \quad m = 2, 3, \ldots, N-2, \\ &\kappa^{N-1} = -z_{N-1}/h_{N-1}\phi'' + (1 + z_{N-1}/h_{N-1}) \sum_{r=2}^{N-2} h_r \zeta^r, \\ &\zeta^{N-1} = 1/h_{N-1}\left(\phi'' - \sum_{r=2}^{N-2} h_r \zeta^r\right) \end{aligned} \tag{5.11}$$

in terms of the $(N-3)$ independent functions of Eq. (5.10).

As before, from Eqs. (2.15) and (5.5), we have

$$\tau^m + (z_m + h_m)\gamma^{(m)} = \tau^{(m+1)} + (z_{m+1} - h_{m+1})\gamma^{(m+1)} \quad \text{on } \mathcal{A}_{m,m+1}XT, \quad m = 1, 2, \ldots, N-1 \tag{5.12}$$

for the continuity of temperature at the interfaces of constituents. Solving this equation for the dependent functions chosen, we get

$$\tau^m = \tau^1 + \sum_{r=1}^{m} z_{rm}\gamma^r, \quad m = 2, 3, \ldots, N \tag{5.13a}$$

in terms of the independent functions of temperature:

$$\tau^1 \in C_{10} \text{ and } \gamma^m \in C_{10}, \quad m = 1, 2, \ldots, N, \tag{5.13b}$$

the number of which is now equal to $(N+1)$ instead of $2N$ in Eq. (5.5).

In closing, we note that the expansions for the fields of mechanical displacement, electric potential, and temperature, (5.3)–(5.5), and the gradient equations, (2.4)–(2.6), obviously imply a distribution of mechanical strain and heat flux,

$$\{S_{ij}, h^i\} = \sum_{r=0}^{R} (\theta^3)^r \{{}_rS_{ij}, {}_rh^i\} \tag{5.14}$$

for each constituent, and that of electric field,

$$E_i = \sum_{r=0}^{R} (\theta^3)^r {}_rE_i \tag{5.15}$$

for each layer. The mechanical strain, electric field, and heat flux of order r, ${}_rS_{ij}$ $(r = 0, 1, 2)$ and ${}_rh^i$ and ${}_rE_i$ $(r = 0, 1)$, are functions of θ^α and t, only. Their explicit forms are obtained in Sec. 7.

6. MECHANICAL, ELECTRICAL, AND THERMAL RESULTANTS

Already in Sec. 5, all the field quantities are considered not to vary widely across the laminae thickness, and hence the linearization of mechanical dis-

placement, electric potential, and temperature fields is introduced with respect to the thickness coordinate. Accordingly, various field quantities needed in the subsequent analysis are now averaged through the thickness of each constituent, and they are introduced below. Thus, we define the two-dimensional stress resultants:

$$[N^{\alpha\beta}, M^{\alpha\beta}, K^{\alpha\beta}] = \int_z \mu[1, \theta^3, (\theta^3)^2] T^{\alpha\beta} d\theta^3,$$
$$(Q^\alpha, R^\alpha) = \int_z \mu(1, \theta^3) T^{\alpha 3} d\theta^3, \tag{6.1}$$
$$N^{33} = \int_z \mu T^{33} d\theta^3, \quad \mathcal{N}^{\alpha\beta} = N^{\alpha\beta} - b^\beta_\nu M^{\alpha\nu},$$
$$\mathcal{M}^{\alpha\beta} = M^{\alpha\beta} - b^\beta_\nu K^{\alpha\nu},$$

body force resultants:

$$[\mathcal{F}^i, \mathcal{G}^i, \mathcal{L}^i] = \int_z \mu[1, \theta^3, (\theta^3)^2] f^i d\theta^3,$$
$$F^i = \mathcal{F}^i - b^\alpha_\beta \mathcal{G}^\beta \delta^i_\alpha, \quad G^i = \mathcal{G}^i - b^\alpha_\beta \mathcal{L}^\beta \delta^i_\alpha, \tag{6.2}$$

gross displacements:

$$[\mathcal{A}^i, \mathcal{B}^i, C^i] = \int_z \mu[1, \theta^3, (\theta^3)^2] u^i d\theta^3, \tag{6.3}$$

acceleration resultants:

$$A^i = \rho(\ddot{\mathcal{A}}^i - b^\alpha_\beta \ddot{\mathcal{B}}^\beta \delta^i_\alpha), \quad B^i = \rho(\ddot{\mathcal{B}}^i - b^\alpha_\beta \ddot{C}^\beta \delta^i_\alpha), \tag{6.4}$$

which, using Eqs. (3.12) and (5.3), are written explicitly as

$$A^i = \rho(\mu_0 \ddot{\alpha}^i + \mu_1 \ddot{\beta}^i), \quad B^i = \rho(\mu_1 \ddot{\alpha}^i + \mu_2 \ddot{\beta}^i),$$

where

$$\mu_n = I_n - 2I_{n+1}H + I_{n+2}K,$$
$$I_n = [(z+h)^{n+1} - (z-h)^{n+1}]/(n+1), \quad n = 0, 1, 2, \cdots, \tag{6.5b}$$

traction resultants:

$$(q^i/p^i) = (+/-)\{(\mu n_3)[T^{3i} - (z +/- h) b^\nu_\alpha T^{3\alpha} \delta^i_\nu]\}_{|\theta^3 = z+/-h} \tag{6.6}$$

and

$$[\mathcal{N}^i, \mathcal{M}^i, K^i] = \int_z \mu[1, \theta^3, (\theta^3)^2] T^i d\theta^3,$$
$$N^i = \mathcal{N}^i - b^\alpha_\beta \mathcal{M}^\beta \delta^i_\alpha, \quad M^i = \mathcal{M}^i - b^\alpha_\beta K^\beta \delta^i_\alpha, \tag{6.7}$$

load resultants:

$$\mathbf{P} = \mathbf{q} - \mathbf{p}, \quad \mathbf{S} = (z+h)\mathbf{q} - (z-h)\mathbf{p} \tag{6.8}$$

and

$$\mathcal{Q}' = \sum_{r=1}^{N} \mathbf{N}^r, \quad \mathcal{R}' = \mathbf{M}' + h' \sum_{r=2}^{N} \mathbf{N}^r,$$
$$\mathcal{R}^m = \mathbf{M}^m + \sum_{r=m+1}^{N} z_{mr} \mathbf{N}^r, \quad m = 2, 3, \ldots, N-1, \tag{6.9}$$
$$\mathcal{R}'' = \mathbf{M}'' - (z'' - h'')\mathbf{N}'',$$

gross electric displacements:

$$(\mathcal{C}^i, \mathcal{D}^i) = \int_z \mu(1, \theta^3) D^i d\theta^3, \tag{6.10}$$

surface charge resultants:

$$e = (\mu n_3 D^3)_{|\theta^3 = z+h'}, \quad d = -(\mu n_3 D^3)_{|\theta^3 = z-h}$$
$$\mathcal{e} = (z+h)e, \quad \mathcal{d} = (z-h)d, \tag{6.11}$$

edge-surface charge resultants:

$$(\Lambda, \P) = \int_z \mu\sigma(1, \theta^3) d\theta^3 \tag{6.12}$$

and

$$\Upsilon_2 = \P_2 - (h_2/h_{N-1})\P_{N-1} + \sum_{r=2}^{N-1} x_{2r}\Lambda_r,$$
$$\Upsilon_m = \P_m - (h_m/h_{N-1})\P_{N-1} + \sum_{r=m}^{N-1} x_{mr}\Lambda_r \tag{6.13}$$
$$\Upsilon_{N-2} = \P_{N-2} - (h_{N-2}/h_{N-1})\P_{N-1} - (z_{N-2} - h_{N-2})\Lambda_{N-2} + (1 + z_{N-1}/h_{N-1})h_{N-2}\Lambda_{N-1}$$

heat flux resultants:

$$(H^i, K^i) = \int_z \mu(1, \theta^3) h^i d\theta^3, \tag{6.14}$$

entropy resultants:

$$(\mathcal{P}, \mathcal{J}) = \int_z \mu(1, \theta^3)\eta d\theta^3, \tag{6.15}$$

surface heat flux resultants:

$$k = (\mu n_3 h^3)_{|\theta^3 = z+h'}, \quad l = -(\mu n_3 h^3)_{|\theta^3 = z-h'}$$
$$\mathcal{k} = (z+h)k, \quad \mathcal{l} = (z-h)l, \tag{6.16}$$
$$(k^*, l^*) = (\mu v^*)_{|\theta^3 = (z+h,\ z-h)},$$

edge-surface heat flux resultants, of functions of θ^α and t, only:

$$(\mathcal{H}, \mathcal{K}) = \int_z \mu(1, \theta^3) v\, d\theta^3 \tag{6.17}$$

and

$$\mathcal{N} = \sum_{r=1}^{N} H^r, \quad \mathcal{M}' = K' + h' \sum_{r=2}^{N} H^r,$$
$$\mathcal{M}'' = K'' - (z'' - h'')H'', \tag{6.18}$$
$$\mathcal{M}^m = K^m + \sum_{r=2}^{N} z_{mr} H^r, \quad m = 1, 2, \ldots, N,$$

where Z is used to designate the interval $[z-h,\ z+h]$.

In view of the definitions in Eqs. (6.6), (6.11), and (6.16), the continuity of tractions, surface charges, and heat fluxes, Eq. (2.15), is expressed by

$$q^m - q^{m+1} = 0, \quad e^m - d^{m+1} = 0, \quad k^m - l^{m+1} = 0 \quad \text{on } \mathcal{A}_{m,m+1}. \tag{6.19}$$

The resultants of stress, heat flux, and entropy and the gross electric displacements are measured per unit length of coordinate curves on the reference surface, $\mathcal{A}$, the resultants of body force, acceleration, and surface load, charge and heat flux per unit area of $\mathcal{A}$, and those of edge-load, and edge-surface charge and heat flux per unit length of the Jordan curve $\mathcal{C}$ of $\mathcal{A}$. Moreover, these resultants are referred to the reference surface; they can, of course, be referred to the midsurface, $\mathcal{A}_m$, of each constituent. That is, the interval $Z = [z-h, z+h]$ is replaced by the interval $\tilde{Z} = [-h, +h]$. The new resultants are defined by

$$[\tilde{N}^{\alpha\beta}, \tilde{M}^{\alpha\beta}, \tilde{K}^{\alpha\beta}] = \int_{\tilde{Z}} \mu[1, \theta^3, (\theta^3)^2] T^{\alpha\beta} d\theta^3,$$
$$(\tilde{Q}^\alpha, \tilde{T}^\alpha) = \int_{\tilde{Z}} \mu(1, \theta^3) T^{\alpha 3} d\theta^3, \tag{6.20}$$
$$\tilde{N}^{33} = \int_{\tilde{Z}} \mu T^{33} d\theta^3$$

and such. They are expressed in terms of those which refer to the reference surface, $\mathcal{A}$, in the form

$$\tilde{N}^{\alpha\beta} = N^{\alpha\beta}, \quad \tilde{M}^{\alpha\beta} = M^{\alpha\beta} - zN^{\alpha\beta},$$
$$\tilde{K}^{\alpha\beta} = K^{\alpha\beta} - 2zM^{\alpha\beta} + z^2 N^{\alpha\beta}, \quad \tilde{Q}^\alpha = Q^\alpha,$$
$$\tilde{T}^\alpha = T^\alpha - zQ^\alpha, \quad \tilde{N}^{33} = N^{33},$$

$$\tilde{C}^i = C^i, \quad \tilde{D}^i = D^i - zC^i, \tag{6.21}$$
$$\tilde{H}^i = H^i, \quad \tilde{K}^i = K^i - zH^i,$$
$$\tilde{p} = p, \quad \tilde{T} = T - zp$$

in which an overtilde represents the resultants which refer to $\mathcal{A}_m$.

7. STRAIN, ELECTRIC FIELD, AND HEAT FLUX DISTRIBUTIONS

To obtain the distributions of strain, electric field, and heat flux in terms of the independent functions of mechanical displacement, electric potential, and temperature, we write the variational integral δI_{12}^{12} of Eq. (4.2) for all the constituents of coated laminae and we also consider Eqs. (5.14) and (5.15). Then integrating with respect to the thickness coordinate and using Eqs. (6.1), (6.10), and (6.14), we obtain

$$\begin{aligned}\delta I_{12}^{12} = \int_A \sum_{r=1}^{N} [&({}_0S_{\alpha\beta} - e_{\alpha\beta})\delta N^{\alpha\beta} + ({}_1S_{\alpha\beta} - \epsilon_{\alpha\beta})\delta M^{\alpha\beta} \\ &+ ({}_2S_{\alpha\beta} - \gamma_{\alpha\beta})\delta K^{\alpha\beta} + ({}_0S_{\alpha 3} - e_{\alpha 3})\delta Q^{\alpha} \\ &+ ({}_1S_{\alpha 3} - \epsilon_{\alpha 3})\delta R^{\alpha} + ({}_0S_{33} - e_{33})\delta N^{33}]^{(r)}\, dA\end{aligned} \tag{7.1a}$$

and

$$\begin{aligned}\delta I_{22}^{22} = &- \int_A \sum_{r=1}^{N} [({}_0E_{\alpha} - e_{\alpha})\delta C^{\alpha} + ({}_0E_3 - e_3)\delta C^3 \\ &+ ({}_1E_{\alpha} - \epsilon_{\alpha})\delta D^{\alpha}]^{(r)}\, dA, \\ \delta I_{32}^{32} = &- \int_A \sum_{r=1}^{N} [({}_0h_i - g_i)\delta H^i + ({}_1h_i - k_i)\delta K^i]^{(r)}\, dA,\end{aligned} \tag{7.1b}$$

where

$$\begin{aligned}&e_{\sigma\nu} = \tfrac{1}{2}(\alpha_{\sigma:\nu} + \alpha_{\nu:\sigma} - 2b_{\sigma\nu}\alpha_3), \\ &\epsilon_{\sigma\nu} = \tfrac{1}{2}(- b_{\sigma}^{\lambda}\alpha_{\lambda:\nu} - b_{\nu}^{\lambda}\alpha_{\lambda:\sigma} + 2c_{\sigma\nu}\alpha_3 + \beta_{\sigma:\nu} + \beta_{\nu:\sigma} - 2b_{\sigma\nu}\beta_3) \\ &\gamma_{\sigma\nu} = \tfrac{1}{2}(- b_{\sigma}^{\lambda}\beta_{\lambda:\nu} - b_{\nu}^{\lambda}\beta_{\lambda:\sigma} + 2c_{\sigma\nu}\beta_3), \\ &e_{\sigma 3} = \tfrac{1}{2}(\alpha_{3,\sigma} + b_{\sigma}^{\nu}\alpha_{\nu} + \beta_{\sigma}), \\ &\epsilon_{\sigma 3} = \tfrac{1}{2}\beta_{3,\sigma}, \\ &e_{33} = \beta_3, \quad \gamma_{\sigma 3} = \epsilon_{33} = \gamma_{33} = 0,\end{aligned} \tag{7.2}$$

and

$$\begin{aligned}&e_{\alpha} = - \kappa_{,\alpha}, \quad \epsilon_{\alpha} = - \zeta_{,\alpha}, \quad e_3 = \zeta, \quad \epsilon_3 = 0, \\ &g^i = - (\kappa^{i\alpha}\tau_{,\alpha} + \kappa^{i3}\gamma), \quad k^i = - \kappa^{i\alpha}\gamma_{,\alpha}.\end{aligned} \tag{7.3}$$

Now, setting δI_{12}^{12} equal to zero and considering Eqs. (5.7), (5.10), and (5.13), the distributions of strain, electric field, and heat flux are found to be

$$\left.\begin{aligned}&{}_0S_{ij} = e_{ij}, \ {}_1S_{ij} = \epsilon_{ij}, \ {}_2S_{ij} = \gamma_{ij} \\ &{}_0E_i = e_i, \ {}_1E_i = \epsilon_i, \ {}_0h^i = g^i, \ {}_1h^i = k^i\end{aligned}\right\} \text{ on } AXT \tag{7.4}$$

with

$$\begin{aligned}&e'_{\sigma\nu} = \tfrac{1}{2}(\alpha_{\sigma:\nu} + \alpha_{\nu:\sigma} - 2b_{\sigma\nu}\alpha_3)', \\ &e_{\sigma\nu}^{(m)} = \tfrac{1}{2}\Big[(\alpha_{\sigma:\nu} + \alpha_{\nu:\sigma} - 2b_{\sigma\nu}\alpha_3)' \\ &\qquad + \sum_{r=1}^{m} z_{rm}(\beta_{\sigma:\nu} + \beta_{\nu:\sigma} - 2b_{\sigma\nu}\beta_3)^{(r)}\Big], \\ &\epsilon'_{\sigma\nu} = \tfrac{1}{2}(- b_{\sigma}^{\lambda}\alpha_{\lambda:\nu} - b_{\nu}^{\lambda}\alpha_{\lambda:\sigma} + 2c_{\sigma\nu}\alpha_3 \\ &\qquad + \beta_{\sigma:\nu} + \beta_{\nu:\sigma} - 2b_{\sigma\nu}\beta_3)', \\ &\epsilon_{\sigma\nu}^{(m)} = \tfrac{1}{2}[(- b_{\sigma}^{\lambda}\alpha_{\lambda:\nu} - b_{\nu}^{\lambda}\alpha_{\lambda:\sigma} + 2c_{\sigma\nu}\alpha_3)' \\ &\qquad + (\beta_{\sigma:\nu} + \beta_{\nu:\sigma} - 2b_{\sigma\nu}\beta_3)^{(m)} \\ &\qquad + \sum_{r=1}^{m} z_{rm}(- b_{\sigma}^{\lambda}\beta_{\lambda:\nu} - b_{\nu}^{\lambda}\beta_{\lambda:\sigma} + 2c_{\sigma\nu}\beta_3)^{(r)}], \\ &\gamma'_{\sigma\nu} = \tfrac{1}{2}(- b_{\sigma}^{\lambda}\beta_{\lambda:\nu} - b_{\nu}^{\lambda}\beta_{\lambda:\sigma} + 2c_{\sigma\nu}\beta_3)', \\ &\gamma_{\sigma\nu}^{(m)} = \tfrac{1}{2}(- b_{\sigma}^{\lambda}\beta_{\lambda:\nu} - b_{\nu}^{\lambda}\beta_{\lambda:\sigma} + 2c_{\sigma\nu}\beta_3)^{(m)}, \\ &e'_{\sigma 3} = \tfrac{1}{2}(\alpha_{3,\sigma} + b_{\sigma}^{\nu}\alpha_{\nu} + \beta_{\sigma})', \\ &e_{\sigma 3}^{(m)} = \tfrac{1}{2}[(\alpha_{3,\sigma} + b_{\sigma}^{\nu}\alpha_{\nu} + \beta_{\sigma})' + \beta_{\sigma}^{(m)} \\ &\qquad + \sum_{r=1}^{m} z_{rm}(\beta_{3,\sigma} + b_{\sigma}^{\nu}\beta_{\nu})^{(r)}], \\ &\epsilon'_{\sigma 3} = \tfrac{1}{2}\beta'_{3,\sigma}, \quad \epsilon_{\sigma 3}^{(m)} = \tfrac{1}{2}\beta_{3,\sigma}^{(m)}, \\ &e'_{33} = \beta'_3, \quad e_{33}^{(m)} = \beta_3^{(m)}, \\ &(\gamma_{\sigma 3} = \epsilon_{33} = \gamma_{33})^{(m)} = 0, \quad m = 1, 2, \ldots, N,\end{aligned} \tag{7.5}$$

and

$$\begin{aligned}&e'_{\alpha} = 0, \quad e_{\alpha}^{(2)} = (z_2 - h_2)\zeta_{,\alpha}^{(2)}, \\ &e_{\alpha}^{(m)} = - \sum_{r=1}^{m} z_{rm}\zeta_{,\alpha}^{(r)}, \quad m = 2, \ldots, N-2, \\ &e^{(N-1)} = - (1 + z_{N-1}/h_{N-1}) \sum_{r=2}^{N-2} h_r\zeta_{,\alpha}^{(r)}, \quad e'' = 0, \\ &\epsilon'_{\alpha} = 0, \quad \epsilon_{\alpha}^{(m)} = - \zeta_{,\alpha}^{(m)}, \quad m = 2, \ldots, N-2, \\ &\epsilon_{\alpha}^{(N-1)} = 1/h_{N-1} \sum_{r=2}^{N} h_{(r)}\zeta_{,\alpha}^{(r)}, \quad \epsilon''_{\alpha} = 0, \\ &e'_3 = 0, \quad e_3^{(m)} = \zeta^{(m)}, \quad m = 2, \ldots, N-2, \\ &e_3^{(N-1)} = 1/h_{N-1}\Big(\phi'' - \sum_{r=2}^{N-2} h_{(r)}\zeta^{(r)}\Big), \quad e''_3 = 0, \\ &\epsilon'_3 = \epsilon_3^{(m)} = \epsilon''_3 = 0,\end{aligned} \tag{7.6}$$

and

$$\begin{aligned}&g'^i = - (\kappa^{i\alpha}\tau_{,\alpha} + \kappa^{i3}\gamma)', \quad \kappa'^i = - (\kappa^{i\alpha}\gamma_{,\alpha})', \\ &g^{(m)i} = - \kappa^{(m)i\alpha}\Big(\tau' + \sum_{r=1}^{m} z_{rm}\gamma^{(r)}\Big)_{,\alpha} - (\kappa^{i3}\gamma)^{(m)}, \\ &k^{(m)i} = - (\kappa^{i\alpha}\gamma_{,\alpha})^{(m)}, \quad m = 2, 3, \ldots, N\end{aligned} \tag{7.7}$$

for the vanishing of the coefficients of arbitrary and independent variations of the stress and heat flux resultants and the gross electric displacements.

8. MACROSCOPIC CONSTITUTIVE RELATIONS

This section deals with the derivation of macroscopic constitutive relations of coated laminae by means of the variational volume integral δI_{11}^{11} of Eq. (4.3). The derived constitutive relations involve the stress, heat flux, and entropy resultants, and gross electric displacement, defined in Sec. 6, and the distributions introduced in Sec. 7. In what follows, we write the variational volume integral for all the constituents of coated laminae, and then carry out integrations with respect to the thickness coordinate. Thus, following some arrangement of terms, we arrive at the variational equation

$$\begin{aligned}\delta I_{11}^{11} = \int_A \sum_{r=1}^{N} [&(N^{\alpha\beta} - N_c^{\alpha\beta})\delta e_{\alpha\beta} + (M^{\alpha\beta} - M_c^{\alpha\beta})\delta\epsilon_{\alpha\beta} \\ &+ (R^{\alpha} - R_c^{\alpha})\delta\epsilon_{\alpha 3} + (K^{\alpha\beta} - K_c^{\alpha\beta})\delta\gamma_{\alpha\beta} \\ &+ (Q^{\alpha} - Q_c^{\alpha})\delta e_{\alpha 3} + (N^{33} - N_c^{33})\delta e_{33}]^{(r)}\, dA,\end{aligned} \tag{8.1a}$$

and

$$\delta I_{21}^{21} = \int_A \sum_{r=1}^{N} [(C^i - C_c^i)\delta e_i + (D^i - D_c^i)\delta \epsilon_i]^{(r)}\, dA,$$

$$\delta I_{31}^{31} = \int_A \sum_{r=1}^{N} [(\rho - \rho_c)\delta n + (T - T_c)\delta m]^{(r)}\, dA, \tag{8.1b}$$

$$\delta I_{32}^{32} = -\int_A \sum_{r=1}^{N} [(H^i - H_c^i)\delta g_i + (K^i - K_c^i)\delta k_i]^{(r)}\, dA,$$

whose Euler equations are the linear constitutive relations of the form

$$N^{\alpha\beta} = N_c^{\alpha\beta}, \quad M^{\alpha\beta} = M_c^{\alpha\beta}, \quad K^{\alpha\beta} = K_c^{\alpha\beta}, \quad Q^\alpha = Q_c^\alpha,$$
$$R^\alpha = R_c^\alpha, \quad N^{33} = N_c^{33}, \quad H^i = H_c^i, \quad K^i = K_c^i, \tag{8.2}$$
$$C^i = C_c^i, \quad D^i = D_c^i, \quad \rho = \rho_c, \quad T = T_c.$$

Here, on account of Eqs. (2.3) and (5.14), we considered a distribution of entropy, $\eta = n + \theta^3 m$, and defined

$$\begin{aligned}
N_c^{\alpha\beta} &= C^{\alpha\beta kl}(\mu_0 e_{kl} + \mu_1 \epsilon_{kl} + \mu_2 \gamma_{kl}) \\
&\quad - C^{k\alpha\beta}(\mu_0 e_k + \mu_1 \epsilon_k) - \lambda^{\alpha\beta}(\mu_0 \tau + \mu_1 \gamma), \\
M_c^{\alpha\beta} &= C^{\alpha\beta kl}(\mu_1 e_{kl} + \mu_2 \epsilon_{kl} + \mu_3 \gamma_{kl}) \\
&\quad - C^{k\alpha\beta}(\mu_1 e_k + \mu_2 \epsilon_k) - \lambda^{\alpha\beta}(\mu_1 \tau + \mu_2 \gamma), \\
K_c^{\alpha\beta} &= C^{\alpha\beta kl}(\mu_2 e_{kl} + \mu_3 \epsilon_{kl} + \mu_4 \gamma_{kl}) \\
&\quad + C^{k\alpha\beta}(\mu_2 e_k + \mu_3 \epsilon_k) - \lambda^{\alpha\beta}(\mu_2 \tau + \mu_3 \gamma), \\
Q_c^\alpha &= C^{\alpha 3kl}(\mu_0 e_{kl} + \mu_1 \epsilon_{kl} + \mu_2 \gamma_{kl}) \\
&\quad - C^{k\alpha 3}(\mu_0 e_k + \mu_1 \gamma_k) - \lambda^{\alpha 3}(\mu_0 \tau + \mu_1 \gamma), \\
R_c^\alpha &= C^{\alpha 3kl}(\mu_1 e_{kl} + \mu_2 \epsilon_{kl} + \mu_3 \gamma_{kl}) \\
&\quad - C^{k\alpha 3}(\mu_1 e_k + \mu_2 \gamma_k) - \lambda^{\alpha 3}(\mu_1 \tau + \mu_2 \gamma), \\
N_c^{33} &= C^{33kl}(\mu_0 e_{kl} + \mu_1 \epsilon_{kl} + \mu_2 \gamma_{kl}) \\
&\quad - C^{k33}(\mu_0 e_k + \mu_1 \epsilon_k) - \lambda^{33}(\mu_0 \tau + \mu_1 \gamma),
\end{aligned} \tag{8.3}$$

on AXT for the stress resultants,

$$\begin{aligned}
C^i &= C^{ijk}(\mu_0 e_{jk} + \mu_1 \epsilon_{jk} + \mu_2 \gamma_{jk}) \\
&\quad + C^{ij}(\mu_0 e_j + \mu_1 \epsilon_j) + \lambda^i(\mu_0 \tau + \mu_1 \gamma), \\
D^i &= C^{ijk}(\mu_1 e_{jk} + \mu_2 \epsilon_{jk} + \mu_3 \gamma_{jk}) \\
&\quad + C^{ij}(\mu_1 e_j + \mu_2 \epsilon_j) + \lambda^i(\mu_1 \tau + \mu_2 \gamma)
\end{aligned} \tag{8.4}$$

on AXT for the gross electric displacements, and

$$\begin{aligned}
\rho &= \lambda^{ij}(\mu_0 e_{ij} + \mu_1 \epsilon_{ij} + \mu_2 \gamma_{ij}) \\
&\quad + \lambda^i(\mu_0 e_i + \mu_1 \epsilon_i) + \alpha(\mu_0 \tau + \mu_1 \gamma), \\
T &= \lambda^{ij}(\mu_1 e_{ij} + \mu_2 \epsilon_{ij} + \mu_3 \gamma_{ij}) + \lambda^i(\mu_1 e_i + \mu_2 \epsilon_i) \\
&\quad + \alpha(\mu_1 \tau + \mu_2 \gamma), \\
H^i &= \mu_0 g^i + \mu_1 k^i, \quad K^i = \mu_1 g^i + \mu_2 k^i,
\end{aligned} \tag{8.5}$$

on AXT for the entropy and heat flux resultants in the case of homogeneous piezoelectric material of constituents. If the material of constituents is heterogenous, we introduce the laminae-elastic constants, -dielectric permittivity, -thermal stress constants, and -pyroelectric constants of the form

$$\begin{aligned}
&(C^{ijkl}, C^{ijk}, C^{ij}, \lambda^{ij}, \lambda^i, \alpha)_R \\
&\quad = \int_Z \mu(\theta^3)^R (C^{ijkl}, C^{ijk}, C^{ij}, \lambda^{ij}, \lambda^i, \alpha)\, d\theta^3
\end{aligned} \tag{8.6}$$

which reduce to

$$\begin{aligned}
&(C^{ijkl}, C^{ijk}, C^{ij}, \lambda^{ij}, \lambda^i, \alpha) \\
&\quad = \mu_R (C^{ijkl}, C^{ijk}, C^{ij}, \lambda^{ij}, \lambda^i, \alpha)
\end{aligned} \tag{8.7}$$

when the material is homogeneous. Thus, to obtain the constitutive relations of heterogenous anisotropic material of laminae, we simply replace $\mu_R C^{ijkl}$ by C_R^{ijkl} and so on, in the foregoing macroscopic constitutive relations.

9. LINEAR THEORY OF THE LAMINAE

Now, we have arrived at the main topic of this study, that is, the derivation of the governing equations of the coated laminae of N constituents on the basis of three-dimensional, linear theory of thermopiezoelectricity. The derivation rests entirely on (i) the fields of displacements, electric potential and temperature, chosen *a priori*; (ii) representing the fields by series expansions in the thickness coordinate; and (iii) employing an averaging procedure of variational type. The first of which is the important choice as a basis for the derivation. The second and, in particular, the linearization in the series expansions are almost compulsory, and together with the third make the derivation comprehensive and tractable. In what follows, by the use of the generalized variational theorem of Sec. 4, we first establish the macroscopic divergence equations of the laminae and then the associated natural boundary and initial conditions.

A. Macroscopic stress equations of motion

To begin with, the volume integral δJ_{11}^{11} of Eq. (4.2) is written in the form

$$\delta J_{11}^{11} = \int_T dt \int_A dA \sum_{r=1}^{N} \int_{z_r - h_r}^{z_r + h_r} \mu\, \delta J_1^r\, d\theta^3 \tag{9.1}$$

with

$$\begin{aligned}
\mu\, \delta J_1 &= [(\mu \mu_\alpha^\nu T^{\beta\alpha})_{;\beta} - \mu \mu_\alpha^\nu (\mu^{-1})_\lambda^\sigma b_\sigma^\lambda T^{3\alpha} - \mu b_\alpha^\nu T^{\alpha 3} \\
&\quad + \mu(\mu_\alpha^\nu T^{3\alpha})_{,3} + \mu \mu_\alpha^\nu (f^\alpha - \rho b^\alpha)](\delta\alpha_\nu + \theta^3 \delta\beta_\nu) \\
&\quad + [(\mu T^{\alpha 3})_{;\alpha} + \mu \mu_\alpha^\nu b_{\nu\beta} T^{\beta\alpha} - \mu(\mu^{-1})_\nu^\alpha b_\alpha^\nu T^{33} \\
&\quad + \mu T^{33}{}_{,3}](\delta\alpha_3 + \theta^3 \delta\beta_3)
\end{aligned} \tag{9.2}$$

for each constituent. Here, we made use of Eqs. (3.9)—(3.11), (5.3), and the identities[60] of the form

$$\begin{aligned}
\mu \mu_\alpha^\nu T^{\beta\alpha}{}_{;\beta} &= (\mu \mu_\alpha^\nu T^{\beta\alpha})_{;\beta} - \mu \mu_\alpha^\nu (\mu^{-1})_\lambda^\sigma b_\sigma^\lambda T^{3\alpha} - \mu b_\alpha^\nu T^{\alpha 3}, \\
\mu T^{\alpha 3}{}_{;\alpha} &= (\mu T^{\alpha 3})_{;\alpha} + \mu \mu_\alpha^\nu b_{\nu\beta} T^{\beta\alpha} - \mu(\mu^{-1})_\beta^\alpha b_\alpha^\beta T^{33}, \\
\mu_\alpha^\nu T^{3\alpha}{}_{;3} &= (\mu_\alpha^\nu T^{3\alpha})_{,3}.
\end{aligned} \tag{9.3}$$

Equation (9.2) is now simplified through the relation which follows from Eqs. (3.6) and (3.12), namely

$$\mu_{,3} = -b_\alpha^\alpha + 2\theta^3 K = -\mu(\mu^{-1})_\beta^\alpha b_\alpha^\beta \tag{9.4}$$

and then it is integrated with respect to θ^3 across the laminae thickness, with the result

$$\delta J_{11}^{11} = \int_T dt \int_A dA \sum_{r=1}^{N} [(\mathbf{V} + \mathbf{P}) \cdot \delta\boldsymbol{\alpha} + (\mathbf{W} + \mathbf{S}) \cdot \delta\boldsymbol{\beta}]^r, \tag{9.5}$$

with

$$
\begin{aligned}
V^{\nu} &= (N^{\beta\nu} - b^{\nu}_{\alpha}M^{\beta\alpha})_{;\beta} - b^{\nu}_{\alpha}Q^{\alpha} + F^{\nu} - A^{\nu},\\
V^{3} &= Q^{\alpha}{}_{;\alpha} + b_{\alpha\beta}N^{\alpha\beta} - c_{\alpha\beta}M^{\alpha\beta} + F^{3} - A^{3},\\
W^{\nu} &= (M^{\beta\nu} - b^{\nu}_{\alpha}K^{\beta\alpha})_{;\beta} - Q^{\nu} + G^{\nu} - B^{\nu},\\
W^{3} &= R^{\alpha}{}_{;\alpha} - N^{33} + b_{\alpha\beta}M^{\alpha\beta} - c_{\alpha\beta}K^{\alpha\beta} + G^{3} - B^{3},
\end{aligned}
\tag{9.6}
$$

for each constituent, where the resultants of stress, load, body force, and acceleration introduced in Sec. 6 are considered. Next, setting the volume integral δJ^{11}_{11} equal to zero, for the arbitrary and independent variations of the displacement functions chosen in Eq. (5.7), we have the $3(N+1)$ macroscopic stress equations of motion for the laminae as follows:

$$
\left.
\begin{aligned}
&\delta\boldsymbol{\alpha}^{1}: \sum_{r=1}^{N} \mathbf{V}^{r} + \mathbf{Y}' = 0\\
&\delta\boldsymbol{\beta}^{1}: \mathbf{W}' + h' \sum_{r=2}^{N} \mathbf{V}^{r} + \mathcal{Y}' = 0\\
&\delta\boldsymbol{\beta}^{m}: \mathbf{W}^{m} + \sum_{r=m}^{N} z_{mr}\mathbf{V}^{r} + \mathcal{Y}^{m} = 0\\
&\delta\boldsymbol{\beta}^{N}: \mathbf{W}'' - (z'' - h'')\mathbf{V}'' + \mathcal{Y}'' = 0
\end{aligned}
\right\} \text{ on } AXT. \tag{9.7a}
$$

Here, the effective loads

$$
\begin{aligned}
&\mathbf{Y}' = \mathbf{q}'' - \mathbf{p}', \quad \mathcal{Y}' = h'(\mathbf{p}' + \mathbf{q}''),\\
&\mathcal{Y}^{m} = 2h_{m}\mathbf{q}'', \quad \mathcal{Y}'' = 2h''\mathbf{q}''
\end{aligned}
\tag{9.7b}
$$

and z_{mr} ($\neq z_{rm}$) by Eq. (5.8b) are used, and also, the continuity of tractions, Eq. (6.19), is taken into consideration in defining the effective surface loads and couples.

B. Macroscopic charge equations of electrostatics

Likewise, we evaluate the second volume integral δJ^{21}_{21} of Eq. (4.2), namely

$$
\delta J^{21}_{21} = -\int_T dt \int_A dA \sum_{r=2}^{N-1} \int_Z \mu\, \delta J^{r}_{2}\, d\theta^{3} \tag{9.8a}
$$

with

$$
\mu\,\delta J_{2} = [(\mu D^{\alpha})_{;\alpha} + (\mu D^{3})_{,3}](\delta\kappa + \theta^{3}\delta\xi) \tag{9.8b}
$$

for each layer. Here, we used Eqs. (3.9)–(3.11), (5.4), and the identity[60]:

$$
\mu D^{\alpha}_{;\alpha} = (\mu D^{\alpha})_{;\alpha} + \mu_{,3}D^{3} \tag{9.9}
$$

which is written in a compact form through Eq. (9.4). We perform integrations with respect to θ^3 in Eq. (9.8), and introduce the gross electric displacements of Eq. (6.10), to obtain

$$
\begin{aligned}
\delta J^{21}_{21} = -\int_T dt \int_A dA \sum_{r=1}^{N} [&(C^{\alpha}{}_{;\alpha} + e - d)\delta\kappa\\
&+ (\mathcal{D}^{\alpha}{}_{;\alpha} - C^{3} + \cdots - \mathcal{d})\delta\xi]^{r}
\end{aligned}
\tag{9.10}
$$

with $\delta\phi' = \delta\phi'' = 0$. By virtue of the arbitrary and independent variations of the electric potential functions, ξ^m ($m = 2, 3, \ldots, N-2$), of Eq. (5.10), we set the volume integral δJ^{21}_{21} equal to zero, to arrive at the macroscopic charge equations of electrostatics in the form

$$
\begin{aligned}
&\delta\xi^{2}: \Upsilon^{(2)\alpha}{}_{;\alpha} + C^{(2)}\\
&\quad + \sum_{r=2}^{N-1} x_{2r}C^{(r)\alpha}{}_{;\alpha} = 0, \quad \text{on } AXT,\\
&\delta\xi^{m}: \Upsilon^{(m)\alpha}{}_{;\alpha} + C^{(m)}\\
&\quad + \sum_{r=m}^{N-1} x_{mr}C^{(r)\alpha}{}_{;\alpha} = 0 \quad \text{on } AXT,\\
&\delta\xi^{(N-2)}: \Upsilon^{(N-2)\alpha}{}_{;\alpha} + C^{(N-2)} - (z_{N-2} - h_{N-2})C^{(N-2)\alpha}{}_{;\alpha}\\
&\quad + (1 + z_{N-1}/h_{N-1})h_{N-2}C^{(N-1)\alpha}{}_{;\alpha} = 0, \quad \text{on } AXT,
\end{aligned}
\tag{9.11}
$$

where

$$
\Upsilon^{\alpha} = \mathcal{D}^{\alpha} - h/h_{N-1}\mathcal{D}^{\alpha}_{N-1}, \quad C = -C^{3} + h/h_{N-1}C^{3}_{N-1} \tag{9.12}
$$

and

$$
x_{mr} = [2 - \delta_{mr} + (-1 + z_{N-1}/h_{N-1})\delta_{N-1,r}]h_{m} - \delta_{mr}z_{m}. \tag{9.13}
$$

In these equations, the continuity of surface charge, Eq. (6.19), across interfaces of layers is taken into account.

C. Macroscopic equations of heat conduction

Similarly, we consider the volume integral δJ^{31}_{31} of Eq. (4.2), given by the form

$$
\delta J^{31}_{31} = -\int_T dt \int_A dA \sum_{r=1}^{N} \mu\,\delta J^{r}_{3}\, d\theta^{3} \tag{9.14a}
$$

with

$$
\mu\,\delta J_{3} = \{\Theta^{-1}[(\mu h^{\alpha})_{;\alpha} + (\mu h^{3})_{,3}] + \dot{\eta}\}(\delta\tau + \theta^{3}\delta\gamma) \tag{9.14b}
$$

for each constituent. Here, Eqs. (3.9)–(3.11), (5.5), and (9.9) are used. Then, the integration of Eq. (9.14) with respect to θ^3 yields

$$
\begin{aligned}
\Theta_{0}\delta J^{31}_{31} = -\int_T dt \int_A dA \sum_{r=1}^{N} [&(H^{\alpha}{}_{;\alpha} + \Theta_{0}\dot{\rho} + k - l)\delta\tau\\
&+ (K^{\alpha}{}_{;\alpha} - H^{3} + \Theta_{0}\dot{I} + \cdots - \mathcal{l})\delta\gamma]^{r}.
\end{aligned}
\tag{9.15}
$$

As before, this equation gives the macroscopic equations of heat conduction:

$$
\begin{aligned}
&\delta\tau': \sum_{r=1}^{N} H^{(r)\alpha}{}_{;\alpha} + \Theta_{0}\sum_{r=1}^{N} \mathring{\rho}^{r} + X' = 0,\\
&\delta\gamma': (K^{\alpha}{}_{;\alpha} - H^{3})' + \Theta_{0}\mathring{\bar{I}}'\\
&\quad + h'\left[\sum_{r=2}^{N} H^{(r)\alpha}{}_{;\alpha} + \Theta_{0}\sum_{r=2}^{N} \mathring{\rho}^{r}\right] + \chi' = 0,\\
&\delta\gamma^{m}: (K^{\alpha}{}_{;\alpha} - H^{3})^{m} + \Theta_{0}\mathring{\bar{I}}^{m}\\
&\quad + \sum_{r=m}^{N} z_{mr}H^{(r)\alpha}{}_{;\alpha} + \Theta_{0}\sum_{r=m}^{N} z_{mr}\mathring{\rho}^{r} + \chi^{m} = 0,\\
&\delta\gamma'': (K^{\alpha}{}_{;\alpha} - H^{3})'' + \Theta_{0}\mathring{\bar{I}}'' - (z'' - h'')\\
&\quad \times (H^{\alpha}{}_{;\alpha} + \Theta_{0}\mathring{\rho})'' + \chi'' = 0, \quad \text{on } AXT
\end{aligned}
\tag{9.16}
$$

with the effective heat fluxes:

$$
\begin{aligned}
&X' = k'' - l'',\\
&\chi' = h'(k' + l''), \quad \chi^{m} = 2h_{m}k'', \quad \chi'' = 2h''k''
\end{aligned}
\tag{9.17}
$$

when we set Eq. (9.15) equal to zero for the unconstrained variations of the temperature functions, τ' and γ^m ($m = 1, 2, \ldots, N$). Herein, the continuity of heat fluxes across interfaces of constituents, Eq. (6.19), is included.

D. Mechanical boundary conditions

We now turn to the boundary conditions associated with the divergence equations for the coated laminae of N constituents. Paralleling the derivation of the macroscopic divergence equations above, we evaluate the

variational surface integrals of Eq. (4.2), introduce the definitions of resultants in Sec. 6, and then obtain the natural boundary conditions for the free and independent variations of the field quantities. The displacements and the temperatures are considered to be prescribed on only part of the edge boundary surface of each constituent, while the tractions and the heat fluxes are specified on the remaining part of edge boundary surface and on the faces of the laminae. The electric potential is applied to the faces and the surface charges are given on the edge boundary. To this end, consider the surface integral δJ_{12}^{12} of Eq. (4.2), of the form

$$\delta J_{12}^{12}=\int_T dt \oint_C ds \sum_{r=1}^{N} \int_Z \mu(\delta J_t^r - \delta J_u^r)\, d\theta^3 + \int_T dt \int_{S_t} \delta I_t\, dA \tag{9.18a}$$

with

$$S_t = S_{1t} \cup S_{ut}, \quad S_t = S_t \cup S_{et}, \quad S_u = S_e \cap S_{et} = S \cap S_t, \tag{9.18b}$$

and

$$\begin{aligned}
\mu\delta J_t &= \mu[(T_*^{\nu} - \nu_\beta T^{\beta\sigma})\mu_\sigma^\nu(\delta\alpha_\nu + \theta^3\delta\beta_\nu) + (T_*^3 - \nu_\sigma T^{\sigma 3})(\delta\alpha_3 + \theta^3\delta\beta_3)],\\
\mu\delta J_u &= \mu\{\mu_\sigma^\nu[u_\nu^* - (\alpha_\nu + \theta^3\beta_\nu)]\delta T^\sigma + [u_3^* - (\alpha_3 + \theta^3\theta_3)]\delta T^3\},\\
\delta I_t &= \mu[(T^{*i} - n_3 T^{3i}) - \theta^3(T^{*\sigma} - n_3 T^{3\sigma})b_\sigma^\nu\delta_\nu^i] \times (\delta\alpha_i + \theta^3\delta\beta_i).
\end{aligned} \tag{9.18c}$$

After integration with respect to θ^3, this equation leads to the natural boundary conditions of tractions:

$$\begin{aligned}
&Q_*^{\prime\alpha} - \nu_\beta \sum_{r=1}^{N} N^{(r)\beta\alpha} = 0, \quad Q_*^{\prime 3} - \nu_\alpha \sum_{r=1}^{N} Q^{(r)\alpha} = 0,\\
&R_*^{\prime\alpha} - \nu_\beta\left(M^{\prime\beta\alpha} + h' \sum_{r=2}^{N} N^{(r)\beta\alpha}\right) = 0,\\
&R_*^{\prime 3} - \nu_\alpha\left(R^{\prime\alpha} + h' \sum_{r=2}^{N} Q^{(r)\alpha}\right) = 0,\\
&\left.\begin{aligned}
&R_*^{(m)\alpha} - \nu_\beta\left(M^{(m)\beta\alpha} + \sum_{r=m}^{N} z_{mr} N^{(r)\beta\alpha}\right) = 0\\
&R_*^{(m)3} - \nu_\alpha\left(R^{(m)\alpha} + \sum_{r=m}^{N} z_{mr} Q^{(r)\alpha}\right) = 0
\end{aligned}\right\} \text{along } CXT,\\
&R_*^{\prime\prime\alpha} - \nu_\beta[M^{\prime\prime\beta\alpha} - (z'' - h'')N^{\prime\prime\beta\alpha}] = 0,\\
&R_*^{\prime\prime 3} - \nu_\alpha[R^{\prime\prime\alpha} - (z'' - h'')Q^{\prime\prime\alpha}] = 0,
\end{aligned} \tag{9.19}$$

and

$$\mathbf{p}^{*\prime} - \mathbf{p}' = 0, \quad \text{on } S_{1t}XT, \qquad \mathbf{q}^{*\prime\prime} - \mathbf{q}'' = 0, \quad \text{on } S_{ut}XT, \tag{9.20a}$$

and those of displacements:

$$\left.\begin{aligned}
&\boldsymbol{\alpha}^{*\prime} - \boldsymbol{\alpha}' = 0\\
&\beta_*^{(m)} - \beta^{(m)} = 0, \quad m = 1, 2, \ldots, N
\end{aligned}\right\} \text{along } CXT. \tag{9.20b}$$

E. Electrical boundary conditions

Substituting Eq. (5.4) into the surface integral δJ_{22}^{22} of Eq. (4.2), of the form

$$\delta J_{22}^{22}=\int_T dt \oint_C ds \sum_{r=1}^{N} \mu\delta J_\sigma^r\, d\theta^3 + \int_T dt \int_{S_t} \mu(\delta I_\phi' + \delta I_\phi'')\, dA \tag{9.21a}$$

with

$$S_\phi = S_t, \quad S_\sigma = S_e \tag{9.21b}$$

and

$$\mu\delta J_\sigma = \mu(\nu_\alpha D^\alpha - \sigma^*)(\delta\kappa + \theta^3\delta\xi), \quad \delta I_\phi = (\phi^* - \phi), \tag{9.21c}$$

and carrying out integrations as in Eq. (9.18), we obtain the natural boundary conditions of surface charge:

$$\begin{aligned}
&\Upsilon_*^2 - \nu_\alpha\left(D^{(2)\alpha} - h_2/h_{N-1}D^{(N-1)\alpha} + \sum_{r=2}^{N-1} x_{2r}C^{(r)\alpha}\right) = 0,\\
&\Upsilon_*^m - \nu_\alpha\left(D^{(m)\alpha} - h_m/h_{N-1}D^{(N-1)\alpha} + \sum_{r=m}^{N-1} x_{mr}C^{(r)\alpha}\right) = 0 \text{ along } CXT,\\
&\Upsilon_*^{(N-2)} - \nu_\alpha[D^{(N-2)\alpha} - h_{N-2}/h_{N-1}D^{(N-1)\alpha} - (z_{N-2} - h_{N-2})C^{(N-2)\alpha} + (1 + z_{N-1}/h_{N-1})h_{N-2}C^{(N-1)\alpha}] = 0,
\end{aligned} \tag{9.22}$$

and those of electric potential:

$$\phi' = \phi_0 \cos\omega t \text{ on } S_{1t}XT, \qquad \phi'' = -\phi_0 \cos\omega t \text{ on } S_{ut}XT, \tag{9.23}$$

where Eqs. (3.9)–(3.12), (5.10), and the recurrence relations (5.11) are taken into consideration. Equation (9.23) clearly implies that an alternative potential difference (ϕ_0 is a constant and ω the circular frequency) is applied to the perfectly conducting electrodes. If the electrodes are shorted, these equations are then replaced by $\phi' = \phi'' = 0$ on $S_{1t}XT$.

F. Thermal boundary conditions

As in the derivation of the mechanical boundary conditions above, the evaluation of the variational surface integral δJ_{32}^{32} of Eq. (4.2) written as

$$\delta J_{32}^{32}=\int_T dt \oint ds \sum_{r=1}^{N} \mu(\delta J_\theta^r + \delta J_h^r)\, d\theta^3 + \int_T dt \int_{S_t} \delta I_h\, dA \tag{9.24a}$$

with

$$S_h = S_t \cup S_{eh}, \quad S_\theta = S_e \cap S_{eh}, \tag{9.24b}$$

and

$$\begin{aligned}
&\mu\delta J_\theta = \nu_\beta(\Theta - \Theta^*)\delta h^\beta,\\
&\Theta_0\mu\delta I_h = \mu(n_3 h^3 - v^*)(\delta\tau + \theta^3\delta\gamma),\\
&\Theta_0\mu\delta J_h = \mu(\nu_\beta h^\beta - v^*)(\delta\tau + \theta^3\delta\gamma),
\end{aligned} \tag{9.24c}$$

yields the natural boundary conditions of heat flux:

$$\begin{aligned}
&N^{*\prime} - \nu_\alpha \sum_{r=1}^{N} H^{(r)\alpha} = 0 \text{ along } CXT,\\
&M^{*\prime} - \nu_\alpha\left(K^{\prime\alpha} + h' \sum_{r=2}^{N} H^{(r)\alpha}\right) = 0 \text{ along } CXT,\\
&M^{*(m)} - \nu_\alpha\left(K^{(m)\alpha} + \sum_{r=m}^{N} z_{mr} H^{(r)\alpha}\right) = 0 \text{ along } CXT,\\
&M^{*\prime\prime} - \nu_\alpha[K^{\prime\prime\alpha} - (z'' - h'')H^{\prime\prime\alpha}] = 0 \text{ along } CXT,
\end{aligned} \tag{9.25a}$$

and

$$l^{*\prime} - l' = 0, \quad \text{on } \mathcal{S}_{1f}XT,$$
$$k^{*\prime\prime} - k'' = 0, \quad \text{on } \mathcal{S}_{uf}XT, \tag{9.25b}$$

and those of temperature:

$$\tau^{*\prime} - \tau' = 0, \quad \gamma_*^{(m)} - \gamma^{(m)} = 0, \quad m = 1, 2, \ldots, N, \quad \text{along } \mathcal{C}XT. \tag{9.26}$$

G. Initial conditions

All that remains now is the specifications of initial conditions corresponding to the fields of mechanical displacement, electric potential, and temperature. To this end, we evaluate the variational integral δI_{13}^{t3} of Eq. (4.2), and, as before, we readily get the natural initial conditions of mechanical displacement:

$$\begin{aligned} &\alpha'(\theta^\sigma, t_0) - \mathbf{v}^{*\prime}(\theta^\sigma) = 0, \quad \dot{\alpha}'(\theta^\sigma, t_0) - \mathbf{w}^{*\prime}(\theta^\sigma) = 0, \\ &\beta^m(\theta^\sigma, t_0) - \nu^{*m}(\theta^\sigma) = 0, \quad \dot{\beta}^m(\theta^\sigma, t_0) - \lambda^{*m}(\theta^\sigma) = 0, \\ &\qquad m = 1, 2, \ldots, N \quad \text{on } A(t_0), \end{aligned} \tag{9.27}$$

those of electric potential:

$$\zeta^m(\theta^\sigma, t_0) - \zeta^{*m}(\theta^\sigma) = 0, \quad m = 2, 3, \ldots, N-2 \quad \text{on } A(t_0), \tag{9.28}$$

and those of temperature:

$$\left.\begin{aligned} &\tau'(\theta^\sigma, t_0) - \tau^{*\prime}(\theta^\sigma) = 0 \\ &\gamma^m(\theta^\sigma, t_0) - \gamma^{*m}(\theta^\sigma) = 0 \end{aligned}\right\} m = 1, 2, \ldots, N \quad \text{on } A(t_0). \tag{9.29}$$

Here, $(\mathbf{v}^{*\prime}, \mathbf{w}^{*\prime}, \nu^{*m}, \lambda^{*m})$, ζ^{*m}, and $(\tau^{*\prime}, \gamma^{*m})$ denote respectively the specified values of mechanical displacements, electric potential, and temperature functions at $t = t_0$.

Thus far, we developed a complete system of two-dimensional, approximate equations of the coated laminae within the scope of three-dimensional theory of thermopiezoelectricity. The system of the mechanical displacement, electric potential, and temperature fields, (5.3)–(5.5), the distributions of strain, electric field, and heat flux, (7.5)–(7.7), the macroscopic constitutive equations of stress and entropy resultants, and gross electric displacements, (8.3)–(8.5), the macroscopic stress equations of motion, charge equations of electrostatics, and equations of heat conduction, (9.7)–(9.16), the natural boundary conditions of tractions, mechanical displacements, surface charge and heat flux resultants, electric potential and temperature, (9.19)–(9.26), and the natural initial conditions of mechanical displacements, electric potential, and temperature, (9.27)–(9.29), constitutes a theory for the coated, thermopiezoelectric laminae of N constituents. The theory accommodates vibrations of single and layered shells (plates) accounting for coupling of mechanical, electrical, and thermal fields.

10. THEOREM OF UNIQUENESS

In this section, sufficient initial and boundary conditions are established in order to assure uniqueness in solutions of the linear governing equations of coated, thermopiezoelectric laminae of N constituents. The classical energy argument and standard Green's identities are used to enumerate the sufficient conditions. The argument relies on the positive definiteness of kinetic and potential energies. Its advantage lies in its easy extensions and applicability to one- and two-dimensional, initial mixed-boundary value problems. Kirchhoff[100] used the energy argument at establishing uniqueness in elastostatics, so did Neumann[101] in elastodynamics. Weiner[85] extended Neumann's result, strictly analogous to Kirchhoff's, to the thermoelastic case. A theorem of uniqueness of Neumann's type is given below for the linear governing equations of coated laminae.

To begin with, we consider two possible sets of solutions to the linear governing equations, namely

$$\Lambda_D^{(\alpha)} = [\alpha_i, \beta_i; e_{ij}, \epsilon_{ij}, \gamma_{ij}; \mathcal{N}^{\alpha\beta}, \mathcal{M}^{\alpha\beta}, Q^\alpha, R^\alpha, N^{33}; \kappa, \zeta; e_i, \epsilon_i; \mathcal{C}^i, \mathcal{D}^i; \tau, \gamma; H^i, K^i; \rho, \mathcal{T}]^{(\alpha)} \tag{10.1}$$

and denote the difference of these solutions by

$$\Lambda_D = \Lambda_D^{(2)} - \Lambda_D^{(1)}. \tag{10.2}$$

By virtue of the linearity of divergence and gradient equations, and constitutive relations of laminae, it is clear that the difference set of functions, Λ_D, satisfies the homogeneous governing equations corresponding to zero body force resultants. Thus, we form the following homogeneous expression from Eqs. (9.7), (9.13), and (9.16), for the difference set of solutions

$$\Gamma = \int_T (\Gamma_M - \Gamma_E - \Gamma_T)\, dt = 0 \tag{10.3a}$$

with

$$\begin{aligned} \Gamma_M = \int_A \Bigg\{ &\left[\sum_{r=1}^{N} (\mathcal{N}^{\beta\nu}{}_{:\beta} - b_\beta^\nu Q^\beta - A^\nu)^{(r)} + Y'^\nu\right]\dot{\alpha}'_\nu \\ &+ \left[\sum_{r=1}^{N} (Q^\alpha{}_{:\alpha} + b_{\alpha\beta}\mathcal{N}^{\alpha\beta} - A^3)^{(r)} + Y'^3\right]\dot{\alpha}'_3 \\ &+ \sum_{m=1}^{N} \Bigg[(\mathcal{M}^{\sigma\alpha}{}_{:\sigma} - Q^\alpha - B^\alpha + \mathcal{Y}^\alpha)^{(m)} \\ &+ \sum_{r=m}^{N} z_{mr}(\mathcal{N}^{\sigma\alpha}{}_{:\sigma} - b_\beta^\alpha Q^\beta - A^\alpha)^{(r)}\Bigg]\dot{\beta}_\alpha^{(m)} \\ &+ \sum_{m=1}^{N} \Bigg[(R^\alpha{}_{:\alpha} - N^{33} + b_{\alpha\sigma}\mathcal{M}^{\alpha\sigma} - B^3 + \mathcal{Y}^3)^{(m)} \\ &+ \sum_{r=m}^{N} z_{mr}(Q^\alpha{}_{:\alpha} + b_{\alpha\sigma}\mathcal{N}^{\alpha\sigma} - A^3)^{(r)}\Bigg]\dot{\beta}_3^{(m)}\Bigg\} dA \end{aligned} \tag{10.3b}$$

and

$$\begin{aligned} \Gamma_E = \int_A \Bigg\{ &\sum_{m=2}^{N-2} \Bigg[\dot{\Upsilon}^{(m)\alpha}{}_{:\alpha} + \dot{C}^{(m)} \\ &+ \sum_{r=m}^{N-1} x_{mr}\dot{C}^{(r)\alpha}{}_{:\alpha}\Bigg]\zeta^{(m)} + \Gamma_\phi\Bigg\} dA, \\ \Gamma_T = \int_A \Bigg\{ &\left[\sum_{r=1}^{N} (H^\alpha{}_{:\alpha} + \Theta_0\dot{\rho})^{(r)} + X'\right]\tau' \\ &+ \sum_{m=1}^{N} \Bigg[(K^\alpha{}_{:\alpha} - H^3)^{(m)} + \Theta_0\dot{\mathcal{T}}^{(m)} \\ &+ \sum_{r=m}^{N} (z_{mr}H^\alpha{}_{:\alpha} + \Theta_0\dot{\rho})^{(r)} + \chi^{(m)}\Bigg]\gamma^{(m)}\Bigg\}\Theta_0^{-1}\, dA, \end{aligned} \tag{10.3c}$$

where Γ_ϕ is the part arising from the unprescribed potentials of (9.10).

A. Kinetic energy

Let k_e be the kinetic energy density, K_e the kinetic energy per unit area of the reference surface, $\mathcal{K}_e$ the laminae-kinetic energy density, and Ω_e the total kinetic energy of laminae. Thus, we write, using Eq. (2.16), the rate of kinetic energy density for the difference solution,

$$\dot{k}_e = \rho b^i \mathring{u}_i \tag{10.4}$$

and

$$K_e = \int_Z \mu k_e \, d\theta^3, \quad \mathcal{K}_e = \sum_{r=1}^{N} K_e^{(r)}, \quad \Omega_e = \int_A \mathcal{K}_e \, dA. \tag{10.5}$$

By inserting Eq. (5.3) into the above and using Eq. (6.4), one finds

$$\dot{K}_e = A^i \dot{\alpha}_i + B^i \dot{\beta}_i \tag{10.6}$$

and, in view of Eqs. (5.8), the rate of laminae-kinetic energy density in the form

$$\dot{\mathcal{K}}_e = \dot{\alpha}'^i \sum_{r=1}^{N} A_i^{(r)} + \sum_{m=1}^{N} \left(B_i^{(m)} + \sum_{r=m}^{N} z_{mr} A_i^{(r)} \right) \dot{\beta}^{(m)i}. \tag{10.7}$$

B. Dissipation function

We define the dissipation function density F_f per unit area of the reference surface, A, by

$$F_f = \int_Z 2\mu F \, d\theta^3, \tag{10.8}$$

the laminae-dissipation function density $\mathcal{F}_f$ and the total dissipation function of laminae Ω_f by

$$\mathcal{F}_f = \sum_{r=1}^{N} F_f^{(r)}, \quad \Omega_f = \int_A \mathcal{F}_f \, dA. \tag{10.9}$$

By the use of Eqs. (2.6), (2.19), (5.5), and (6.14), we obtain the dissipation function density and that of laminae as

$$F_f = -\Theta_0^{-1} (H^\alpha \tau_{,\alpha} + H^3 \gamma + K^\alpha \gamma_{,\alpha}) \tag{10.10}$$

and

$$\mathcal{F}_f = -\Theta_0^{-1} \left\{ \tau'_{,\alpha} \sum_{r=1}^{N} H^{(r)\alpha} + \sum_{m=1}^{N} \left(K^{(m)\alpha} + \sum_{r=m}^{N} z_{mr} H^{(r)\alpha} \right) \gamma_{,\alpha}^{(m)} + \sum_{r=1}^{N} (H^3 \gamma)^{(r)} \right\} \tag{10.11}$$

for the difference set of solutions. Here, Eq. (5.13) is considered.

C. Biot's generalized free energy

Similarly, we define Biot's generalized free energy density B_e per unit area of the reference surface, A, the laminae-Biot's energy density $\mathcal{B}_e$ and the total Biot's generalized free energy of laminae Ω_B by

$$B_e = \int_Z \mu B \, d\theta^3, \quad \mathcal{B}_e = \sum_{r=1}^{N} B_e^{(r)}, \quad \Omega_B = \int_A \mathcal{B}_e \, dA. \tag{10.12}$$

The rate of Biot's generalized free energy density follows readily from Eqs. (2.17) and (2.18) in the form

$$\dot{B} = T^{ij} \dot{S}_{ij} + E_i \dot{D}^i + \Theta \dot{\eta}. \tag{10.13}$$

Hence, we have the rate of energy density $\mathcal{B}_e$,

$$\mathring{\mathcal{B}}_e = (\mathcal{B}_{\mathcal{M}} + \mathcal{B}_{\mathcal{E}} + B_T)^\circ \tag{10.14a}$$

with

$$(\mathcal{B}_{\mathcal{M}}, \mathcal{B}_{\mathcal{E}}, \mathcal{B}_T) = \sum_{r=1}^{N} (B_M, B_E, B_T)^{(r)} \tag{10.14b}$$

and

$$(\dot{B}_M, \dot{B}_E, \dot{B}_T) = \int_Z \mu (T^{ij} \mathring{S}_{ij}, E_i \mathring{D}^i, \theta \dot{\eta}) \, d\theta^3 \tag{10.14c}$$

and the rate of total Biot's energy of laminae,

$$\mathring{\Omega}_B = (\Sigma_M + \Sigma_E + \Sigma_T)^\circ \tag{10.15a}$$

with

$$(\Sigma_M, \Sigma_E, \Sigma_T) = \int_A (\mathcal{B}_{\mathcal{M}}, \mathcal{B}_{\mathcal{E}}, \mathcal{B}_T) \, dA. \tag{10.15b}$$

With the aid of the stress, heat flux, and entropy resultants, and gross electric displacements of Sec. 6, and the distributions of strain and electric field of Sec. 7, one obtains the rates of densities of the reference surface in the form

$$\begin{aligned} \dot{B}_M &= \mathcal{N}^{\sigma\nu} \dot{\alpha}_{\nu:\sigma} + \mathcal{M}^{\sigma\nu} \dot{\beta}_{\nu:\sigma} - b_{\sigma\nu} (\mathcal{N}^{\sigma\nu} \dot{\alpha}_3 + \mathcal{M}^{\sigma\nu} \dot{\beta}_3) \\ &\quad + Q^\sigma (\dot{\alpha}_{3,\sigma} + b_\sigma^\nu \dot{\alpha}_\nu + \dot{\beta}_\sigma) + R^\sigma \beta_{3,\sigma} + N^{33} \dot{\beta}_3, \\ \dot{B}_E &= -(\mathring{C}^\alpha \kappa_{,\alpha} + \mathring{D}^\alpha \xi_{,\alpha} + \mathring{C}^3 \xi), \\ \dot{B}_T &= \dot{P} \tau + \dot{T} \gamma. \end{aligned} \tag{10.16}$$

And, upon use of Eqs. (5.8), (5.11), and (5.13), those of laminae take the form

$$\begin{aligned} \dot{\mathcal{B}}_{\mathcal{M}} &= (\dot{\alpha}_{\sigma:\nu} - b_{\sigma\nu} \dot{\alpha}_3)' \sum_{r=1}^{N} \mathcal{N}^{(r)\nu\sigma} + \sum_{r=1}^{N} (N^{33} \dot{\beta}_3)^{(r)} \\ &\quad + \sum_{r=1}^{N} (Q^\sigma \dot{\beta}_\sigma)^{(r)} + \sum_{m=1}^{N} \left\{ [(\dot{\beta}_{\sigma\cdot\nu} - b_{\sigma\nu} \dot{\beta}_3) \mathcal{M}^{\sigma\nu}]^{(m)} \right. \\ &\quad \left. + (\dot{\beta}_{\sigma:\nu} - b_{\sigma\nu} \dot{\beta}_3)^{(m)} \sum_{r=m}^{N} z_{mr} \mathcal{N}^{\nu\sigma}_{(r)} \right\} \\ &\quad + \sum_{m=1}^{N} \left[(R^\sigma \dot{\beta}_{3,\sigma})^{(m)} + (\dot{\beta}_{3,\sigma} + b_\sigma^\nu \dot{\beta}_\nu)^{(m)} \right. \\ &\quad \left. \times \sum_{r=m}^{N} z_{mr} Q^{(r)\sigma} \right] + (\dot{\alpha}_{3,\sigma} + b_\sigma^\nu \dot{\alpha}_\nu) \sum_{r=1}^{N} Q^{(r)\sigma}, \\ \dot{\mathcal{B}}_{\mathcal{E}} &= -\sum_{m=2}^{N-2} \left[\left(\Upsilon^\alpha + \sum_{r=m}^{N-1} \varkappa_{mr} C^{(r)\alpha} \right) \xi_{,\alpha} + C\xi \right]^{(m)}, \\ \dot{\mathcal{B}}_T &= \tau' \sum_{r=1}^{N} P^{(r)} + \sum_{m=1}^{N} \left[\dot{T}^{(m)} + \sum_{r=m}^{N} z_{mr} \dot{P}^{(r)} \right] \gamma^{(m)} \end{aligned} \tag{10.17}$$

for the difference set of solutions.

D. Sufficient conditions

We now turn to the homogeneous equation, (10.3), stated for the difference set of solutions. After using Green's integral identities, a comparison of this equation with the laminae-kinetic energy, -dissipation function and -Biot's generalized free energy densities, Eqs. (10.7), (10.11), and (10.17), leads to

$$\begin{aligned} \Gamma_M &= -\int_A (\mathring{\mathcal{K}}_e + \mathring{\mathcal{B}}_{\mathcal{M}}) \, dA + \oint_A J_{MA} \, dA + \oint_C J_{MC} \, ds, \\ \Gamma_E &= \int_A \mathring{\mathcal{B}}_{\mathcal{E}} \, dA + \int_A J_{EA} \, dA + \oint_C J_{EC} \, ds, \\ \Gamma_T &= \int_A (\mathring{\mathcal{B}}_T + \mathcal{F}_f) \, dA + \int_A J_{TA} \, dA + \oint_C J_{TC} \, ds. \end{aligned} \tag{10.18}$$

Here, we introduced the integrands of line and area integrals of the form

$$J_{MA}=(q''-p')^{i}\dot{\alpha}_i'+2q''^{i}\sum_{m=1}^{N}(h\dot{\beta}_i)^{(m)},$$

$$J_{MC}=\nu_\sigma\Big[\dot{\alpha}_\lambda'\sum_{r=1}^{N}N^{(r)\sigma\lambda}+\dot{\alpha}_3'\sum_{r=1}^{N}Q^{(r)\sigma}+\sum_{m=1}^{N}\Big(M^{(m)\sigma\lambda}+\sum_{r=m}^{N}z_{mr}N^{(r)\sigma\lambda}\Big)\dot{\beta}_\lambda^{(m)}+\sum_{m=1}^{N}\Big(R^{(m)\sigma}+\sum_{r=m}^{N}z_{mr}Q^{(r)\sigma}\Big)\dot{\beta}_3^{(m)}\Big], \tag{10.19a}$$

and

$$J_{EA}=(-d_2+e_{N-2})\phi'+(e_{N-1}+d_{N-1})\phi'',$$

$$J_{EC}=\nu_\alpha\Big[\sum_{m=2}^{N-1}\Big(D^{(m)\alpha}-h_{(m)}/h_{N-1}D^{(N-1)\alpha}+\sum_{r=m}^{N-1}x_{mr}C^{(r)\alpha}\Big)\zeta^{(m)}\Big],$$

$$J_{TC}=\nu_\alpha\Big[\tau'\sum_{r=1}^{N}H^{(r)\alpha}+\sum_{m=1}^{N}\Big(K^{(m)\alpha}+\sum_{r=m}^{N}z_{mr}H^{(r)\alpha}\Big)\gamma^{(m)}\Big], \tag{10.19b}$$

$$J_{TA}=(k''-l')\tau'+2k''\sum_{m=1}^{N}(h\gamma)^{(m)},$$

which evidently vanish due to the homogeneous mechanical, electrical, and thermal boundary conditions, Eqs. (9.19)–(9.26), and initial conditions, Eqs. (9.27)–(9.29), for the difference set of solutions. Thus, inserting all the results above, and then integrating by parts and assembling them in Eq. (10.3), we finally arrive at the equation

$$\Gamma=-\Omega_e(t_1)-\Omega_B(t_1)+\Omega_e(t_0)+\Omega_B(t_0)-\int_T\Omega_f\,dt=0. \tag{10.20a}$$

Since, as already stated, the kinetic energy and dissipation function, and Biot's generalized free energy density are positive definite, by definition, and initially zero; so that the total kinetic energy, dissipation function and Biot's generalized free energy, Ω_e, Ω_B, and Ω_f, calculated by integration from the difference set of solutions for the laminae, have the same properties in the absence of discontinuities and singularities of any type. Consequently, guided by the usual arguments based on the positive definiteness of these quantities, it follows from Eq. (10.20a) that

$$\Omega_e(t_1)=\Omega_e(t_0)=\Omega_B(t_1)=\Omega_B(t_0)=\Omega_f=0 \tag{10.20b}$$

which implies a trivial solution for the difference set of solutions, Λ_D, for the case when the line and area integrands in Eq. (10.19) vanish, and hence it assures the uniqueness of solutions of the linear governing equations of coated laminae. We then conclude the following theorem of uniqueness.

Theorem: Given a regular region of finite laminae space $B+\partial B$, with its entire boundary surface ∂B ($S_t\cup S_u=S_\sigma\cup S_\phi=S_h\cup S_\theta=\partial B$ and $S_t\cap S_u=S_\sigma\cap S_\phi=S_h\cap S_\theta=\emptyset$) in the Euclidean 3-space Ξ, then there exists at most one set of single-valued solutions, Λ_D, namely

$$\Lambda_D=[\alpha_i \text{ and } \beta_i\in C_{12};\ e_{ij},\epsilon_{ij} \text{ and } \gamma_{ij}\in C_{00};$$
$$N^{\alpha\beta},M^{\alpha\beta},Q^\alpha,R^\alpha \text{ and } N^{33}\in C_{10};\ \kappa \text{ and } \zeta\in C_{10};$$
$$e_i \text{ and } \epsilon_i\in C_{00};\ C^i \text{ and } D^i\in C_{10};\ \tau \text{ and } \gamma\in C_{10};$$
$$H^i \text{ and } K^i\in C_{10};\text{ and } P \text{ and } T\in C_{01}] \tag{10.21}$$

which satisfies the macroscopic divergence and gradient equations, (9.7)–(9.16) and (9.19)–(9.26), distributions of strain, electric field, and heat flux, (7.3)–(7.5), and linear constitutive relations (8.3)–(8.5), of coated, thermopiezoelectric laminae of N constituents, provided that the kinetic and Biot's generalized free energies, and the dissipation function are positive definite, and the following are specified:

(i) one member of each of the products in the integrands J_{MA}, J_{EA}, and J_{TA} at each point of the reference surface, A,

(ii) one member of each of the products in the integrands J_{MC}, J_{EC}, and J_{TC} at each point on the Jordan curve, C, and

(iii) the initial values of independent mechanical displacement, electric potential and temperature functions, that is, α', $\dot{\alpha}'$, $\beta^{(m)}$, $\dot{\beta}^{(m)}$ ($m=1,2,\ldots,N$), $\zeta^{(m)}$ ($m=2,3,\ldots,N-1$), and τ' and $\gamma^{(m)}$ ($m=1,2,\ldots,N$) on the reference surface.

11. CONCLUDING REMARKS

Presented herein is a two-dimensional theory for dynamic problems of coated laminae of N constituents, accounting for coupling of mechanical, electrical, and thermal fields within the scope of three-dimensional theory of thermopiezoelectricity. A generalized variational theorem is formulated so as to describe all the three-dimensional fundamental equations of thermopiezoelectricity as appropriate Euler equations. Then, the theory is systematically derived by means of the variational theorem together with a linearized chosen field of mechanical displacements, electric potential, and temperature. In the derivation, the classical Kirchhoff–Love hypotheses on mechanical displacements of shells are abrogated in each constituent of the laminae. Hence, the transverse contractions and shear deformations and the translational and rotational inertia of, as well as all the dynamic interactions among, constituents are included in the formulation. The continuity of tractions, mechanical displacements, electric charge and potential, heat fluxes, and temperature is maintained in accordance with perfectly bonding constituents of the laminae. Further, the results are established in an invariant form, and hence applicable to an arbitrarily shaped laminae using a particular coordinate system most suitable for its geometrical configuration, and they accommodate the extensional, thickness, and flexural type of motions of coated laminae.

The only alternative variational formulation to the theory of thermopiezoelectricity was due to Mindlin.[28] His theorem is suitably modified to establish the generalized variational theorem of Sec. 4, which has the advantage of yielding the constitutive relations, gradient equations as well as initial conditions, in addition to the divergence equations and boundary conditions of laminae. A direct method of integration in lieu of the variational theorem may be used in the derivation, but it is not suitable for the laminated type of structure as already noted in Ref. 102. Moreover, the equivalence

of $\delta I(\Lambda_D)=0$ to the initial-mixed boundary value problems of the theory of thermopiezoelectricity is shown in Sec. 4. Though it is not necessary herein, the explicit form of I is not given. Thus, Mindlin's as well as the present variational theorem can be considered quasitype. A variational theorem with I instead of δI can be derived by transforming the initial-mixed boundary value problem into an equivalent boundary value problem through Laplace transformation, which includes the initial conditions implicitly, and then constructing the theorem for the equivalent problem; this will be reported elsewhere.

By means of the classical energy arguments, a theorem of uniqueness is rigorously proved for the governing equations of the linear theory presented. The theorem enumerates the mechanical, electrical and thermal boundary and initial conditions sufficient for uniqueness in solutions of the linear governing equations of coated, thermopiezoelectric laminae. The energy arguments[103] are by no means the only way of establishing the conditions sufficient for uniqueness. It is worth remarking that the methods involving reflection principle, logarithmic convexity arguments, and Holmgren's theorem can be also used to examine uniqueness in solutions of initial-mixed boundary value problems. Though they were illustrated for three-dimensional elastodynamics, they can, of course, be used for one- and two-dimensional cases as well.[104]

The results describe the dynamic behavior of both uncoated and coated, thermopiezoelectric laminae of N constituents, and contain those of certain earlier studies as special cases. We begin by noting that when only one layer is considered, that is, $N=3$, a theory of coated, thermopiezoelectric shell is established, and also dropping out the effects of electrodes and temperature, it recovers the higher order theory of piezoelectric crystal surfaces, due to the author,[60] up to the first order. In the absence of curvature effects, our results reduce to those of coated, thermopiezoelectric crystal multilayer plates, in which $b_\beta^\alpha=0$ and μ_β^α is then equal to the Kronecker delta. By considering again only one layer, the results can be specialized to yield those of Mindlin,[28] but including the effect of elastic stiffness and inertia of electrodes. In addition, if the terms involving the upper and lower electrodes are discarded, the results do agree completely with those of Mindlin.[28] Further, we note that neglecting the electrical effect, an effective stiffness theory[105] of laminated composite shells (plates) is obtained. Also, omitting the thermal effect, this theory of composite shells recovers the one due to the author.[106]

Although piezoelectricity is basically a linear electromechanical phenomena, nonlinear effects in piezoelectric substances have also been demonstrated experimentally.[107] To construct a nonlinear theory of the coated, thermopiezoelectric laminae is beyond the scope of this paper, and it is left for a future study.

The type of thermal boundary conditions in Eq. (2.12) is prescribed in the analysis. However, various types of boundary conditions arising from thermal interactions with the exterior of laminae region may be prescribed as well. A type of considerable interest is the radiation condition, namely

$$n_i h^i - C_r\Theta^* = 0 \quad \text{on } SXT \tag{11.1}$$

in lieu of Eq. (2.12). Here, C_r denotes a positive constant and C_e the value of C_r at the edge of laminae, and it takes the values between zero (i.e., adiabatic boundary) and infinity (i.e., $\Theta=0$, isothermal boundary). Now replacing v^* by $C_r\Theta^*$ in the variational theorem (4.1), the thermal boundary conditions of coated laminae, corresponding to Eq. (11.1), are obtained through Eqs. (4.1), (5.5), (5.13), and (6.16), in the form

$$\begin{aligned} &k'' - C_r[\tau'' + (2H-h')\gamma'']_* = 0 \quad \text{on } S_{uf}XT,\\ &l' - C_r(\tau' - h'\gamma')_* = 0 \quad \text{on } S_{1f}XT, \end{aligned} \tag{11.2}$$

and

$$\nu_\alpha \sum_{r=1}^{N} (C_e^{-1}H^\alpha)^{(r)} - T_*' = 0,$$

$$(\nu_\alpha C_e^{-1}K^\alpha)^{(m)} + \nu_\alpha \sum_{r=m}^{N} z_{mr}(C_e^{-1}H^\alpha)^{(r)} - T_*^{(m)} = 0, \tag{11.3a}$$

$$\nu_\alpha C_e^{-1}[K^\alpha - (z-h)H^\alpha]'' - T_*'' = 0, \quad \text{along } CXT$$

with

$$T' = \tau' \sum_{r=1}^{N} \mu_0^{(r)} + \sum_{m=1}^{N} \left[\mu_1^{(m)} + \sum_{r=m}^{N} z_{mr}\mu_0^{(r)}\right]\gamma^{(m)},$$

$$\begin{aligned} T^{(m)} = {} & \left(\mu_1^{(m)} + \sum_{r=m}^{N} z_{mr}\mu_0^{(r)}\right)\tau' + \sum_{r=1}^{m-1}\left(\mu_1^{(m)}\right.\\ & \left. + \sum_{r=m}^{N} z_{mr}\mu_0^{(r)}\right)(2-\delta_{1r})h_r\gamma^{(r)}\\ & + \left[\mu_2^m - 2(z_m - h_m)\mu_1^{(m)} + \sum_{r=m}^{N} z_{mr}^2\mu_0^{(r)}\right]\gamma^{(m)}\\ & + \sum_{r=m+1}^{N}\left(2h_m\mu_1^{(r)} + \sum_{r=m+1}^{N} z_{m+1,r}\mu_0^{(r)}\right)\gamma^{(r)}, \end{aligned} \tag{11.3b}$$

$$\begin{aligned} T'' = {} & [\mu_1 - (z-h)\mu_0]''\tau' + \sum_{r=1}^{N}\{[\mu_1 - (z-h)\mu_0]''z_{rm}\\ & + [\mu_2 - (z-h)\mu_1]''\delta_{Nr}\}\gamma^{(r)} \end{aligned}$$

which lead to the uniqueness of solutions, as before.

Further, the fields chosen in Eqs. (5.3)—(5.5) can be expanded into a series of simple thickness modes instead of the thickness coordinate as a starting point. And then, a set of two-dimensional, approximate

governing equations, immensely more tractable than a three-dimensional one, can be constructed by means of the present method of analysis as in the case of purely elastic plates.[108] The present method is employed to establish the governing equations of the type of laminae composed of N distinct elastic constituents and coated completely with electrodes on both its faces. Using the governing equations, a number of specific problems solved by recent numerical techniques[109,110] will be topics of future study. Finally, it is concluded that the method of analysis can be fruitfully extended to be useful in the analysis of a composite layered structure of one piezoelectric and another magnetoelastic media,[111] and, in addition, of multilayer structures comprising any number and layered arrangements of electrodes and piezoelectric constituents, which have valuable acceptance in multistage transducers for both sonar and delay line applications.

ACKNOWLEDGMENT

The author wishes to thank Dean T. Özalp, Prof. Dr. K. Kafali, and Mr. Sükrü Kapucu for their kind interest and encouragement, and Dr. Vedia Arpaci and Mr. M. Remzi for their kind hospitality and patience during this investigation. He gratefully acknowledgs the financial support provided by I. T. U. Gemi Insaati Fakültesi and TÜBİTAK, and also thanks Professor N. Jones of M. I. T. and the Department of Ocean Engineering of Massachusetts Institute of Technology where part of this study has been carried out.

[1]W.G. Cady, Piezoelectricity (Dover, New York, 1964), Vols. I and II.
[2]I.S. Zheludev, *Physics of Crystalline Dielectrics*, Vol. II-*Electrical Properties* (Plenum, New York, 1971).
[3]P.W. Forsbergh, Jr., in *Handbuch der Physik, Dielektrika* (Springer-Verlag, Berlin, 1956), Vol. XVII, pp. 264—392.
[4]W. Voigt, *Lehrbuch der Kristallphysik* (Teubner, Leipzig, 1928; Reprint, Edwards, Ann Arbor, 1946).
[5]J. Curie and P. Curie, Bull. Soc. Min. France 3, 90—3 (1880).
[6]P. Curie, *Oeuvres de Pierre Curie* (Gauthier-Villars and Cie, Paris, 1908).
[7]G. Lippmann, An. Chim. Phys. Ser. 5, 24, 145—78 (1881).
[8]J. Curie and P. Curie, Compt. Rend. 93, 1137 (1881).
[9]J.F. Nye, *Physical Properties of Crystals: Their Representation by Tensors and Matrices* (Clarendon, London, 1960).
[10]B. Jaffe, W.R. Cook, Jr., and H. Jaffe, *Piezoelectric Ceramics* (Academic, New York, 1971).
[11]J. van Randeraat, Ed., *Piezoelectric Ceramics* (Philips' Gloeilampenfabrienken, Eindhoven, The Netherlands, 1968).
[12]E. Fukada, Biorheology 5, 199—208 (1968).
[13]M.H. Shamos and L.S. Lavine, Nature 213, 267—9 (1967).
[14]V.A. Bazhenov, *Piezoelectric Properties of Wood* (Consultants Bureau, New York, 1961).
[15]E. Fukada and I. Yasuda, J. Phys. Soc. Jpn. 12, 1158—62 (1957).
[16]W.P. Mason, *Piezoelectric Crystals and Their Applications to Ultrasonics* (Van Nostrand, New York, 1950).
[17]R.M. White, Proc. IEEE 58, 1238—76 (1970).
[18]F.L. N-Nagy and G.C. Joyce, in *Physical Acoustics* (Academic, New York, 1972), Vol. IX, pp. 129—65.
[19]D.A. Berlincourt, D.R. Curran, and H. Jaffe, in *Physical Acoustics* (Academic, New York, 1964), Vol. 1, Part A, pp. 169—270.
[20]R.D. Mindlin, *An Introduction to the Mathematical Theory of Vibrations of Elastic Plates* (U.S. Army Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey, 1955).
[21]A.L. Cauchy, Exercises Math. 3, 328—355 (1828).
[22]S.D. Poisson, Mém. Acad. Sci. 8, 357—70 (1829).
[23]G. Kirchhoff, Crelles Jl, 40, 51—88 (1850).
[24]R.D. Mindlin, "Theory of Beams and Plates" (lecture notes at Columbia University, 1955).
[25]H.F. Tiersten and R.D. Mindlin, Q. Appl. Math. 20, 107—19 (1962).
[26]R.D. Mindlin, in *Progress in Applied Mechanics* (Macmillan, New York, 1963), Prager Anniversary Volume, pp. 73—84.
[27]R.D. Mindlin, Int. J. Solids Struct. 8, 895—906 (1972).
[28]R.D. Mindlin, Int. J. Solids Struct. 10, 625—37 (1974).
[29]H.F. Tiersten, *Linear Piezoelectric Plate Vibrations* (Plenum, New York, 1969).
[30]P.C.Y. Lee and D.W. Haines, in R.D. Mindlin, *Applied Mechanics* (Pergamon, New York, 1974), pp. 227—53.
[31]R. Holland and E.P. Eer Nisse, *Design of Resonant Piezoelectric Devices* (M.I.T. Press, Cambridge, Massachusetts, 1969), Res. Monog. No. 56.
[32]R. Holland, J. Acoust. Soc. Am. 43, 988—97 (1968).
[33]E.P. Eer Nisse, J. Acoust. Soc. Am. 40, 1045—55 (1966).
[34]J.L. Bleustein, J. Acoust. Soc. Am. 45, 614—20 (1969).
[35]H.S. Paul, J. Acoust. Soc. Am. 44, 478—82 (1968).
[36]Y. Kagawa and T. Yamabuchi, IEEE Trans. Sonics Ultrason. SU-23, 263—72 (1976).
[37]G.H. Schmidt, J. Eng. Math. 6, 133—42 (1972).
[38]D.H. Keuning, J. Eng. Math. 6, 143—54 (1972).
[39]H.S. Paul, Appl. Sci. Res., Sec. B 9, 359—73 (1961).
[40]I.A. Vekovishcheva, Izv. Akad. Nauk SSSR Mekh. Tverd. Tela 9, 108—13 (1974) [Mechanics of Solids 9, 93—7 (1974)].
[41]I.A. Vekovishcheva, Izv. Akad. Nauk SSSR Mekh. Tverd. Tela 10, 168—71 (1975) [Mechanics of Solids 10, 146—9 (1975)].
[42]R.R. Rizzo and J.L. Nowinski, in *Recent Advances in Engineering Science* (Gordon and Breach, New York, 1970), Vol. 5, Part I, pp. 91—115.
[43]N.I. Muskhelishvili, *Some Basic Problems of the Mathematical Theory of Elasticity* (Noordhoff, Groningen, The Netherlands, 1953).
[44]S.G. Lekhnitskii, *Theory of Elasticity of an Anisotropic Elastic Body* (Holden-Day, San Francisco, 1963).
[45]B.A. Kudriavtsev, V.Z. Parton, and V.I. Rakitin, Prikl. Mat. Mekh. 39, 149—59 (1975) [Sov. Appl. Math. Mech. 39, 136—46 (1975)].
[46]R.A. Toupin, J. Acoust. Soc. Am. 31, 315—18 (1959).
[47]C.V. Stephenson, J. Acoust. Soc. Am. 28, 51—6 (1956).
[48]D.S. Drumheller and A. Kalnins, J. Acoust. Soc. Am. 47, 1343—53 (1970).
[49]J.F. Haskins and J.L. Walsh, J. Acoust. Soc. Am. 29, 729—34 (1957).
[50]H.S. Paul, J. Sci. Eng. Res. 6, 249—256 (1962).
[51]G.E. Martin, J. Acoust. Soc. Am. 35, 510—20 (1963).
[52]D.B. Dianov and A.G. Kuz'menko, Akust. Zh. 16, 42—8 (1970) [Sov. Phys. Acoust. 16, 34—8 (1970)].
[53]N.T. Adelman, Y. Stavsky, and E. Segal, J. Sound Vib. 38, 245—54 (1975).
[54]N.T. Adelman, Y. Stavsky, and E. Segal, J. Acoust. Soc. Am. 57, 356—60 (1975).
[55]V.N. Lazutkin, Akust. Zh. 17, 588—92 (1971) [Sov. Phys. Acoust. 17, 496—9 (1972)].
[56]V.N. Lazutkin and A.I. Mikhailov, Akust. Zh. 20, 414—9 (1974) [Sov. Phys. Acoust. 20, 253—5 (1974)].
[57]V.N. Lazutkin and Yu.V. Tsyganov, Akust. Zh. 17, 394—9 (1971) [Sov. Phys. Acoust. 17, 330—4 (1972)].
[58]V.N. Lazutkin and A.I. Mikhailov, Akust. Zh. 20, 582—6 (1974) [Sov. Phys. Acoust. 20, 350—2 (1975)].
[59]I.A. Viktorov, Akust. Zh. 20, 123—7 (1974).
[60]M.C. Dökmeci, J. Math. Phys. 15, 2248—52 (1974).
[61]M.C. Dökmeci, J. Acoust. Soc. Am. 54, 292(A) (1973).
[62]M. Greenspan and R.M. Wilmotte, J. Acoust. Soc. Am. 30, 528—32 (1958).
[63]W.E. Newell, Proc. IEEE 53, 575—81, 1305—9 (1965).
[64]W.J. Fry and F. Dunn, J. Acoust. Soc. Am. 34, 188—92 (1962).

[65]E.K. Sittig, IEEE Trans. Sonics Ultrason. SU-14, 167—174 (1967).
[66]E.K. Sittig, in *Physical Acoustics* (Academic, New York, 1972), Vol. IX, pp. 221—75.
[67]R.V. Schmidt and F.W. Voltmer, IEEE Trans. Microwave Theory Tech. MTT-17, 920—6 (1969).
[68]A.H. Fahmy and E.L. Adler, Appl. Phys. Lett. 22, 495—7 (1973).
[69]A.H. Fahmy and E.L. Adler, Proc. IEEE 122, 470—2 (1975).
[70]N.C. Cheng and C.T. Sun, J. Acoust. Soc. Am. 57, 632—8 (1975).
[71]K.E. Pauley and S.B. Dong, Wave Electron. 1, 265—85 (1976).
[72]B.A. Finlayson and L.E. Scriven, Int. J. Heat Mass Transfer 10, 799—821 (1967).
[73]P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vols. I and II.
[74]C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1960).
[75]S.G. Mikhlin, *Variational Methods in Mathematical Physics* (Macmillan, New York, 1964).
[76]E.P. Eer Nisse, IEEE Trans. Sonics Ultrason. SU-14, 153—60 (1967).
[77]R. Holland and E.P. Eer Nisse, IEEE Trans. Sonics Ultrason. SU-15, 119—32 (1968).
[78]I.A. Vekovishcheva, Prikl. Mekh. 7, 129—33 (1971) [Sov. Appl. Mech. 7, 1049—52 (1974)].
[79]H.F. Tiersten, J. Math. Phys. 9, 1445—50 (1968).
[80]M.C. Dökmeci, Bull. Am. Phys. Soc., Ser. II, 18, 778A (1973); Lett. Nuovo Cimento 7, 449—54 (1973).
[81]M.A. Biot, *Variational Principles in Heat Transfer; A Unified Lagrangian Analysis of Dissipative Phenomena* (Clarendon, London, 1970).
[82]O.D. Kellogg, *Foundations of Potential Theory* (Springer, Berlin, 1929 and Ungar, New York, 1946).
[83]R.D. Mindlin, in *Problems of Continuum Mechanics* (Soc. Ind. Appl. Math., Philadelphia, 1961), Muskhelishvili Anniversary Volume, pp. 282—90.
[84]P.M. Naghdi, in *Handbuch der Physik, Mechanics of Solids* (Springer, New York, 1972), Vol. VIa/2, pp. 425—640.
[85]J.H. Weiner, Q. Appl. Math. 15, 102—5 (1957).
[86]J.L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1949).
[87]J.A. Schouten, *Tensor Analysis for Physicists* (Oxford U.P., London, 1951).
[88]G. Venkataraman, L.A. Feldkamp, and V.C. Sahni, *Dynamics of Perfect Crystals* (M.I.T., Cambridge, Massachusetts, 1975).
[89]H.F. Tiersten, in *Recent Advances in Engineering Science* (Gordon and Breach, New York, 1970), Vol. 5, Part 1, pp. 63—90.
[90]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. 1.
[91]B. Fraeijs De Veubeke, Bull. Serv. Tech. Aeron. (Bruxelles, 1951), No. 24.
[92]B. Fraeijs De Veubeke, in *Fluid Dynamics Transactions* (Polish Sci., Warszawa, 1965), Vol. 3, pp. 111—26.
[93]K. Washizu, *Variational Methods in Elasticity and Plasticity* (Pergamon, New York, 1974), 2nd ed.
[94]P. Rosen, J. Appl. Phys. 25, 336—8 (1954).
[95]P. Glandsdorff and I. Prigogine, Phys. Amsterdam 30, 351—74 (1954).
[96]M.E. Gurtin, Arch. Rat. Mech. Anal. 13, 179—91 (1963).
[97]M.E. Gurtin, in *Handbuch der Physik, Mechanics of Solids* (Springer, New York, 1972), Vol. VIa/2, pp. 1—295.
[98]M.C. Dökmeci, in *Recent Advances in Engineering Science* (Gordon and Breach, New York, 1970), Vol. 5, Part 2, pp. 189—207.
[99]I.N. Vekua, in *Theoretical and Applied Mechanics* (Springer, New York, 1973), Proc. XIIIth Int. Congr. Theor. Appl. Mechs., Moscow, pp. 322—39.
[100]G. Kirchhoff, J. Reine Angew. Math. 56, 285—313 (1859).
[101]F. Neumann, *Vorlesungen über die Theorie der Elasticität der festen Körper und des Lichtäthers* (Teubner, Leipzig, 1885).
[102]M.C. Dökmeci and MG. AlpD., Rheol. Acta 12, 106—13 (1973).
[103]R.J. Knops and L.E. Payne, *Uniqueness Theorems in Linear Elasticity* (Springer, New York, 1972).
[104]M.C. Dökmeci, J. Eng. Math. 9, 311—22 (1975).
[105]J.D. Achenbach and G. Herrmann, in *Developments in Mechanics, Proc. 10th Midwestern Mech. Conf.* (Johnson, Boulder, Colorado, 1968), Vol. 4, pp. 91—106.
[106]M.C. Dökmeci, in *2nd Inter. Conf. on Struct. Mechs. in Reactor Technology* (Comm. European Communities, *Brussels*, 1973), Vol. V, Part M, Paper M5/9, pp. 1—17.
[107]J.J. Gagnepain and R. Besson, in *Physical Acoustics* (Academic, New York, 1975), Vol. XI, pp. 245—88.
[108]P.C.Y. Lee and Z. Nikodem, Int. J. Solids Structs. 8, 581—612 (1972).
[109]J.T. Oden and B.E. Kelley, Int. J. Numer. Meth. Eng. 3, 161—179 (1971).
[110]R.H. Gallagher, *Finite Element Analysis: Fundamentals* (Prentice-Hall, Englewood Cliffs, New Jersey, 1975).
[111]T. Bhattacharyya, M. Tsutsumi, and N. Kumagai, IEEE Trans. Microwave Theory Tech. MTT-24, 226—30 (1976).

Fluctuation theories and Gaussian stochastic processes

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 16 May 1977)

The theory of fluctuations for systems near equilibrium has given rise to two developments which generalize the theory in two distinct ways. One of these developments is focused on the theory of fluctuations far from equilibrium where the dynamics is nonlinear. The other development has focused on extending the class of fluctuating forces to include forces with non-delta function correlations. The near equilibrium theory corresponds with the theory of *stationary, Gaussian, Markov processes*; the nonlinear, far from equilibrium theory corresponds to the theory of *nonstationary, Gaussian, Markov processes*; and the non-delta function, force correlation theory corresponds to the theory of *stationary, Gaussian, non-Markovian* processes in one form, and to the theory of *nonstationary, Gaussian, non-Markovian* processes in another form. The common feature found in all these theories is Gaussianness.

I. INTRODUCTION

Over an extended period of time, Langevin's[1] theory of Brownian motion has developed into a remarkably general theory for nonequilibrium processes. The early work of Uhlenbeck and Ornstein,[2] which applied Langevin's theory to the harmonic oscillator, provided the basis upon which Onsager and Machlup[3] were able to establish a stochastic foundation for irreversible thermodynamics. Subsequently, Landau and Lifshitz suggested how to extend this theory even further when they presented equations for hydrodynamic[4] fluctuations and for electromagnetic[5] fluctuations. These suggestions gave rise to a general formulation,[6] which Fox and Uhlenbeck[7] showed was equivalent to the theory of *stationary, Gaussian, Markov processes.*

Although the theory is quite general, it is not universal, and is limited in two distant physical ways. First of all, the applications mentioned above, as well as all others, are confined to systems which are dynamically close to full equilibrium. This implies that the equations are linear. Secondly, on the time scale determined by the relaxation time for the system, the correlation time for the fluctuating forces is negligibly short, as is manifested mathematically by a fluctuating force autocorrelation formula involving a delta function of the time variables. The general situation is exhibited by the equations for a *stationary, Gaussian, Markov* process.[7]

$$\frac{d}{dt}\mathbf{a}(t) = -\mathbf{A}\mathbf{a}(t) - \mathbf{S}\mathbf{a}(t) + \tilde{\mathbf{F}}(t), \tag{1}$$

with

$$\langle\tilde{\mathbf{F}}(t)\rangle = 0 \text{ and } \langle\tilde{F}(t)\tilde{F}^{\dagger}(s)\rangle = 2Q(t-s), \tag{2}$$

in which $\mathbf{a}(t)$ is a vector valued, N-component process, $\mathbf{A}$ is an $N\times N$ antisymmetric matrix, $\mathbf{S}$ is an $N\times N$ symmetric matrix with nonnegative eigenvalues, $\tilde{\mathbf{F}}(t)$ is a vector valued, N-component fluctuating force, and $\langle\cdots\rangle$ denotes stochastic averaging. $\mathbf{Q}$ is an $N\times N$ symmetric matrix of correlations. Near equilibrium, the entropy of the system is given by the quadratic form,[3,7]

$$S(t) = S_0 - \tfrac{1}{2}\mathbf{a}^{\dagger}(t)\mathbf{E}\mathbf{a}(t), \tag{3}$$

in which $\mathbf{E}$ is the $N\times N$, symmetric entropy matrix. A fluctuation–dissipation relation exists and is expressed by

$$2\mathbf{Q} = (\mathbf{A}+\mathbf{S})\mathbf{E}^{-1} + \mathbf{E}^{-1}(\mathbf{A}^{\dagger}+\mathbf{S}^{\dagger}). \tag{4}$$

In Eqs. (1)–(4), † denotes the matrix adjoint. The entire statistical description of these processes is given by the solution to an associated Fokker–Planck equation.[7]

The limitations of the theory have prompted two developments which are aimed at extending the domain of applicability. The first development concerns establishment of a theory of fluctuations for systems far from equilibrium in which case the dynamics become nonlinear. van Kampen[8] initiated this development using a master equation basis which has recently been clarified by Kubo *et al.*[9] In addition, an equivalent, Langevin-like, fluctuating force theory has been proposed by Keizer,[10] Nitzan *et al.*,[11] Akcasu,[12] and Portnow and Kitahara.[13] The second development concerns the linear, near equilibrium theory in the case in which the fluctuating force autocorrelation formula involves an extended time dependence which can not properly be represented by a delta function. Zwanzig's[14] projection operator method is the key to this development, although it was Mori[15] who first achieved a general formulation. An alternative basis for these "generalized Langevin" equations involving "contraction of the description"[16] has been presented by Hauge and Martin-Löf.[17] Their theory explicitly uses Gaussian forces whereas the Mori theory usually leaves the characterization of the higher than second-order moments uspecified. Here the forces will be taken as Gaussian in the Mori theory as well.

What emerges as common among these two developments and the original theory is Gaussianness. Because 1977 is the 200th anniversary of Gauss' birth, it seems especially appropriate to present a unified treatment of these Gaussian theories.

II. FLUCTUATIONS FAR FROM EQUILIBRIUM

The description of fluctuations far from equilibrium is based upon taking the "thermodynamic limit" of a master equation description of the underlying microscopic process. In both van Kampen's[8] pioneering work

 0022-2488/78/1901-0127$1.00

and Kubo's more recent extention, this limit is the limit of a large "system-size" parameter. Regardless of how one expresses it, the result is that in the limit of a very large system, the master equation description goes over into a Fokker–Planck equation description. The resulting Fokker–Planck equation represents a *nonstationary, Gaussian, Markov process.*

For the special case of a system of chemical reactions obeying the McQuarrie[18] master equation, Kurtz[19] has proved, with all the rigor one would like, that the limit of a very large volume yields a Fokker–Planck equation. In a general setting, this result is not so well established, and Fox and Kac[20] have recently explored another special case in which delicate questions regarding initial distributions and scaling have been broached. In any case, the resulting Fokker–Planck equation has the form

$$\frac{\partial}{\partial t}P(\mathbf{n},t) = -\frac{\partial}{\partial \mathbf{n}}\cdot(\mathbf{H}(t)\mathbf{n}P(\mathbf{n},t)) + \frac{\partial}{\partial \mathbf{n}}\cdot\boldsymbol{\gamma}(t)\cdot\frac{\partial}{\partial \mathbf{n}}P(\mathbf{n},t), \tag{5}$$

in which $\mathbf{n}$ is an N-component vector valued process, $\mathbf{H}(t)$ is an $N\times N$ matrix, and $\boldsymbol{\gamma}(t)$ is an $N\times N$ matrix. This Fokker–Planck equation describes the time evolution of the distribution function for the *deviations around the deterministic solution,* $\bar{\mathbf{n}}(t)$. The equations which determine $\bar{\mathbf{n}}(t)$ are also deduced from the large system-size limit of the underlying master equation and are generally nonlinear:

$$\frac{d}{dt}\bar{\mathbf{n}}(t) = \mathbf{K}[\bar{\mathbf{n}}(t)], \tag{6}$$

in which $\mathbf{K}$ is a vector valued functional of $\bar{\mathbf{n}}(t)$. The matrix elements of $\mathbf{H}(t)$, which appears in (5), are determined from $\mathbf{K}[\bar{\mathbf{n}}(t)]$ by

$$H_{ij}(t) \equiv \frac{\partial K_i}{\partial \bar{n}_j}. \tag{7}$$

The derivation of (5) from the underlying master equation also shows that $\boldsymbol{\gamma}(t)$ is a functional of $\bar{\mathbf{n}}(t)$. Therefore, both $\mathbf{H}(t)$ and $\boldsymbol{\gamma}(t)$ acquire their time dependence indirectly, through $\bar{\mathbf{n}}(t)$, the deterministic solution to (6). In this kind of theory, the deterministic solution carries the fluctuations "on its back", so to speak, and there is no feedback on the deterministic solution by the fluctuations. Therefore, there is no possibility of "renormalization"[21] of dissipative parameters in the deterministic equation (6) by fluctuations which are determined from (5).

Even though van Kampen[22] has repeatedly expressed the view that there does not exist an equivalent fluctuating force formulation of this theory, Nitzan[11] *et al.* and Portnow and Kitahara[13] have each shown that a Langevin-like formulation for fluctuations near steady states which are far from equilibrium is possible, and Akcasu[12] and Keizer[10] have independently presented theories of Langevin-like structure for the entire nonequilibrium domain. Each theory can be represented by an equation like

$$\frac{d}{dt}\mathbf{n}(t) = \mathbf{H}(t)\mathbf{n}(t) + \tilde{\mathbf{g}}(t), \tag{8}$$

with $\tilde{\mathbf{g}}(t)$ and N-component, Gaussian fluctuating force satsifying

$$\langle\tilde{\mathbf{g}}(t)\rangle = 0 \quad \text{and} \quad \langle\tilde{\mathbf{g}}(t)\tilde{\mathbf{g}}^{\dagger}(s)\rangle = 2\boldsymbol{\gamma}(t)\delta(t-s). \tag{9}$$

The solution to (8) may be expressed in terms of the time-ordered exponential, which is denoted by $\underset{\sim}{T}\exp\{\int_0^t ds\cdots\}$

$$\mathbf{n}(t) = \underset{\sim}{T}\exp\Big\{\int_0^t ds\,\mathbf{H}(s)\Big\}\mathbf{n}(0) + \int_0^t \underset{\sim}{T}\exp\Big\{\int_s^t \mathbf{H}(s')\,ds'\Big\}\tilde{\mathbf{g}}(s)\,ds \tag{10}$$

and this result has average value

$$\langle\mathbf{n}(t)\rangle = \underset{\sim}{T}\exp\Big\{\int_0^t \mathbf{H}(s)\,ds\Big\}\mathbf{n}(0) \tag{11}$$

which agrees with what would be obtained from (5). Similarly, if $\Sigma^{-1}(t)$ denotes the covariance matrix which is defined by

$$\Sigma^{-1}(t) \equiv \langle(\mathbf{n}(t) - \langle\mathbf{n}(t)\rangle)(\mathbf{n}(t) - \langle\mathbf{n}(t)\rangle)^{\dagger}\rangle, \tag{12}$$

then (8) implies

$$\frac{d}{dt}\Sigma^{-1}(t) = \mathbf{H}(t)\Sigma^{-1}(t) + \Sigma^{-1}(t)\mathbf{H}^{\dagger}(t) + \boldsymbol{\gamma}(t) + \boldsymbol{\gamma}^{\dagger}(t) \tag{13}$$

which also agrees with what would be obtained from (5). Because the process is Gaussian, (11) and (12), or (11) and (13) completely determine the process and are clearly equivalent with (5).

From (9) it follows that $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\dagger}$. Consequently, at equilibrium *or* at a steady state where $(d/dt)\Sigma^{-1}(t) = 0$, we get a fluctuation-dissipation relation from (13):

$$2\boldsymbol{\gamma} = \mathbf{H}\Sigma^{-1} + \Sigma^{-1}\mathbf{H}^{\dagger} \tag{14}$$

in which $\boldsymbol{\gamma}$, $\mathbf{H}$, and Σ^{-1} assume their equilibrium, or steady state, values. Thus, there is a parallel to (4) even at steady states.[23]

III. GENERALIZED LANGEVIN EQUATIONS

The description of fluctuations near equilibrium with fluctuating forces which have a non-delta function correlation can be given in a form closely resembling the formulation of Mori[15] and the representation used by Hauge and Martin-Löf.[17] The differences are merely a matter of notational convention. As pointed out in the introduction, Gaussian fluctuating forces are assumed.

The equations take the form

$$\frac{d}{dt}\mathbf{a}(t) = \Omega\mathbf{a}(t) - \int_0^t \phi(t-s)\mathbf{a}(s)\,ds + \tilde{\mathbf{f}}(t) \tag{15}$$

with $\tilde{\mathbf{f}}(t)$ satisfying

$$\langle\tilde{\mathbf{f}}(t)\rangle = 0 \quad \text{and} \quad \langle\tilde{\mathbf{f}}(t)\tilde{\mathbf{f}}^{\dagger}(s)\rangle = \phi(|t-s|)\{\mathbf{a}(0)\mathbf{a}^{\dagger}(0)\} \tag{16}$$

in which $\{\mathbf{a}(0)\mathbf{a}^{\dagger}(0)\}$ denotes an equilibrium average of the initial $\mathbf{a}$ values. The $\mathbf{a}$ variables are defined such that $\{\mathbf{a}(0)\} = 0$. The matrix Ω represents nondissipative coupling and satisfies the identity[24]

$$\{\mathbf{a}(0)\mathbf{a}^{\dagger}(0)\}\Omega^{\dagger} = -\Omega\{\mathbf{a}(0)\mathbf{a}^{\dagger}(0)\}. \tag{17}$$

The matrix kernel, $\phi(t-s)$, is the memory kernel which makes the process *non-Markovian*. Following Uhlenbeck and Ornstein,[2] we will use both $\langle\cdots\rangle$ and $\{\cdots\}$ averages in the following analysis. This will lead to the consequence that the equations describe a *stationary, Gaussian, non-Markovian process*.

The solution to (15) is

$$\mathbf{a}(t) = \mathbf{M}(t)\mathbf{a}(0) + \int_0^t \mathbf{M}(t-s)\tilde{\mathbf{f}}(s)\,ds, \tag{18}$$

in which $\mathbf{M}(t)$ is defined through its Laplace transform, $\hat{\mathbf{M}}(z)$, by

$$\hat{\mathbf{M}}(z) \equiv [z\mathbf{1} - \Omega + \hat{\phi}(z)]^{-1}, \tag{19}$$

where $\hat{\phi}(z)$ is the Laplace transform of $\phi(|t|)$. From (16) and (18) it follows that

$$\{\langle \mathbf{a}(t)\rangle\} = 0 \tag{20}$$

and that the autocorrelation matrix is, for $t \geq s$

$$\begin{aligned}\{\langle \mathbf{a}(t)\mathbf{a}^\dagger(s)\rangle\} &= \mathbf{M}(t)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(s) \\ &\quad + \int_0^t dt' \int_0^s ds'\, \mathbf{M}(t-t')\phi(|t'-s'|) \\ &\quad \times \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(s-s').\end{aligned} \tag{21}$$

In the Appendix, we prove that

$$\begin{aligned}&\int_0^t dt' \int_0^s ds'\, \mathbf{M}(t-t')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(s-s') \\ &= \tfrac{1}{2}\left(\mathbf{M}(|t-s|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} + \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(|t-s|)\right) \\ &\quad - \mathbf{M}(t)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(s).\end{aligned} \tag{22}$$

Together, (21) and (22) imply the *stationary* result

$$\begin{aligned}&\{\langle \mathbf{a}(t)\mathbf{a}^\dagger(s)\rangle\} \\ &= \tfrac{1}{2}\left(\mathbf{M}(t-s)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} + \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(t-s)\right).\end{aligned} \tag{23}$$

From (18) and (19), we have $\mathbf{M}(0) = \mathbf{1}$, so that (23) implies for $t = s$ that

$$\{\langle \mathbf{a}(t)\mathbf{a}^\dagger(t)\rangle\} = \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}, \tag{24}$$

which exhibits *stationarity* another way.

It should be emphasized that it is also possible to analyze the equations (15) and (16) without performing the second, $\{\cdots\}$-averaging, while computing the autocorrelation matrix. The results then correspond with a *nonstationary, Gaussian, non-Markovian process*.

Clearly, (16) parallels (4) and (14) as the fluctuation–dissipation relation. The parallel is most easily seen if it is noted that $\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}$ is the natural analog of $\mathbf{E}^{-1}$ in (4) and of Σ^{-1} in (14), and if Eqs. (17) and (A4) are used.

IV. DISCUSSION OF RESULTS

The common feature found in all the fluctuation theories discussed here has been Gaussianness. For the two Markovian theories, all of the statistics can be determined by the Fokker–Planck equations or equivalently by their solutions, the conditional two time probability distribution. The non-Markovian theories are also determined completely by a two-time correlation matrix because these theories are Gaussian.[25] Therefore, as far as being able to totally determine the process is concerned, Gaussianness appears to be more general than Markovianness. Moreover, even a non-Markovian case is still fully described by a *single*, two-time correlation matrix.[25] In view of this circumstance, it seems appropriate to enquire whether or not it is possible to build a foundation for stochastic processes based upon Gaussian fluctuations while relinquishing the desire for Markovianness. This would enable one to avoid the technical inconsistencies of delta correlated, Markovian equations which prompted Doob[26] to recast the Langevin equation, and which has since lead to the Ito and Stratonovich calculi.[27] In a subsequent paper we will show how Gaussian, non-Markovian stochastic processes may be formulated with the ordinary Riemann integral, without the "stochastic" integral,[27,28] and in close parallel with the equations of microscopic dynamics with which physics is usually presented.

ACKNOWLEDGMENT

This work has been supported by a grant from the National Science Foundation, #PHY77-07372.

APPENDIX: PROOF OF (23)

The proof utilizes the double Laplace transform

$$\begin{aligned}&\int_0^\infty dt \exp(-zt) \int_0^\infty ds \exp(-z's) \int_0^t dt' \int_0^s ds'\, \mathbf{M}(t-t')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(s-s') \\ &= \int_0^\infty dt' \int_0^\infty ds' \int_{t'}^\infty dt \int_{s'}^\infty ds \exp[-z(t-t')]\,\mathbf{M}(t-t')\exp(-zt')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\exp(-z's')\exp[-z'(s-s')]M^\dagger(s-s') \\ &= \int_0^\infty dt' \int_0^\infty ds' \int_0^\infty d\tau \int_0^\infty d\sigma \exp(-z\tau)\mathbf{M}(\tau)\exp(-zt')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\exp(-z's')\exp(-z'\sigma)M^\dagger(\sigma) \\ &= \hat{\mathbf{M}}(z) \int_0^\infty dt' \int_0^\infty ds' \exp(-zt')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\exp(-z's')\hat{\mathbf{M}}^\dagger(z').\end{aligned} \tag{A1}$$

Now,

$$\begin{aligned}&\int_0^\infty dt' \int_0^\infty ds' \exp(-zt')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\exp(-z's') = \int_0^\infty dt' \int_0^\infty ds' \exp[-z(t'-s')]\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\exp[-(z+z')s'] \\ &= \int_0^\infty ds' \int_{-s'}^\infty d\sigma \exp(-z\sigma)\phi(|\sigma|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\exp[-(z+z')s']\end{aligned}$$

$$= \int_0^\infty ds'(\hat{\phi}(z) + \int_{-s'}^0 d\sigma \exp(-z\sigma)\phi(|\sigma|))\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} \exp[-(z+z')s']$$

$$= \frac{\hat{\phi}(z)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}}{z+z'} - \int_0^\infty ds' \left(\frac{1}{z+z'}\frac{d}{ds'} \exp[-(z+z')s']\right) \int_{-s'}^0 d\sigma \exp(-z\sigma)\phi(|\sigma|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}$$

$$= \frac{\hat{\phi}(z)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}}{z+z'} + \frac{1}{z+z'} \int_0^\infty ds' \exp[-(z+z')s' \exp(zs')\phi(s')\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} = \frac{\hat{\phi}(z)+\hat{\phi}(z')}{z+z'}\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}. \tag{A2}$$

From (16) it follows that

$$\langle \tilde{\mathbf{f}}(t)\mathbf{f}^\dagger(s)\rangle^\dagger = \langle \tilde{\mathbf{f}}(s)\tilde{\mathbf{f}}^\dagger(t)\rangle = \phi(|s-t|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} = \langle \tilde{\mathbf{f}}(t)\tilde{\mathbf{f}}^\dagger(s)\rangle. \tag{A3}$$

Therefore,

$$\phi(|t-s|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} = \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\phi^\dagger(|t-s|). \tag{A4}$$

Using (A2) and (A4), the last line of (A1) can be written

$$\hat{\mathbf{M}}(z) \int_0^\infty dt' \int_0^\infty ds' \exp(-zt')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} \exp(-z's')\hat{\mathbf{M}}^\dagger(z')$$

$$= \hat{\mathbf{M}}(z) \frac{\hat{\phi}(z)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} + \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\hat{\phi}^\dagger(z')}{z+z'} \hat{\mathbf{M}}^\dagger(z') \tag{A5}$$

and (A2) can be rewritten

$$\tfrac{1}{2} \int_0^\infty dt' \int_0^\infty ds' \exp(-zt')(\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} + \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\phi^\dagger(|t'-s'|)) \exp(-z's')$$

$$= \frac{\hat{\phi}(z)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} + \{\tilde{\mathbf{a}}(0)\mathbf{a}^\dagger(0)\}\hat{\phi}^\dagger(z')}{z+z'}. \tag{A6}$$

From (20) and its adjoint we get

$$\hat{\mathbf{M}}(z)\hat{\phi}(z) = \mathbf{1} - \hat{\mathbf{M}}(z)(z\mathbf{1} - \Omega) \text{ and } \hat{\phi}^\dagger(z')\hat{\mathbf{M}}^\dagger(z') = \mathbf{1} - (z'\mathbf{1} - \Omega^\dagger)\hat{\mathbf{M}}^\dagger(z'). \tag{A7}$$

Using (A7) in (A5) gives

$$\hat{\mathbf{M}}(z) \int_0^\infty dt' \int_0^\infty ds' \exp(-zt')\phi(|t'-s'|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} \exp(-z's')\hat{\mathbf{M}}^\dagger(z')$$

$$= \frac{\hat{\mathbf{M}}(z)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} + \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\hat{\mathbf{M}}^\dagger(z')}{z+z'} - \hat{\mathbf{M}}(z)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\hat{\mathbf{M}}^\dagger(z'), \tag{A8}$$

in which (17) has been used. From (A6) it is seen that this is the same as the double Laplace transform of

$$\tfrac{1}{2}(\mathbf{M}(|t-s|)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\} + \{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\hat{\mathbf{M}}^\dagger(|t-s|)) - \mathbf{M}(t)\{\mathbf{a}(0)\mathbf{a}^\dagger(0)\}\mathbf{M}^\dagger(s),$$

which completes the proof of (23).

[1]P. Langevin, C.R. Acad. Sci. (Paris) **146**, 530 (1908).
[2]G.E. Uhlenbeck and L.S. Ornstein, Phys. Rev. **36**, 823 (1930).
[3]L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).
[4]L.D. Landau and E.M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959), Chap. 17.
[5]L.D. Landau and E.M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960), Chap. 13.
[6]S.R. DeGroot and P. Mazur, *Non-equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
[7]R.F. Fox and G.E. Uhlenbeck, Phys. Fluids **13**, 1893, 2881 (1970).
[8]N.G. van Kampen, Can. J. Phys. **39**, 551 (1961).
[9]R. Kubo, K. Matsuo, and K. Kitahara, J. Stat. Phys. **9**, 51 (1973).
[10]J.E. Keizer, J. Chem. Phys. **63**, 398, 5037 (1975).
[11]A. Nitzan, P. Ortoleva, J. Deutch, and J. Ross, J. Chem. Phys. **61**, 1056 (1974).
[12]Z.A. Akcasu, J. Stat. Phys. **16**, 33 (1977).
[13]J. Portnow and K. Kitahara, J. Stat. Phys. **14**, 501 (1976).
[14]R. Zwanzig, Phys. Rev. **124**, 983 (1961).
[15]H. Mori, Prog. Theor. Phys. **33**, 423 (1965).
[16]G.E. Uhlenbeck and G.W. Ford, *Lectures in Statistical Mechaics* (American Mathematics Society, Providence, R.I., 1963), Chap. 1.
[17]E.H. Hauge and A. Martin-Löf, J. Stat. Phys. **7**, 259 (1973).
[18]D.A. McQuarrie, J. Appl. Prob. **4**, 413 (1967).
[19]T.G. Kurtz, J. Chem. Phys. **57**, 2976 (1972).
[20]R.F. Fox and M. Kac, BioSystems **8**, 187 (1977).
[21]P. Mazur, in *Fundamental Problems in Statistical Mechanics III* (North-Holland, Amsterdam, 1975), Chap. 11.
[22]N.G. van Kampen, Adv. Chem. Phys. **34**, 245 (1976).
[23]J.E. Keizer (personal communication, 1974); also see J. Chem. Phys. **65**, 4431 (1976).
[24]See Ref. 15; Mori uses $i\Omega$ where I have used Ω.
[25]R.F. Fox, J. Math. Phys. **18**, 2331 (1977).
[26]J.L. Doob, Ann. Math. **43**, 351 (1942).
[27]L. Arnold, *Stochastic Differential Equations* (Wiley-Interscience, New York, 1974).
[28]H.P. McKean, *Stochastic Integrals* (Academic, New York, 1969).

Transforming Gaussians into Wannier functions [a)]

Gregory H. Wannier

Physics Department, University of Oregon, Eugene, Oregon 97403
(Received 25 July 1977)

It is clear *a priori* that equal Gaussian functions, spread over a lattice, can be transformed into Wannier functions. The transformation is carried out here analytically in one dimension, with the help of the theory of theta functions. The results confirm and illustrate the properties commonly assigned to these functions, with one startling exception.

1. INTRODUCTION

In the last few years it has been realized that Wannier functions can simplify the calculation of some properties of solids, and correspondingly, a beginning has been made in their calculation.[1–5] This article represents a contribution to this endeavor. In my study of diamagnetic properties of electrons I realized accidentally that a set of Gaussian functions, distributed over a regular lattice, can be transformed analytically into Wannier functions. The occasion arises therefore to exhibit them both analytically and numerically in this particular case. The analytical procedure furnishes proofs for things that have usually been inferred numerically, and the curves generally confirm the observations made by others. One striking departure shows up; it will be discussed in the text.

2. CONSTRUCTION OF THE WANNIER FUNCTIONS

Gaussian functions in three dimensions are the product of one-dimensional Gaussians. They have therefore the great advantage that if one orthonormalizes them in one dimension orthonormalization in three dimensions is automatic. Therefore, only the one-dimensional pro problem will be discussed below, namely how to transform into Wannier functions a set of Gaussians lying on a straight line a fixed distance apart. The problem so formulated contains one variable parameter, namely the ratio of the lattice spacing to the Gaussian spread. The end results and the exhibited curves will be functions of that parameter.

The procedure we shall use is the one oulined at the time Wannier functions were invented.[6] We construct Bloch functions out of the Gaussians in the form of Fourier series in k, normalize the Bloch functions, and then Fourier invert back to obtain the Wannier functions. This procedure brings in a square root, the square root of the norm. Both the Bloch function and the norm happen to be theta functions. The rich treasure of internal and mutual relationships for these functions makes an analytic procedure possible. The notation and formulas for theta functions employed here are the ones of Whittaker and Watson.[7]

We adopt 1 as the interval of periodicity and treat the Gaussians as having a variable spread ϕ. Therefore, our starting functions read

$$\psi(x-n)=\exp[-\tfrac{1}{2}\pi\phi(x-n)^2], \tag{1}$$

where n runs over the set of integers. The Bloch functions derived therefrom then read

$$b(x,k)=\sum_{n=-\infty}^{+\infty}\exp[2\pi ikn-\tfrac{1}{2}\pi\phi(x-n)^2]. \tag{2}$$

Here the "Dutch" k, equal to the reciprocal wavelength, has been employed for convenience.

In the notation of Whittaker and Watson,[7] expression (2) equals

$$b(x,k)=\exp[-\tfrac{1}{2}\pi\phi x^2]\,\vartheta_3[\pi k-\tfrac{1}{2}\pi i\phi x,\ \exp(-\pi\phi/2)],$$

The Jacobi transformation can therefore be applied to it, yielding the alternate form

$$b(x,k)=(2/\phi)^{1/2}\sum_{n=-\infty}^{+\infty}\exp[2\pi i(n+k)x-(2\pi/\phi)(n+k)^2]. \tag{3}$$

It is seen that the passage from (2) to (3) gets us without difficulty from the one common representation of Bloch functions to the other.

The prescription for forming Wannier functions[6] now demands that the Bloch functions above be normalized over the primitive cell; this is here the unit interval. The normalization is immediate on the form (3),

$$\mathrm{NORM}=\int_0^1|b(x,k)|^2dx=(2/\phi)\sum_{n=-\infty}^{+\infty}\exp\left(-\frac{4\pi}{\phi}(n+k)^2\right). \tag{4}$$

The norm is again a theta function.[3] Its alternate form reads

$$\mathrm{NORM}=\phi^{-1/2}\sum_{n=-\infty}^{+\infty}\exp[2\pi ikn-\tfrac{1}{4}\pi\phi n^2]. \tag{5}$$

To follow the procedure further, the reciprocal of the square root of this expression has to be found. We pass for this purpose to Jacobi's expression of the function ϑ_3 as an infinite product.[7] We find with the abbreviation

$$q=-\exp[-\tfrac{1}{4}\pi\phi], \tag{6}$$

$$\mathrm{NORM}=\phi^{-1/2}\prod_{\nu=1}^{\infty}(1-q^{2\nu})\{1-2q^{2\nu-1}\cos 2\pi k+q^{4\nu-2}\}. \tag{7}$$

The curly bracket can be further split, yielding

$$\mathrm{NORM}=\phi^{-1/2}\prod_{\nu=1}^{\infty}(1-q^{2\nu})\{1-q^{2\nu-1}\exp(2\pi ik)\}\times\{1-q^{2\nu-1}\exp(-2\pi ik)\}. \tag{8}$$

The two curly brackets in (8) are conjugate complex, that is each has an absolute value which is the square root of the absolute value of the curly bracket in (7). The square root of the norm therefore equals

$$(\mathrm{NORM})^{1/2}=\phi^{-1/4}\prod_{\nu=1}^{\infty}(1-q^{2\nu})^{1/2}\,|1-q^{2\nu-1}\exp(-2\pi ik)|. \tag{9a}$$

We now take advantage of the freedom which we have in

[a)]This work was supported in part by the National Science Foundation.

 0022-2488/78/1901-0131$1.00

the formation of Wannier functions. The Bloch functions must be normalized, but are allowed to contain an arbitrary phase factor depending periodically on k. Different phase assignments yield different, but equally valid Wannier functions. We use this freedom to adopt (9a) without absolute bars as the definition of the norm:

$$(\mathrm{NORM})^{1/2}=\phi^{-1/4}\prod_{\nu=1}^{\infty}(1-q^{2\nu})^{1/2}\times[1-q^{2\nu-1}\exp(-2\pi ik)]. \tag{9b}$$

We now need the reciprocal of this which means essentially the reciprocal R of the last set of factors in (9b),

$$R=\prod_{\nu=0}^{\infty}(1+q^{2\nu+1}\exp(-2\pi ik)+q^{4\nu+2}\exp(-4\pi ik)+q^{6\nu+3}\exp(-6\pi ik)+\cdots). \tag{10a}$$

The product R should be available as a Fourier series in k; this means ordering it with respect to powers of $\exp(-2\pi ik)$,

$$R=1+C_1\exp(-2\pi ik)+C_2\exp(-4\pi ik)+C_3\exp(-6\pi ik)+\cdots. \tag{10b}$$

Each coefficient C_n is a sum of terms consisting of n factors. Each factor is an odd power of q; the same factor may occur repeatedly, but no combination of powers occurs more than once. C_n is the sum of all possible such combinations.

I owe to Milton Wannier[8] the evaluation of this coefficient. One starts out with the observation that the powers in each term may be assumed ordered, starting with the smallest. Thus if we are to find the coefficient C_4 of $\exp(-8\pi ik)$ we write down the expression

$$C_4=\sum_{n_1}\sum_{n_2}\sum_{n_3}\sum_{n_4}q^{2n_1+1}q^{2n_2+1}q^{2n_3+1}q^{2n_4+1} \tag{10c}$$

with the restrictions

$$0\leq n_1\leq n_2\leq n_3\leq n_4<\infty. \tag{10d}$$

We liberate ourselves from the restriction (10d) by the substitutions

$$n_1=m_1,$$
$$n_2=m_1+m_2,$$
$$n_3=m_1+m_2+m_3,$$
$$n_4=m_1+m_2+m_3+m_4.$$

(10b) then takes the form

$$C_4=q^4\sum_{m_1}\sum_{m_2}\sum_{m_3}\sum_{m_4}q^{2(m_4+2m_3+3m_2+4m_1)}=q^4\sum_{m_4}q^{2m_4}\sum_{m_3}q^{4m_3}\sum_{m_2}q^{6m_2}\sum_{m_1}q^{8m_1}.$$

The summation indices m_i range over all nonnegative integers and do so independently of each other. We can therefore sum over each index separately and obtain

$$C_4=q^4(1-q^2)^{-1}(1-q^4)^{-1}(1-q^6)^{-1}(1-q^8)^{-1} \tag{10e}$$

and so on for the other C_n's. Inserting this into (10b) we end up with

$$R=1+\sum_{\nu=1}^{\infty}q^{\nu}\exp(-2\pi i\nu k)/\prod_{\mu=1}^{\nu}(1-q^{2\mu}). \tag{10f}$$

Substitution of (10f) into (9) thus yields the reciprocal of the square root of the norm

$$(\mathrm{NORM})^{-1/2}=\phi^{1/4}\prod_{\nu=1}^{\infty}(1-q^{2\nu})^{-1/2}\times\{1+\sum_{\nu=1}^{\infty}q^{\nu}\exp(-2\pi i\nu k)/\prod_{\mu=1}^{\nu}(1-q^{2\mu})\}. \tag{11}$$

By the procedure outlined the Wannier function centered at the origin equals

$$a(x)=\int_0^1(\mathrm{NORM})^{-1/2}b(x,k)\,dk \tag{12}$$

and all other Wannier functions are obtained from this one by simple displacement. If we take $(\mathrm{NORM})^{-1/2}$ in the form (11) and the Bloch function in the form (2), then the two factors are both in the form of Fourier series in k. The integral is therefore trivial and yields, with (6)

$$a(x)=\phi^{1/4}\prod_{\nu=1}^{\infty}\{1-\exp(-\tfrac{1}{2}\pi\phi\nu)\}\{\exp(-\tfrac{1}{2}\pi\phi x^2)+\sum_{\mu=1}^{\infty}[-\exp(-\tfrac{1}{4}\pi\phi)]^{\mu}\exp[-\tfrac{1}{2}\pi\phi(x-\mu)^2]/\prod_{\rho=1}^{\mu}\{1-\exp(-\tfrac{1}{2}\pi\phi\rho)\}\}. \tag{13}$$

3. DISCUSSION AND FUTURE OUTLOOK

Equation (13) fulfills the promise made at the beginning of the paper, namely to exhibit Wannier functions explicitly in a special case. The formula does this by yielding an explicit expression for the coefficient with which a given Gaussian participates in a Wannier function. If the one parameter of the problem, $\phi^{1/2}$, the ratio of the lattice spacing to the Gaussian spread, is known the expressions will be numbers. The infinite

TABLE I. Table of $I(h)$ [Eq. (15)] versus h for $\phi=2$: proof that we deal with a Wannier function.

h	$I(h)$
0.0	1.000000
0.1	0.979600
0.2	0.920281
0.3	0.827436
0.4	0.709240
0.5	0.575600
0.6	0.436937
0.7	0.303026
0.8	0.182035
0.9	0.079891
1.0	0.000000
1.2	−0.091353
1.4	−0.106788
1.6	−0.077548
1.8	−0.034898
2.0	0.000000
2.2	0.018571
2.4	0.021938
2.6	0.016019
2.8	0.007230
3.0	0.000000
3.2	−0.003857
3.4	−0.004558
3.6	−0.003329
3.8	−0.001503
4.0	0.000000
4.5	0.000852
5.0	0.000000
5.5	−0.000177
6.0	0.000000

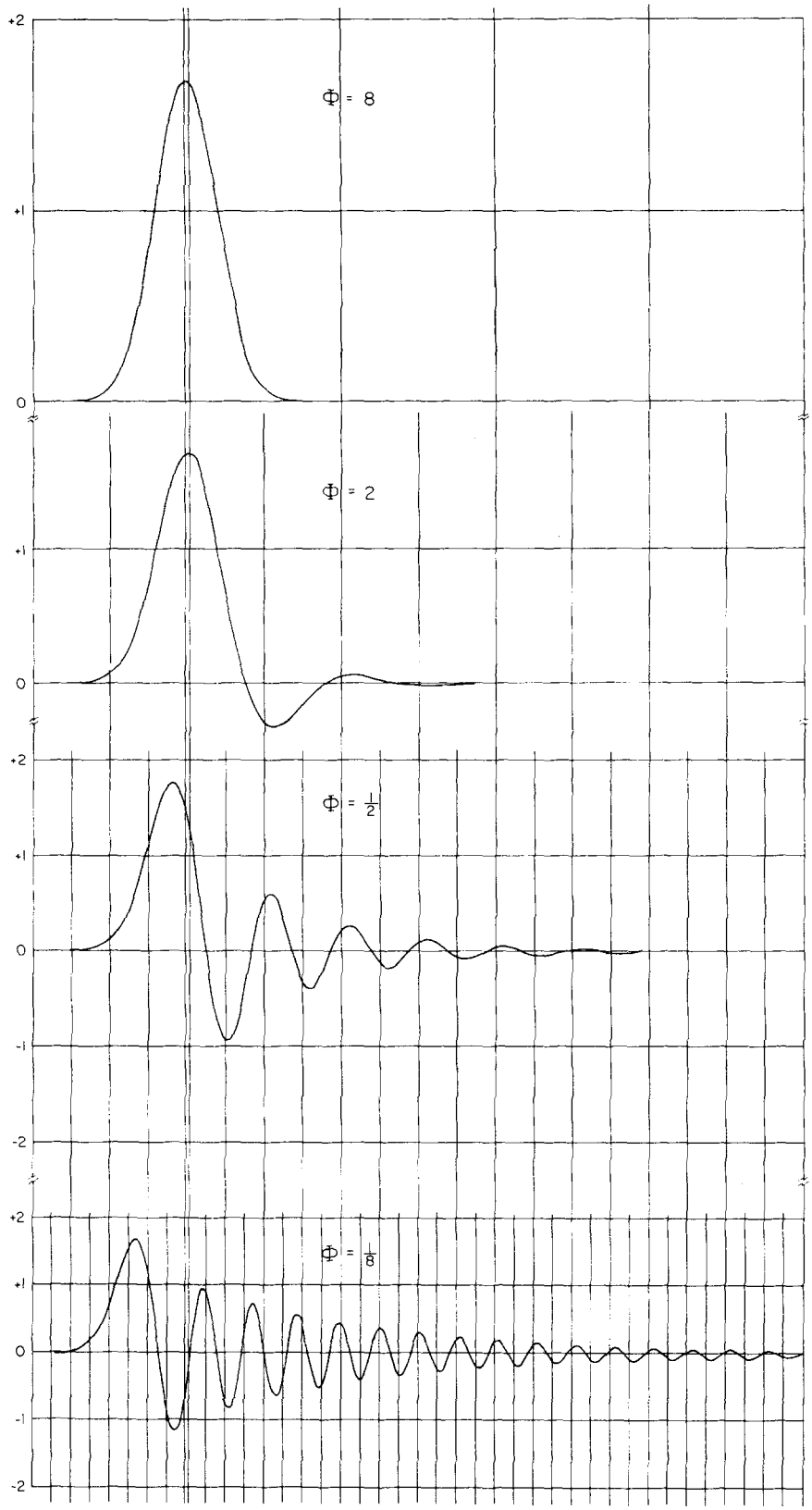

FIG. 1. The Wannier function going with Gaussians for successively smaller ϕ. In the picture the Gaussians were kept the same (essentially undistinguishable from the first graph), and the crystalline period was reduced by a factor $\frac{1}{2}$ from one graph to the next. The period is outlined by vertical bars each time, and the vertical double line indicates the origin. It is seen that the Gaussian gets gradually chewed up except at its left edge, and that an exponentially damped oscillation becomes the predominant feature. The unsymmetric structure of the function is discussed in the text.

product in (13) is $\vartheta_1'(0)$, and has, like other theta functions, a second form[7]

$$\prod_{\nu=1}^{\infty}\{1-\exp(-\tfrac{1}{2}\nu\pi\phi)\}=2\phi^{-1/2}\exp\left[\frac{\pi}{12}\left(\frac{\phi}{4}-\frac{4}{\phi}\right)\right]\times\prod_{\nu=1}^{\infty}\{1-\exp(-8\pi\nu/\phi)\} \tag{14}$$

The product is thus not only convergent, but can be made rapidly convergent under all circumstances. Calculation of the finite product can also be accelerated by (14) if ϕ is small and the number of terms in the products large.

Equation (13) also gives a good deal of qualitative

insight into the function $a(x)$. We deal with an entire function because it is a sum of Gaussians with exponentially decreasing coefficients. It is square integrable, and the distribution represented by it has moments of all orders. The appearance of the function is illustrated in Fig. 1 for $\phi = 8, 2, \frac{1}{2}, \frac{1}{8}$. For convenience of illustration the basic Gaussian is kept constant in the picture (essentially indistinguishable from the first curve), and the lattice period is reduced by a factor $\frac{1}{2}$ from one picture to the next. We see that this amounts to a gradual chopping up of the basic Gaussian, and a progressively more dominant oscillating tail; the period of the oscillation is twice the lattice period. The damping of the oscillation diminishes on either type of scale as ϕ becomes smaller and the root mean square value of x increases. By a numerical accident the mean value of x remains very close to zero throughout; only the last graph shows a small negative value. That the functions constructed are indeed Wannier functions is shown in Table I for the case $\phi = 2$. The Table shows $I(h)$ versus h where $I(h)$ is defined as

$$I(h) = \int_{-\infty}^{+\infty} a(x)\, a(x+h)\, dx. \tag{15}$$

The integral equals 1 for $h = 0$ and equals 0 for all other integer values of h as it is supposed to.

One predominant characteristic of Wannier functions which shows up here has already been exhibited in Ref. 1. Wannier functions tail away from the center in the form of damped oscillations. The oscillation is antiperiodic in the lattice period (this poses some in interesting questions for f.c.c. lattices). The damping is exponential; this important feature, which might be inferred from the graphs is proved here explicitly by the factor $[-\exp(-\frac{1}{4}\pi\phi)]^{\mu}$ in the sum (13); the finite product that acts in the opposite direction only delays this convergence because it finally approaches the value of the corresponding infinite product which is fixed.

We must now say something about the obviously surprising feature of the function, namely that it is asymmetric in x, Gaussian for negative x, and oscillating for positive x. This is not at all in agreement with the qualitative picture most of us have had regarding this function. One generally thinks of it as a function (or a set of functions) which reproduces the point symmetry of the lattice. This picture is violated in the curves of Fig. 1. We may wish to dismiss this feature as an artifact of the calculation, arising from the way the norm was calculated. This is of course true, but the functions computed converge more strongly in the direction in which they behave abnormally, as they preserve there the convergence of the original Gaussians. Now it is clear that if one had kept the norm real the function would have decayed in both directions in the form of a damped oscillation; it would therefore have been less concentrated than the one shown. It seems therefore in this case that the function with the correct symmetry property and the function with the least spread are not the same function. One should also mention that this feature is not necessarily limited to this example where it arises from a property of the theta functions. One might take the norm in other cases, find all its complex roots in k, and suppress all the roots on one side of the real axis to extract the square root. This would create a different set of Wannier functions, not yet investigated.

I may conclude this with the remark that I think it highly likely that Wannier functions can also be constructed for higher harmonic oscillator states; this is usually the case for properties discovered for the ground state.[9]

In the present context this calculation is nothing more than an illustration which should perhaps have been given forty years ago[6]; the excuse may be offered that it was not so clear then that exponential convergence would become such a crucial debating point.

The context in which the author became aware of these properties is the theory of electronic diamagnetism. For some years now, a group of us have been trying to elucidate this subject by starting with free electrons in a magnetic field, and letting these electrons be acted upon by a crystalline lattice potential. Gaussian functions are unperturbed eigenfunctions in such a problem, and the use of Wannier-like basis functions has been discussed by several authors. It seems highly likely that one can generalize the preceding calculation to construct such functions from Gaussians. What is missing in the present calculation is the inclusion of the Peierls phase.[10] It has been mentioned elsewhere[11] that the eigenstates have a nesting property which makes the same difference equation appear in the Onsager method and in a method using Landau functions. Only the exponent of certain exponentials is replaced by π^2 times its reciprocal. Equivalences of this type arise quite commonly in the theory of theta functions. There is therefore some real hope that the calculation presented here can be extended to magnetism, and that when this is done the status on the Onsager hypothesis will be clarified.

[1] F. A. Butler and R. D. Zwicker, Solid State Commun. **13**, 837 (1973).
[2] D. Pugh, Mol. Phys. **22**, 773 (1971).
[3] M. Kertesz and G. Biczo, Phys. Stat. Sol. B **60**, 249 (1973).
[4] J. R. Smith, Phys. Rev. Lett. **32**, 774 (1974).
[5] W. Andreoni, Phys. Rev. B **14**, 4247 (1976).
[6] G. H. Wannier, Phys. Rev. **52**, 153 (1937).
[7] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1935), Chapter XXI.
[8] M. Wannier, private communication.
[9] G. H. Wannier, Phys. Stat. Sol. B **70**, 727 (1975), Part 3.
[10] R. Peierls, Z. Phys. **80**, 763 (1933).
[11] D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).

A derivation of the virial expansion with application to Euclidean quantum field theory[a)]

R. Menikoff and D. H. Sharp

Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545
(Received 26 April 1977)

In this paper we give a derivation of the virial expansion and some of its generalizations. Our derivation is based on the generating functional which defines a representation of the density operator $\rho(\mathbf{x})$ in a nonrelativistic local current algebra. The virial expansion results from solving a functional differential equation for this quantity. We exploit the well-known analogy between quantum field theory and classical statistical mechanics to explore the use of the virial expansion in Euclidean quantum field theory. Specifically, we show that the virial expansion can be used to derive Feynman's rules and to provide a perturbation expansion about a static ultralocal model. The latter is worked out in detail in the case of a free neutral scalar model, and outlined in the case of a $\lambda\phi^4$ model.

I. INTRODUCTION

The virial expansion plays an important role in statistical mechanics, and it has, of course, been derived by several different methods. In this paper, we give a derivation of the virial expansion for correlation functions by a method which lends itself naturally to applications in Euclidean quantum field theory, and we discuss, in a very preliminary way, some of those applications. In this derivation, a key role is played by the generating functional $L(f)$ for the density correlations $R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, defined by

$$L(f) = \sum_{n=0}^{\infty} \frac{1}{n!} \int d\mathbf{x}_1 \cdots \int d\mathbf{x}_n \times \prod_{j=1}^{n} \{\exp[if(\mathbf{x}_j)] - 1\} R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n). \tag{1.1}$$

Exactly the same quantity is employed when one uses local currents to describe nonrelativistic quantum mechanics in the N/V limit,[1,2] and $L(f)$ is also closely related to the generating functional introduced by Bogoliubov[3] and to the familiar partition function of statistical mechanics.

The particular importance of the generating functional (1.1) follows from the fact that it can be used to define a representation of the density operator $\rho(\mathbf{x})$ on a Hilbert space[1,4] (via the GNS construction) and also to determine a measure which defines a state on the C^*-algebra consisting, essentially, of the bounded functions of the density operator.[5] The main facts about generating functionals which we use in this paper are summarized briefly in Sec. II.

The density correlation functions for a system of particles interacting via a two-body potential $U(\mathbf{x})$ satisfy the BBGKY hierarchy of coupled equations

$$\nabla_{\mathbf{x}_1} R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = -\beta \sum_{j=2}^{n} \nabla_{\mathbf{x}_1} U(\mathbf{x}_1 - \mathbf{x}_j) R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) - \beta \int d^3x_{n+1} \nabla_{\mathbf{x}_1} U(\mathbf{x}_1 - \mathbf{x}_{n+1}) R_{n+1}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}), \tag{1.2}$$

where $\beta = (1/kT)$. These are equivalent to a single functional equation[2] for $L(f)$:

$$[\nabla - i\nabla f(\mathbf{x})] \frac{1}{i} \frac{\delta L(f)}{\delta f(\mathbf{x})} = -\beta \int d^3y \, \nabla U(\mathbf{x} - \mathbf{y}) : \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} \frac{1}{i} \frac{\delta}{\delta f(\mathbf{y})} : L(f). \tag{1.3}$$

In Sec. III, we show that Eq. (1.3) has the formal solution

$$L(f) = Z(f)/Z(0), \tag{1.4}$$

where

$$Z(f) = \exp\left[-\frac{\beta}{2} \int\int d^3x d^3y \, U(\mathbf{x} - \mathbf{y}) : \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} \frac{1}{i} \frac{\delta}{\delta f(\mathbf{y})} :\right] L_0(f)$$

and $L_0(f)$ is the generating functional for a free gas ($U = 0$ or $T = \infty$) in the thermodynamic limit. For spinless bosons, this is given by[6,7]

$$L_0(f) = \exp(\rho_0 \int d^3x \{\exp[if(\mathbf{x})] - 1\}), \tag{1.5}$$

where ρ_0 is the average density of the system. Equation (1.4) can be thought of as expressing the effect of the interaction as a perturbation about the representation of the thermodynamic limit of the free theory. Similar formal solutions to functional equations have been used for many years in quantum field theory.[8]

In Sec. IV we show that a formal expansion for $L(f)$ in powers of β can be developed by expanding the exponential in Eq. (1.4) and using graphical techniques. The series can be resummed in powers of ρ_0 with the interaction entering only through the quantity $[\exp(-\beta U) - 1]$. This leads to the fugacity expansion for the correlation functions when ρ_0 is identified with the activity of the interacting system. A further resummation can be performed which results in an expansion for the correlation functions in powers of the average density of the interacting system, $\bar{\rho}$. This is the virial expansion. The last resummation is similar to a charge renormalization in quantum field theory, and it can be shown to result in a representation of the density operator which is inequivalent to the free representation. It should be noted, however, that for a suitable class of potentials the virial expansion has been shown to have a nonzero

[a)]Work supported by the Energy Research and Development Administration.

 0022-2488/78/1901-0135$1.00

radius of convergence,[9,10] while the Feynman series is believed to be asymptotic. At the end of Sec. IV, we compare our derivation of the virial expansion to others that have been given previously.

In Sec. V, we sketch generalizations of the virial expansion which are applicable when one works with external potentials, many-body potentials, or potentials which can usefully be decomposed into two parts (e.g., potentials with short range and long range pieces).

Numerous authors have discussed the close analogy between classical statistical mechanics and Euclidean quantum field theory.[11] In particular, the application of the expansions used in statistical mechanics to quantum field theory has been the subject of several important papers. Symanzik,[12] for example, has exploited this analogy to motivate a resummation of the Feynman series into the loop expansion. More recently Glimm, Jaffe, and Spencer[13] have studied the particle structure of the $P(\phi)_2$ model in a rigorous manner by using expansions similar to the high temperature expansions of classical statistical mechanics. In Sec. VI of this paper, we consider the example of an interacting scalar field and explore ways in which the virial expansion can be applied in Euclidean quantum field theory.

In Euclidean space, the generating functional for an interacting scalar field is formally given by[14]

$$L(f) = Z(f)/Z(0),$$

where

$$Z(f) = \int \prod_{x,y\in R^4} \exp[-\tfrac{1}{2}q_x(x,(-\nabla^2+m^2)y)q_y] \times \prod_{x\in R^4} \exp[-:P(q_x):]\exp(if_x q_x)\,dq_x \qquad (1.6)$$

and $P(q_x)$ is a polynomial in q_x whose form specifies the interaction. The formal expansion of (1.6), treating $P(q_x)$ as a perturbation on the Gaussian term, yields the familiar Feynman series for the n-point functions. In Sec. VIA, we show that the functional (1.6) can be rearranged in such a way that the Feynman series results directly when the virial expansion of Sec. IV is applied to $L(f)$.

An alternative approach to the study of Eq. (1.6) has been advocated by Klauder,[15] who suggests starting with the static ultralocal model, treating the off-diagonal terms of the Gaussian factor in Eq. (1.6) as perturbations. In the static ultralocal model all space–time points decouple, leading to a generating functional of the form[16]

$$L_{\mathrm{SUL}}(f) = \exp\left(\int d^n x \int \{\exp[ikf(x)]-1\}\,c(k)\,dk\right), \qquad (1.7)$$

where $c(k)$ is a suitable distribution. In Sec. VIB, we show that the generating functional (1.7) has the same form as one which would be associated, in classical statistical mechanics, with a five-dimensional system (four Euclidean dimensions and a fictitious fifth dimension corresponding to the variable k) in an external potential, with average density $\rho_0(x,k) = c(k)$. Moreover, the off-diagonal terms in the Gaussian factor in $L(f)$ couple space–time points together in the same manner as does a two-body potential in statistical mechanics. Thus the generating functional (1.6) has a form to which the virial expansion can be applied. However, in this case the "potential" $(x,(-\nabla^2+m^2)y)$ is quite singular and must be regularized. One way that this can be done is by working on a space–time lattice, but then it is necessary to show the convergence of the theory as the lattice spacing goes to zero.

A simple example where all this can be worked out explicitly is that of the free neutral scalar field, treated in Sec. VIC. In this section we show, starting on a lattice, that the free neutral scalar theory can be obtained by applying the virial expansion to the free static ultralocal theory. We also show that in this case a resummation of the series must be performed before letting the lattice spacing go to zero.

For canonical theories with polynomial interactions, the formal procedure of obtaining a static ultralocal generating functional by neglecting the off-diagonal terms in Eq. (1.6) leads to the trivial result $L_{\mathrm{SUL}}(f) = 1$. In Sec. VID we indicate how this result might be circumvented by working on a lattice and using the generalization of the virial expansion given in Sec. VB. Finally, we note that this complication does not arise in the noncanonical theories of the kind recently considered by Klauder,[17] in which the static ultralocal generating functional has the form:

$$L_{\mathrm{SUL}}(f) = \exp\left(\int d^n x \int \frac{dk}{k}\{\exp[ikf(x)]-1\} \times \exp(-\tfrac{1}{2}m^2k^2)\exp[-P(k)]\right). \qquad (1.8)$$

The virial expansion can be applied directly to (1.8) once the kernel has been regularized.

In closing, we stress that our investigation of the virial expansion in field theory is at a very preliminary stage. We have as yet no results on the possible convergence of the formal virial expansion in the case of interacting theories, nor have we investigated the resummations that may have to be carried out in order to effect such convergence.

II. PRELIMINARIES

In later sections we will utilize the facts about generating functionals which are summarized below. We will be considering systems of identical spinless particles interacting via a two-body potential U.

A. Correlation functions

The nth correlation function $R_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ is the probability density for finding n particles at the points $\mathbf{x}_1,\ldots,\mathbf{x}_n$. For a canonical ensemble of N particles in a volume V they are defined as follows:

$$R_n^{(N)}(\mathbf{x}_1,\ldots,\mathbf{x}_n) = [N!/(N-n)!]Z_N^{-1} \times \int_V d^3x_{n+1}\cdots\int_V d^3x_N \exp(-\beta U_N), \qquad (2.1)$$

where U_N is the potential for N particles,

$$U_N = \tfrac{1}{2}\sum_{i\neq j=1}^{N} U(\mathbf{x}_i - \mathbf{x}_j), \qquad (2.2)$$

Z_N is the partition function,

$$Z_N = \int_V d^3x_1 \cdots \int_V d^3x_N \exp(-\beta \mathcal{U}_N), \tag{2.3}$$

and $\beta = (kT)^{-1}$.

In the grand canonical ensemble the correlation functions are given by

$$R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \Xi^{-1} \sum_{N \geq n} (z^N/N!) Z_N R_n^{(N)}(\mathbf{x}_1, \ldots, \mathbf{x}_n), \tag{2.4}$$

where Ξ is the grand canonical partition function,

$$\Xi = \sum_{n=0}^{\infty} (z^n/n!) Z_n, \tag{2.5}$$

z is the activity,

$$z = \lambda^{-3} \exp[\beta \mu], \tag{2.6}$$

μ is the chemical potential [$\exp(\beta\mu)$ is often referred to as the fugacity], and λ is the thermal wavelength

$$\lambda = (h^2/2\pi mkT)^{1/2}. \tag{2.7}$$

B. The density operator $\rho(\mathbf{x})$

Correlation functions are also used in nonrelativistic quantum mechanics. For N particles in a volume V, the ground state correlation functions are defined as in Eq. (2.1), but with $\exp(-\beta \mathcal{U}_N)$ replaced by $|\Omega|^2$, where Ω is the ground state wavefunction. The correlation functions can also be interpreted as the n-point functions of the number density operator $\rho(\mathbf{x}) = \psi^\dagger(\mathbf{x})\psi(\mathbf{x})$,

$$R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = (\Omega, : \rho(\mathbf{x}_1) \cdots \rho(\mathbf{x}_n) : \Omega), \tag{2.8}$$

where : : stands for normal ordering.[2] Under equal-time commutation, $\rho(\mathbf{x})$ and the particle flux density operator $\mathbf{J}(\mathbf{x})$ form a local current algebra which can be used to formulate nonrelativistic quantum mechanics in terms of local observables.[18,2,19] For a system with N particles, the local currents can be represented on the usual quantum mechanical Hilbert space of square integrable functions.[20,21]

C. The generating functional $L(f)$

In the thermodynamic limit, one can no longer represent the local currents using wavefunctions, just as in statistical mechanics one can no longer write the Boltzmann distribution function as $\exp[-\beta \mathcal{U}_N(\mathbf{x}_1, \ldots, \mathbf{x}_N)]$. Instead, a representation of the density operator can be defined by means of the generating functional

$$L(f) = (\Omega, \exp[i\rho(f)]\Omega), \tag{2.9}$$

where f is an element of a space of test functions, often taken to be Schwartz' space $\mathcal{S}$.

The generating functional (2.9) is a continuous functional having the following properties:

(i) $L(f) = L(-f)^*$,

(ii) $L(0) = 1$,

(iii) $|L(f)| \leq 1$,

(iv) $L(f)$ is positive functional, i.e., $\sum_{i,j=1}^n c_i c_j^* L(f_k - f_j) \geq 0, \forall\ c_i \in \mathbb{C},\ f_i \in \mathcal{S}$ and $n = 1, 2, \cdots$.

It can be shown[22] that $L(f)$ is the Fourier transform of a positive measure on $\mathcal{S}'$, the real continuous dual of $\mathcal{S}$,

$$L(f) = \int_{\mathcal{S}'} d\mu(F) \exp[i(F, f)]. \tag{2.10}$$

Finally, we note that the generating functional can be expressed in terms of the correlation functions as follows:

$$L(f) = \sum_{n=0}^{\infty} \frac{1}{n!} \int d^3x_1 \cdots \int d^3x_n \times \prod_{j=1}^{n} \{\exp[if(\mathbf{x}_j)] - 1\} R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n). \tag{2.11}$$

One can think of Eq. (2.11) as combining the correlation functions together to give a representation of the density operator[4] on the Hilbert space $\mathcal{H} = L^2_\mu(\mathcal{S}')$. Alternatively, one can think of the generating functional as determining a measure, $d\mu(F)$, which defines a state on the C^*-algebra which describes the statistical mechanical system.[5]

A different generating functional, defined as

$$L_B(f) = \sum_{n=0}^{\infty} \frac{1}{n!} \int d^3x_1 \cdots \int d^3x_n \times \prod_{i=1}^{n} f(\mathbf{x}_i)(\bar{\rho})^{-n} R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n), \tag{2.12}$$

where $\bar{\rho} = R_1 =$ the average density, was introduced by Bogoliubov in his important study of statistical mechanics.[3] In terms of the density operator, Eq. (2.12) is equivalent to

$$L_B(f) = (\Omega, : \exp[\rho(f/\bar{\rho})] : \Omega).$$

This functional does not have the properties listed above and, consequently, it neither defines a representation of the density operator nor a state on a C^*-algebra.

A functional of physical interest which is closely related to $L(f)$ is the partition function, considered as a functional of an external potential $\phi(\mathbf{x})$. This is defined as in Eqs. (2.3) and (2.5), but with

$$\mathcal{U}_N = \tfrac{1}{2} \sum_{i \neq j=1}^{N} U(\mathbf{x}_i - \mathbf{x}_j) + \sum_{i=1}^{N} \phi(\mathbf{x}_i).$$

The generating functional $L(f)$ is then related to the partition functional $\Xi(\phi)$ by

$$L(f) = \Xi(\phi = -i\beta^{-1}f)/\Xi(0).$$

D. Cluster functions

We will find it convenient later to use the Ursell cluster functions $T_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. These are given in terms of the correlation functions by

$$T_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_G (-1)^{m-1}(m-1)! \prod_{j=1}^{m} R_{G_j}(\mathbf{x}_k \in G_j), \tag{2.13}$$

where G is a partition of $(1, 2, \ldots, n)$ into distinct subsets $(G_1, G_2, \ldots, G_m)$. The inverse relation is

$$R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_G \prod_{j=1}^{m} T_{G_j}(\mathbf{x}_k \in G_j). \tag{2.14}$$

The generating functional can be expressed in terms of the cluster functions. One finds

$$L(f)=\exp\Big(\sum_{n=1}^{\infty}\frac{1}{n!}\int d^3x_1\cdots\int d^3x_n \times\prod_{j=1}^{n}\{\exp[if(\mathbf{x}_j)]-1\}T_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)\Big). \tag{2.15}$$

Thus, the cluster functions are particularly useful when dealing with $\ln L(f)$. They correspond to the truncated n-point functions used in field theory.

E. Functional equation for $L(f)$

A coupled set of equations for the correlation functions can be obtained by taking the gradient of R_n in Eq. (2.1) or Eq. (2.4). This gives the BBGKY hierarchy of equations:

$$\begin{aligned}&\nabla_{\mathbf{x}_1}R_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)\\&=-\beta\sum_{j=2}^{n}\nabla_{\mathbf{x}_1}U(\mathbf{x}_1-\mathbf{x}_j)R_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)\\&\quad-\beta\int d^3x_{n+1}\nabla_{\mathbf{x}_1}U(\mathbf{x}_1-\mathbf{x}_{n+1})R_{n+1}(\mathbf{x}_1,\ldots,\mathbf{x}_{n+1}).\end{aligned} \tag{2.16}$$

Since this set of equations holds for all N and V, we assume they are true in the thermodynamic limit as well. The BBGKY hierarchy is equivalent to the following functional equation[2,23] for $L(f)$:

$$\begin{aligned}&[\nabla_{\mathbf{x}}-i\nabla f(\mathbf{x})]\frac{1}{i}\frac{\delta L(f)}{\delta f(\mathbf{x})}\\&=-\beta\int d^3y\nabla_{\mathbf{x}}U(\mathbf{x}-\mathbf{y}):\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:L(f),\end{aligned} \tag{2.17}$$

where by normal ordering of the functional derivatives we mean

$$:\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:=\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\Big(\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}-\delta(\mathbf{x}-\mathbf{y})\Big).$$

The prescription for normal ordering the functional derivatives follows from that used to normal order products of the density operator. For a more detailed discussion, see Ref. 2.

Equation (2.17) has many solutions. For example, it is easy to check that if the test functions are restricted to have support in a volume V, then the generating functional for the canonical ensemble

$$L^{(N)}(f)=\int_V d^3x_1\cdots\int_V d^3x_N \times\prod_{j=1}^{N}\exp[if(\mathbf{x}_j)]\exp(-\beta U_N)/Z^{(N)}), \tag{2.18}$$

and that for the grand canonical ensemble

$$L_Z(f)=\sum_{n=0}^{\infty}(z^n/n!)Z^{(N)}L^{(N)}(f)/\Xi \tag{2.19}$$

satisfy Eq. (2.17).

Here we are interested in solutions of Eq. (2.17) which are valid in the thermodynamic limit. To study these, we must specify a suitable space of test functions and appropriate boundary conditions. The space of test functions can be taken to be either Schwartz' space of infinitely differentiable functions of fast decrease or the space of infinitely differentiable functions of compact support. In Ref. 7, boundary conditions were identified which uniquely determined the solution to Eq. (2.17) in the free case, $U=0$. We do not have a precise analog of this result for the interacting case, but we note that, on physical grounds, we would expect the boundary conditions to include the following:

(i) translation invariance

$$L(f)=L(f_{\mathbf{a}}),$$

where $f_{\mathbf{a}}(\mathbf{x})=f(\mathbf{x}-\mathbf{a})$;

(ii) cluster decomposition property,

$$\lim_{\lambda\to\infty}[L(f+g_{\lambda\mathbf{a}})-L(f)L(g_{\lambda a})]=0;$$

(iii) $\frac{1}{i}\frac{\delta L(f)}{\delta f(\mathbf{x})}\Big|_{f=0}=\bar{\rho}$, the average density.

In Sec. III, we shall develope a formal solution to Eq. (2.17) which is valid in the thermodynamic limit and which satisfies the three conditions listed above.

III. FORMAL SOLUTION OF THE FUNCTIONAL EQUATION FOR $L(f)$

We can use the analogy with quantum mechanics to motivate the form of the solution to Eq. (2.17). Let Ω_0 be the free ground state and suppose the interacting ground state Ω has the form

$$\Omega=\exp[-\tfrac{1}{4}\beta\int\int d^3x\,d^3y U(\mathbf{x}-\mathbf{y}):\rho(\mathbf{x})\rho(\mathbf{y}):]\Omega_0. \tag{3.1}$$

Then, from Eq. (2.9), one finds

$$\begin{aligned}L(f)&=(\Omega_0,\exp[-\tfrac{1}{2}\beta\int\int d^3x d^3y U(\mathbf{x}-\mathbf{y}):\rho(\mathbf{x})\rho(\mathbf{y}):]\\&\quad\times\exp[i\rho(f)]\Omega_0)/(\Omega,\Omega)\\&=Z(f)/Z(0)\end{aligned} \tag{3.2}$$

where

$$Z(f)=\exp\Big[-\tfrac{1}{2}\beta\int\int d^3x d^3y U(\mathbf{x}-\mathbf{y})\times:\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:\Big]L_0(f)$$

and $L_0(f)=(\Omega_0,\exp[i\rho(f)]\Omega_0)$ is the generating functional for the free theory. In the thermodynamic limit, this is given by[6,7]

$$L_0(f)=\exp(\rho_0\int d^3x\{\exp[if(\mathbf{x})]-1\}). \tag{3.3}$$

Next, we show formally that the expression for $L(f)$ given in Eq. (3.2) satisfies the functional equation (2.17). We begin by computing the left-hand side of Eq. (2.17)

$$\begin{aligned}&[\nabla_{\mathbf{x}}-i\nabla f(\mathbf{x})]\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\Big\{\exp\Big[-\tfrac{1}{2}\beta\iint d^3x d^3y U(\mathbf{x}-\mathbf{y})\\&\times:\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:\Big]L_0(f)\Big\}\\&=\exp\Big[-\tfrac{1}{2}\beta\iint d^3x d^3y U(\mathbf{x}-\mathbf{y}):\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:\Big]\end{aligned}$$

$$\times[\nabla_{\mathbf{x}} - i\nabla f(\mathbf{x})]\frac{1}{i}\frac{\delta L_0(f)}{\delta f(\mathbf{x})} + \left[[(\nabla_{\mathbf{x}} - i\nabla f(\mathbf{x})]\right.$$

$$\times\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})},\ \exp\left(-\tfrac{1}{2}\beta\int\int d^3x d^3y U(\mathbf{x}-\mathbf{y})\right.$$

$$\left.\left.\times :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:\right)\right] L_0(f). \tag{3.4}$$

The first term on the right-hand side of Eq. (3.4) vanishes since the free generating functional satisfies

$$(\nabla_{\mathbf{x}} - i\nabla f(\mathbf{x}))\frac{1}{i}\frac{\delta L_0(f)}{\delta f(x)} = 0. \tag{3.5}$$

The second term can be simplified by using the identity

$$[A, \exp B] = [A, B]\exp B,$$

valid if $[[A, B], B] = 0$. This gives

$$\left[[\nabla_{\mathbf{x}} - i\nabla f(\mathbf{x})]\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}, \exp\left(-\tfrac{1}{2}\beta\int\int d^3x d^3y U(\mathbf{x}-\mathbf{y})\right.\right.$$

$$\left.\left.\times :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:\right)\right]$$

$$= -\beta\int d^3y \nabla_{\mathbf{x}} U(\mathbf{x}-\mathbf{y}) :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:$$

$$\times\exp\left(-\tfrac{1}{2}\beta\int\int d^3x d^3y U(\mathbf{x}-\mathbf{y})\right.$$

$$\left.\times :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:\right).$$

Thus, Eq. (3.4) reduces to Eq. (2.17) when $L(f)$ is given by Eq. (3.2).

Similar techniques have been used for many years in quantum field theory.[8] For example, in the case of a neutral scalar field with a $\lambda\phi^4$ interaction one can use the equations of motion to show that the generating functional

$$L(j) = (\Omega, T\exp[i\phi(j)]\Omega) \tag{3.6}$$

satisfies the functional equation

$$(\Box + m^2)\frac{1}{i}\frac{\delta L(j)}{\delta j(x)} = j(x)L(j) - \lambda\left(\frac{1}{i}\frac{\delta}{\delta j(x)}\right)^3 L(j), \tag{3.7}$$

which has the formal solution

$$L(j) = \frac{Z(j)}{Z(0)} = \exp\left[-\frac{i}{4}\lambda\int d^4x\left(\frac{1}{i}\frac{\delta}{\delta j(x)}\right)^4\right]\frac{Z_0(j)}{Z(0)}. \tag{3.8}$$

Here $Z_0(j)$ is the generating functional for the free theory, given by:

$$Z_0(j) = \exp\left(\frac{i}{2}\int d^4k\tilde{j}(k)[k^2 + m^2 - i\epsilon]^{-1}\tilde{j}(k)\right). \tag{3.9}$$

Expanding $Z(j)$ in powers of λ leads to the Feynman perturbation series. We show in the next section that expanding $L(f)$ leads to the virial expansion. In Sec. VIA, we will show that in Euclidean space the Feynman perturbation series can be obtained directly from the virial expansion.

IV. THE VIRIAL EXPANSION

A. Graphical expansion of Eq. (3.2)

In this section we obtain the virial expansion for $L(f)$, Eq. (3.2), by developing a diagrammatic expansion for

$$Z(f) = \exp\left[-\tfrac{1}{2}\beta\int\int d^3x d^3y U(\mathbf{x}-\mathbf{y})\right.$$

$$\left.\times :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})}:\right] L_0(f), \tag{4.1}$$

where

$$L_0(f) = \exp[\rho_0\int d^3x(\exp[if(\mathbf{x})] - 1)]. \tag{4.2}$$

By expanding the exponential in Eq. (4.1) in powers of β and performing the functional derivatives the resulting series can be written in a concise form using graphs[24-26] as follows:

$$Z(f) = \left[\sum_{N=2}^{\infty}(1/N!)\sum_{G_N} W(G_N)\right] L_0(f), \tag{4.3}$$

where $\sum_{G_N}$ is the sum over all allowed graphs with N labeled and directed lines and $W(G_N)$ is the weight assigned to each graph. The allowed graphs may be connected or disconnected, with any number of lines between pairs of distinct vertices, but have no isolated vertices. The weight assigned to a graph G with N directed and labeled lines is given by

$$W(G_N) = 2^{-N}\int d^3x_1 \cdots \int d^3x_m \prod_{q=1}^{m}\rho_0\exp[if(\mathbf{x}_q)]$$

$$\times\prod_{l=1}^{N}[-\beta U(\mathbf{x}_{l_1} - \mathbf{x}_{l_2})], \tag{4.4}$$

where the vertices have been arbitrarily numbered $1, 2, \ldots, m$ and l stands for a directed line connecting the vertex l_1 to l_2. Thus, we associate the factor $\rho_0\exp[if(\mathbf{x})]$ and an integral with each vertex and a factor $-\beta U(\mathbf{x}_1 - \mathbf{x}_2)$ with a line. Graphs with a line connecting one vertex to itself do not occur due to the normal ordering of the functional derivatives in Eq. (4.1). For a graph with a given structure the weight is independent of the labels on the lines and the directions if one considers central potentials, $U(\mathbf{x}) = U(|\mathbf{x}|)$, as we shall do here. The labels and directions on the lines are necessary, however, in order to obtain the correct counting factors. The labels take account of the different ways in which the functional derivatives can act when the exponential in Eq. (4.1) is expanded, while the directions serve to distinguish between the two functional derivatives in the exponent.

For physical potentials that are repulsive and singular at the origin (e.g., hard core potentials) the weights associated with the graphs, as given above, are ill-defined. This difficulty can be overcome if one performs a partial summation of the series for $Z(f)$ so that the quantity $[\exp(-\beta U) - 1]$, which is not singular at the origin, enters rather than $(-\beta U)$. We define a simple graph as a connected or disconnected graph with at most one line between any pair of vertices. A partial summation can now be performed over all graphs that can be obtained from a given simple graph by replacing any single line by multiple lines between the same pair of vertices. At the same time, we switch from graphs with labeled and directed lines to those with labeled vertices. The result, derived in Appendix A, is given by Eq. (4.3)

where now G_N is a simple graph with N labeled vertices none of which are isolated and the weight assigned to a graph is

$$W(G_N) = \int d^3x_1 \cdots \int d^3x_N \prod_{q=1}^{N} \rho_0 \exp[if(\mathbf{x}_q)] \times \prod_l \{\exp[-\beta U(\mathbf{x}_{l_1} - \mathbf{x}_{l_2})] - 1\}. \quad (4.5)$$

In Eq. (4.5), the vertices are labeled $1, 2, \ldots, N$ and l stands for a line. Thus, the effect of the partial summation is to restrict the allowed graphs to simple graphs and to associate the factor $[\exp(-\beta U) - 1]$ rather than $(-\beta U)$ with a line. The factor $\rho_0 \exp[if(\mathbf{x})]$ and an integral are still associated with each vertex.

In order to properly normalize the generating functional, we will need to express $Z(f)$ as a product. To do this, we use a standard theorem in graph theory which is stated precisely in Appendix B. Applying the theorem to $Z(f)$, we obtain

$$Z(f) = \exp\left[\sum_{N=1}^{\infty} (1/N!) \sum_{G_N} W(G_N)\right], \quad (4.6)$$

where G_N is a connected simple graph with N labeled vertices and $W(G_N)$ is given by Eq. (4.5). Here we define $\exp[W(G_1)] = L_0(f)$.

To obtain the cluster functions, we need the generating functional in the form given by Eq. (2.15). Thus we need to express $Z(f)$ in terms of the quantity $\{\exp[if(\mathbf{x})] - 1\}$. The weight assigned to a graph depends on f through the factor $\exp[if(\mathbf{x})]$ associated with each vertex. We expand these factors as follows:

$$\prod_{q=1}^{N} \exp[if(\mathbf{x}_q)] = \prod_{q=1}^{N} (\{\exp[if(\mathbf{x}_q)] - 1\} + 1) = \sum_G \prod_{j \in G} \{\exp[if(\mathbf{x}_j)] - 1\}, \quad (4.7)$$

where G is a subset of $\{1, 2, \ldots, N\}$.

We can incorporate this expansion into $Z(f)$ by introducing graphs with two types of vertices: a γ_1 vertex associated with the factor $\rho_0\{\exp[if(\mathbf{x})] - 1\}$ and a γ_2 vertex associated with the factor ρ_0. For each labeled graph with N vertices in the expansion of $Z(f)$ we substitute 2^N new graphs obtained by adding to every vertex an additional label of either γ_1 or γ_2. The old graphs with the same structure get counted $(1/N!) \times$ (the number of different ways to label the vertices). The number of new graphs with the same structure and having $n\gamma_1$ vertices and $(N-n)\gamma_2$ vertices is thus $N!/n!(N-n)!$. To obtain the correct counting factor, we need to label the γ_1 vertices and the γ_2 vertices independently. Thus a new graph with $n\gamma_1$ vertices and $(N-n)\gamma_2$ vertices enters the expansion with the counting factor

$(1/n!) \times (1/(N-n)!)$

$\times$ (the number of different ways to label the γ_1 vertices)

$\times$ (the number of different ways to label the γ_2 vertices),

In terms of these new graphs we can write

$$Z(f) = \exp \sum_{n=0}^{\infty} (1/n!) \sum_{m=0}^{\infty} (1/m!) \sum_{G_{n,m}} W(G_{n,m}), \quad (4.8)$$

where $G_{n,m}$ is a simple connected graph with $n\gamma_1$ vertices labeled $1, 2, \ldots, n$ and $m\gamma_2$ vertices labeled $n+1, \ldots, n+m$ and

$$W(G_{n,m}) = \int d^3x_1 \cdots \int d^3x_{n+m} \rho_0^{n+m} \prod_{q=1}^{n} \{\exp[if(\mathbf{x}_q)] - 1\} \times \prod_l \{\exp[-\beta U(\mathbf{x}_{l_1} - \mathbf{x}_{l_2})] - 1\}. \quad (4.9)$$

The generating functional $L(f)$ is obtained by normalizing $Z(f)$:

$$L(f) = Z(f)/Z(0) = \exp\left[\sum_{n=1}^{\infty} (1/n!) \sum_{m=0}^{\infty} (1/m!) \sum_{G_{n,m}} W(G_{n,m})\right]. \quad (4.10)$$

By comparing Eqs. (4.9)—(4.10) with Eq. (2.15), the cluster functions are seen to be given by

$$T_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{m=0}^{\infty} (1/m!) \sum_{G_{n,m}} \tilde{W}(G_{n,m}) \quad (4.11)$$

where

$$\tilde{W}(G_{n,m}) = \rho_0^{n+m} \int d^3x_{n+1} \cdots \int d^3x_{n+m} \prod_l \{\exp[-\beta U(\mathbf{x}_{l_1} - \mathbf{x}_{l_2})] - 1\}. \quad (4.12)$$

Equation (4.11) is the usual fugacity expansion for the cluster functions when ρ_0 is identified with the activity z of the interacting system, Eq. (2.6). The average density $\bar{\rho} = T_1$ of the interacting system is not the same as the average density ρ_0 of the free system which we perturbed about. However, a further partial summation can be performed to obtain an expansion for the cluster functions in terms of $\bar{\rho}$. This is known as the virial expansion, and the result of the partial summation is that $L(f)$ is given by Eq. (4.10) but with the allowed graphs now further restricted to be 1-irreducible and with the weight assigned to a graph given by Eq. (4.9) with ρ_0 replaced by $\bar{\rho}$.[27] A graph is 1-irreducible if it has no articulation vertices. A vertex is an articulation vertex if upon its removal the graph splits into disconnected pieces one of which contains no γ_1 vertices. This procedure is similar to charge renormalization in quantum field theory since we now associate the factor $\bar{\rho}$ with a vertex instead of ρ_0. Furthermore, by using the representation theory of the local nonrelativistic current algebra referred to in Sec. IIB, it can be shown[1] that, as a consequence of translation invariance and the cluster decomposition property, the representation of the density operator for the interacting system determined by the generating functional $L(f)$ is unitarily inequivalent to the free representation corresponding to the generating functional $L_0(f)$. This is an example of the folk theorem in field theory that a renormalization results in an inequivalent representation of the fields. If we were dealing with a system having a finite number of degrees of freedon corresponding to N particles in a volume V, then the perturbation expansion of the generating functional resulting from Eq. (4.1) would lead to an equivalent representation of the density operator and the average density would not get renormalized. This can be shown by using the fact that for N particles the density operator can be represented on the

usual quantum mechanical Hilbert space of square integrable wave functions.

It is interesting that β enters the expression for the generating functional, Eq. (4.1), in the same manner as the coupling constant λ enters in the field theoretic generating functional, Eq. (3.8). As a result of partial summations, $L(f)$ was expressed as a power series in $\bar{\rho}$ while β appeared only in the quantity $[\exp(-\beta U) - 1]$. The replacement of β by $\bar{\rho}$ as an expansion parameter leads to a series whose convergence can be proved for certain classes of potentials. In field theory, the perturbation expansion in powers of λ is believed to lead to an asymptotic series. The analogy with statistical mechanics suggests that in field theory partial summation may lead to other useful and possibly convergent expansions. Indeed, precisely such considerations were advanced by Symanzik[42] to motivate the transformation from the perturbation expansion to the loop expansion in Euclidean quantum field theory. Somewhat different applications of these ideas will be considered in Sec. VI of this paper.

B. Comparison with other derivations of the virial expansion

In this subsection we compare and contrast our derivation of the virial expansion with previous ones. The fugacity expansion was first derived by Mayer and Montroll.[28] They started by considering a given correlation function in the canonical ensemble, Eq. (2.1). A cluster expansion for this correlation function was developed and then the thermodynamic limit was taken. In taking the thermodynamic limit, the activity is introduced as follows

$$Z(N-n, V)/Z(N, V) \to z^n \quad \text{as } N \to \infty, \ V \to \infty,$$
$$\text{and } N/V \to \bar{\rho}.$$

Thus, the same technique is used to treat each correlation function individually. In Sec. IVA all the correlation functions were combined into a single generating functional $L(f)$. The thermodynamic limit for a free system was taken first to obtain the generating functional $L_0(f)$, and then a cluster expansion was developed to express the generating functional for the interacting system, $L(f)$, about $L_0(f)$.

A second method[26] for deriving the virial expansion is to start with the grand canonical partition function, treated as a functional of an external potential as in Eq. (2.5). A cluster expansion is developed for $\ln\Xi$ and then a thermodynamic limit is taken such that $V \to \infty$ with the activity z fixed. In this method of derivation the activity enters from the start because one works with the grand canonical ensemble, and all the correlation functions are treated together by introducing an external potential into the partition function. The expression for Ξ is the same as Eq. (4.6) with $\rho_0 \exp(if)$ replaced by $z\exp(-\beta\phi)$, where ϕ is the external potential. In contrast, the generating functional we use defines a representation of the density operator, and by taking the thermodynamic limit first and performing the cluster expansion second, ρ_0 starts out having an interpretation as the average density of the free system and then, after the interaction is added, one finds that the average density needs to be renormalized and that ρ_0 plays the role of the activity.

Another way of obtaining the fugacity expansion is to go back to the functional equation for $L(f)$, Eq. (2.17). We know it has solutions corresponding to N particles in a volume V, Eq. (2.18). Linear combinations of these are also solutions since the functional equation is linear, and the idea is to try to piece these N-particle solutions together in such a way that the resulting solution satisfies the boundary conditions corresponding to the thermodynamic limit. Motivated by the cluster property, we first consider the expression

$$L(f) = Z(f)/Z(0),$$

where

$$Z(f) = \sum_{n=0}^{\infty} (1/n!) \int_V d^3x_1 \cdots \int_V d^3x_n \times \prod_{i=1}^{n} \exp[if(\mathbf{x}_i)] R_n^{(0)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \tag{4.13}$$

and

$$R_n^{(0)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \rho_0^n \prod_{j>k} \exp[-\beta U(\mathbf{x}_j - \mathbf{x}_k)].$$

This is the generating functional for the grand canonical ensemble, Eq. (2.19). It is the analog of the grand canonical partition function. If we were to extend the integrals in Eq. (4.13) to infinity to obtain translation invariance, they would diverge. Instead, we introduce the cluster expansion

$$L(f) = \exp\left[\sum_{n=1}^{\infty} (1/n!) \int d^3x_1 \cdots \int d^3x_n \left(\prod_{i=1}^{n} \exp[if(\mathbf{x}_i)] - 1\right) \times T_n^{(0)}(\mathbf{x}_1, \ldots, \mathbf{x}_n)\right], \tag{4.14}$$

where the $T_n^{(0)}$ are the cluster functions of $R_n^{(0)}$, Eq. (2.13). Since the cluster functions approach zero for large values of their arguments, the integrals can be extended to infinity. This form for the $L(f)$ is equivalent to the graphical expression given in Eq. (4.8). The cluster functions of the correlation functions are then given by

$$T_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = T_n^{(0)}(\mathbf{x}_1, \ldots, \mathbf{x}_n) + \sum_{m=1}^{\infty} (1/m!) \int d^3x_{n+1} \cdots \times \int d^3x_{n+m} T_{n+m}^{(0)}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+m}), \tag{4.15}$$

which is equivalent to the graphical expression given in Eqs. (4.11) and (4.12). The graphical expressions are easier to manipulate, thus facilitating efforts at partial summation. This argument is due, essentially, to Zumino.[29]

The last method for deriving the fugacity expansion which we shall discuss is based on the Kirkwood–Salzburg equations. These can be derived starting with the grand canonical ensemble and are given by[30,31]

$$R_1(\mathbf{x}_1) = z\left[1 + \sum_{n=1}^{\infty} (1/n!) \int d^3y_1 \cdots \int d^3y_n\right.$$

$$\times K(\mathbf{x}_1; \mathbf{y}_1, \ldots, \mathbf{y}_n) R_n(\mathbf{y}_1, \ldots, \mathbf{y}_n)],$$

$$R_m(\mathbf{x}_1, \ldots, \mathbf{x}_m) = z \exp[-\beta W'(\mathbf{x}_1, \ldots, \mathbf{x}_m)] \Big[R_{m-1}(\mathbf{x}_2, \ldots, \mathbf{x}_m) + \sum_{n=1}^{\infty} (1/n!) \int d^3y_1 \cdots \int d^3y_n K(\mathbf{x}_1; \mathbf{y}_1, \ldots, \mathbf{y}_n) \times R_{m+n-1}(\mathbf{x}_2, \ldots, \mathbf{x}_m, \mathbf{y}_1, \ldots, \mathbf{y}_n) \Big], \tag{4.16}$$

where

$$K(\mathbf{x}; \mathbf{y}_1, \ldots, \mathbf{y}_n) = \prod_{j=1}^{n} \{\exp[-\beta U(\mathbf{x} - \mathbf{y}_j)] - 1\}$$

and

$$W'(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \sum_{i=2}^{m} U(\mathbf{x}_1 - \mathbf{x}_i).$$

Setting $R_1^{(0)} = z$ and $R_n^{(0)} = 0$ $(n > 1)$ on the right-hand side of Eq. (4.16) and iterating results in the fugacity expansion, the Kirkwood–Salzburg equations can be used in place of the BBGKY equations which were employed in Sec. IVA. The main difference between the two sets of equations is that, whereas the BBGKY equations are integro–differential equations which couple ∇R_n to R_n and R_{n+1}, the Kirkwood–Salzburg equations are integral equations which couple R_n to R_{n-1} and all higher correlation functions.

The correlation functions can be thought of as elements in a Banach space having norm

$$\|\{R_n\}\|_\xi = \sup_{n \geq 1} [\xi^{-n} \text{ess sup} |R_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)|]. \tag{4.17}$$

The Kirkwood–Salzburg equations can be considered as a fixed point equation

$$\{R_n\} = \{R_n^{(0)}\} + zK\{R_n\} \tag{4.18}$$

in this Banach space. By showing that K is a contraction operator one can prove that the iteration procedure mentioned above converges. This method has been used by Ruelle[10] to obtain a nonzero lower bound for the radius of convergence of the fugacity expansion for a certain class of potentials, thus proving the convergence of the fugacity expansion for small z and obtaining a bound on the correlation functions:

$$\text{ess sup} |R_n| < (\text{const}) \xi^n. \tag{4.19}$$

This bound is sufficient to show that the series for $L(f)$ given by Eq. (4.10) is convergent.

In Sec. IVA, the correlation functions were combined to form a generating functional which in turn defined a representation of the density operator on a Hilbert space and the virial expansion was interpreted as perturbing the free representation so as to obtain a representation for the density operator on a new Hilbert space which accommodates the interaction. In the method just described, the correlation functions are considered to define an element of a Banach space, and the fugacity expansion is interpreted as describing the convergence of the iterative solution of the Kirkwood–Salzburg equations to an element of the same Banach space. Finally note that the correlation functions, besides defining an element of a Banach space, must satisfy other physical conditions. For example, in order to obtain the correct probability interpretation for the correlation functions one needs $R_n > 0$, and to obtain a representation of the density operator we need the generating functional to be positive in the Bochner sense.

V. GENERALIZATIONS OF THE VIRIAL EXPANSION

In this section we outline several generalizations of the graphical expansion developed in Sec. IV. Some of these will be used in the next section.

A. External potentials

It is simple to extend the virial expansion so as to include the effect of an external potential $U_1(\mathbf{x})$. For N particles, the correlation functions are defined as in Eq. (2.1) with U_N replaced by

$$U_N = \tfrac{1}{2} \sum_{i \neq j=1}^{N} U_2(\mathbf{x}_i - \mathbf{x}_j) + \sum_{i=1}^{N} U_1(\mathbf{x}_i). \tag{5.1}$$

It can be shown that the associated generating functional satisfies the functional equation

$$(\nabla_{\mathbf{x}} - i\nabla f(\mathbf{x})) \frac{1}{i} \frac{\delta L(f)}{\delta f(\mathbf{x})} = -\beta \left((\nabla U_1)(\mathbf{x}) \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} + \int d^3y (\nabla U_2)(\mathbf{x} - \mathbf{y}) : \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} \frac{1}{i} \frac{\delta}{\delta f(\mathbf{y})} : \right) L(f). \tag{5.2}$$

In the thermodynamic limit, this equation has the formal solution

$$L(f) = Z(f)/Z(0),$$

where

$$Z(f) = \exp\left[-\tfrac{1}{2}\beta \iint d^3x d^3y U_2(\mathbf{x} - \mathbf{y}) : \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} \frac{1}{i} \frac{\delta}{\delta f(\mathbf{y})} : \right] \times \exp\left[-\beta \int d^3x U_1(\mathbf{x}) \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} \right] L_0(f) \tag{5.3}$$

and, as before,

$$L_0(f) = \exp(\rho_0 \int d^3x \{\exp[if(\mathbf{x})] - 1\}).$$

Since $\exp[\delta/\delta f(x)]$ acts a translation operator on $f(\mathbf{x})$, we have

$$L_1(f) = Z_1(f)/Z_1(0) = \exp\left[-\beta \int d^3x U_1(\mathbf{x}) \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} \right] \frac{L_0(f)}{Z_1(0)} = \exp \int \left(d^3x \rho_0 \exp[-\beta U_1(\mathbf{x})] \{\exp[if(\mathbf{x})] - 1\} \right), \tag{5.4}$$

and thus Eq. (5.3) can be written as

$$L(f) = \exp\left[-\tfrac{1}{2}\beta \iint d^3x d^3y \, U_2(\mathbf{x} - \mathbf{y}) \times : \frac{1}{i} \frac{\delta}{\delta f(\mathbf{x})} \frac{1}{i} \frac{\delta}{\delta f(\mathbf{y})} : \right] \frac{L_1(f)}{Z(0)}. \tag{5.5}$$

It is easily checked that $L_1(f)$ is the N/V limit of the generating functional for a system of bosons interacting with an external potential.[32] Furthermore, the expansion for $L(f)$ developed in the previous section goes through in the same manner except that we now associate the factor $\rho_0 \exp[-\beta U_1(\mathbf{x})]$ instead of just ρ_0 with a vertex in assigning a weight to graph.

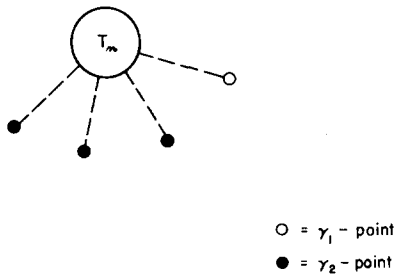

FIG. 1. A generalized vertex T_n.

B. Generalized vertices

Suppose one has an interaction potential which can be divided into two parts, U and W. Suppose also that by some means one can solve for the generating functional $L_U(f)$ corresponding to the interaction U. One can then perturb about $L_U(f)$ with the interaction W to obtain the generating functional corresponding to the total interaction $U+W$. To do this, we note that if $L_U(f)$ satisfies the functional equation:

$$(\nabla_x - i\nabla f(\mathbf{x}))\frac{1}{i}\frac{\delta L_U(f)}{\delta f(\mathbf{x})} = -\int d^3y\, \nabla_x U(\mathbf{x}-\mathbf{y}) : \frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})} : L_U(f), \tag{5.6}$$

then

$$L_{U+W}(f) = \frac{Z(f)}{Z(0)} = \exp\left[-\iint d^3x\, d^3y\, W(\mathbf{x}-\mathbf{y}) \times : \frac{1}{i}\frac{\delta}{\delta f(\mathbf{x})}\frac{1}{i}\frac{\delta}{\delta f(\mathbf{y})} : \right]\frac{L_U(f)}{Z(0)} \tag{5.7}$$

formally satisfies the functional equation

$$[\nabla_x - i\nabla f(\mathbf{x})]\frac{1}{i}\frac{\delta L_{U+W}(f)}{\delta f(\mathbf{x})} = -\int d^3y\, \nabla_x (U+W)(\mathbf{x}-\mathbf{y}) : \frac{\delta}{\delta f(\mathbf{x})}\frac{\delta}{\delta f(\mathbf{y})} : L_{U+W}(f), \tag{5.8}$$

which determines the generating functional for the interaction $U+W$.

By writing the generating functional in the form

$$L_U(f) = \exp\left(\sum_{n=1}^{\infty}(1/n!)\int d^3x_1 \cdots \int d^3x_n \times \prod_{i=1}^{n}\{\exp[if(\mathbf{x}_i)]-1\}T_n^{(U)}(\mathbf{x}_1,\ldots,\mathbf{x}_n)\right), \tag{5.9}$$

one can develope a graphical expansion for $L_{U+W}(f)$. To do this, one needs to define graphs with "generalized vertices." A generalized vertex can be pictured as a circle labeled T_n with n dotted lines extending out of it, each ending either in a γ_1 point or a γ_2 point (Fig. 1). A graph is now formed by connecting pairs of points in the generalized vertices with solid lines, with the points labeled and restricted by the following conditions (Fig. 2):

(1) At most one line joins a pair of points;

(2) a γ_2 point must be connected by at least one line;

(3) a γ_1 point may be connected by no lines.

We assign a weight to a graph by the following prescription:

(1) We associate with each labeled γ_1 point the factor $\int d^3x_i\{\exp[if(\mathbf{x}_i)]-1\}$;

(2) we associate with each labeled γ_2 point the factor $\int d^3x_i$;

(3) we associate with each generalized vertex the factor $T_n^{(U)}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$;

(4) we associate with each line the factor $\{\exp[-W(\mathbf{x}_1-\mathbf{x}_2)]-1\}$.

The expansion developed in Sec. IV can be extended to the case at hand by replacing the graphs used there by graphs having generalized vertices.

This type of expansion can be applied in several ways:

(1) In many cases of interest the potential has both a long range and a short range part. Typically, the virial expansion works reasonably well for short range potentials but not very well for long range potentials. Thus other methods may be required to calculate the effect of the long range potential, but when this has been done, the above form of the virial expansion can be used to take account of the short range part of the potential perturbatively. An example of this kind of procedure has been given by Lebowitz, Stell, Baer, and Theumann,[33] who have developed a graphical expansion which enables one to calculate the correlation functions while treating the short range and long range pieces of the potential differently.

(2) In nonrelativistic many-body quantum mechanics the Jastrow approximation, in which the ground state is approximated by a wavefunction of the form $\Omega_0 = \Pi_{i>j}\exp[\frac{1}{2}\phi(\mathbf{x}_i-\mathbf{x}_j)]$, is frequently employed for Bose systems. In this approximation, the ground state correlation functions are the same as those for a classical statistical mechanical system with a two-body interaction potential $-\beta U_2(\mathbf{x}) = \phi(\mathbf{x})$. For spinless fermions, the analog of the Jastrow approximation is to write the ground state wavefunction as $\Omega_0 = \Pi_{i>j}\exp[\frac{1}{2}\phi(\mathbf{x}_i-\mathbf{x}_j)]\Omega_F$, where Ω_F is the ground state wavefunction of the free Fermi system. The generating functional for a free Fermi gas is known.[32] Therefore, in the Jastrow approximation, the correlation functions of an interacting Fermi system can be obtained by using the virial expansion to perturb about the correlation functions for a free Fermi system.

(3) In Section VID, we show formally that the Euclidean, lattice, $\lambda\phi^4$ theory can be obtained by applying the virial expansion in the form given in this section to the free static ultralocal lattice theory.

FIG. 2. Example of a graph with generalized vertices.

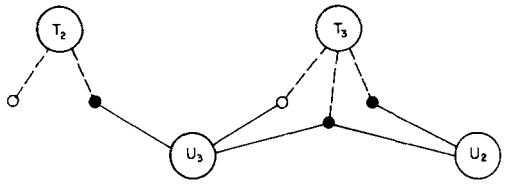

FIG. 3. Example of a graph with generalized vertices and generalized lines.

C. Many-body potentials

Finally, we consider the generalization of the virial expansion to the case of many-body potentials. These do not usually occur in statistical mechanics problems. However, in nonrelativistic many-body quantum mechanics, corrections to the Jastrow wavefunction of the form

$$\Omega_0 = \prod_{i>j} \exp[\tfrac{1}{2}\phi_2(\mathbf{x}_i - \mathbf{x}_j)] \prod_{i>j>k} \exp[\tfrac{1}{2}\phi_3(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)] \quad (5.10)$$

have been considered,[34–37] whose effect on the correlation functions is the same as including a three-body potential in a classical statistical mechanics problem. If the original Jastrow wavefunction is a good first approximation in a given problem, then contributions from the ϕ_3 terms can be included perturbatively using the virial expansion.

The generating functional for a system of particles interacting via a many-body potential U_n, $n = 1, 2, 3, \cdots$, satisfies the functional equation:

$$[\nabla_{\mathbf{x}_1} - i\nabla f(\mathbf{x}_1)]\frac{1}{i}\frac{\delta L(f)}{\delta f(\mathbf{x}_1)} = -\sum_{n=1}^{\infty}\frac{1}{(n-1)!}\int d^3x_2 \cdots \int d^3x_n \nabla_{\mathbf{x}_1} U_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) \times :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x}_1)} \cdots \frac{1}{i}\frac{\delta}{\delta f(\mathbf{x}_n)}: L(f). \quad (5.11)$$

In the thermodynamic limit, Eq. (5.11) has the formal solution

$$L(f) = \frac{Z(f)}{Z(0)} = \exp\left[-\sum_{n=1}^{\infty}\frac{1}{n!}\int d^3x_1 \cdots \quad d^3x_n U_n \times :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x}_1)} \cdots \frac{1}{i}\frac{\delta}{\delta f(\mathbf{x}_n)}:\right]\frac{L_0(f)}{Z(0)}. \quad (5.12)$$

Since we may want to solve first for part of the potential, treating the rest as a perturbation, we consider an expansion for

$$L(f) = \exp\left[-\sum_{n=1}^{\infty}\frac{1}{n!}\int d^3x_1 \cdots \int d^3x_n U_n \times :\frac{1}{i}\frac{\delta}{\delta f(\mathbf{x}_1)} \cdots \frac{1}{i}\frac{\delta}{\delta f(\mathbf{x}_n)}:\right] \times \exp\left[\sum_{n=1}^{\infty}(1/n!)\int d^3x_1 \cdots \int d^3x_n \times \prod_{i=1}^{n}\{\exp[if(\mathbf{x}_i)] - 1\} T_n / Z(0). \quad (5.13)$$

A graphical expansion of Eq. (5.13) can be obtained by generalizing the results of Sec. VB as follows. In the graphs used previously, the lines represented two-body interactions. We shall now need "generalized lines" to describe the many-body interactions. We represent a generalized line pictorially by a circle labeled U_n with n solid lines extending out of it, and we consider graphs formed by connecting the points of the generalized vertices with generalized lines (Fig. 3). To assign a weight to a graph we may follow the prescription given in the previous subsection, except that in place of rule (4) of that prescription we now associate the factor $\{\exp[-U_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)] - 1\}$ with each generalized line.

It is our conjecture that the expansion developed in Sec. VB can be extended to the present case simply by replacing the graphs used there with the generalized graphs defined here.

Remarks: (1) Clearly, if $T_n = 0$ for $n \geq 2$ and $U_n = 0$ for $n \neq 2$, then the expansion described here reduces to the one given in Sec. VB.

(2) Similar generalized diagrams have been introduced in connection with many-body quantum mechanics by Campbell.[38]

(3) The number of generalized graphs increases rapidly and soon becomes unmanageable. This expansion is likely to be useful only if it facilitates partial summations of the series or if T_n and U_n fall off sufficiently fast that the series converges rapidly, so that it is necessary to deal only with the first few terms.

VI. APPLICATIONS TO EUCLIDEAN QUANTUM FIELD THEORY

A. Relationship between the virial expansion and Feynman diagrams

The virial expansion can be applied to the formal expression for the generating functional in Euclidean quantum field theory to obtain the usual Feynman perturbation series. To illustrate this point, we consider the example of a neutral scalar field with a $\lambda\phi^4$ interaction. The generating functional for this model is given by[11,12]

$$L(f) = Z(f)/Z(0),$$

where

$$Z(f) = \exp\left[\frac{\lambda}{4!}\int d^4x :\left(\frac{1}{i}\frac{\delta}{\delta f(x)}\right)^4:\right] \times \exp[-\tfrac{1}{2}\iint d^4x d^4y f(x)\Delta(x-y)f(y)] \quad (6.1)$$

and

$$\Delta(x)\int \frac{d^4k}{(2\pi)^4}\frac{\exp(ik\cdot x)}{k^2 + m^2}.$$

It is well known that a formal expansion of this expression for $L(f)$ in powers of λ leads to the Feynman perturbation series for the n-point functions. We shall show that this result can also be obtained using the virial expansion. To do this, we will need to use the following:

Lemma[39] (*reciprocity relation*): Let $F(z)$ be a limit of

polynomials in z, $f(x)$ a test function and $\Delta(x)$ a kernel, such as given in Eq. (6.1). Then

$$F\left(\frac{1}{i}\frac{\delta}{\delta f(x)}\right)\exp\left[\iint d^4x\,d^4y\,f(x)\Delta(x-y)f(y)\right]$$
$$=\exp\left[-\tfrac{1}{2}\iint d^4x\,d^4y\,f(x)\Delta(x-y)f(y)\right]$$
$$\times\left\{\exp\left[-\tfrac{1}{2}\iint d^4x\,d^4y\,\Delta(x-y)\frac{1}{i}\frac{\delta}{\delta j(x)}\right]\right.$$
$$\left.\times\frac{1}{i}\frac{\delta}{\delta j(y)}\right]F(j)\Bigg\}_{j(x)=i\int d^4y\,\Delta(x-y)f(y)} \tag{6.2}$$

is a formal identity. Applying the Lemma to Eq. (6.1), we obtain

$$L(f)=Z(f)/Z(0)$$
$$=\exp\left[-\tfrac{1}{2}\iint d^4x\,d^4y\,f(x)\Delta(x-y)f(y)\right]$$
$$\times\left\{\exp\left[-\tfrac{1}{2}\iint d^4x\,d^4y\,\Delta(x-y)\frac{1}{i}\frac{\delta}{\delta j(x)}\frac{1}{i}\frac{\delta}{\delta j(y)}\right]\right.$$
$$\left.\times\exp\left[\frac{\lambda}{4!}\int d^4x:j^4(x):\right]\right\}_{j(x)=i\int d^4y\Delta(x-y)f(y)}\Bigg/Z(0). \tag{6.3}$$

We consider the expression

$$\exp\left[-\tfrac{1}{2}\iint d^4x\,d^4y\,\Delta(x-y)\frac{1}{i}\frac{\delta}{\delta j(x)}\frac{1}{i}\frac{\delta}{\delta j(y)}\right]$$
$$\times\exp\left[\frac{\lambda}{4!}\int d^4x:j(x)^4:\right]\Bigg/Z(0) \tag{6.4}$$

in more detail. This can be written as

$$\exp\left[-\tfrac{1}{2}\iint d^4x\,d^4y\,\Delta(x-y):\frac{1}{i}\frac{\delta}{\delta j(x)}\frac{1}{i}\frac{\delta}{\delta j(y)}:\right]$$
$$\times\exp\left[\frac{\lambda}{4!}\int d^4x\int dk(\exp[ikj(x)]-1)\left(i\frac{d}{dk}\right)^4\delta(k)\right]\Bigg/Z(0)$$
$$=\left\{\exp\left[-\tfrac{1}{2}\int d^4x_1\int dk_1\int d^4x_2\int dk_2k_1k_2\Delta(x_1-x_2)\right.\right.$$
$$\left.\times:\frac{1}{i}\frac{\delta}{\delta j(x_1,k_1)}\frac{1}{i}\frac{\delta}{\delta j(x_2,k_2)}:\right]$$
$$\times\exp\left[\frac{\lambda}{4!}\int d^4x\int dk(e^{ij(x,k)}-1)\right.$$
$$\left.\left.\times\left(i\frac{d}{dk}\right)^4\delta(k)\right]\Bigg/Z(0)\right\}_{j(x,k)=kj(x)}. \tag{6.5}$$

To derive Eq. (6.5), we have shifted the normal ordering in Eq. (6.4) from the $j^4(x)$ terms to the functional derivatives. This can be justified by examining the terms in the expansion for each expression. Equation (6.5) is in a form to which the virial expansion can be applied. [Compare with Eq. (4.1).] Note that to write the expression in this way we have introduced a fictitious fifth dimension corresponding to the variable k.

To apply the virial expansion, we make the following identifications:

(1) $k_1k_2\Delta(x_1-x_2)$ corresponds to the two-body potential $U(x_1,k_1;x_2,k_2)$;

(2) $(\lambda/4!)(id/dk)^4\delta(k)$ corresponds to the activity $\rho_0(x,k)$;

(3) $\beta=1$.

Then, using Eqs. (4.1), (4.9), and (4.10), the right-hand side of Eq. (6.5) can be written in the form

$$\exp\left[\sum_{n=1}^{\infty}(1/n!)\sum_{m=0}^{\infty}(1/m!)\sum_{G_{n,m}}W(G_{n,m})\right], \tag{6.6}$$

where $G_{n,m}$ is a simple connected graph with $n\gamma_1$ vertices labeled from $1,2,\ldots,n$ and $m\gamma_2$ vertices labeled from $n+1,\ldots,n+m$, and the weight assigned to a graph is

$$W(G_{n,m})=\int d^4x_1\int dk_1\cdots\int d^4x_{n+m}\int dk_{n+m}$$
$$\times\prod_{i=1}^{n}\{\exp[ik_ij(x_i)]-1\}\times\prod_{i=1}^{m+n}\left[\frac{\lambda}{4!}\left(i\frac{d}{dk_i}\right)^4\delta(k_i)\right]$$
$$\times\prod_{l}\{\exp[-k_{l_1}k_{l_2}\Delta(x_{l_1}-x_{l_2})]-1\}. \tag{6.7}$$

This expression can be simplified by doing the integrals over k_i. Because of the factor $(id/dk_i)^4\delta(k_i)$, Eq. (6.7) can be written in terms of connected graphs with four lines meeting at every γ_2 vertex and one, two, or three lines meeting at every γ_1 vertex. To compute the weight assigned to a graph, we include a factor $[\Delta(x_1-x_2)]^n/(n!)^2$ if there are n lines ($n=1,2,3$) connecting the points x_1 and x_2, a factor λ for each γ_2 vertex, and a factor $\lambda j(x)^{4-n}/(4-n)!$ for each γ_1 vertex which has $n=1,2,$ or 3 lines connected to it. The normal ordering in expressions (6.4) and (6.5) has the effect of eliminating graphs in which a line connects one vertex to itself. To enhance the resemblance to field theory diagrams, we treat all vertices as γ_2 vertices and associate an external line (a line in which only one end is connected to a vertex) with each factor of $j(x)$. Thus, we can expand the generating functional $L(f)$ as

$$L(f)=\exp\left(-\tfrac{1}{2}\int\int d^4x\,d^4y\,f(x)\Delta(x-y)f(y)\right.$$
$$\left.+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}(1/m!)\sum_{G_{n,m}}W(G_{n,m})\right), \tag{6.8}$$

where $G_{n,m}$ is a connected graph with n external lines and m labeled vertices in which four lines meet at every vertex. The weight is given by

$$W(G_{n,m})=\lambda^m\int d^4x_1\cdots\int d^4x_m\prod_{(i,j)}\frac{[\Delta(x_i-x_j)]^{n_{ij}}}{(n_{ij}!)^2}$$
$$\times\prod_{l}[i\int d^4y_l\Delta(x_l-y_l)f(y_l)]^{n_l}, \tag{6.9}$$

where n_{ij} is the number of lines between the vertices x_i and y_j and n_l is the number of external lines at the vertex x_l.

Equations (6.8) and (6.9) determine all the n-point functions. If the reduction formulas are used to calculate the scattering amplitude, then in computing the weight of a graph we would associate with an external line the wavefunction of an incoming or outgoing particle instead of the factor $j(x)=[i\int d^4y\Delta(x-y)f(y)]$, and we would have the usual Feynman rules for scattering.

Thus, the virial expansion for the correlation functions can be thought of as a statistical mechanical analog of the Feynman series for the field theory scattering amplitude. In order to get finite results, the formal perturbation expansion discussed here would have to be renormalized or resummed. This procedure is believed to result in an asymptotic series in the coupling constant.

B. Perturbing about the static ultralocal model

We consider a relativistic scalar field ϕ associated with the Lagrangian

$$L = \int d^4x[\tfrac{1}{2}\partial^\mu\phi\partial_\mu\phi - \tfrac{1}{2}m^2\phi^2 - P(\phi)], \tag{6.10}$$

where $P(\phi)$ is a polynomial in ϕ which defines the interaction and $d^4x = d\mathbf{x}\,dt$. In Euclidean space, the generating functional for this model can be written as a formal functional integral in the following manner[14]:

$$\begin{aligned} L(f) &= Z(f)/Z(0) \\ &= \int \prod_{x,y\in R^4} \exp[-\tfrac{1}{2}q_x(x,(-\nabla^2+m^2)y)q_y] \\ &\quad\times \prod_{x\in R^4} \exp[-:P(q_x):]\exp(if_xq_x)\,dq_x/Z(0). \end{aligned} \tag{6.11}$$

The expansion of Eq. (6.11) treating $P(q_x)$ as a perturbation on the Gaussian term leads to the Feynman series for the n-point functions, as discussed in Sec. VIA. In the remainder of the paper we shall be considering an alternative expansion for $L(f)$ which results from treating the off-diagonal terms in the Gaussian factor in Eq. (6.11) as perturbations about the remaining terms in the integral. Thus in this approach the idea is to perturb about the static ultralocal form of the theory, rather than the free field theory. Static ultralocal models have been discussed by several physicists, among them Klauder,[15] Caianiello and Scarpetta,[40] and Kovesi-Domokos.[41] These models have the property that, since all gradient terms are dropped from the theory, different space–time points are completely decoupled from one another. It will be seen that by using the virial expansion the off-diagonal terms in Eq. (6.11), whose effect is to couple together different space–time points, can be treated perturbatively in a manner similar to that used to include the effects of two-body potentials in classical statistical mechanics.

Equation (6.11) can be written in the form

$$\begin{aligned} L(f) &= \exp\left[-\tfrac{1}{2}\iint d^4x\,d^4y\,\omega(x-y):\frac{1}{i}\frac{\delta}{\delta f(x)}\frac{1}{i}\frac{\delta}{\delta f(y)}:\right] \\ &\quad\times L_{\mathrm{SUL}}(f)/Z(0), \end{aligned} \tag{6.12}$$

where $\omega(x-y)$ is the kernel of the Laplacian operator $-\nabla^2 = -\sum_i \partial^2_{x_i}$ and $L_{\mathrm{SUL}}(f)$ is the static ultralocal generating functional:

$$\begin{aligned} L_{\mathrm{SUL}} &= Z_{\mathrm{SUL}}(f)/Z_{\mathrm{SUL}}(0) \\ &= \int \prod_x \exp[-\tfrac{1}{2}[m^2+\omega(0)]q_x^2 - :P(q_x):] \\ &\quad\times \exp(if_xq_x)\,dq_x/Z_{\mathrm{SUL}}(0). \end{aligned} \tag{6.13}$$

Assuming that the formal expression (6.13) exists and defines an infinitely divisible process, one can use the Lévy–Khinchine formula of probability theory to show that L_{SUL} has the general form[15,16]

$$\begin{aligned} L_{\mathrm{SUL}}(f) = \exp\Big\{ &\int d^4x[-f^2(x)/m + i\alpha(x)f(x) \\ &+ \int_{-\infty}^{\infty} d\sigma(k)(\exp(ikf(x)-1)]\Big\}. \end{aligned} \tag{6.14}$$

Equation (6.14), in turn, can be written as

$$L_{\mathrm{SUL}}(f) = \exp\Big\{\int d^4x \int dk[\exp(ikf(x)) - 1]c(k)\Big\}, \tag{6.15}$$

where

$$c(k) = m^{-1}d^2\delta(k)/dk^2 - \alpha d\delta(k)/dk + d\sigma(k)/dk.$$

Thus, combining Eqs. (6.12)–(6.15), we obtain

$$\begin{aligned} L(f) = \Bigg\{ &\exp\left[-\tfrac{1}{2}\int d^4x_1\int dk_1\int d^4x_2\int dk_2 k_1k_2\omega(x_1-x_2)\right. \\ &\left.\times:\frac{1}{i}\frac{\delta}{\delta f(x_1,k_1)}\frac{1}{i}\frac{\delta}{\delta f(x_2,k_2)}:\right] \\ &\times\exp\left[\int d^4x\int dk[\exp(if(x,k))-1]c(k)\right]\Big|_{f(x,k)=kf(x)}\Bigg\}\Big/ Z(0). \end{aligned} \tag{6.16}$$

The generating functional is now in a form to which the virial expansion can be applied. Note that, as in Sec. VIA, to achieve this, it is necessary to integrate over a fifth dimension corresponding to the variable k. Note also that the kernel $k_1k_2w(x_1-x_2)$ corresponds to a two-body potential and $c(k)$ to the activity.

In order for the virial expansion to be applicable to Eq. (6.16) in any but the most formal sense, however, two problems must be overcome. The first is that the kernel $\omega(x-y)$ is quite singular and needs to be regularized. This can be done by working on a lattice, as illustrated by the examples discussed in the following two subsections. The second problem is that for many types of interactions [such as $P(\phi) = \lambda\phi^4$] the formal expression (6.13) leads to the trivial result $L_{\mathrm{SUL}}(f) = 1$. In Sec. VIID, we indicate how this difficulty might be overcome by working on a lattice and using the generalization of the virial expansion given Sec. VB. Finally, we note that Klauder[17] has suggested a modification of the measure dq_x in Eq. (6.11) which leads to a noncanonical theory having a static ultralocal generating functional of the simple form:

$$\begin{aligned} L_{\mathrm{SUL}}(f) = \exp\Bigg(&\int d^4x\int\frac{dk}{k}\{\exp[ikf(x)]-1\} \\ &\times\exp[-\tfrac{1}{2}[m^2+\omega(0)]k^2 - P(k)\}\Bigg). \end{aligned} \tag{6.17}$$

The above-mentioned difficulty does not apply to this type of noncanonical theory.

C. Example: The free neutral scalar field theory

Here we illustrate the ideas of the previous subsection by showing that the free neutral scalar theory can be obtained by applying the virial expansion to the free static ultralocal theory.

To regularize the kernel $\omega(x-y)$ of the operator $-\nabla^2$, we introduce a lattice as follows. Let n = the number of space–time dimensions, δ = the lattice spacing, and let $i=(i_1,\ldots,i_n)$ be an ordered n-tuple of integers. The points $x\in R^n$ are then replaced by the lattice of points labeled by $x_i=(\delta i_1,\ldots,\delta i_n)$ and the test functions are replaced by lattice fields $f_i=f(x_i)$.

The static ultralocal generating functional for a free scalar field can be written, on the lattice, as

$$L^\delta_{\mathrm{SUL}}(f)=Z^\delta_{\mathrm{SUL}}(f)/Z^\delta_{\mathrm{SUL}}(0)$$
$$=\prod_i\int_{-\infty}^{\infty}dq_i\exp(iq_if_i\delta^n)\exp(-\delta^n m^2q_i^2/2)/Z^\delta_{\mathrm{SUL}}(0)$$
$$=\exp\left(-\delta^n\sum_i f_i^2/2m^2\right). \tag{6.18}$$

The kernel of the operator $-\nabla^2$ can be defined on the lattice either by using a finite difference approximation[14]

$$\omega^\delta_{ij}=\left[2^n\delta_{ij}-\sum_{|j-j'|=1}\delta_{ij'}\right]/\delta^{2+n} \tag{6.19a}$$

or by using Fourier transforms[42]

$$\omega^\delta_{ij}=\int_{-\pi/\delta}^{\pi/\delta}\frac{dp_1}{2\pi}\cdots\int_{-\pi/\delta}^{\pi/\delta}\frac{dp_n}{2\pi}$$
$$\times(p_1^2+\cdots+p_n^2)\exp[ip\cdot(x_i-x_j)]. \tag{6.19b}$$

Thus, on a lattice, the free field generating functional [see Eqs. (6.12) and (6.18) with $P(q_x)=0$] can be written as

$$L^\delta(f)=Z^\delta(f)/Z^\delta(0)$$
$$=\exp\left[-\tfrac{1}{2}\int dk_1\int dk_2\,\delta^{2n}\sum_{i\neq j}k_1k_2\omega^\delta_{ij}\right.$$
$$\left.\times:\frac{1}{i\delta^n}\frac{\delta}{\delta f_i(k_1)}\frac{1}{i\delta^n}\frac{\delta}{\delta f_j(k_2)}:\right]$$
$$\times\exp\left[\delta^n\sum_l\int dk\{\exp[if_l(k)]-1\}\right.$$
$$\left.\times\left(\frac{d^2\delta(k)}{dk^2}\Big/2(m^2+\delta^n\omega_{ll})\right)\right]\Big|_{f_1(k)=kf_i}\Big/Z^\delta(0). \tag{6.20}$$

Applying the virial expansion to Eq. (6.20), we obtain

$$L^\delta(f)=\exp\left[\sum_{p=1}^{\infty}(1/p!)\sum_{q=0}^{\infty}(1/q!)\sum_{G_{p,q}}W(G_{p,q})\right], \tag{6.21}$$

where $G_{p,q}$ is a simple connected graph with $p\,\gamma_1$ vertices labeled from $1,2,\ldots,p$ and $q\,\gamma_2$ vertices labeled from $p+1,\ldots,p+q$ and the weight assigned to a graph is

$$W(G_{p,q})=\delta^n\sum_{i_1}\int dk_1\cdots\delta^n\sum_{i_{p+q}}\int dk_{p+q}$$
$$\times\prod_{j=1}^{p}[\exp(ik_jf_{i_j})-1]\times\prod_{r=1}^{p+q}\left(\frac{d^2\delta(k_r)}{dk_r^2}\Big/2(m^2+\delta^n\omega_{rr})\right)$$
$$\times\prod_l(\exp[-k_{l_1}k_{l_2}\omega_{i_{l_1}i_{l_2}}]-1). \tag{6.22}$$

Upon carrying out the integrals over k_j one finds that, owing to the occurance of the factor $d^2\delta(k)/dk^2$, the only nonvanishing graphs are $G_{1,0}$ and $G_{2,q}$, $q=0,1,\cdots$. As a result, $L^\delta(f)$ simplifies and takes the form

$$L^\delta(f)=\exp\left\{-\delta^{2n}\sum_{i,j}\frac{f_if_j}{2(m^2+\delta^n\omega_{ii})}\right.$$
$$\left.\times\left[I+\sum_{p=1}^{\infty}\left(-\frac{\tilde\omega}{m^2+\delta^n\omega_{ii}}\right)^p_{i,j}\right]\right\}, \tag{6.23}$$

where $I=\delta^{-n}\delta_{ij}$, $\tilde\omega_{ij}=\omega_{ij}-\omega_{ii}\delta_{ij}$, $(\tilde\omega)^2_{ij}=\delta^n\sum_k\tilde\omega_{ik}\tilde\omega_{kj}$, etc. The series $\sum_p(\cdot)^p_{ij}$ in Eq. (6.23) is the expansion of

$$\left(I+\frac{\tilde\omega}{m^2+\delta^n\omega_{ii}}\right)^{-1}_{i,j}.$$

Therefore, Eq. (6.23) gives

$$L^\delta(f)=\exp\left[-\tfrac{1}{2}\delta^{2n}\sum_{i,j}f_i[m^2I+\omega]^{-1}_{ij}f_j\right], \tag{6.24}$$

and, taking the limit $\delta\to 0$, we obtain the generating functional for the free neutral scalar theory,

$$L(f)=\exp[-\tfrac{1}{2}\int d^nx\int d^ny\,f(x)(m^2-\nabla^2)^{-1}f(y)]. \tag{6.25}$$

Remarks: (1) Note that since the kernel ω^δ is singular in the limit $\delta\to 0$, each individual term in the series $\sum_p(\cdot)^p_{ij}$ in Eq. (6.23) is ill-defined as $\delta\to 0$. However, the limit $\delta\to 0$ of the sum of the series is well defined. Thus the sum must be performed before the limit is taken.

(2) In this example, we have essentially used the virial expansion with lattice cutoff to derive a special case of the formal Gaussian combinatoric identity[43]:

$$\exp\left[-\tfrac{1}{2}i\int d^4x\int d^4y\frac{\delta}{\delta j(x)}A(x-y)\frac{\delta}{\delta j(y)}\right]$$
$$\times\exp\left[\tfrac{1}{2}i\int d^4x\int d^4y\,j(x)B(x-y)j(y)\right]$$
$$=\exp\left\{\tfrac{1}{2}\int d^4x\int d^4y\,j(x)[B(1-AB)^{-1}](x-y)j(y)\right\}$$
$$\times\exp[\tfrac{1}{2}\mathrm{Tr}\ln(1-AB)^{-1}].$$

(3) Scarpetta[44] has used a different method to derive the free neutral scalar theory by perturbing about the static ultralocal theory.

D. Example: Neutral scalar field with $\lambda\phi^4$ interaction

To get a better idea of what is involved in this approach, we briefly examine a neutral scalar field with a $\lambda\phi^4$ interaction in this section. One might expect from looking at the formal expression (6.13) that the static ultralocal generating functional for this model would be given by Eq. (6.15), with $c(k)$ determined by the equation

$$\int dq\exp[iqf(x)]\exp\{-\tfrac{1}{2}(m^2+\omega(0))q^2-\lambda q^4\}$$
$$=\exp(\int dk\{\exp[ikf(x)]-1\}c(k)). \tag{6.26}$$

However, a theorem of Newman[45] can be adapted to show that a function $c(k)$ satisfying Eq. (6.26) does not exist. Indeed, a direct calculation of the functional integral in Eq. (6.13) leads to the trivial result $L_{\mathrm{SUL}}(f)=1$.

This difficulty might be overcome if one starts with the static ultralocal generating functional defined on a lattice. For the case of a $\lambda\phi^4$ interaction, Eq. (6.13) gives

$$L^{\delta}_{\mathrm{SUL}}(f) = Z^{\delta}_{\mathrm{SUL}}(f)/Z^{\delta}_{\mathrm{SUL}}(0) = \prod_i \int_{-\infty}^{\infty} dq \exp(iqf_i\delta^n) \exp\{-\tfrac{1}{2}\delta^n[m^2+\omega(0)]q^2\} \exp(-\lambda\delta^n q^4)/Z^{\delta}_{\mathrm{SUL}}(0)$$

$$= \exp\left\{\sum_i \ln\left[1 + \frac{\int dq[\exp(iqf_i\delta^n) - 1](\exp\{-\frac{1}{2}\delta^n[m^2+\omega(0)]q^2\}) \exp(-\lambda\delta^n q^4)}{\int dq \exp\{-\frac{1}{2}\delta^n[m^2+\omega(0)]q^2\} \exp(-\lambda\delta^n q^4)}\right]\right\}$$

$$= \exp\left\{\sum_{p=1}^{\infty} (1/p!) \sum_{i_1} \int dq_1 \cdots \sum_{i_p} \int dq_p \prod_{j=1}^{p} [\exp(iq_j f_j \delta^n) - 1] T_p(q_1, i_1; \cdots; q_p, i_p)\right\}, \tag{6.27}$$

where

$$T_p(q_1, i_1; \cdots; q_p, i_p) = (p-1)!(-1)^{p-1}\delta_{i_1 i_2} \cdots \delta_{i_1 i_p} \prod_{j=1}^{p} \left[\frac{\exp\{-\frac{1}{2}\delta^n[m^2+\omega(0)]q_j^2\} \exp(-\lambda\delta^n q_j^4)}{\int dq \exp\{-\frac{1}{2}\delta^n[m^2+\omega(0)]q^2\} \exp(-\lambda\delta^n q^4)}\right]. \tag{6.28}$$

Thus, the full generating functional for a $\lambda\phi^4$ field theory on a lattice is given by

$$L^{\delta}(f) = Z^{\delta}(f)/Z^{\delta}(0) = \left\{\exp\left[-\tfrac{1}{2}\int dq_1 \int dq_2 \delta^{2n} \sum_{i\neq j} q_1 q_2 \omega^{\delta}_{ij} : \frac{1}{i\delta^n} \frac{\delta}{\delta f_i(q_1)} \frac{1}{i\delta^n} \frac{\delta}{\delta f_j(q_2)} :\right] L^{\delta}_{\mathrm{SUL}}(f)\Big|_{f_i(q)=qf_i}\right\} \Big/ Z^{\delta}(0), \tag{6.29}$$

and one can see that it is in a form to which the generalization of the virial expansion described in Sec. VB can be applied.

In the free theory case discussed in the previous subsection it was necessary to carry out a resummation of the series before taking the limit $\delta \to 0$. Most likely, similar resummations will have to be carried out in the interacting case as well before one is able to take the $\delta \to 0$ limit. Of course, additional resummations corresponding to renormalizations of m and λ will be necessary as well. We will not pursue these points in this paper; however, we believe they merit further investigation.

Remarks: (1) One can also show that Eq. (6.15) and the generalization of Eq. (6.26) to the case of an arbitrary polynomial interaction of degree greater than 2 does not define an ultralocal generating functional. However, it may be possible to define a generating functional in this manner for nonpolynomial interactions such as the sine–Gordon interaction. Also, it may be possible to define the static ultralocal generating function by using Eq. (6.13) on a lattice and adjusting the mass and coupling constant as the limit of zero lattice spacing is taken, in order to obtain an expression of the form (6.15). This procedure could be interpreted as taking into account renormalization in L_{SUL}. It would also simplify the use of the virial expansion to obtain the full $L(f)$.

(2) Caianiella and Scarpetta[40] have determined the n-point functions for the static ultralocal $\lambda\phi^4$ model by solving the hierarchy of coupled equations for the Green's functions. These may be used to define a formal generating functional. However, we do not know whether it would possess positivity and the other essential properties of generating functionals.

ACKNOWLEDGMENTS

It is a pleasure for the authors to thank Dr. D.K. Campbell and Dr. J. Kiskis for their valued comments on the manuscript.

APPENDIX A: RESUMMATION OF EQ. (4.3)

The expression for the generating functional in Eq. (4.3) contains a summation over allowed graphs with labeled and directed lines. The allowed graphs (connected or disconnected but with no isolated vertices) may have any number of lines between pairs of vertices. In this section we outline the resummation which reduces the allowed graphs to simple graphs with labeled vertices. The first step is to switch from graphs with directed and labeled lines to those with labeled vertices. To accomplish this, we need to introduce the symmetry factor.[26] The symmetry factor, σ, of a graph is the order of the group of permutations of vertices that leave the graph invariant, times the factor $\Pi_{i,j} n_{ij}!$, where n_{ij} is the multiplicity of lines between vertices i and j. For a graph G with N_V vertices and N_l lines, it can be shown that the symmetry factor is given by

$$\frac{1}{\sigma(G)} = \prod_{i,j}\left(\frac{1}{n_{ij}!}\right)\left(\frac{\text{the number of ways to label vertices}}{N_V!}\right)$$

$$= \left(\frac{\text{the number of different ways to label lines}}{N_l!}\right)$$

$$\times \begin{pmatrix}\text{factor of } \frac{1}{2} \text{ for each connected}\\ \text{subgraph with two vertices}\end{pmatrix}$$

$$= \left(\frac{\text{number of ways to label and direct lines}}{N_l!\, 2^{N_l}}\right). \tag{A1}$$

Thus, by including the factor $\Pi_{i,j}(1/n_{ij}!)$ in the weight, we can switch from graphs with labeled and directed lines to those with labeled vertices and still get the same counting factors.

The resummation leading to simple graphs is performed as follows. Every graph can be reduced to a unique simple graph by replacing multiple lines between every pair of vertices by a single line. Next we consider the family of graphs $\{G\}$ which reduces to a given simple graph G. We denote a member of this family by $G(n_1, \ldots, n_{N_l})$, where n_i is the multiplicity of the ith line. The symmetry factor of $G(n_1, \ldots, n_{N_l})$ is

$$\frac{1}{\sigma(n_1, \ldots, n_{N_l})} = \left(\frac{1}{\sigma(G)}\right)\frac{P}{\pi n_i!}, \tag{A2}$$

where P is the number of nontrivial permutations of the multiplicities which leave the graph invariant, i.e., the

permutations π having the properties that $G(n_1, \ldots, n_{N_I}) = G(n_{\pi_1}, \ldots, n_{\pi_{N_I}})$ and $n_{\pi_i} \neq n_i$ whenever $\pi_i \neq i$. As a result, the counting factor is just right to vary each multiplicity independently. Therefore, the sum of the weights for graphs in $\{G\}$ with labeled vertices is

$$\frac{1}{N_V!}\sum_{\{G\}}\int d^3x_1 \cdots \int d^3x_{N_V}\left[\prod_{i=1}^{N_V}\rho_0 \exp[if(\mathbf{x}_i)]\right] \times \prod_{I=1}^{N_I}\frac{1}{n_I!}[-\beta U(\mathbf{x}_{I_1}-\mathbf{x}_{I_2})]^{n_I}$$

$$= \frac{1}{\sigma(G)}\sum_{n_1=1}^{\infty}\cdots\sum_{n_{N_I}=1}^{\infty}\int d^3x_1 \cdots \int d^3x_{N_V} \times\left(\prod_{i=1}^{N_V}\rho_0 \exp[if(\mathbf{x}_i)]\right)\prod_{I=1}^{N_I}\frac{1}{n_I!}\,[-\beta U(\mathbf{x}_{I_1}-\mathbf{x}_{I_2})]^{n_I}$$

$$= \frac{1}{\sigma(G)}\int d^3x_1 \cdots \int d^3x_{N_V}\left[\prod_{i=1}^{N_V}\rho_0 \exp[if(\mathbf{x}_i)]\right] \times \prod_{I=1}^{N_I}(\exp[-\beta U(\mathbf{x}_{I_1}-\mathbf{x}_{I_2})]-1)\,. \tag{A3}$$

Thus the sum over a family of graphs with the same simple graph results in a modification of the weight associated with a simple graph $[-\beta U$ is replaced by $(\exp(-\beta U)-1)]$. The summation over all allowed graphs can thus be performed by first summing over the family $\{G\}$ and then over all simple graphs G.

APPENDIX B: A THEOREM ON GRAPHS

In Sec. IV we used the following:

Theorem: Let C_N be a connected graph with N labeled vertices, G_N a graph, either connected or disconnected, with N labeled vertices and $X(G_N)$ the weight assigned to G_N such that

(1) $W(G_N)$ is independent of the labeling G_N,

(2) $W(G_N)$ is the product of the weights for each disjoint connected part of G_N.

Then,

$$1+\sum_{N=1}^{\infty}\frac{x^N}{N!}\sum_{G_N}W(G_N) = \exp\left(\sum_{N=1}^{\infty}\frac{x^N}{N!}\sum_{C_N}W(C_N)\right).$$

For a proof see Ref. 25.

[1]R. Menikoff, J. Math. Phys. 15, 1138 (1974).
[2]R. Menikoff and D.H. Sharp, J. Math. Phys. 16, 2341 (1975).
[3]N.N. Bogoliubov, J. Phys. (USSR) 10, 256 (1946). A translation of this paper appears as "Problems of a Dynamical Theory in Statistical Physics," in *Studies in Statistical Mechanics, Vol. I*, edited by J. de Boer and G. Uhlenbeck (North-Holland, Amsterdam, 1962). pp. 11—50.
[4]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), Sec. 4.7.
[5]Ruelle, Ref. 4, Sec. 7.1.2. The use of C^*-algebras in statistical mechanics is also reviewed in G.G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).
[6]G.A. Goldin and D.H. Sharp, "Lie Algebras of Local Currents and Their Representations," in *Group Representations in Mathematics and Physics, Battelle Seattle* 1969 *Rencontres*, edited by V. Bargmann (Springer, New York, 1970), pp. 300—11.
[7]G.A. Goldin, J. Grodnik, R.T. Powers, and D.H. Sharp, J. Math. Phys. 15, 88 (1974).
[8]For a review of this subject, see H.M. Fried, *Functional Methods and Models in Quantum Field Theory* (MIT Press, Cambridge, 1972).
[9]O. Penrose, J. Math. Phys. 4, 1312 (1963).
[10]D. Ruelle, Ann. Phys. (N.Y.) 25, 109 (1963).
[11]A. review of recent developments in this subject, and extensive references to the literature, are to be found in B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, N.J., 1974).
[12]K. Symanzik, "Euclidean Quantum Field Theory," in *Local Quantum Theory*, edited by R. Jost (Academic, New York, 1969), pp. 152—226.
[13]J. Glimm, A. Jaffe, and T. Spencer, "The Particle Structure of The Weakly Coupled $P(\phi)_2$ Model and Other Applications of High Temperature Expansion," in *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973), pp. 132—242; Ann. Math. 100, 585 (1974).
[14]See Ref. 11, Chap. VIII.
[15]J.R. Klauder, Commun. Math. Phys. 18, 307 (1970); Acta. Phys. Austr., Suppl. VIII, 227 (1971).
[16]C. Newman, Ph.D. thesis (Princeton University, 1971); Commun. Math. Phys. 26, 169 (1972).
[17]J.R. Klauder, Acta. Phys. Austr. 41, 237 (1975); Augmented Quantum Field Theory: A Proposal to Extend Conventional Formulations", Phys. Rev. D (to be published).
[18]R.F. Dashen and D.H. Sharp, Phys. Rev. 165, 1857 (1968).
[19]For reviews see Ref. (6) and D.H. Sharp, "What We Have Learned About Representing Local Nonrelativistic Current Algebras," in *Local Currents and Their Applications*, edited by D.H. Sharp and S.S. Wightman (North-Holland, Amsterdam, 1974), pp. 85—98.
[20]J. Grodnik and D.H. Sharp, Phys. Rev. D 1, 1531 (1970).
[21]G.A. Goldin, J. Math. Phys. 12, 462 (1971).
[22]I. Gel'fand and N. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4.
[23]Bogoliubov has derived a similar functional equation for his generating functional $L_B(f)$:

$$\nabla_{\mathbf{x}}\frac{\delta L_B(f)}{\delta f(\mathbf{x})} = -\beta\int d^3y\,\nabla_{\mathbf{x}}U(\mathbf{x}-\mathbf{y})[f(\mathbf{y})+\bar{\rho}]\,\delta^2 L_B(f)/\delta f(\mathbf{x})\delta f(\mathbf{y}).$$

An equation similar to Eq. (2.17), generalized to the case of N-body potentials, also enters into the determination of the representations of nonrelativistic current algebras, as discussed in Ref. 2.
[24]Our notation for graphs is similar to that used in Refs. 25 and 26.
[25]G.E. Uhlenbeck and G.W. Ford, "The Theory of Linear Graphs with Applications to the Theory of the Virial Development of the Properties of Gases," in *Studies in Statistical Mechanics, Vol. I*, edited by J. De Boer and G. Uhlenbeck (North-Holland, Amsterdam, 1962), pp. 123—211.
[26]G. Stell, "Cluster Expansions for Classical Systems in Equilibrium" in *Classical Fluids*, edited by H.L. Frisch and J.L. Lebowitz (Benjamin, New York, 1964), pp. 171—266.
[27]For details see Refs. 25 and 26.
[28]J.E. Mayer and E.W. Montroll, J. Chem. Phys. 9, (1941); see also Ref. 25.
[29]B. Zumino, Phys. Fluids 2, 20 (1959).
[30]See Ruelle, Ref. 4, Sec. 4.2.
[31]Symanzik, Ref. 12, has shown how to write the Kirkwood—Salzburg equations as a functional equation for Bogoluibov's generating functional $L_B(f)$. In terms of the generating functional used in this paper, Symanzik's equation reads

$$c^{-if(\mathbf{x})}\frac{1}{i}\frac{\delta L(f)}{\delta f(\mathbf{x})} = zL(f(\cdot)+i\beta U(\cdot-\mathbf{x})).$$

Note, however, that to use this equation one must consider functionals with complex argument.
[32]R. Menikoff, J. Math. Phys. 15, 1394 (1974).
[33]J.L. Lebowitz, G. Stell, and S. Baer, J. Math. Phys. 6, 1282 (1965); G. Stell, J.L. Lebowitz, S. Baer, and W. Theumann, J. Math. Phys. 7, 1532 (1966).

[34]C. Woo, Phys. Rev. A **6**, 2312 (1972).
[35]C.E. Campbell, Phys. Lett. A **44**, 471 (1973).
[36]E. Feenberg, Ann. Phys. (N.Y.) **48**, 128 (1974).
[37]R. Menikoff and D.H. Sharp, J. Math. Phys. **16**, 2353 (1975).
[38]C.E. Campbell, J. Math. Phys. **16**, 1076 (1975).
[39]See Fried, Ref. 8, Appendix H.
[40]E.R. Caianiello and G. Scarpetta, Nuovo Cimento A **22**, 448 (1974).
[41]S. Kovesi-Domokos, Nuovo Cimento A **33**, 769 (1976).
[42]See, for example, S.D. Drell, M. Weinstein and S. Yankielowicz, Phys. Rev. D **14**, 487 (1976).
[43]See Fried, Ref. 8, Appendix F.
[44]G. Scarpetta, Lett. Nuovo Cimento **13**, 302 (1975).
[45]See Theorem 6.6 of Newman's thesis, Ref. 16.

Multiple steady states in a simple reaction-diffusion model with Michaelis-Menten (first-order Hinshelwood-Langmuir) saturation law: The limit of large separation in the two diffusion constants

J. L. Ibañez and M. G. Velarde[a)]

Departamento de Física—C-3, Universidad Autónoma de Madrid, Cantoblanco (Madrid), Spain
(Received 21 March 1977)

The admissible multiple nonuniform steady states of a model bimolecular autocatalytic reaction-diffusion system with Michaelis-Menten (first-order Hinshelwood-Langmuir) saturation law are constructed in the case of large scale separation in the two diffusion constants. Both the Dirichlet and the Neumann problems are discussed in a one-dimensional geometry, and the corresponding bifurcation pictures are given.

1. INTRODUCTION

In a recent note[1] (see also Ref. 2 and for motivation Refs. 3, 4) we have discussed the existence and stability of limit cycle and nonlinear spatially ordered structures in a simple bimolecular autocatalytic reaction—diffusion model with Michaelis—Menten (first-order Hinshelwood—Langmuir, which corresponds to Holling's law in ecology) saturation law. The model is

$$A \to Y, \tag{1.1a}$$

$$X + Y \to 2X, \tag{1.1b}$$

$$X \xrightarrow{S} P, \tag{1.1c}$$

in which S accounts for the saturation process whose law is given below. Under various simplifying assumptions, which are specified fully in Ref. 1, the differential problem that describes (1) in terms of dimensionless quantities has the form

$$\frac{\partial X}{\partial t} = XY - \frac{X}{1+qX} + D_X \Delta X, \tag{1.2a}$$

$$\frac{\partial Y}{\partial t} = A - XY + D_Y \Delta Y, \tag{1.2b}$$

in which X, Y, and A are positive concentrations; A and $q > 0$ are treated as parametric constants, and D_X and D_Y are diffusion coefficients and are also constant.

In Ref. 1 (and in Ref. 2 for a spherically shaped surface problem) we have discussed the formation of nonlinear structures when $D_Y \sim D_X$, though $D_Y > D_X$. Both Dirichlet's (fixed concentrations on boundaries) and Neumann's (fixed fluxes on boundaries) conditions were considered in Ref. 1. Recently Boa[5] has discussed the multiple steady states having finite deviation from the fixed point of a similar model to (1) in the limit $D_Y \to \infty$. We are referring to Boa's work on the (Brusselator) trimolecular model of Prigogine and co-workers.[6] On the other hand, Koppell and Howard[7] have argued that such a limiting case ($D_Y \to \infty$, $D_X < \infty$) should be of interest by analogy with the double-diffusive (thermohaline) problem,[8] in which the large separation of scales of heat diffusivity (κ) and mass diffusivity (D), $\kappa \gg D$, is at the origin of convective instability. It seems to us[1,2] that such an analogy has no firm foundations as it has also been pointed out by Auchmuty and Nicolis.[9] However, the change in the boundary value problem of (2) introduced by the limit $D_Y = \infty$ is so dramatic that we decided to explore the salient features of (2) with $D_Y = \infty$, for which (2) becomes a "conservative" system. Besides, the discussion that follows gives a more complete picture of bifurcation with two largely different diffusion constants than the brief account provided by Boa.[5]

2. STABILITY OF THE FIXED POINT IN THE LIMIT $D_Y = \infty$

The one-dimensional boundary value problem for steady states of Eqs. (1.2) in the limit $D_Y = \infty$ is

$$0 = XY - \frac{X}{1+qX} + D_X \frac{d^2}{dr^2} X \tag{2.1a}$$

$$0 = \frac{d^2}{dr^2} Y \tag{2.1b}$$

with *either* (Dirichlet problem)

$$X(0) = X(1) = \text{const}, \tag{2.2a}$$

$$Y(0) = Y(1) = \text{const}, \tag{2.2b}$$

or (Neumann problem)

$$\left.\frac{dX}{dr}\right|_{r=0} = \left.\frac{dX}{dr}\right|_{r=1} = \text{const}, \tag{2.3a}$$

$$\left.\frac{dY}{dr}\right|_{r=0} = \left.\frac{dY}{dr}\right|_{r=1} = \text{const}, \tag{2.3b}$$

where the constant can be set vanishing.

The solution of (2.1b) is

$$Y = K, \tag{2.4a}$$

in which K is a positive constant. Thus, in the time scale considered we cannot see any Y fluctuation. We may choose for K the value corresponding to the steady solution at $D_Y < \infty$, $D_Y > D_X$. We have

$$Y = 1 - qA, \tag{2.4b}$$

and Eq. (2.1a) becomes

[a)] Also at Laboratoire de Dynamique et Thermophysique des Fluides (CNRS No. 72), Université de Provence, Centre de Saint-Jerome, 13397—Marseille—Cedex 4 (France).

 0022-2488/78/1901-0151$1.00

$$0 = X(1 - qA) - \frac{X}{1 + qX} + D_X \frac{d^2}{dr^2} X. \tag{2.5}$$

Equation (2.5) has the nontrivial steady solution

$$X = A/(1 - qA). \tag{2.6}$$

The stability of (2.5) in the neighborhood of (2.6) is decided by the following eigenvalue problem:

$$L(q, A, D_X)\psi \equiv \left[qA(1 - qA) + D_X \frac{d^2}{dr^2}\right]\psi = \lambda\psi, \tag{2.7}$$

with functions ψ defined in $[0, 1]$ that possess continuous second-differentials and satisfy the b.c. We define the following inner product:

$$\langle \psi, \phi \rangle = \int_0^1 \psi\phi^* \, dr \tag{2.8}$$

for ψ and ϕ belonging to the space defined above.

For the Dirichlet problem (2.7) yields

$$\psi_k = a_k \sin \pi k r \tag{2.9}$$

with a_k real and $k = 1, 2, 3, \cdots$.

For the Neumann problem from (2.7) we have

$$\psi_k = b_k \cos \pi k r \tag{2.10}$$

with b_k real and $k = 0, 1, 2, 3, \cdots$. In both cases the wave-dependent eigenvalue λ_k corresponding to ψ_k is given by

$$qA(1 - qA) - D_X \pi^2 k^2 = \lambda_k \tag{2.11}$$

and the fixed point (2.6) is asymptotically stable if for all allowed values of k the real part of the eigenvalue is negative.

From (2.11) we get:

(*i*) *Neumann problem*: The fixed point (2.6) is unstable for all values of q and A such that $qA < 1$. This comes from the fact that

$$\lambda_0 = qA(1 - qA) > 0. \tag{2.12}$$

Thus, the fixed point is always unstable for perturbations of infinite wavelength.

(*ii*) *Dirichlet problem*: There is the following alternative. *Either*

$$D_X > 1/4\pi^2 \tag{2.13}$$

and the fixed point (2.6) is asymptotically stable for q and A such that $0 < qA < 1$, *or*

$$D_X < 1/4\pi^2 \tag{2.14}$$

and the fixed point (2.6) is asymptotically stable for q and A restricted to the region

$$qA \in (0, A_1^-) \cup (A_1^+, 1), \tag{2.15a}$$

$$A_1^\pm = 0.5 \pm 0.5(1 - 4\pi^2 D_X)^{1/2} \tag{2.15b}$$

and unstable otherwise.

(*iii*) *In either case (Dirichlet or Neumann problem)*: At q and D_X fixed, there exists a set of values of A, $A_{k_c}^\pm$, for which the normal mode corresponding to k_c becomes unstable. We have

$$L(q, A_{k_c}^\pm, D_X)\psi_{k_c} = 0, \tag{2.16a}$$

$$A_{k_c}^\pm = (1/q)(0.5 \pm 0.5[1 - 4\pi^2 D_X k_c^2]^{1/2}), \tag{2.16b}$$

where if

$$A \in (0, A_{k_c}^-) \cup (A_{k_c}^+, q^{-1}), \tag{2.17}$$

the fixed point (2.6) is asymptotically stable to perturbations of wavelength $2\pi/k_c$. For $A \in (0, q^{-1})$ there is an upper bound to normal modes m that may become unstable. We have

$$1/4\pi^2(m + 1)^2 < D_X < 1/4\pi^2 m^2. \tag{2.18}$$

Thus, with $D_X < \infty$ the fixed point (2.6) is asymptotically stable to perturbations of short wavelength.

Lastly, we note that the following transversality condition is satisfied,

$$\langle \psi_{k_c}^a, L'(q, A_{k_c}^\pm, D_X)\psi_{k_c} \rangle \neq 0, \tag{2.19}$$

in which L' comes from differentiating L with respect to A and $\psi_{k_c}^a$ is an eigenfunction of the adjoint operator of L, with vanishing eigenvalue. In our case here we merely have $\psi_{k_c}^a = \psi_{k_c}$.

3. NONLINEAR STEADY SOLUTIONS

We shall now construct the branches that bifurcate at $A_{k_c}^\pm$. We use the Poincarê–Linstedt method, and define the series expansions

$$A_c^\pm = A_{k_c}^\pm + \sum_{n=1}^{\infty} \epsilon^n A_{k_c}^{(n)\pm}, \tag{3.1a}$$

$$X_{k_c}^\pm = X_S + \sum_{n=1}^{\infty} \epsilon^n U_{k_c}^{(n)\pm}. \tag{3.1b}$$

Introducing (3.1) in (2.5) a hierarchy of equations is generated in powers of the new unknown ϵ. $A_{k_c}^{(n)}$ and $U_{k_c}^{(n)}$ are to be determined.

(i) Dirichlet problem

The results found are different according to the character, even or odd, of the critical mode.

(*ia*) *If k_c is even*, we have

$$A_{k_c}^{(1)\pm} = 0 \tag{3.2a}$$

and

$$A_{k_c}^{(2)\pm} = \{\tfrac{3}{4} q(1 - qA_{k_c}^\pm) - [128q(1 - qA_{k_c}^\pm)^6 k_c^4/\pi^2 D_X]S_1\}(1 - 2qA_{k_c}^\pm)^{-1}, \tag{3.2b}$$

in which

$$S_1 = \sum_{\substack{k=1 \\ (k \text{ odd})}}^{\infty} [k^2(k^2 - 4k_c^2)^2(k^2 - k_c^2)]^{-1} = \frac{-5\pi^2}{2^7 \cdot 6 \cdot k_c^6}. \tag{3.2c}$$

(*ib*) *If k_c is odd*, we have

$$A_{k_c}^{(1)\pm} = -8(1 - qA_{k_c}^\pm)^3 q/3k_c\pi(1 - 2qA_{k_c}^\pm). \tag{3.3}$$

It is to be noted that with k_c even

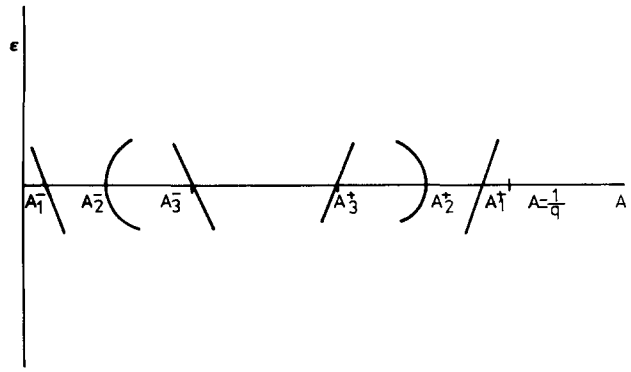

FIG. 1. Qualitative sketch of the various bifurcating branches that appear with Dirichlet b.c. The quantity ϵ denotes an amplitude and $\epsilon = 0$ accounts for the fixed point (the homogeneous steady solution).

$$\mathrm{sgn}(A_{k_c}^{(2)+}) \neq \mathrm{sgn}(A_{k_c}^{(2)-}) \tag{3.4}$$

whereas with k_c odd we have

$$\mathrm{sgn}(A_{k_c}^{(1)+}) \neq \mathrm{sgn}(A_{k_c}^{(1)-}). \tag{3.5}$$

Figure 1 depicts the various possible bifurcations for values of D_X such that only the first three modes $k_c = 1, 2, 3$ can be destabilized, with

$$1/36\pi^2 > D_X > 1/64\pi^2. \tag{3.6}$$

(ii) Neumann problem

Whatever the values of k_c are, we have

$$A_{k_c}^{(1)\pm} = 0 \tag{3.7a}$$

and

$$A_{k_c}^{(2)\pm} = [\tfrac{3}{4}(1 - qA_{k_c}^{\pm})^4 q + (5/6A_{k_c}^{\pm})(1 - qA_{k_c}^{\pm})^5](1 - 2qA_{k_c}^{\pm})^{-1}. \tag{3.7b}$$

We also have

$$\mathrm{sgn}(A_k^{(2)+}) \neq \mathrm{sgn}(A_{k_c}^{(2)-}). \tag{3.8}$$

Figure 2 describes the bifurcations that appear for values of m in the range (3.6).

4. FURTHER DETAILS ABOUT THE NONUNIFORM STEADY STATES

For later convenience we introduce the new variable $V(r) = DdX/dr$. Then Eq. (2.5) can be recast in the following matrix form:

$$\frac{d}{dr}\begin{pmatrix} X \\ V \end{pmatrix} = \begin{pmatrix} X/D_X \\ -(1 - qA)X + \dfrac{X}{1 + qX} \end{pmatrix} = \begin{pmatrix} \partial E/\partial V \\ -\partial E/\partial X \end{pmatrix}, \tag{4.1}$$

in which $E(X, V)$ denotes a first integral of (2.5)

$$E(X, V) = \frac{V^2}{2D_X} + \frac{X^2(1 - qA)}{2} - \frac{X}{q} + \log\left(\frac{|1 + qX|}{q^2}\right). \tag{4.2}$$

As a matter of fact, the existence of a conserved quantity like E to Eq. (2.5) follows immediately from the type of equation that we are dealing with. A generic equation like

$$\frac{d^2X}{dr^2} + f(X) = 0 \tag{4.3}$$

has the following first integral:

$$E = \frac{1}{2}\left(\frac{dX}{dr}\right)^2 + \int f(X)\, dX. \tag{4.4}$$

This corresponds to the differential equation also found in the stability analysis of catalytic wires.[4,10]

The system (4.1) has two fixed points, namely

$$X^{(1)} = V^{(1)} = 0 \tag{4.5a}$$

and

$$X^{(2)} = A(1 - qA)^{-1}, \quad V^{(2)} = 0. \tag{4.5b}$$

The stability of these two fixed points is related to the eigenvalues of the linearized operator (for given b.c.), $\mathcal{L}$,

$$\mathcal{L}\begin{pmatrix} \partial E/\partial V \\ -\partial E/\partial X \end{pmatrix}\Bigg|_{X^{(i)},\, V^{(i)},\ i=1,2} = \begin{pmatrix} 0 & D_X^{-1} \\ [1 - qA] + [1 + qX^{(i)}]^{-2} & 0 \end{pmatrix}. \tag{4.6}$$

It is to be noted that the existence of a first integral like E forbids asymptotic stability to the fixed points (4.5). Nor can they be nodes or focuses.[11] It is found that (4.5a) is a saddle and (4.5b) is a center. Figure 3 gives some of the admissible trajectories for the values of E. S_1 and S_2 correspond respectively to (4.5a) and (4.5b). There is an obvious symmetry arising from the invariance of E under the change V into $-V$.

It is clear that not all trajectories in Fig. 3 are realized, for the admissible trajectories to be realized they must satisfy the b.c.

(*i*) *Dirichlet problem*: The only trajectories that are solutions of the problem are those for which, if at the origin, we have $X(r = 0) = X^{(2)}$ and $V(0)$, then at the end point the value is $X(r = 1) = X^{(2)}$ with $V(r = 1)$. These are the trajectories that after any number of turns in a unit distance end at $X = X^{(2)}$, though V may be different from $V(0)$.

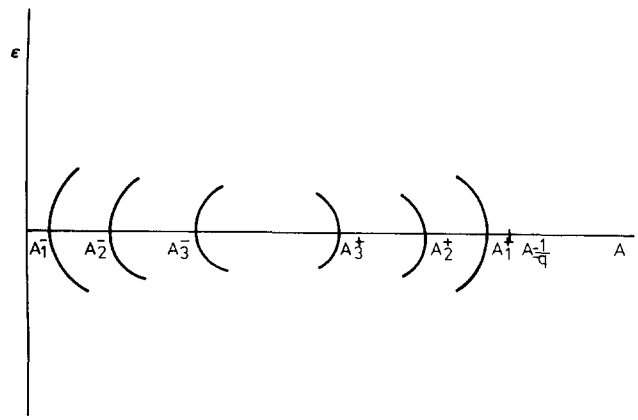

FIG. 2. Qualitative sketch of the bifurcating branches with Neumann b.c. and ϵ as above.

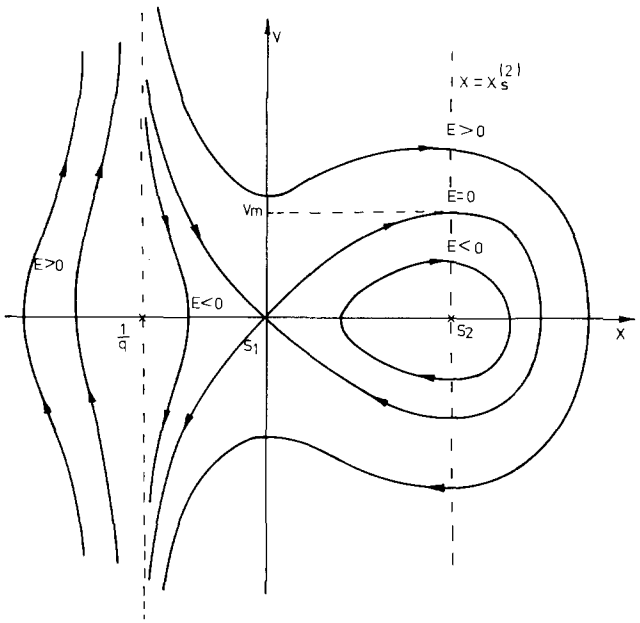

FIG. 3. Phase trajectories of the system (4.1). E is a *conserved* quantity. S_1 and S_2 correspond respectively to the two fixed points of the system.

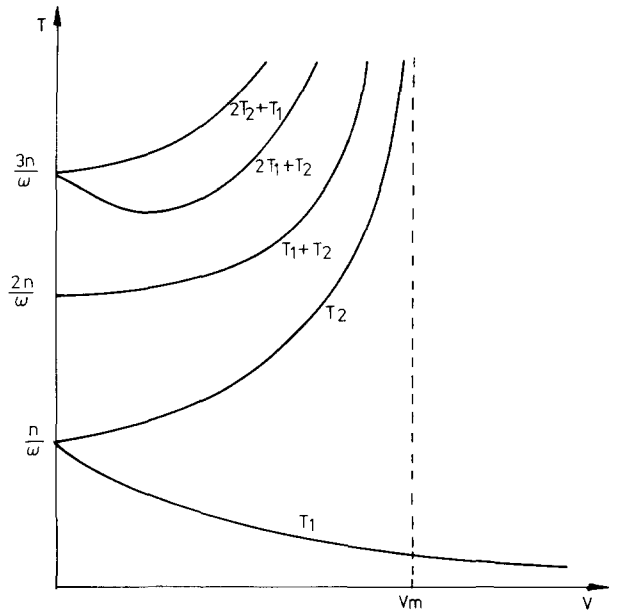

FIG. 4. "Spatial periodicity" of (4.1) as function of the condition at the origin.

(*ii*) *Neumann problem*: The admissible trajectories must satisfy the following conditions: $X(r=0)$ and $X(r=1)$ are both real and positive and $V(r=0)=V(r=1)=0$. They start and end after a unit distance on the axis $V=0$.

Figure 4 gives a qualitative sketch of the period in r-space of trajectories of (4.1) as function of the conditions at the origin. $T_1(V)$ denotes the distance from $(X^{(2)}, V)$ to $(X^{(2)}, -V)$, with V real and positive. $T_2(V)$ corresponds to the passage from $(X^{(2)}, -V)$ to $(X^{(2)}, V)$. They do not need to be equal. V_m gives the value of V for which the point $(X^{(2)}, \pm V_m)$ lies on the separatrix, $E=0$. We have

$$V_m = \left\{ -2D_X \left[\frac{A}{1-qA} \left(\frac{A}{2} - \frac{1}{q} \right) - \frac{\log|1-qA|}{q^2} \right] \right\}^{1/2}. \tag{4.7}$$

Note that $T_2(V_m)=\infty$ for the origin (0, 0) is a solution of (4.1). On the other hand, at $V=0$ [(4.5b)], T_1 and T_2 are equal:

$$T_1(0) = T_2(0) = \pi/\omega, \tag{4.8}$$

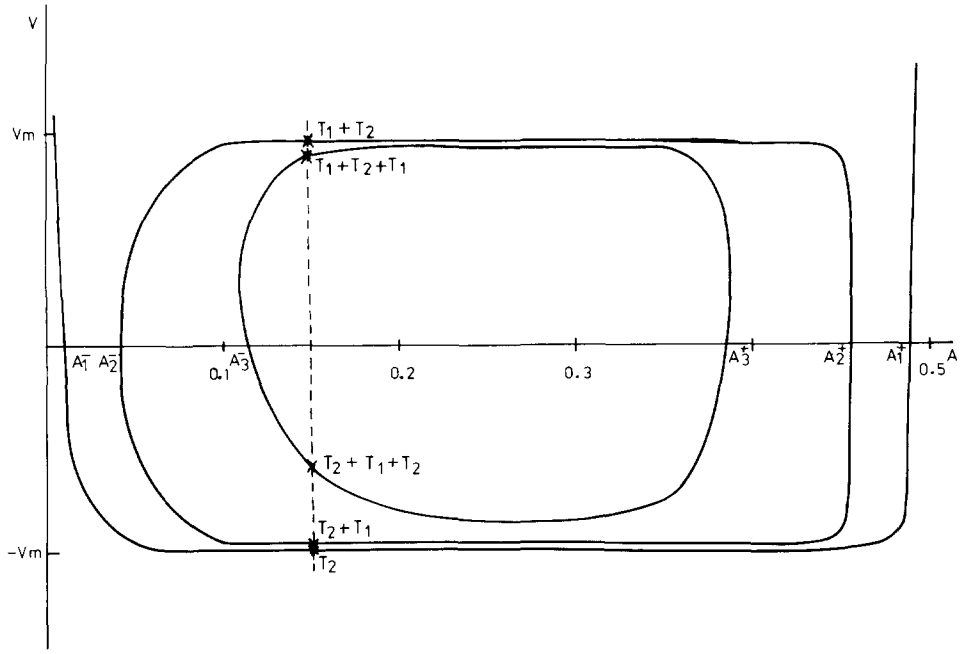

FIG. 5. Actual bifurcation picture with Dirichlet b.c. in the case $q=2.0$ and $D_x=0.002$.

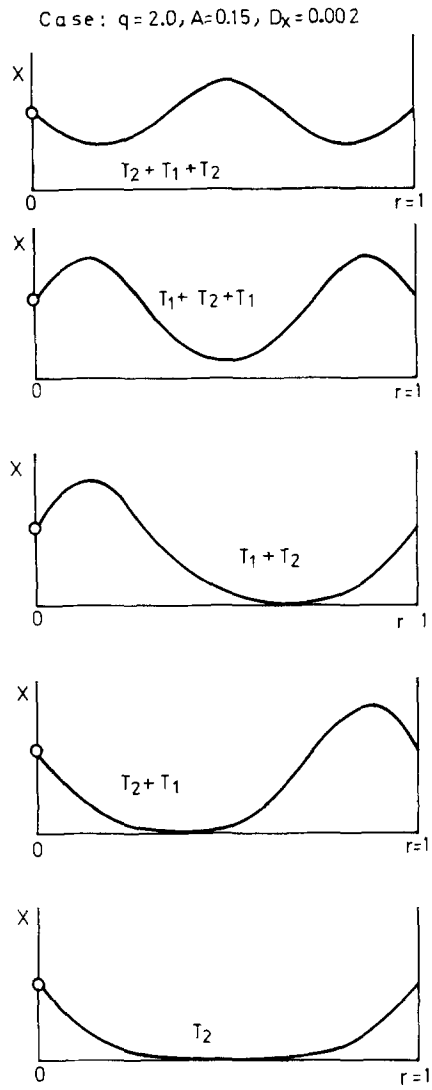

FIG. 6. The five nonuniform steady states available for $q=2.0$, $D_x=0.002$, and $A=0.15$ in the preceding figure.

with ω given by the purely imaginary eigenvalue of (4.6) with $i=2$. We have

$$\omega=[qA(1-qA)/D_X]^{1/2}. \tag{4.9}$$

For Dirichlet b.c. a solution to (4.1) corresponds to the intersect of any of the curves of Fig. 4 with the ordinate $T=1$. With values of D_X in the region $1/36\pi^2 > D_X > 1/64\pi^2$ the following results are found:

(a) For values of A such that $A\in(0, A_1^-)\cup(A_1^+, 1/q)$ in which $A_n^\pm$ are given by (2.16b), it follows that $\pi/\omega>1$. Thus, only the curve T_1 intersects the ordinate $T=1$ and yields a solution to (4.1).

(b) For values of A, $A\in(A_1^-, A_2^-)\cup(A_2^+, A_1^+)$, it follows that $\pi/\omega<1<2\pi/\omega$. Thus, there is only steady solution corresponding to an intersect of $T=1$ with the T_2 curve.

(c) With $A\in(A_2^-, A_3^-)\cup(A_3^+, A_2^+)$ we have $2\pi/\omega<1<3\pi/\omega$. Thus, for every value of A there are at least three satationary solutions to (4.1): One corresponds to T_2 and two belong to the intersects of $T=1$ with the curve (T_1+T_2), respectively depending on the initial condition $(X^{(2)}, V)$ or $(X^{(2)}, -V)$. On the other hand, the minimum on the curve $(2T_1+T_2)$ is correlated to the eventual existence in the neighborhood of A_3^- and A_3^+ of two other solutions, which respectively correspond to the initial conditions $(X^{(2)}, V_1)$ and $(X^{(2)}, V_2)$, with V_1 and V_2 positive constants.

(d) Lastly for $A\in(A_3^-, A_3^+)$ we have $3\pi/\omega<1$, and then every value of A branches five solutions: One with T_2, two with (T_1+T_2) and two with $(2T_1+T_2)$ and (T_1+2T_2), respectively.

The whole bifurcation picture is depicted in Fig. 5 for the case $q=2.0$ and $D_X=0.002$. To the vanishing value of V belongs the fixed point (homogeneous steady solution). There is a clear correspondence of the branches drawn in Fig. 5 with the above constructed solutions by the Poincaré–Linstedt method. In particular, it is to be noted the close relationship between the existence of a minimum on the $(2T_1+T_2)$ curve and the kind of bifurcation that appears at $k_c=3$. More generally one is led to say that the existence of a minimum in the function $(n+1)T_1(V)+nT_2(V)$, with n an integer and $V\neq 0$, corresponds to bifurcation with k_c odd.

To the generic case of Fig. 5 and for $A=0.15$ correspond the five steady nonuniform concentration profiles drawn in Fig. 6. The intersects of $A=0.15$ with the curves in Fig. 5 give the values V to be reported to Fig. 3. Thus, we have the value of X, together with the slope at the origin, and E. Then with these initial values the profiles are constructed by direct computer integration of (4.1).

Lower values of D_X ($D_X<0.002$) merely yield more critical values $A_{k_c}^\pm$ but qualitatively do not change the picture. For the Neumann b.c., as the trajectories must start and end at $V=0$, their existence depends on the behavior of the functions $(n/2)[T_1(V)+T_2(V)]$ of

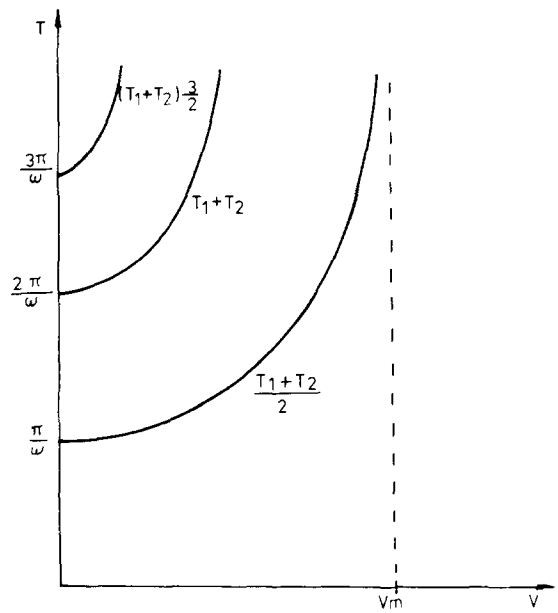

FIG. 7. The function $\frac{1}{2}n(T_1+T_2)$ for bifurcation with Neumann b.c.

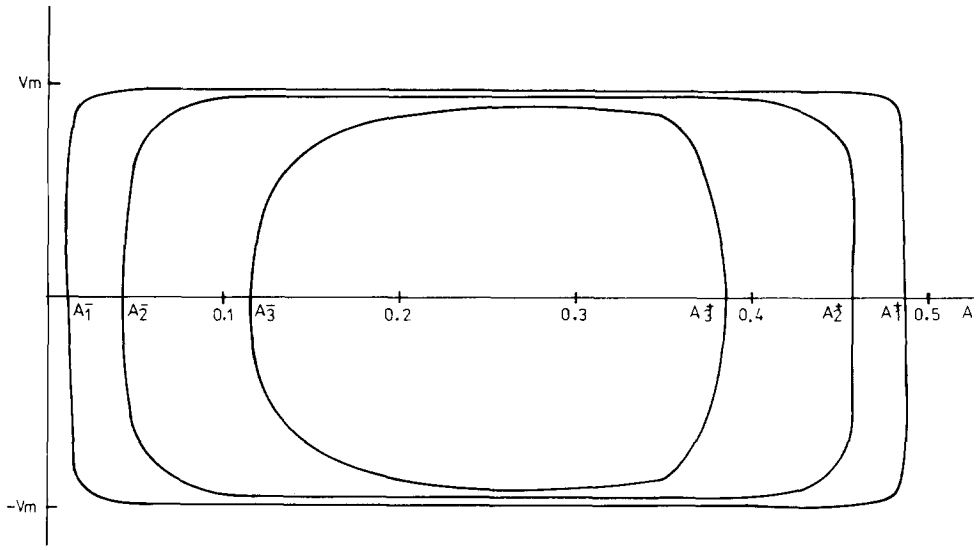

FIG. 8. Actual bifurcation picture with Neumann b.c. in the case $q=2.0$ and $D_x=0.002$.

which a qualitative sketch is given in Fig. 7. Every intersection with the ordinate $T=1$ yields two solutions belonging respectively to the conditions at the origin $(X_1, 0)$ and $(X_2, 0)$ where $X_1 < X^{(2)} < X_2$. They respectively cross the points $(X^{(2)}, V)$ and $(X^{(2)}, -V)$. Figure 8 gives a qualitative bifurcation picture for the case $q=2.0$ and $D_X=0.002$.

We have completed our aim of providing a classification of all *admissible* spatially nonuniform steady states that are expected in the limit $D_Y \gg D_X$, $D_Y \to \infty$, $D_X < \infty$, for the system (1.2), with either Dirichlet or Neumann b.c. All the new solutions bifurcate from a homogeneous fixed point when the parameters A, q, and D_X have values for which this homogeneous solution is unstable. Whether or not they appear in a "realizable" experiment is a matter of their stability. The stability analysis of the nonuniform solutions though is simple of conception yields a very difficult task, and at present lies outside the scope of this note. What we can safely say, however, is that all branches described in Figs. 4 and 6 that have $(\partial T/\partial V)_{V=0} < 0$ are unstable.[10,12]

Note added in proof: Shortly before receipt of the galley proofs we learned of the work by N. Chafee and E. F. Infante, Appl. An. **4**, 17 (1974), where a different though related problem is discussed. Following their method, the stability of the nonuniform states T_2, T_1+T_2, and $2T_2+T_1$ (see Figs. 4 and 6) can be decided with the use of the Lyapunov function

$$V(\phi)=\int_0^1\{\tfrac{1}{2}(\nabla\phi)^2-\int_0^{\phi(x)}f(\xi)\,d\xi\}\,dx,$$

in which ϕ describes an initial, though not merely infinitessimal disturbance upon the steady state whose stability is being tested, and

$$f(\xi)=\frac{1}{D_x}\left(A+(1-qA)\,\xi-\frac{A+(1-qA)\,\xi}{A+q(1-qA)\,\xi}\right).$$

Using $V(\phi)$, it is shown that the states T_1+T_2 and $2T_2+T_1$ are unstable, whereas the profile corresponding to T_2 is the only stable one. Computer runs confirmed the analytical results.

ACKNOWLEDGMENTS

The results reported in this note were obtained in the course of research sponsored by the Instituto de Estudios Nucleares (Spain). The authors acknowledge the cooperation and fruitful discussions with V. Fairén in the early stages of this work. The authors are also grateful to the anonymous referee for useful remarks and references that have been incorporated in the paper.

[1] J. L. Ibañez, V. Fairen, and M. G. Velarde, Phys. Letters A **58**, 364 (1976).
[2] J. L. Ibañez and M. G. Velarde, J. Non-equil. Thermodyn., in press.
[3] I. Balslev and H. Degn, J. Theoret. Biol. **49**, 173 (1975).
[4] R. Aris, *The Mathematical Theory of Diffusion and Reaction in Permeable Catalysts* (Clarendon, Oxford, 1976), Vol. I, Sec. 4.7.3, Vol. II, Secs. 7.9 and 8.8.
[5] J. Boa, Stud. Appl. Math. **54**, 9 (1975).
[6] See, for instance, G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Interscience-Wiley, New York, 1977).
[7] N. Kopell and L. N. Howard, Stud. Appl. Math. **52**, 291 (1973).
[8] See for a review R. S. Schechter, M. G. Velarde and J. K. Platten, Advan. Chem. Phys. **26**, 265 (1974).
[9] J. Auchmuty and G. Nicolis, Faraday Symp. Chem. Soc. **9**, 265 (1974).
[10] D. Luss and M. A. Ervin, Chem. Eng. Sci. **27**, 315 (1972).
[11] L. Meirovitch, *Methods of Analytical Dynamics* (McGraw-Hill, New York, 1970), Chap. VI.
[12] G. Gavalas, *Nonlinear Differential Equations of Chemically Reacting Systems* (Springer, New York, 1968), Sec. 2.9.

A characterization of Lorentz transformations

A. Lenard

Departments of Mathematics and Physics, Indiana University, Bloomington, Indiana 47401
(Received 26 April 1977)

If a one-to-one correspondence of Minkowski space–time onto itself is such that timelike lines, and only timelike lines, map onto timelike lines, then the correspondence is an inhomogeneous Lorentz transformation combined with a dilation.

In an early issue of this journal, Zeeman[1] gave a remarkable characterization of Lorentz transformations. Namely, if a one-to-one correspondence of Minkowski space—time onto itself is such that the relation of temporal precedence is preserved in both directions, then the mapping is necessarily an inhomogeneous Lorentz transformation combined with a dilation. The purpose of the present note is to point out that there is a characterization which is similar in spirit, one whose proof reduces in a very simple manner to two theorems of geometry. The first of these is the Theorem of Desargues. The second is the fact that every collineation in a real affine space is induced by a nonsingular inhomogeneous linear transformation in terms of coordinates.

We propose to characterize Lorentz transformations by the property that they map timelike lines into timelike lines. As in Zeeman's theorem, no hypothesis of continuity is needed, or indeed any other regularity assumption. Nor is it assumed that the order of the points on a timelike line is preserved, or otherwise related in any particular way, to the order of their images under the given mapping. Also, unlike Zeeman's theorem, ours is valid without the restriction that the dimension of space—time exceed 2.

Theorem: If a one-to-one correspondence of Minkowski space—time onto itself is such that it, as well as its inverse, maps timelike lines onto timelike lines, then the correspondence is an inhomogeneous Lorentz transformation combined with a dilation.

Consider three distinct points in space—time, say A, B, and C. Let t be a timelike line such that the given points do not lie on one line parallel to t. Suppose that six points X, Y, Z, X', Y', and Z' exist such that the triplets AYZ, BZX, CXY, as well as the triplets $AY'Z'$, $BZ'X'$, and $CX'Y'$ are collinear along six timelike lines not parallel with t, but the lines XX', YY', and ZZ' are parallel to t. It follows then from Desargues' theorem that the triplet ABC is collinear. Conversely, if that triplet is collinear, the rest of the points with the above properties can be found. Note that parallelism of two timelike lines can be characterized this way: They are parallel if and only if they are distinct, do not meet, and there are two more timelike lines meeting each other as well as the two given lines. These considerations show that collinearity of three points is a relation that can be defined entirely in terms of the weaker relation of collinearity *along timelike* lines. Suppose now that a one-to-one correspondence of Minkowski space—time onto itself, satisfying the hypotheses of the theorem, is given. Then, by what we just said, it and its inverse preserve the relation of collinearity of triplets of points in general (such mappings are called collineations in geometry). But then, according to a basic theorem in geometry[2] the mapping is necessarily affine (inhomogeneous linear) in terms of coordinates. Its homogeneous part preserves the sign of the Minkowski quadratic form. It follows immediately that this must be a nonzero multiple of a Lorentz transformation. This proves the theorem.

[1] E. C. Zeeman, J. Math. Phys. 5, 490 (1964).
[2] O. Veblen and J. W. Young, *Projective Geometry* (Blaisdell, New York, 1910), esp. Sec. 71 of Vol. I and Sec. 99 of Vol. II.

 0022-2488/78/1901-0157$1.00

Expanding shearfree spatially homogeneous universes with a nonsynchronous time coordinate and anisotropy of the universe

A. J. Fennelly

Department of Physics and Astronomy, Western Kentucky University, Bowling Green, Kentucky 42101 [a]
(Received 17 January 1977)

The apparent isotropy of the microwave background radiation on all angular scales and regions thus far observed is usually accepted as evidence of the isotropy of the universe. Current applications of the Einstein theory of gravitation in cosmology couples all rotation or peculiar velocity to anisotropy in the background radiation. Rotation in particular is coupled to shear. These are based on a number of results that apply to spatially homogeneous cosmologies with a synchronous time coordinate. While the models allow the past of our universe to be rather interesting, they constrain the present to be quite boring. We introduce a nonsynchronous time coordinate and show that the present could allow more interesting fluid motions, including rotation and peculiar velocity in spite of the microwave background's isotropy. We use the formalism to construct rotating Friedman universes and study observations in them. Almost all other cosmological data are consistent with a rotation of the universe.

1. INTRODUCTION

It is generally accepted that the universe we presently observe is highly isotropic on scales 200 Mpc or greater.[1] This is based mainly on the apparent isotropy of the microwave background radiation.[2] This interpretation of the observations rests heavily on cosmological models with a synchronous time coordinate and spatially homogeneous 3-spaces,[25] allowing fluid motions with nonzero shear, vorticity, and acceleration of the fluid motions.[3]

A number of theorems have been generated in the synchronous time anisotropic formalism coupling shear with vorticity, and evolution of anisotropy, and the allowed peculiar velocities.[4,5] Other theorems connect the Bianchi classifications of the homogeneous 3-spaces to the allowed fluid motions.[6] The theorems assume a synchronous time coordinate, which means that the existence of vorticity requires *a priori* the existence of a shear tensor with nondiagonal components. The $G_{\mu 0}$ components of the Einstein equations require that the shape matrix determining the metric anisotropy[4] not commute with its time derivative. These conditions are satisfied only if the shape matrix determining the metric anisotropy[4] not commute with its time derivative. These conditions are satisfied only if the shape matrix has nonzero nondiagonal components; hence the shear tensor also has nonzero nondiagonal components. These $G_{\mu 0} = -KT_{\mu 0}$ equations also require that the group structure constants $C^{\alpha}_{\beta\gamma}$ be nonzero in synchronous systems. Hence the connection of Bianchi types and allowed fluid motions follows.

The assumption of these theorems are very restrictive. Their conditions obviously strongly imply the results of the theorems. Only when the most general set of conditions has been investigated to the same result can it be said that the theorems are generally true. This enlargement of the physics is the introduction of nonsynchronous time, of metrics with nonzero shift between spatial hypersurfaces.[5] One also hopes to bypass the present belief that boring observations (isotropic background) imply boring fluid dynamics (isotropic expanding fluid). A further motivation are the possible large-scale inhomogeneity in the universe,[6,7] the distributions of quasars and of galaxy ellipticities and position angles,[8,9] and the anomalous slope of the number-count relation for radio sources.[10] These data are definitely not consistent with isotropy and their successful interpretation solely in such a context seems unlikely. We look for physical understanding that will fit all the data simultaneously and allow the universe to be a bit more interesting place to live in.

Glass[11] has derived an identity which supports these arguments. With vorticity vector ω^{α}, density ρ, projection tensor $h_{\alpha\beta} = g_{\alpha\beta} + u_{\alpha}u_{\beta}$, magnetic type gravitational field $H_{\alpha\beta}$ (defined below), and gravitational constant K, he finds

$$3\omega^{\alpha}H_{\alpha\beta} = \frac{K}{3}h_{\alpha\beta}\nabla^{\alpha}\rho\,, \tag{1}$$

where ∇^{α} is the gradient. In the metrics discussed below we have

$$ds^2 = -(dt - \Omega_{\mu}\omega^{\mu})^2 + \omega_{\mu}\omega_{\nu}\delta_{\mu\nu} \tag{2}$$

where $\omega_{\mu} = e^{\alpha}e^{\beta}_{\mu\nu}\sigma^{\nu}$ with α a function of time and β a symmetric traceless 3×3 matrix. The σ^{ν} are a basis of differential forms on 3-spaces of constant time and Ω_{μ} is the fluid peculiar velocity. Then in such frames $\nabla^{\alpha}\rho = (\partial\rho/\partial t)\Omega^{\alpha}$ and one finds

$$3\omega^{\alpha}H_{\alpha\beta} = \left(\frac{K}{3}\right)h_{\alpha\beta}\Omega^{\beta}(\partial\rho/\partial t). \tag{3}$$

Although there are homogeneous space sections[25] over which the σ^{ν} forms define a group of motions, the models are nevertheless inhomogeneous to observers as long as there is evolution with the time. It is possible to find perfect fluid models for $\Omega_{\beta}\alpha e^{-\alpha}$, $(\partial\rho/\partial t)\alpha e^{-4\alpha}$, $\omega_{\alpha}\alpha e^{-2\alpha}$, and $H_{\alpha\beta}\alpha e^{-3\alpha}$ so that Glass'[11] identity is consistent.

[a] Guest Summer Research Fellow, May—August 1975, Mathematics Department, University of London King's College, Strand WC2R 2LS, U. K.

 0022-2488/78/1901-0158$1.00

Hence shearfree cosmologies can have expansion and rotation.

In further confirmation of this discussion, Shikin[12] has shown that cosmological models with a synchronous metric ($\Omega_\mu = 0$) are not complete. They do not allow spacelike trajectories of matter. This point was discussed in my dissertation. Causality restrictions implied by a synchronous coordinate system *a priori* prohibit continuation of the fluids into spacelike regions. Hence the manifold created by Lie-dragging geodesics as generators of the spacetime is incomplete. But, as Shikin[12] shows, the complete system is inhomogeneous. His metric is still diagonal, but the time coordinate is not homogeneous. The same applies to the metrics in this paper as the transformation

$$du = dt - \Omega_\mu \omega^\mu \tag{4}$$

shows. The $\alpha = \alpha(t)$ and $\beta = \beta(t)$ become $\alpha = \alpha(u + \int \Omega_\mu \omega^\mu)$ and $\beta = \beta(u + \int \Omega_\mu \omega^\mu)$. The terms involving $\int \Omega_\mu \omega^\mu$ clearly will cause inhomogeneity. This is akin to the stationary inhomogeneous spacelike regions in the whimper singularities of Ellis and King[13] and Collins.[14] These metrics allow (unusual) complete spacetimes and the behavior implied by Glass' identity is therefore not surprising. The real choice of a complete metric is then between a nonsynchronous spatially homogeneous system or a synchronous inhomogeneous one.

2. METRIC AND FIELD EQUATIONS

We express an isotropic Friedman metric in null form decomposed into a set of null tetrad vectors[15] $g_{ab} = 2l_{(a}n_{b)} - 2m_{(a}\overline{m}_{b)}$, where l_a and n_a are tangent to a null geodesic and m_a is the complex vector spanning the celestial sphere. Then l_a is tilted to have components in the m_a directions. The full set of tetrad vectors is transformed accordingly, the tilted metric composed from them, and then returned to the time/space form

$$ds^2 = -(dt - \Omega_\mu \omega^\mu)^2 + \omega^\mu \omega^\nu, \tag{5}$$

where $\omega^\mu = e^\alpha \sigma^\nu$, Ω_μ is the peculiar velocity, $\alpha = \alpha(t)$, $\Omega_\mu = \Omega_\mu(t, X^\mu)$, and the σ^μ are basis forms obeying the curl relations $d\sigma^\mu = \frac{1}{2} C^\mu_{\nu\rho} \sigma^\nu \wedge \sigma^\rho$ with $C^\mu_{\nu\rho}$ the structure constants of their isometry group. The vorticity $\omega_{\mu\nu}$, acceleration u_μ, and expansion $\theta_{\mu\nu}$ are $\omega_{\mu\nu} = \dot{\Omega}_{[\mu}\Omega_{\nu]} - \Omega_K e^{-\alpha} C^K_{\mu\nu}$, $a_\mu = -(\dot{\Omega}_\mu + \Omega_\mu \dot{\alpha})$, $\theta_{\mu\nu} = \dot{\alpha}\delta_{\mu\nu}$). (There is no shear).

The affine connection is $\Gamma^\mu_{00} = -\Gamma^0_{0\mu} = -a_\mu$, $\Gamma^\mu_{\lambda 0} = -\Gamma^0_{\lambda\mu} = \theta_{\mu\nu} + \omega_{\mu\nu}$, $\Gamma^\lambda_{0\lambda} = -\Gamma^\lambda_{0\mu} = \omega_{\mu\lambda}$, and $\Gamma^\mu_{\rho\lambda} = -\Gamma^\lambda_{\rho\mu} = -e^{-\alpha} C^\rho_{\lambda\mu}$. We accept that the Friedman models are correct to zero order and treat $\omega_{\mu\nu}$ and u_μ as first-order quantities. neglecting all terms of higher order in Einstein's equations.[3] Under these conditions the field equations are

$$3\dot{\alpha}^2 + \tfrac{1}{2}R^* = K\rho, \tag{6a}$$

$$-2C^\rho_{\lambda\rho}\omega_{\mu\nu}e^{-\alpha} + e^{-\alpha}C^\rho_{\mu\lambda}\omega_{\rho\lambda} + 2\dot{\alpha}a_\mu = 0, \tag{6b}$$

$$-6\ddot{\alpha} - 9\dot{\alpha}^2 - \tfrac{1}{2}R^* = 3K\gamma\rho, \tag{6c}$$

$$\dot{\omega}_{\mu\nu} + 2\dot{\alpha}\omega_{\mu\nu} + a_\lambda e^{-\alpha}(C^\mu_{\lambda\nu} - \tfrac{1}{3}\delta_{\mu\nu}C^\rho_{\rho\lambda}) = 0, \tag{6d}$$

with R^* = the isotropic curvature of the space sections for the Bianchi types chosen, $I: C^\mu_{\nu\rho} = 0$, $V: C^2_{[21]} = C^3_{[31]} = 1$, $IX: C^\mu_{\nu\rho} = \epsilon^\mu_{\gamma\rho}$. With equations of state $p = \gamma\rho$ the conservation equations are

$$\dot{\rho} = -3(1+\gamma)\rho\dot{\alpha}, \tag{7a}$$

$$a_a = -(\delta^b_a + u_a u^b)\frac{p_{1b}}{(\rho + p)}, \tag{7b}$$

$$a_\mu = -3\dot{\alpha}\gamma u_\mu. \tag{7c}$$

The geodesic equations split into the zero-order (unperturbed) set for q^a and the correction set for the correction factors δ^a (then $\overline{q}^a = q^a + \delta^a$),

$$\frac{d\delta^0}{dt} - (\dot{\Omega}_\lambda + \Omega_\lambda\dot{\alpha})q^\lambda + \dot{\alpha}\delta_{\mu\lambda}\frac{(q^\mu\delta^\lambda + q^\lambda\delta^\mu)}{q^0} + \frac{\omega_{\mu\nu}q^\mu q^\nu}{q^0} - \dot{\alpha}\delta_{\mu\lambda}\frac{q^\mu q^\lambda \delta^0}{q_0^2} = 0, \tag{8a}$$

$$\frac{d\delta^\mu}{dt} + (\dot{\Omega}_\mu + \Omega_\mu\dot{\alpha})q^0 + \dot{\alpha}\delta_{\mu\nu}\delta^\nu + 2\omega_{\mu\nu}q^\nu - e^{-\alpha}C^\nu_{\rho\mu}\frac{q^\nu\delta^\rho + q^\rho\delta^\nu}{q^0} - \frac{q^\nu q^\rho \delta^0}{q_0^2} = 0. \tag{8b}$$

The perturbation of the expansion factor in the metric ϵ is described by terms in Eqs. (6a) and (6c) with the usual Friedman parts of those equations assumed valid [we use $R^* = Ke^{-2\alpha}e^{-2\epsilon} \approx Ke^{-2\alpha}(1 + 2\epsilon)$]

$$\dot{\epsilon}(6\dot{\alpha}) + R^* = 0, \tag{9a}$$

$$-6\ddot{\epsilon} - 18\dot{\alpha}\dot{\epsilon} - 2\dot{u}_\lambda e^{-\alpha}C^\mu_{\mu\lambda} - \epsilon R^*(\epsilon) = 0. \tag{9b}$$

For the unperturbed Friedman models with dustlike matter we have the following. For type I, $e^\alpha = t^{2/3}$, $q^i = \text{const}$. For type IX, $e^\alpha = \alpha_0 \sin^2\tau/4$, $\dot{\tau} = e^{-\alpha}$, $t = \frac{1}{2}e^{\alpha_0}[\tau - 2\sin\tau/2]$, $q^c = \text{const}$. In type V, $e^\alpha = (8\pi M/3) \times \sinh^2\tau/2$, $\dot{\tau} = e^{-\alpha}$, $M = \mu_R[3\dot{\alpha}_R^2 - 8\pi\rho_R)/3]^{-3/2}$, $t = (e^{\alpha_0}/2) \times (8\pi M/3)[\tau - 2\sinh(\tau/2)]$, $q^1 = \cos\theta$, $q^2 = \sin\theta\cos\varphi$; $q^3 = \sin\theta\sin\varphi$, $\cot\theta/2 = Ke^\tau$, and $\varphi = \text{const}$. The perturbed solutions are as follows.

For type I, $\epsilon = \text{const}$, $\omega_{\mu\nu} = 0$, $\dot{u}_\mu = 0$, and hence Ω_μ $\Omega^0_\mu \exp[-(\alpha - \alpha_0)]$. The space curvatures are the same in all directions. Equation (7b) gives a necessary counterexample. With $v_\mu = \Omega_\mu$, $v_0 = -1$, $v^0 = 1$, $g_{0\mu} = \Omega_\mu e^\alpha$,

$$\frac{\partial v_\mu}{\partial t} - \dot{\alpha}v_\mu e^{-\alpha} = -\frac{v_\mu p_{10}}{(\rho + p)}. \tag{10}$$

In type I with nonvanishing pressure this solves to $v_\mu = v^0_\mu(t/t_0)^{3\gamma} \exp[-t^{2/3(1+\gamma)}]/(K\rho_0)^{1/2}$. Hence neither the acceleration nor the peculiar velocity vanish in type I.[25]

For type IX

$$\epsilon = \epsilon_0 \exp\{[\tan(\tau/4)]^{-1/3}\}, \tag{11a}$$

$$\Omega_\mu = \Omega^0_\mu e^{-\alpha}(\tan\tau/4)^{-1/4}, \tag{11b}$$

$$\delta^0 = \delta^0_0 + \tfrac{1}{2}\Omega^0_\mu q^\mu \chi \ln\left[\frac{1-\alpha}{1-\alpha_0}\right], \tag{12a}$$

and

$$\delta^\mu = \delta^\mu_0 + \tfrac{1}{2}\Omega^0_\mu q^0 \chi \ln\left[\frac{1+\alpha}{1+\alpha_0}\right], \tag{12b}$$

where $\chi = e^\alpha + e^{-\alpha}$. For type V

$$\epsilon = \epsilon_0\{\exp[\tan(\tau/4)]^{1/3}\}, \tag{13a}$$

$$\Omega_1 = \Omega^0_1 e^{-\alpha}, \tag{13b}$$

$$\Omega_A = \Omega^0_A e^{-\alpha}[\tan(\tau/4)]^{-3/4}, \tag{13c}$$

$$\delta^0 = \delta_0^0 + \tfrac{3}{2}\chi(\Omega_2^0 q^2 + \Omega_3^0 q^3)(v - v_0), \tag{14a}$$

$$\delta^1 = \delta_0^1 + 2\chi(\Omega_2^0 q^2 + \Omega_3^0 q^3)(u - u_0), \tag{14b}$$

$$\delta^A = \delta_0^A + \Omega_A^0 \chi\left(\frac{3q^0}{2} - 2q^1\right)(u - u_0), \tag{14c}$$

where $v - v_0 = \log[(1-\alpha)/(1-\alpha_0)]^e + e(\alpha - \alpha_0)$ and $u - u_0 = \log[(1+\alpha)/(1+\alpha_0)]^e - e(\alpha - \alpha_0)$ (e is the base of the natural logs).

3. OBSERVATIONS. I. THE NUMBER COUNTS AND REDSHIFTS

We first discuss redshifts. The relevant equation from Kristian and Sachs,[16] appropriately truncated, is

$$z = \frac{\theta r}{3} + \frac{r^2}{2}(v + v_\mu e^\mu), \tag{15a}$$

$$v = \frac{\rho}{6} + \frac{\theta^2}{3}, \tag{15b}$$

$$v_\mu = \tfrac{2}{5}\theta_{|\mu} + \tfrac{1}{5}(\theta\delta_{\mu\gamma} + \omega_{\mu\gamma})_{|\gamma}. \tag{15c}$$

Then

$$z = \frac{\theta r}{3} + \frac{r^2}{2}\left[\frac{\rho}{6} + \frac{\theta^2}{3} + \frac{3}{5}\dot{\theta}\Omega_\mu e^\mu\right] \tag{16a}$$

in type I, and

$$z = \frac{\theta r}{3} + \frac{r^2}{2}\left[\frac{\rho}{6} + \frac{\theta^2}{3} + \left\{\frac{3}{5}\dot{\theta}\Omega_\mu + \frac{e^{-\alpha}}{5}(C^\gamma_{\gamma\delta}\omega_{\mu\delta} + C^\gamma_{\mu\delta}\omega_{\delta\gamma})\right\}e^\mu\right] \tag{16b}$$

in type V, and

$$z = \frac{\theta r}{3} + \frac{r^2}{2}\left[\frac{\rho}{6} + \frac{\theta^2}{3} + \left\{\frac{3}{5}\dot{\theta}\Omega_\mu + \frac{e^{-\alpha}}{5}C^\gamma_{\mu\delta}\omega_{\delta\gamma}\right\}e^\mu\right] \tag{16c}$$

in type IX. There will be an affect on the number counts of radio sources due to rotation alone. Since the velocity vector is not orthogonal to the homogeneous spatial hypersurfaces, an observer's rest space will be tilted, slicing through spatial hypersurfaces of the time coordinate t.[5] Therefore, the light cones will be tilted. We therefore find two rotation-induced effects on the number counts.

First, because geodesic path are skewed, the luminosity distance increases as a power of the intrinsic luminosity whose power is also dependent on the luminosity distance itself. We can see this for a purely rotating model since the rotational contribution makes the difference in the counts. Using Kristian and Sachs results,[16] we have

$$dN = \left[1 + \frac{r^2}{2}\left(\frac{n}{6} - \frac{\omega^2}{3}\right)\right]\left\{n + \frac{r^2}{2}(n^2 - \omega^2 n)\right\}r^2 dr dO \tag{17}$$

where dO is the element of angular measure on the celestial sphere and n is the intrinsic number density. In a Gödel universe[17] n and ω are constants, and we find

$$N = \frac{4\pi}{3}nr^3 + \left(\frac{7}{12}n - \frac{2\omega^2}{3}\right)\frac{4\pi}{5}nr^5. \tag{18}$$

Since there is no focussing, the received flux S is related to the intrinsic power of source P by $S = P/4\pi r^2$ and Eq. (18) becomes

$$N = \frac{nP^{3/2}}{3(4\pi)^{1/2}}S^{-3/2} + \frac{nP^{5/2}}{60(4\pi)^{1/3}}(7n - 8\omega^2)S^{-5/2} \tag{19}$$

in a form similar to the usual $N = N_0(S^{-\beta}/S_0)$ for the integral number counts. If the Eq. (19) is constrained to that form, then obviously $1.5 \leq \beta \leq 2.5$.

Second, an apparent spatial density gradient will exist,[4] leading to an anisotropy in the number counts (on the celestial sphere). This depends on the fluid motions. If the velocity is tilted with respect to the unit normal to the spatial hypersurfaces and a quantity changes through the evolution of the hypersurfaces (i.e., has a nonvanishing time derivative, where the time labels succeeding hypersurfaces), then that quantity will possess a nonvanishing spatial gradient in the 3-surfaces orthogonal to the fluid velocity (in the observer's rest space). The expression for this is

$$f_{1\mu} = f_{10}v^0 v_\mu = f_{1t}\Omega_\mu \tag{20}$$

so that the direction of the apparent gradient is in the direction of Ω_μ. The orientation of the anisotropy is therefore in the Ω_μ direction and its relation to ω_μ (rotation vector) varies as Ω_μ does. The gradient can thus be parallel to or orthogonal to the rotation axis.

Gödel[18] gives a result for a class of rotating models To first order, integrated over two hemispheres of the celestial sphere split by a plane orthogonal to ω^a, the difference in number of galaxies from one hemisphere to the other is

$$\frac{|N_1 - N_2|}{N_1 + N_2} = \frac{9}{8}\frac{|\omega| r e^\alpha \dot{\alpha}}{C^2}. \tag{21}$$

A similar result has been given by Batakis[19] and Fennelly.[20] In all cases r is the radial distance to the limits of the integration.

For the number counts our main formula is

$$dN = (1+z)r^2 dr dO\{n + r(n\theta + n_{1\mu}e^\mu)\}, \tag{22}$$

where dO is the element of angular measure on the celestial sphere. With the formulas given above, Eq. (22) reads (n is the number density)

$$dN = \left[1 + \frac{\theta r}{3} + \frac{r^2}{2}\left\{\frac{\rho}{6} + \frac{\theta^2}{3} + \left(\frac{3\dot{\theta}}{5}\Omega_\mu + \frac{e^{-\alpha}}{5}C^\gamma_{\gamma\delta}\omega_{\mu\delta} + \frac{e^{-\alpha}}{5}C^\gamma_{\mu\delta}\omega_{\delta\gamma}\right)e^\mu\right\}\right] \times [n + r(n\theta + \dot{n}\Omega_\mu e^\mu)]r^2 dr dO. \tag{23}$$

The different group types alter Eq. (23) as in Eqs. (16a)—(16c). In type I, the redshift distance relation reads

$$z = \frac{\theta r}{3} + \frac{r^2}{2}\left(\frac{\rho}{6} + \frac{\theta^2}{3}\right) + \frac{3r^2}{10}\dot{\theta}\Omega_\mu e^\mu. \tag{24}$$

The term second order in the distance containing $\dot{\theta}\Omega_\mu e^\mu$ will vary with the angle on the celestial sphere. However, the effect is second order in the luminosity distance and so will show only at large distances. The anisotropy would be dipole in nature. Inserting the solutions the anisotropy term is $0.9\ddot{\alpha}\Omega^0 t_0^{2/3}\cos\delta t^{-8/3}r^2$ which for $e^{-\alpha} \propto t^{-2/3}$ for dust, is $-0.9\Omega^0 t_0^{2/3}\cos\delta t^{-8/3}r^2$ with δ the angle between e^μ and Ω_μ. This may be rewritten in more common astronomical notation[21] $0.9\Omega^0(q_0 - 1)h^2 r^2$

$\times\exp(\alpha_0-\alpha)\cos\delta$, where q_0 is the deceleration parameter and h is the Hubble parameter. Since q_0 may not differ significantly from zero,[22,23] we may write, the large-scale anisotropy of the microwave background as $-1.8(h^2r^3)\exp(\alpha_0-\alpha)\Omega^0$. Observations put this number at 0.1%.[2] If the radiation were last scattered at a redshift of 1,000, then $\Omega^0\sim10^{-3}$ and $\Omega_\mu\sim10^{-1}$ (at $z=1{,}000$).

The number-count relation becomes

$$dN=\left[1+\frac{\theta r}{3}+\frac{r^2}{2}\left(\frac{\rho}{6}+\frac{\theta^2}{3}+\frac{3\theta}{5}\Omega\cos\delta\right)\right]\times[n+r(n\theta+\dot{n}\Omega\cos\delta)]r^2\,dr\,dO. \tag{25}$$

The anisotropy term, after neglecting all terms to order r^3 or higher, is

$$dN=\left[\dot{n}r+\left(\frac{\dot{n}\theta}{3}+\frac{3\dot{\theta}n}{10}\right)r^2\Omega\cos\delta\right]r^2drdO. \tag{26}$$

We must be careful of the terms involving $\dot{n}$: The number density can change because of volume expansion. It can also change because of evolution of the sources in both luminosity (cutoffs in survey used) and actual number of sources. A recent derivation[24] has no mention of this. Let n be of the form $n_0e^{-3\alpha}e^{-\lambda t}t^m$, where λ and m are positive numbers describing the evolution. Since, as we look backward in time along the lightcone $r\alpha-t$, for small volumes we may write n as $n_0\exp(-3\alpha+\lambda r)/r^{+m}$ in Eq. (26), and Eq. (26) reads

$$dN=\left[-\theta r+\left(\frac{3\mathring{\theta}}{10}-\frac{\theta^2}{3}\right)r^2-(\lambda+mr^{-1})r-\frac{\theta}{3}(\lambda+mr^{-1})r^2\right]r^2\,dr\,dO. \tag{27}$$

Equation (27) integrates to give the following relation:

$$|N_1-N_2|(N_1+N_2)^{-1}=\tfrac{3}{8}\Omega r\left(\theta+\lambda+\frac{m}{r}\right)+\left(\frac{\theta^2}{8}+\theta\left[\lambda+\frac{m}{r}\right]-\frac{9\mathring{\theta}}{80}\right)\Omega r^2. \tag{28}$$

With strong enough evolution we can now set Ω at its value of $\Omega^0\sim10^{-3}$ found for the microwave background and allow the rest of any anisotropy to be caused by a suitably chosen evolution parameter. Note also that the insertion of the evolution term in the isotropic parts of the number-count relation will change the slope of the relation, as discussed previously, to zeroth order.

The redshift distance relation reads from Eq. (16b) in type V,

$$z=\frac{\theta r}{3}+\frac{r^2}{2}\left(\frac{\rho}{6}+\frac{\theta^2}{3}\right)+r^2\left(\frac{3\dot{\theta}\Omega_\mu e^\mu}{10}+\frac{2e^{-\alpha}}{5}\omega_{\mu1}e^\mu\right), \tag{29}$$

and from Eq. (16c) in type IX,

$$z=\frac{\theta r}{3}+\frac{r^2}{2}\left(\frac{\rho}{6}+\frac{\theta^2}{3}\right)+r^2\left(\frac{3\dot{\theta}\Omega_\mu e^\mu}{10}+\frac{e^{-\alpha}}{5}\epsilon_{\delta\gamma\mu}\omega_{\delta\gamma}e^\mu\right). \tag{30}$$

The number-count relation is altered in each case. The anisotropy term is

$$dN=\left\{\left[\dot{n}r+\left(\frac{\dot{n}\theta}{3}+\frac{3\dot{\theta}n}{10}\right)r^2\right]\Omega_\mu e^\mu+\frac{2ne^{-\alpha}}{5}\omega_{\mu1}e^\mu\right\}r^2drdO \tag{31}$$

in type V, which becomes

$$dN=\left\{\left[\dot{n}r+\left(\frac{\dot{n}\theta}{3}+\frac{3\dot{\theta}n}{10}\right)r^2\right]\Omega+\frac{2ne^{-\alpha}}{5}\omega\right\}\cos\delta\,r^2drdO. \tag{32}$$

We find for type IX that the number-count anisotropy term is the same to the first order.

From the redshift relation we find limits on Ω_μ similar similar to those for type I. The results are similar to the number counts, and evolution is required as discussed above for type I.

4. OBSERVATIONS. II. DISTORTION AND PROPER MOTION

There are the Kristian–Sachs distortion and proper motion effects.[16] The distortion may have been measured[9] and the proper motion could contribute a transverse component to the quasar redshifts. The relevant equations are

$$e=1+\tfrac{2}{3}\max(m^\mu m^\nu e^\gamma\epsilon_{\gamma\mu\beta}H_{\beta\nu}) \tag{33a}$$

for the distortion of the ellipticity e and

$$\frac{de^\mu}{dt}=h^{\mu\nu}\{e^\beta\omega_{\nu\beta}+re^\beta\omega_{\gamma\beta}u_{\nu||\gamma}+\tfrac{1}{2}e^\beta e^\gamma\epsilon_{\beta\nu\lambda}H_{\gamma\lambda}\}, \tag{33b}$$

where m^μ spans the celestial sphere, $H_{\beta\nu}$ is the gravimagnetic field,[3,16] e^μ is the direction of the observer's telescope in his local cartesian axes, and a different projection operator, $h^{\mu\nu}=\delta^{\mu\nu}-e^\mu e^\nu$, is used. We can estimate the size of $H_{\mu\nu}$ to first order from equations derived from the Bianchi identity[3,11]:

$$H^{||b}_{ab}=(\rho+p)\omega_a, \tag{34a}$$

$$H_{ab||c}u^c=-\theta H_{ab}, \tag{34b}$$

and the constraint equation

$$H_{ab}=-h^f_{(a}n_{b)cde}u^c\omega^{d||e}_f. \tag{34c}$$

Using the metric and equations as developed for the Bianchi I, V, and IX models above, we may determine $H_{\mu\nu}$ in each case. In the ω^μ basis of Eq. (1), Eq. (34b) reads $H_{ab||0}=-\theta H_{ab}$ and Eq. (34c) reads $H_{ab}=-\delta^f_{(a}n_{b)0de}\omega^{d||e}_f$.

In type I, $H_{ab}=H^0_{ab}e^{-\alpha}$ because $H_{ab||0}=-\theta H_{ab}$ implies

$$\dot{H}_{ab}+\mathring{\alpha}H_{ab}=\omega^c_bH_{ac} \tag{35}$$

and the rotation vanishes in these models for type I so H_{ab} only exists if there is gravitational radiation.

In types V and IX we find [from Eq. (34c)]

$$H_{11}=-H_{33}=\omega_{12||3},\quad H_{12}=\omega_{23||2},$$

$$H_{32}=-\omega_{21||2},\quad H_{22}=H_{13}=0.$$

For initially dust models we may use the unperturbed values for e^α, $\mathring{\alpha}$, and ρ and then use $\omega_{\mu\nu}=\mathring{\Omega}_{[\mu}\Omega_{\nu]}$

$-\Omega_\kappa e^{-\alpha} C^\kappa_{\mu\nu}$ for ω to determine the H_{ab} components. The driving term $H_{ac}\omega^c_b$ in Eq. (35) is second order. To first order only gravitational radiation $H_{ab} = H^0_{ab} e^{-\alpha}$ and a rotation-generated induction field will appear in types V and IX. In type V we have

$$H_{12} = \omega_{13} e^{-\alpha}, \tag{36}$$

and in type IX

$$H_{12} = e^{-\alpha}\omega_{21}, \tag{37a}$$

$$H_{23} = e^{-\alpha}\omega_{23} \tag{37b}$$

for the induction fields.

The distortion is then maximized at

$$e = 1 + \tfrac{2}{3}\{\max[e^3 H_{12}(m^{1^2} - m^{2^2})]\} \tag{38}$$

in type V, and at

$$e = 1 + \tfrac{2}{3}\{\max[e^3 H_{12}(m^{1^2} = m^{2^2}) + e^1 H_{23}(m^{2^2} - m^{3^2})]\} \tag{39}$$

in type IX.

In type I there is no distortion as the induction field $H_{ab} = 0$.

In types V and IX there is distortion. Assume that the major axis of the ellipsoidal image is along m^1, then Eq. (38) for type V is

$$e = 1 + \tfrac{2}{3} e^3 H_{12}(m^{1^2} - m^{2^2}) \tag{40}$$

and in the direction of maximum distortion, e^3, we have $e = 1 + \frac{2}{3} H_{12}\delta$ where δ is the difference $(m^{1^2} - m^{2^2})$. In type IX this is

$$e = 1 + \tfrac{2}{3}[e^3 H_{12}\delta + e^1 H_{23}\epsilon], \tag{41}$$

where ϵ is the difference $(m^{2^2} - m^{3^2})$. Insertion of the solutions in Sec. 2 gives 10^{-11}/yr for the ellipticity change in any elliptical segment cut from the microwave background temperature ellipsoid. Detecting this seems much too difficult with present technology. This distribution of ellipticities of distant galaxies photographed in a given field changes approximately as $e - e_0 = \frac{2}{3}\Omega^0 e^{-3\alpha}$. If emission is at redshifts of about 0.25, 0.5, 0.75, and 1, then respectively the values of Δe are 0.2×10^{-3}, 0.8×10^{-3}, 2.4×10^{-3}, and 9.6×10^{-3}. These are not extreme changes. The effect has been claimed to be detected by a few investigators,[9] but the matter is still not concluded one way or the other.[3]

To first order in dynamical quantities, the proper-motion equation reads

$$\frac{de^\mu}{dt} = \omega_{\mu\beta} e^\beta + r(\tfrac{1}{3}\theta\delta_{\mu\gamma}\omega_{\gamma\beta}e^\beta + \tfrac{1}{2}e^\beta e^\gamma \epsilon_{\beta\mu\lambda} H_{\gamma\lambda}) - e^\mu r\left(\frac{\theta}{3} e^\beta \omega_{\gamma\beta} e^\gamma + \tfrac{1}{2} e^\beta e^\gamma e^\nu \epsilon_{\beta\gamma\lambda} H_{\gamma\lambda}\right). \tag{42}$$

The terms in $H_{\gamma\lambda}$, when there is only the induction field, are an order of magnitude lower than the other terms, and as they will involve terms containing the square and the cube of the e^μ components, we will neglect them in hopes of simplifying the equations. Section 2 suggests that $e^\beta \omega_{\gamma\beta} e^\gamma$ will angle-average to $(\omega_{\gamma\beta}\omega^{\gamma\beta})^{1/2}$ which is $\Omega e^{-\alpha} = (\Omega_\mu \Omega^\mu)^{1/2} e^{-\alpha}$, and Eq. (42) becomes

$$\frac{d}{dt}\begin{pmatrix} e^1 \\ e^2 \\ e^3 \end{pmatrix} = \begin{pmatrix} -r\dot{\alpha}\Omega e^{-\alpha} & x\omega_{12} & x\omega_{13} \\ -x\omega_{12} & -r\dot{\alpha}\Omega e^{-\alpha} & x\omega_{23} \\ -x\omega_{13} & -x\omega_{23} & -r\dot{\alpha}\Omega e^{-\alpha} \end{pmatrix}\begin{pmatrix} e^1 \\ e^2 \\ e^3 \end{pmatrix} \tag{43}$$

with $x = (1 + r\dot{\alpha})$. In type I the proper motion de_μ/dt vanishes and $\hat{e} = \hat{e}_0$.

Taking e^1 as a typical component of the position vector, the components precess according to Eq. (39) as

$$\frac{de^1}{dt} = e^1_0 r\ \dot{\alpha}\ \Omega e^{-\alpha} + [1 + r\ \dot{\alpha}][\omega_{12} e^2_0 + \omega_{13} e^3_0] \tag{44}$$

in both types V and IX. For the proper-motion effect, solve Eq. (44) by inserting the results of Sec. 2 in it, and then differentiate to give the precession rate in observer time t; this gives a precession of 10^{-5} rad/yr for the microwave background temperature ellipsoid and 10^{-8} rad/yr for the quasars of their distances are indicated by their redshift. These are presently not detectable.

5. CONCLUSIONS

We have exhibited and examined a set of cosmological models based on a formalism which does not *a priori* couple the fluid motions of shear, vorticity, and peculiar velocity in a nonsynchronous system that allows an inhomogeneity like Shikin's[12] complete manifold models.[25] From this we were able to obtain tilted shear-free first-order models of the universe with expansion, peculiar motion, and vorticity in Bianchi types V and IX, and a tilted model with expansion and peculiar velocity in type I in confirmation of Glass' identity.[11] None of the models seriously conflicts with any observations. In fact, they are all consistent with the microwave background and also with the number-count, redshift, proper-motion, and distortion data. We have therefore exhibited what may be called rotating Friedman universes. The peculiar velocity in the models was limited to $\Omega_\mu < 10^{-3}$. The limits on ω in type V and IX models are 10^{-9}. The ratio $\omega/\dot{\alpha}$ is then 10^{-5}, greater than that previously found.[3,5] These observations need to be considered in more detail, with shear of course included, if fully realistic models are to be examined in light of the data.

A heuristic discussion showed that rotation could seriously affect the number-count relation, both in shape and isotropy. Subsequent calculation confirmed the heuristic argument.

It is not claimed that any model shown here presents the real universe, only that such models exist in general relativity. Realistic models probably must exhibit shear and the spin of an observer's Fermi-transported reference axis (the dragging of inertial

frames). The formalism developed here will next be applied in those directions.

ACKNOWLEDGMENTS

I thank my mentors Graham Frye, Arthur Komar, David Finkelstein, Remo Ruffini, and E. L. Schüking for their advice and guidance. Portions of this are contained in my doctoral dissertation, or were presented at meetings of the A. P. S. The encouragement—editing and typing—of my wife is most appreciated.

[1]C. B. Collins and S. W. Hawking, Astrophys. J. **180**, 317 (1973).
[2]R. B. Partridge and D. T. Wilkinson, Phys. Rev. Lett. **18**, 557 (1967).
[3]S. W. Hawking, Astrophys. J. **145**, 544 (1966); Mon. Not. R. Astron. Soc. **142**, 129 (1969).
[4]A. R. King and G. F. R. Ellis, Commun. Math. Phys. **31**, 209 (1973).
[5]C. B. Collins and S. W. Hawking, Mon. Not. R. Astron. Soc. **162**, 307 (1973); I. Osvåth, J. Math. Phys. **12**, 1078 (1971); R. A. Matzner, Ann. Phys. (N. Y.) **65**, 438 (1971).
[6]Refs. 4 and 5, and J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968); T. E. Perko, R. A. Matzner, and L. C. Shepley, Phys. Rev. D **6**, 969 (1972).
[7]P. J. E. Peebles, Astrophys. J. **183**, 1081 (1973).
[8]G. K. Miley, Mon. Not. R. Astron. Soc. **152**, 477 (1971).
[9]M. Rheinhardt, Mon. Not. R. Astron. Soc. **156**, 151 (1972).
[10]I. J. K. Pauling-Toth and K. J. Kellermann, Astron. J. **77**, 797 (1972).
[11]E. M. Glass, J. Math. Phys. **16**, 2361 (1975).
[12]I. S. Shikin, Zh. Eksp. Teor. Fiz. **68**, 1583 (1975) [Sov. Phys. JETP **41**, 794 (1975)].
[13]G. F. R. Ellis and A. R. King, Commun. Math. Phys. **38**, 119 (1974).
[14]C. B. Collins, Commun. Math. Phys. **39**, 131 (1974).
[15]S. W. Hawking, J. Math. Phys. **9**, 598 (1968).
[16]J. Kristian and R. K. Sachs, Astrophys. J. **143**, 379 (1966).
[17]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[18]K. Gödel, Proc. Int. Cong. Math. (6th) **1**, 175 (1950).
[19]N. Batakis, thesis, City College of C. U. N. Y. (1974).
[20]A. J. Fennelly, thesis, Belfer Graduate School of Science, Yeshiva University (1975); B. A. P. S. **19**, (1974); A. J. Fennelly and G. Frye, B. A. P. S. **19**, 555A (1974).
[21]A. Sandage, Astrophys. J. **133**, 355 (1961).
[22]A. Sandage and G. A. Tammann, Astrophys. J. **197**, 265 (1975).
[23]A. Sandage, G. A. Tammann, and E. Hardy, Astrophys. J. **172**, 253 (1972).
[24]N. Batakis and J. M. Cohen, Phys. Rev. D **12**, 1544 (1975).
[25]*Note added in proof*: In the models introduced in this paper, "spatially homogeneous" or "homogeneous" spaces should be understood to be homogeneous submanifolds orthogonal to the fluid 4-velocity, not 3-spaces labeled by the time coordinate. The group acts on these submanifolds and the models are inhomogeneous.

A new approach to the eigenvalues of the Gel'fand invariants for the unitary, orthogonal, and symplectic groups

S. A. Edwards

Department of Mathematical Physics, University of Adelaide, Adelaide, S. Australia 5001
(Received 10 May 1977)

The expressions for the Gel'fand invariants (Casimir operators) of U(n), O(n), and Sp(n) in terms of the IR labels are derived by relating them to the *trace* of a suitably defined operator P^k. The method has unity and simplicity, since trace (P^k) can be related directly to the IR labels using the Weyl dimension function—without having to determine the eigenvectors of P.

1. INTRODUCTION

The Gel'fand invariants of the groups U(n), O(n), and Sp($2h$) have eigenvalues which can be expressed as functions of the irreducible representation labels in a simple and elegant way.[1–4] Two distinct methods for deriving these functions have evolved: One, due to Louck and Biedenharn, has been applied only to U(n)[1,2,4–6]; the other, first enunciated by Perelomov and Popov, has been applied in addition to the orthogonal and symplectic groups.[3,7–10] Both methods, however, proceed by a somewhat complicated path to expressions comparatively simple in form. In this paper an alternative technique is presented for deriving the expressions, one which (it is felt) preserves throughout the simplicity apparent in the final results. The framework upon which this technique is built is provided essentially by Green's work[11,12,4] on characteristic identities for the generators of U(n), O(n), and Sp($2h$), and particularly as modified by Hannabuss[13] and O'Brien, Cant, and Carey.[14]

In Sec. 2, we derive the form of the expression for the Gel'fand invariants common to all the groups, and in Sec. 3, we use this to derive the particular formulas that appear in Refs. 1–4, 6 and 10. Appendices A and B contain some details of the calculation.

2. GENERAL RESULTS

Our procedure is to set up a matrix P the *trace* of whose kth power is directly related to the eigenvalue of the kth-degree Gel'fand invariant. The eigenvalues of P and their multiplicities are simple to evaluate, and from them we can immediately write down an expression for $\mathrm{tr}P^k$. As distinct from Perelemov and Popov's method, it is not necessary to know the eigenvectors of P, either right or left.

The summation convention for contraction between upper and lower indices will be used throughout. The generators of U(n) satisfy the commutation relations

$$[E^i_j, E^k_l] = \delta^k_j E^i_l - \delta^i_l E^k_j, \quad i,j,k,l = 1,\ldots,n. \tag{1}$$

To define the symplectic and orthogonal groups, we introduce a metric tensor g_{ij}; no generality is lost in defining it explicitly by

$$g_{ij} = \begin{cases} \delta_{ij} & \text{for O}(n), \quad i,j = 1,\ldots,n, \\ \delta_{i\,j+h} - \delta_{i+h\,j} & \text{for Sp}(2h), \quad i,j = 1,\ldots,n = 2h. \end{cases} \tag{2}$$

The group action on vectors v^i and w^i is required in each case to leave the form $g_{ij}v^iw^j$ invariant. For n even ($=2h$), the contravariant form of the metric tensor for Sp(n) exists and is

$$g^{ij} = \delta^{j\,i+h} - \delta^{j+h\,i}. \tag{3}$$

For convenience we define the quantity η by

$$g_{ij} = \eta g_{ji},$$

i.e.,

$$\eta = \begin{cases} 1 & \text{for O}(n), \\ -1 & \text{for Sp}(n). \end{cases} \tag{4}$$

Also, let $\eta = 0$ for U(n). The orthogonal and symplectic generators may be written in terms of the U(n) generators:

$$\tilde{E}_{ij} = g_{ik}E^k_j - g_{kj}E^k_i, \quad i,j,k = 1,\ldots,n. \tag{5}$$

($\tilde{E}$ is used here to avoid confusion with E, but will be dropped below to enable a unified treatment.) The commutation relations become

$$[\tilde{E}_{ij}, \tilde{E}_{kl}] = g_{kj}\tilde{E}_{il} - g_{il}\tilde{E}_{kj} - g_{ik}\tilde{E}_{jl} + g_{lj}\tilde{E}_{ki}. \tag{6}$$

The generators we use for the definition of the Gel'fand invariants are of mixed variance, given by

$$\tilde{E}^i_j = g^{ik}\tilde{E}_{kj}. \tag{7}$$

The Gel'fand invariants for U(n), O(n), and Sp(n) are in each case now defined by

$$I_k = E^{i_1}_{i_2}E^{i_2}_{i_3}\cdots E^{i_k}_{i_1}, \tag{8}$$

or alternatively

$$\bar{I}_k = E^{i_2}_{i_1}E^{i_3}_{i_2}\cdots E^{i_1}_{i_k}, \tag{9}$$

[where E replaces $\tilde{E}$ for O(n) and Sp(n)].

In the following, we shall restrict ourselves to evaluating the I_k and the $\bar{I}_k$ on an arbitrary (but fixed) finite-dimensional irreducible representation with highest weight $[m]$ ($=[m_1,\ldots,m_n]$ for U(n), but $[m_1,\ldots,m_h]$ for O($2h$), O($2h+1$), and Sp($2h$), and $m_1 \geq m_2 \geq \cdots \geq m_{n\,\text{or}\,h}$). The labeling is the usual one for U(n); for O(n) we take the Cartan subalgebra to be the set $\{-iE_{2r-1\,2r}: r = 1,\ldots,h\}$ and for Sp($2h$) the set $\{E_{r+h\,r}: r = 1,\ldots,h\}$. We denote the carrier space of the representation $[m]$ by $V_{[m]}$, and use E^i_j for the generators on this space; we introduce also a second representation $[l]$ (which we shall later choose to be $[1\dot{0}]$ or $[\dot{0}-1]$) and

 0022-2488/78/1901-0164$1.00

denote by F_j^i the generators on $V_{[l]}$. We may then define a representation on $V_{[l]}\otimes V_{[m]}$ by the generators $G_j^i = F_j^i\otimes 1 + 1\otimes E_j^i$. Consider now the operator on $V_{[l]}\otimes V_{[m]}$

$$P' = F_j^i \otimes E_i^j. \tag{10}$$

P' intertwines with the generators, as can be seen from

$$\begin{aligned} I_2([l]\otimes[m]) &= (F_j^i\otimes 1 + 1\otimes E_j^i)(F_i^j\otimes 1 + 1\otimes E_i^j) \\ &= I_2([l])\otimes 1 + 2P' + 1\otimes I_2([m]). \end{aligned} \tag{11}$$

Thus on each irreducible component, $[n]$, of $[l]\otimes[m]$, P' reduces to the constant

$$P'_{[n]} = \tfrac{1}{2}\{I_2([n]) - I_2([l]) - I_2([m])\}. \tag{12}$$

We have therefore immediately

$$\mathrm{tr}(P')^k = \sum_{[n]\in[l]\otimes[m]} (p'_{[n]})^k \times D([n]), \tag{13}$$

where $D([n])$ is the dimension of $V_{[n]}$.

We now choose $[l]=[\mathring{1}\mathring{0}]$, the fundamental representation for all of the groups, and choose a basis for $V_{[\mathring{1}\mathring{0}]}$ such that the U(n) generators have matrix elements

$$\langle\alpha|F_j^i|\beta\rangle = \delta_{i\alpha}\delta_{j\beta}, \tag{14}$$

and the O(n) and Sp($2h$) generators are related to the F_j^i in (14) by Eqs. (5) and (7). Taking an arbitrary basis $\{|r\rangle : r = 1, \ldots, D([m])\}$ for $V_{[m]}$, and writing $|\alpha; r\rangle$ for $|\alpha\rangle\otimes|r\rangle$, we have

$$\begin{aligned} \langle\alpha; r|P'_{[\mathring{1}\mathring{0}]}|\beta; s\rangle &= \langle\alpha|F_j^i|\beta\rangle\langle r|E_i^j|s\rangle \\ &= \begin{cases} \langle r|E_\alpha^\beta|s\rangle & \text{for U}(n), \\ 2\langle r|E_\alpha^\beta|s\rangle & \text{for O}(n),\ \text{Sp}(2h). \end{cases} \end{aligned} \tag{15}$$

The details of this calculation are in Appendix A. The commutation relations for each group are left invariant under the transformation of the operators

$$F_j^i \to \bar{F}_j^i = -F_i^j.$$

The $[\mathring{1}\mathring{0}]$ representation of U(n) now becomes $[\mathring{0}-1]$, but with O(n) and Sp($2h$), $[\mathring{1}\mathring{0}]$ remains as $[\mathring{1}\mathring{0}]$. In the same basis ($|\alpha; r\rangle$) as above, but with respect to the new generators $\bar{F}_j^i$, the matrix elements of P' (written $P'_{[\mathring{0}-1]}$) become (see Appendix A)

$$\langle\alpha; r|P'_{[\mathring{0}-1]}|\beta; s\rangle = \begin{cases} -\langle r|E_\beta^\alpha|s\rangle & \text{for U}(n), \\ -2\langle r|E_\beta^\alpha|s\rangle & \text{for O}(n),\ \text{Sp}(2h). \end{cases} \tag{16}$$

For O(n) and Sp($2h$), the transformation $F\to\bar{F}$ has the same effect as a simple change of basis in $[\mathring{1}\mathring{0}]$; accordingly $P'_{[\mathring{1}\mathring{0}]}$ and $P'_{[\mathring{0}-1]}$ may be regarded as equal, the matrix elements of (15) and (16) being related by the change of basis; this "equality" will finally appear as the equality of $|I_k|$ and $|\bar{I}_k|$ for O(n) and Sp($2h$).

We now define

$$P = \begin{cases} -P'_{[\mathring{0}-1]} & \text{for U}(n), \\ -\tfrac{1}{2}P'_{[\mathring{0}-1]} & \text{for O}(n),\ \text{Sp}(2h), \end{cases}$$

$$\bar{P} = \begin{cases} P'_{[\mathring{1}\mathring{0}]} & \text{for U}(n), \\ \tfrac{1}{2}P'_{[\mathring{1}\mathring{0}]} & \text{for O}(n),\ \text{Sp}(2h). \end{cases} \tag{17}$$

Then

$$\begin{aligned} \mathrm{tr}P^k &= \sum_{\alpha_1, r_1} \langle\alpha_1; r_1|P^k|\alpha_1; r_1\rangle \\ &= \sum_{\substack{\alpha_1,\ldots,\alpha_k \\ r_1,\ldots,r_k}} \langle\alpha_1; r_1|P|\alpha_2; r_2\rangle\langle\alpha_2; r_2|P|\alpha_3; r_3\rangle\cdots \\ &\quad \times\langle\alpha_k; r_k|P|\alpha_1; r_1\rangle \\ &= \sum_{r_1,\ldots,r_k} \langle r_1|E_{\alpha_2}^{\alpha_1}|r_2\rangle\langle r_2|E_{\alpha_3}^{\alpha_2}|r_3\rangle\cdots\langle r_k|E_{\alpha_1}^{\alpha_k}|r_1\rangle \\ &= \mathrm{tr}E_{\alpha_2}^{\alpha_1}E_{\alpha_3}^{\alpha_2}\cdots E_{\alpha_1}^{\alpha_k} \\ &= I_k \times D([m]). \end{aligned} \tag{18}$$

Similarly

$$\mathrm{tr}\bar{P}^k = \bar{I}_k \times D([m]). \tag{19}$$

Hence, a general formula for the Gel'fand invariants is

$$\begin{aligned} D([m])\,I_k &= \sum_{[n]\in[\mathring{0}-1]\otimes[m]} p_{[n]}^k D([n]), \\ D([m])\,\bar{I}_k &= \sum_{[n]\in[\mathring{1}\mathring{0}]\otimes[m]} \bar{p}_{[n]}^k D([n]), \end{aligned} \tag{20}$$

where the $p_{[n]}$ and $\bar{p}_{[n]}$ are the eigenvalues of P and $\bar{P}$:

$$\begin{aligned} p_{[n]} &= \begin{cases} -\tfrac{1}{2}\{I_2([n]) - I_2([m]) - I_2([\mathring{0}-1])\} & \text{for U}(n), \\ -\tfrac{1}{4}\{I_2([n]) - I_2([m]) - I_2([\mathring{1}\mathring{0}])\} & \text{for O}(n)\ \text{and Sp}(2h), \end{cases} \\ \bar{p}_{[n]} &= \begin{cases} \tfrac{1}{2}\{I_2([n]) - I_2([m]) - I([\mathring{1}\mathring{0}])\} & \text{for U}(n), \\ \tfrac{1}{4}\{I_2([n]) - I_2([m]) - I_2([\mathring{1}\mathring{0}])\} & \text{for O}(n)\ \text{and Sp}(2h). \end{cases} \end{aligned} \tag{21}$$

Evaluation of the $\overset{(-)}{p}_{[n]}$ in terms of the m_i, and Weyl's dimension formula, provide a more explicit form of Eqs. (20) and (21). It will be clear from what follows that we lose no generality by imposing on $[m]$ the strict inequalities

$$m_1 > m_2 > \cdots > m_n. \tag{22}$$

This is because we may *define*, for $m_i = m_{i+1}$,

$$\begin{aligned} &D([m_1, \ldots, m_i - 1, m_i, \ldots, m_n]) \\ &\quad = D([m_1, \ldots, m_i, m_{i+1} + 1, \ldots, m_n]) = 0, \end{aligned} \tag{23}$$

[this turns out to be compatible with the Weyl dimension formulas for U(n), O(n), and Sp($2h$)]. If Eq. (22) holds, there is a one-to-one correspondence between the irreducible components of $[\mathring{1}\mathring{0}]\otimes[m]$ and the weights of $[\mathring{1}\mathring{0}]$; for U($n$), this is also true of $[\mathring{0}-1]$. Hence we may label the irreducible components, $[n]$, of $[\mathring{1}\mathring{0}]\otimes[m]$ (and so also the eigenvalues of $\bar{P}$), by the weights of $[\mathring{1}\mathring{0}]$, and similarly with $[\mathring{0}-1]$ and the eigenvalues of P. We write

$$[n]_i = [m] + [\bar{\Delta}]_i([\Delta]_i),$$

and $\bar{p}_i$ (p_i) for the eigenvalue of $\bar{P}(P)$ in $[n]_i$, where

$$[\bar{\Delta}]_i = [0, \ldots, 0, 1, 0, \ldots, 0] \quad (1 \text{ in the } i\text{th position})$$

and

$$[\Delta]_i = [0, \ldots, 0, -1, 0, \ldots, 0] \quad (-1 \text{ in the } i\text{th position})$$

when

$i=1,\ldots,n$ for U(n)

$i=1,\ldots,h$ for O($2h$), O($2h+1$), and Sp($2h$),

and

$$[\bar{\Delta}]_{n+1-i}=[\Delta]_i,$$

when

$i=1,\ldots,n$ $[i\neq(n+1)/2]$ for O($n=2h$), O($n=2h+1$), and Sp($n=2h$); for O($2h+1$) we set

$$\overset{(-)}{[\Delta]}_{(n+1)/2}(=\overset{(-)}{[\Delta]}_{h+1})=[0,\ldots,0].$$

The expression for I_2 in terms of the m_i can be straightforwardly determined, and from it we can write down the eigenvalues p_i and $\bar{p}_i$:

$$\left.\begin{aligned} p_i&=m_i+n-\eta-i \\ \bar{p}_i&=m_i+1-1 \end{aligned}\right\} \begin{array}{l} \text{with } i=1,\ldots,n \text{ for U}(n), \\ i=1,\ldots,h \text{ for O}(2h),\ \text{O}(2h+1), \\ \qquad\qquad \text{Sp}(2h), \end{array}$$

$$p_{n+1-i}=-\bar{p}_i \qquad \text{with } i=1,\ldots,n \text{ for O}(n=2h), \quad \text{O}(n=2h+1),\ \text{Sp}(n=2h),$$

$$p_{(n+1)/2}=-\bar{p}_{(n+1)/2}=(n-1)/2=h \text{ for O}(n=2h+1). \tag{24}$$

(details in Appendix B). For U(n), the p_i and $\bar{p}_i$ are respectively the partial hooks p_{in} and p_{i1}. Equations (20) and (21) thus reduce to the formula (B12) of Ref. 1 and its siblings:

$$\begin{aligned} D([m])I_k&=\sum_{i=1}^{n} p_i^k D([m]+[\Delta]_i), \\ D([m])\bar{I}_k&=\sum_{i=1}^{n} \bar{p}_i^k D([m])+[\bar{\Delta}]_i). \end{aligned} \tag{25}$$

3. FULLY EXPLICIT FORMULAS

Weyl's dimension functions are given in Ref. 15 in the form

$$D([m])=R([q])/\{R([q])|_{[m]=[0]}\},$$

with

$$q_i=\begin{cases} m_i+n-i, & i=1,\ldots,n \text{ for U}(n) \\ m_i+\frac{1}{2}(n-\eta+1)-i, & i=1,\ldots,h \text{ for O}(n),\ \text{Sp}(n), \end{cases}$$

$$R([q])=\begin{cases} \prod_{i<j}(q_i-q_j) & \text{for U}(n), \\ \prod_{1\leq i<j\leq h}(q_i^2-q_j^2) & \text{for O}(2h), \\ \prod_{1\leq i<j\leq h}(q_i^2-q_j^2)\prod_{1\leq i\leq h} q_i & \text{for O}(2h+1),\ \text{Sp}(2h). \end{cases} \tag{26}$$

From the relations

$$\begin{aligned} p_i-p_j&=q_i-q_j \qquad \text{(all groups)}, \\ p_i-p_{n+1-j}&=q_i+q_j \quad [\text{O}(n) \text{ and Sp}(n)], \end{aligned} \tag{27}$$

we may write

$$R([q])=R([p])=\begin{cases} \prod_{i<j}(p_i-p_j) & \text{for U}(n), \\ \prod_{1\leq i<j\leq h}(p_i-p_j)(p_i-p_{n+1-j}) & \text{for O}(2h), \\ (1/2^h)\prod_{1\leq i<j\leq h}(p_i-p_j)(p_i-p_{n+1-j}) & \\ \quad\times\prod_{1\leq i\leq h}(p_i-p_{n+1-i}) & \text{for O}(2h+1),\ \text{Sp}(2h), \end{cases} \tag{28}$$

from which

$$\frac{D([m]+[\Delta]_i)}{D([m])}=\begin{cases} \prod_{\substack{j=1\\ j\neq i}}^{n}\dfrac{(p_j-p_i+1)}{(p_j-p_i)} & \text{for U}(n), \\ \prod_{\substack{j=1\\ j\neq i,\,n+1-i}}^{n}\dfrac{(p_j-p_i+1)}{(p_j-p_i)} & \text{for O}(n=2h), \\ \prod_{\substack{j=1\\ j\neq i,\,(n+1)/2,\,n+1-i}}^{n}\dfrac{(p_j-p_i+1)}{(p_j-p_i)} & \\ \quad\times\dfrac{(p_i-p_{n+1-i}+2)}{(p_i-p_{n+1-i}+2\delta_{i(n+1)/2}} & \text{for O}(n=2h+1),\ \text{Sp}(n=2h). \end{cases} \tag{29}$$

In obtaining (29) from (28) we have used the relations

$$\begin{aligned} &p_j-p_i=p_{n+1-i}-p_{n+1-j} \qquad [i\neq(n+1)/2 \text{ for O}(2h+1)], \\ &D([m]+[\Delta]_{(n+1)/2})/D([m])=1 \text{ for O}(2h+1), \\ &p_j-p_{(n+1)/2}+1=-(p_{n+1-j}-p_{(n+1)/2}) \text{ for O}(2h+1) \end{aligned} \tag{30}$$

[the $2\delta_{i\,(n+1)/2}$ in the denominator of the expression in (29) for O($2h+1$) being needed to accommodate the second equation of (30)]. The equations (29) can be equally well written by replacing throughout p with $\bar{p}$, Δ with $\bar{\Delta}$, and the "+" signs in the rhs numerators (and in front of $2\delta_{i(n+1)/2}$) with "−" signs; for we have

$$\begin{aligned} &[\bar{\Delta}]_i=-[\Delta]_i, \\ &\bar{p}_j-\bar{p}_i=p_j-p_i \ [i=1,\ldots,n,\ i\neq(n+1)/2 \text{ for O}(2h+1)], \\ &\bar{p}_j-\bar{p}_{(n+1)/2}-1=-(\bar{p}_{n+1-j}-\bar{p}_{(n+1)/2}) \text{ for O}(2h+1). \end{aligned} \tag{31}$$

Thus finally we express the I_k, $\bar{I}_k$ in the form

$$I_k=\sum_{i=1}^{n} p_i^k \prod_{\substack{j=1\\ j\neq i}}^{n}\frac{(p_j-p_i+1+\epsilon_{ji})}{(p_j-p_i)}, \tag{32}$$

$$\bar{I}_k=\sum_{i=1}^{n} \bar{p}_i^k \prod_{\substack{j=1\\ j\neq i}}^{n}\frac{(\bar{p}_j-\bar{p}_i-1-\epsilon_{ji})}{(\bar{p}_j-\bar{p}_i)}, \tag{33}$$

where

$$\epsilon_{ji}=\begin{cases} 0 & \text{for U}(n), \\ -\delta_{jn+1-i} & \text{for O}(n=2h), \\ +\delta_{jn+1-i}-\delta_{i\,(n+1)/2} & \text{for O}(n=2h+1), \\ +\delta_{jn+1-i} & \text{for Sp}(n=2h). \end{cases}$$

[A comparison of Eq. (24) and Eq. (2.14) of Ref. 3 confirms the equivalence of Eq. (32) and Eq. (3.6) of that paper.] One can readily verify from (30), (31), (32), and (33) that for O(n) and Sp(n)

$$I_k = (-)^k \bar{I}_k.$$

ACKNOWLEDGMENTS

The author would like to thank Professor H. S. Green for the inspiration of the characteristic identities, and Dr. A. L. Carey and A. Cant for some valuable help and enlightenment. He would also like to acknowledge the financial support of a Commonwealth Postgraduate Research Award.

APPENDIX A: MATRIX ELEMENTS OF $P'_{[1\dot{0}]}$ AND $P'_{[\dot{0}-1]}$

For U(n)

$$\begin{aligned}\langle\alpha;r|P'_{[1\dot{0}]}|\beta;s\rangle &= \langle\alpha|F^i_j|\beta\rangle\langle r|E^j_i|s\rangle \\ &= \delta^i_\alpha\delta_{j\beta}\langle r|E^j_i|s\rangle \\ &= \langle r|E^\beta_\alpha|s\rangle. \end{aligned} \tag{A1}$$

For O(n) and Sp($2h$)

$$\begin{aligned}\langle\alpha;r|P'_{[1\dot{0}]}|\beta;s\rangle &= \langle\alpha|F^i_j|\beta\rangle\langle r|E^j_i|s\rangle \\ &= g^{ik}\langle\alpha|F_{kj}|\beta\rangle g^{jl}\langle r|E_{li}|s\rangle \\ &= g^{ik}(g_{km}\delta^m_\alpha\delta_{j\beta} - g_{mj}\delta^m_\alpha\delta_{k\beta})g^{jl}\langle r|E_{li}|s\rangle \\ &= (g^{ik}g_{k\alpha}g^{\beta l} - g^{i\beta}g_{\alpha j}g^{jl})\langle r|E_{li}|s\rangle \\ &= (\delta^i_\alpha g^{\beta l} - g^{i\beta}\delta^l_\alpha)\langle r|E_{li}|s\rangle \\ &= \langle r|E^\beta_\alpha + g^{\beta i}\delta^l_\alpha E_{il}|s\rangle \\ &= 2\langle r|E^\beta_\alpha|s\rangle. \end{aligned} \tag{A2}$$

The calculation for $P'_{[\dot{0}-1]}$ is identical except for the substitution $\langle\alpha|F^i_j|\beta\rangle = -\delta^i_\beta\delta_{j\alpha}$. It is notable that the result (A2) is independent of the particular choice of g_{ij}, provided only that $g_{ij} = \eta g_{ji}$.

APPENDIX B: THE EIGENVALUES OF P AND $\bar{P}$

$I_2([m])$ is evaluated by letting it act on a highest weight vector $|{}^{[m]}_{(m)}\rangle$, and using the commutation relations and the fact that $E|{}^{[m]}_{(m)}\rangle = 0$ whenever E is a generator that increases the weight. This method is standard and the results are[11,12] for U(n)

$$I_2 = \sum_{i=1}^{n} m_i(m_i + n + 1 - 2i),$$

for O(n), Sp($2h$)

$$I_2 = 2\sum_{i=1}^{h} m_i(m_i + n + 1 - \eta - 2i). \tag{B1}$$

Thus

$$I_2([1\dot{0}]) = \begin{cases} n = I_2([\dot{0}-1]) & \text{for U}(n) \\ 2(n-\eta) & \text{for O}(n),\ \text{Sp}(2h), \end{cases} \tag{B2}$$

and, from (21), with $1 \le i \le h$ for O(n) and Sp($2h$),

$$\begin{aligned} p_i &= \tfrac{1}{2}\sum_{j=1}^{n\ \text{or}\ h}[-(m_j - \delta_{ji})(m_j - \delta_{ji} + n + 1 - \eta - 2i) \\ &\quad + m_j(m_j + n + 1 - \eta - 2i)] + \tfrac{1}{2}(n-\eta) \\ &= \tfrac{1}{2}(m_i + m_i + n + 1 - \eta - 2i - 1 + n - \eta) \\ &= m_i + n - \eta - i, \end{aligned} \tag{B3}$$

and

$$\begin{aligned} \bar{p}_i &= \tfrac{1}{2}\sum_{j=1}^{n\ \text{or}\ h}[(m_j + \delta_{ji})(m_j + \delta_{ji} + n + 1 - \eta - 2i) \\ &\quad - m_j(m_j + n + 1 - \eta - 2i)] - \tfrac{1}{2}(n-\eta) \\ &= \tfrac{1}{2}(m_i + m_i + n + 1 - \eta - 2i + 1 - n + \eta)] \\ &= m_i + 1 - i. \end{aligned} \tag{B4}$$

That $p_{n+1-i} = -\bar{p}_i$ for O(n), Sp($2h$) ($1 \le i \le n$) is clear from (21) and the relation $[\Delta]_{n+1-i} = [\bar{\Delta}]_i$. For O($2h+1$),

$$p_{(n+1)/2} = \tfrac{1}{4}I_2([1\dot{0}]) = \tfrac{1}{2}(n-\eta) = -\bar{p}_{(n+1)/2}.$$

[1] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970) (particularly Appendix B).
[2] J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[3] C. O. Nwachuku and M. A. Rashid, J. Math. Phys. **17**, 1611 (1976).
[4] H. S. Green, to appear J. Aust. Math. Soc. B **19**; also University of Adelaide preprint, "Spectral Resolution of the Identity of Elements of a Lie Algebra."
[5] R. L. Hudson, J. Math. Phys. **15**, 1067 (1974).
[6] M. C. K. Aguilera-Navarro and V. C. Aguilera-Navarro, J. Math. Phys. **17**, 1173 (1976).
[7] A. M. Perelomov and V. S. Popov, Yad. Fiz. **3** (a) 924, (b) 1127 (1966) [Sov. J. Nucl. Phys. **3** (a) 676, (b) 819 (1966)].
[8] V. S. Popov and A. M. Perelomov, Yad. Fiz. **5**, 693 (1967) [Sov. J. Nucl. Phys. **5**, 489 (1967)].
[9] M. K. F. Wong and H.-Y. Yeh, J. Math. Phys. **16**, 1243 (1975).
[10] S. Okubo, J. Math. Phys. **16**, 528 (1975).
[11] H. S. Green, J. Math. Phys. **12**, 2106 (1971).
[12] A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971).
[13] K. C. Hannabuss, unpublished.
[14] D. M. O'Brien, A. Cant, and A. L. Carey, to appear Ann. Inst. Henri Poincaré; also University of Adelaide preprint, "On Characteristic Identities for Lie Algebras."
[15] Willard Miller Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972), pp. 339, 350, 361–2.

On the self-induced transparency effect of the three-wave resonance process[a)]

S. C. Chiu

The Rockefeller University, New York, New York 10021
(Received 12 July 1977)

The self-induced transparency effect of the three-wave resonance is studied by the inverse scattering method. By transforming to the characteristic coordinates of the background wave, the inverse-scattering theory becomes greatly simplified. With the presence of a constant background wave, the three-wave process is dispersive, and the solitons and continuum behave in a different way from those of spatially bounded wavepackets. The continuum decays away, depositing energy to the background. The solitons have velocities which are amplitude dependent.

I. INTRODUCTION

The effect of self-induced transparency of a powerful ultrashort coherent light pulse through a resonant two-level medium was first predicted and experimentally verified by McCall and Hahn.[1] In this effect, the light pulse has sufficient intensity to cause total population inversion of the medium, and yet its duration is short compared with the relaxation time of the excited medium, so that the medium returns the energy to the pulse by coherent stimulated emission due to the excess intensity, and lossless propagation of coherent pulses becomes possible. The subsequent discovery of the solvability of the of the equations of this effect by Bäcklund transformations and the inverse-scattering method[2-4] allows a thorough study of the analytical solutions of this effect. It is now well known that arbitrary incident pulses will eventually break up into solitary pulses above a threshold. Interactions among solitons eventually only cause phase shifts without affecting the solitons' shapes. The so called "radiation" or continuous portion of the pulse is gradually absorbed by the medium.[4]

Armstrong *et al.*[5] first showed that there is an analogous effect in the three-wave resonance process, in which three interacting waves satisfy the resonance conditions $\pm\omega_1 \pm \omega_2 \pm \omega_3 = 0$, $\pm k_1 \pm k_2 \pm k_3 = 0$. More recently, Nozaki *et al*[6-7] studied this effect in the three-wave interactions in plasmas. Here, one of the waves acts as a background whose spatial extent is much larger than those of the other two waves, so that with respect to the latter two waves, the former may be considered to be infinite in extent. Nonlinear resonant interaction is such that the background feeds energy to the leading edge of the two short pulses which, in turn, return the energy to the background at the trailing edge.[6,7]

The discovery of an inverse-scattering transform of the equations of the three-wave process by Zakharov and Manakov[8] greatly facilitates the understanding of this process. Using this transform, Zakharov and Manakov[9] and Kaup[10] have independently studied the solutions in which all three waves are spatially bounded. Recently, using the Bäcklund transformation, Case and the author[11] found generalizations of the solutions of the three-wave self-induced-transparency of Nozaki *et al*,[6-7] and new solutions which showed that, in general, solitons are composed of two components, denoted as fast and slow according to their velocities; depending on which wave acts as the background, only one of the components is stable. The speeds of the solitons now depends on the width and amplitude of the pulses, in contrast to solutions studied in Refs. 8—10.

In this paper, we propose to study the three-wave self-induced transparency effect by the inverse-scattering transform. By transforming to the characteristic coordinates of the background wave, the "potential matrix" in the linear operator of inverse scattering transform contains only the two spatially bounded waves. The eigenfunctions of this operator has a much simpler analyticity structure than those of the original operator of Zakharov and Manakov[8] as far as the self-induced transparency effect is concerned, however, the "initial" scattering data are not initial conditions at time $t = 0$, but at the characteristic coordinate $\tau = 0$, which is presumed to be the leading or trailing edge of the background.

The inverse scattering transform allows a more complete analytical study. One can now investigate multi-solitons. They can chase through each other, due to the amplitude dependence of velocities, in contrast to previously studied cases in which there are only three fundamental velocities, the group velocities C_i. The effect of the solitary pulses on the background is a phase change which is the sum of those of the stable branch of each individual solitary pulse. The solitons chase through each other without changing their shapes, but cause phase shifts. The background causes dispersion so that the "continuum" portion of the short wave decay as long as they are in the medium of the background, giving up energy to the background.

In Sec. II, we present the inverse scattering theory of the relevant linear operator. In Secs. III and IV, we discuss soliton solutions and the behavior of the continuum. Only the case in which v_{12} (or q_3) acts as the background, and $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1$ (i.e., $\omega_2 = \omega_1 + \omega_3$) is studied in detail. Other cases are briefly discussed towards the end. As expected, in explosive instability, a singularity can develop. We determine the condition under which this could happen. For other possibilities, initially regular pulses will stay regular.

[a)]The work was supported in part by the Air Force Office of Scientific Research, Grant 76-3085.

 0022-2488/78/1901-0168$1.00

II. INVERSE-SCATTERING TRANSFORM

Following Kaup,[10] we write the one-dimensional three-wave equations in the form

$$\begin{aligned} &q_{1t}+C_1q_{1x}=i\gamma_1q_2^*q_3^*,\\ &q_{2t}+C_2q_{2x}=i\gamma_2q_1^*q_3^*,\\ &q_{3t}+C_3q_{3x}=i\gamma_3q_2^*q_1^*. \end{aligned} \tag{II.1}$$

Here, γ_i are constants with $\gamma_i^2=1$; q_i are the envelope wave amplitudes with group velocities C_i, whose central frequencies and wavenumbers satisfy

$$\begin{aligned} &\pm\omega_1\pm\omega_2\pm\omega_3=0,\\ &\pm k_1\pm k_2\pm k_3=0. \end{aligned} \tag{II.2}$$

The case when all signs are the same pertains to explosive instability; correspondingly, γ_i have the same sign. Other possibilities represent decay instability, with one of the γ's having a different sign. We also assume that

$$C_1<C_2<C_3. \tag{II.3}$$

For simplicity, we consider the case when q_3 acts as the background. In the characteristic coordinates of q_3,

$$\begin{aligned} &Z=(2C_3)^{-1}(x+C_3t),\\ &\tau=(2C_3)^{-1}(x-C_3t). \end{aligned} \tag{II.4}$$

Equations (II.1) allow an inverse scattering transform via the linear eigenvalue problem[12]

$$-i\frac{\partial}{\partial z}\begin{matrix}\psi_1\\ \psi_2\\ \psi_3\end{matrix}+\begin{bmatrix}-\frac{\zeta}{\alpha} & 0 & v_{13}\\ 0 & -\frac{\zeta}{\alpha} & v_{23}\\ v_{31} & v_{32} & +\frac{\zeta}{\alpha}\end{bmatrix}\begin{bmatrix}\psi_1\\ \psi_2\\ \psi_3\end{bmatrix}=0. \tag{II.5}$$

Here

$$\frac{1}{\alpha}=\tfrac{1}{2}\beta_{13}\beta_{23}, \tag{II.6}$$

$$\beta_{ij}=C_j-C_i=\beta_i-\beta_j \tag{II.7}$$

[β_i, introduced below in Eq. (II.11), are thus defined up to a constant whose value does not affect the results]. V_{ij} are related to q_i by

$$\begin{aligned} &v_{31}=(\beta_{23}/\beta_{12})^{1/2}q_2,\\ &v_{23}=(\beta_{13}/\beta_{12})^{1/2}q_1,\\ &v_{12}=(\beta_{13}/\beta_{23})^{1/2}q_3, \end{aligned} \tag{II.8}$$

and

$$v_{ij}=v_{ji}^*\epsilon_i\epsilon_j, \tag{II.9}$$

where

$$(\epsilon_1,\epsilon_2,\epsilon_3)=(\gamma_1,-\gamma_2,\gamma_3). \tag{II.10}$$

The dependence is given by

$$i\frac{\partial}{\partial\tau}\begin{bmatrix}\psi_1\\ \psi_2\\ \psi_3\end{bmatrix}=\begin{bmatrix}-\zeta C_3(C_2+\beta_2+\beta_1) & 2C_3\beta_{13}^{-1}v_{12} & (C_2+C_3)\beta_{23}^{-1}v_{13}\\ 2C_3\beta_{13}^{-1}v_{21} & -\zeta C_3(C_1+2\beta_2) & (C_1+C_3)\beta_{13}^{-1}v_{23}\\ (C_2+C_3)\beta_{23}^{-1}v_{31} & (C_1+C_3)\beta_{13}^{-1}v_{32} & -\zeta(C_1C_2+C_3\beta_2+C_3\beta_3)\end{bmatrix}\begin{bmatrix}\psi_1\\ \psi_2\\ \psi_3\end{bmatrix}. \tag{II.11}$$

All waves except v_{12} vanish at $Z\to\pm\infty$ and finite τ. Cross differentiating Eqs. (II.5) and (II.11) shows that q_i satisfies Eqs. (II.1) and leaves ζ constant. Equation (II.5) is of the same form as the linear eigenvalue problem studied by Manakov with respect to a two-component nonlinear Schrödinger equation.[13] Using this operator for the inverse scattering transform has the advantage over that used in Refs. 8–10 in that the background v_{12}, which does not vanish at $x=-\infty$, occurs only in the τ dependence. The analytic properties of the scattering matrix and the wavefunctions becomes greatly simplified. However, the connection between the inverse-scattering transform and the initial value problem is no longer so simple. Presumably, the solution via inverse-scattering transform can be obtained as follow: Knowing $v_{13}(Z,\tau_0)$ and $v_{23}(Z,\tau_0)$ for a fixed τ_0 and all Z, we can find the scattering datum of Eq. (II.5) at $\tau=\tau_0$; the knowledge of v_{12} at $Z=\pm\infty$ and appropriate intervals of τ then allows us to find the τ-dependence of the scattering datum; $v_{ij}(Z,\tau)$ is then obtained for the corresponding intervals of τ by inverse-scattering, i.e., the Gel'fand–Levitan equations.

The initial values are, however, $v_{ij}(x,0)$ at all x and $t=0$. Taking $\tau_0=0$, we see that $v_{ij}(x,0)$ for $x\geq 0$ completely determines $v_{ij}(Z,\tau)$ for all $Z>0$, $\tau\geq 0$. (At $Z=+\infty$, $v_{13}=v_{23}=0$, but v_{12} need not vanish.) Similarly, knowing $v_{ij}(x,0)$ for $x\leq 0$ completely determines $v_{ij}(Z,\tau)$ for $Z<0$, $\tau\leq 0$. (Again, at $Z=-\infty$, $v_{13}=v_{23}=0$, but v_{12} need not vanish.) The knowledge of $v_{13}(Z,0)$ and $v_{23}(Z,0)$ for $Z>0$, $\tau=0$, and $v_{ij}(Z,0)$ for $x\leq 0$ determines $v_{ij}(Z,\tau)$ for all $Z>0$, $\tau<0$; and similarly, $v_{13}(Z,0)$, $v_{23}(Z,0)$ for $Z<0$, $\tau=0$ and $v_{ij}(x,0)$ for $x\geq 0$ determines $v_{ij}(Z,\tau)$ for all $Z<0$, $\tau>0$. There is a certain degree of independence between $v_{ij}(Z,\tau>0)$ and $v_{ij}(Z,\tau<0)$. Given $v_{13}(Z,\tau=0)$, $v_{23}(Z,\tau=0)$, and $v_{12}(Z=-\infty,\tau)$ for $\tau\leq 0$, the values of $v_{ij}(Z,\tau<0)$ are completely determines, but there exists infinitely many sets of values of $v_{ij}(Z,\tau>0)$ which will give the same $v_{13}(Z,\tau=0)$ and $v_{23}(Z,\tau=0)$. There are various ways of defining initial-boundary values in Z, τ that gives a one-to-one correspondence to some initial problem in x, t. To conveniently study the self-induced transparency effect, we set

$$\lim_{Z\to-\infty}v_{12}\to\begin{cases}\beta_{12}k, & \tau\leq 0,\\ 0, & \tau>0.\end{cases} \tag{II.12}$$

From what we have said, it is clear that one should not be too concerned about what happens at $\tau>0$, except for the values of $v_{13}(Z,0)$ and $v_{23}(Z,0)$ they produce at $\tau=0$.

Indeed, we shall ignore the region $\tau > 0$ for most of our discussions.

The analytic properties of the solutions and the scattering matrix of Eq. (II. 5) for v_{ij} with or without the symmetry, Eq. (II. 9), remain the same as those in Ref. 12. Thus, defining the Jost solutions $\phi^{(\pm)i}{}_j(Z,\zeta)$ by

$$\lim_{k\to\pm\infty}\phi^{(\pm)3}(Z,\zeta)\to\begin{bmatrix}0\\0\\1\end{bmatrix}\exp\left(-i\frac{\zeta}{\alpha}Z\right), \tag{II. 13}$$

$$\lim_{Z\to\pm\infty}\phi^{(\pm)\gamma}{}_j(Z,\zeta)\to\delta^{\gamma}{}_j\exp\left(i\frac{\zeta}{\alpha}Z\right),\quad \gamma=1,2,\quad j=1,2,3, \tag{II. 14}$$

where $\delta^{\gamma}{}_j$ is the Kronecker delta, one finds that $\phi^{(+)\gamma}(Z,\zeta)\exp(-i\zeta Z/\alpha)$, $\gamma=1,2$, and $\phi^{(-)3}(Z,\zeta)\exp(i\zeta Z/\alpha)$ are analytic in the upper half ζ plane, while $\phi^{(-)\gamma}(Z,\zeta)\exp(-i\zeta Z/\alpha)$ and $\phi^{(+)3}(Z,\zeta)\exp(i\zeta Z/\alpha)$ are analytic in the lower half ζ plane, provides that $|v_{\gamma 3}|$ and $|v_{3\gamma}|$ are integrable. These solutions are thus uniquely defined in a half ζ plane up to and including the real line. Within the domains of definition,

$$\lim_{|\zeta|\to\infty}\phi^{(\pm)\gamma}{}_j(Z,\zeta)\exp\left(-i\frac{\zeta}{\alpha}Z\right)\to\delta^{\gamma}{}_j+O\left(\frac{1}{\zeta}\right), \tag{II. 15}$$

and

$$\lim_{|\zeta|\to\infty}\phi^{(\pm)3}{}_j(Z,\zeta)\exp\left(i\frac{\zeta}{\alpha}Z\right)\to\delta^{3}{}_j+O\left(\frac{1}{\zeta}\right). \tag{II. 16}$$

We have used Greek indices γ, σ to denote those ranging from 1 to 2, and Roman indices i, j for those ranging from 1 to 3. This will be the convention adopted in the future. On the real line in $\zeta=0$, one may also define the scattering matrices $a_{ij}(\zeta)$ and $b_{ij}(\zeta)$ by

$$\phi^{(+)i}(Z,\zeta)=\sum_{I=1}^{3}b_{iI}(\zeta)\phi^{(-)I}(Z,\zeta), \tag{II. 17}$$

$$\phi^{(-)i}(Z,\zeta)=\sum_{I=1}^{3}a_{iI}(\zeta)\phi^{(+)I}(Z,\zeta). \tag{II. 18}$$

Clearly,

$$\sum_{I=1}^{3}a_{iI}(\zeta)b_{Ij}(\zeta)=\delta_{ij}. \tag{II. 19}$$

It can also be shown that

$$\det(a)=\det(b)=1. \tag{II. 20}$$

From the mentioned analytic properties of $\phi^{(\pm)i}(Z,\zeta)$ and their integral reprentations in terms of Green's functions [see Eq. (II. 35)], one can easily show that $b_{\gamma\sigma}(\zeta)$ and $a_{33}(\zeta)$ are analytic in the upper half ζ plane, and $a_{\gamma\sigma}(\zeta)$, $b_{33}(\zeta)$ are analytic in the lower half ζ plane. We remind that all functions depend on τ which has not been explicitly indicated, but are obvious from Eq. (II. 11).

The bound states are given by:

(a) at ζ_k, $k=1,\ldots,N$, $\mathrm{Im}(\zeta_k)>0$,

$$\det[b_{\gamma\sigma}(\zeta_k)]=a_{33}(\zeta_k)=0; \tag{II. 21}$$

then there exists constants $C_{\gamma 3}(\zeta_k)$ such that

$$b_{22}(\zeta_k)\phi^{(+)1}(Z,\zeta_k)-b_{12}(\zeta_k)\phi^{(+)2}(Z,\zeta_k)=C_{13}(\zeta_k)\phi^{(-)3}(Z,\zeta_k),$$
$$-b_{21}(\zeta_k)\phi^{(+)1}(Z,\zeta_k)+b_{11}(\zeta_k)\phi^{(+)2}(Z,\zeta_k)=C_{23}(\zeta_k)\phi^{(-)3}(Z,\zeta_k); \tag{II. 22}$$

(b) at $\tilde{\zeta}_l$, $l=1,\ldots,M$, $\mathrm{Im}(\tilde{\zeta}_l)<0$,

$$\det[a_{\gamma\sigma}(\tilde{\zeta}_l)]=b_{33}(\tilde{\zeta}_l)=0; \tag{II. 23}$$

then there exists constants $\tilde{C}_{3\gamma}(\tilde{\zeta}_l)$ such that

$$\phi^{(+)3}(Z,\tilde{\zeta}_l)=\sum_{\gamma=1}^{2}\tilde{C}_{3\gamma}(\tilde{\zeta}_l)\phi^{(-)\gamma}(Z,\tilde{\zeta}_l). \tag{II. 24}$$

For simplicity, we assume that all zeros are simple and do not lie on the real line. Although this is not the most general possibility, the presence of other possibilities presents no theoretical difficulty and merely complicates presentation. We also note that depending on symmetries of particular potentials, there may be further restrictions on bound states.

When symmetry, Eq. (II. 9), is imposed on the potentials, we have

$$a_{ij}(\zeta)=\epsilon_i\epsilon_j b^{*}_{ji}(\zeta^{*}). \tag{II. 25}$$

The numbers of bound states in the upper and lower half ζ planes are now equal, and $\zeta_k=\tilde{\zeta}^{*}_k$. The bound state constants $C_{\gamma 3}(\zeta_l)$ and $\tilde{C}_{3\gamma}(\zeta^{*}_l)$ are related by

$$C_{\gamma 3}(\zeta_l)=\epsilon_\gamma\epsilon_3\tilde{C}^{*}_{3\gamma}(\zeta^{*}_l). \tag{II. 26}$$

When $\epsilon_1=\epsilon_2=\epsilon_3$, there are no extra restrictions on the bound states. For all other cases, further restrictions are necessary for regular initial conditions.

We shall now present the Gel'fand—Levitan inverse scattering theory.

The wavefunctions $\phi^{(-)i}{}_j(Z,\zeta)$ allow the following Volterra representations:

$$\phi^{(-)\gamma}{}_j(Z,\zeta)=\delta^{\gamma}{}_j\exp\left(i\frac{\zeta}{\alpha}Z\right)+\int_{-\infty}^{Z}K^{(-)\gamma}{}_j(Z,s)\exp\left(i\frac{\zeta}{\alpha}s\right)ds, \tag{II. 27}$$

$$\phi^{(-)3}{}_j(Z,\zeta)=\delta^{3}{}_j\exp\left(-i\frac{\zeta}{\alpha}Z\right)+\int_{-\infty}^{Z}K^{(-)3}{}_j(Z,s)\exp\left(-i\frac{\zeta}{\alpha}s\right)ds, \tag{II. 28}$$

where $K^{(-)i}{}_j(Z,s)$ are independent of ζ. This fact can be verified by substituting Eqs. (II. 27) and (II. 28) into Eq. (II. 5). One finds from Eq. (II. 27) that for $Z>s$,

$$-i\left(\frac{\partial}{\partial Z}+\frac{\partial}{\partial s}\right)K^{(-)\gamma}{}_\sigma(Z,s)+v_{\sigma 3}(Z)K^{(-)\gamma}{}_3(Z,s)=0,$$
$$-i\left(\frac{\partial}{\partial Z}-\frac{\partial}{\partial s}\right)K^{(-)\gamma}{}_3(Z,s)+\sum_{\sigma=1}^{2}v_{3\sigma}(Z)K^{(-)\gamma}{}_\sigma(Z,s)=0, \tag{II. 29}$$

with boundary conditions

$$K^{(-)\gamma}{}_j(Z,-\infty)=0, \tag{II. 30}$$

and

$$K^{(-)\gamma}{}_3(Z,Z)=-iv_{3\gamma}(Z)/2. \tag{II. 31}$$

The fact that Eqs. (II. 29)—(II. 31) can be uniquely solved for $K^{(-)\gamma}{}_j(Z,s)$ for $s<Z$ follows from the theory

of characteristics of first order partial differential equations. Similarly, for $K^{(-)3}{}_j(Z,s)$,

$$-i\left(\frac{\partial}{\partial Z}-\frac{\partial}{\partial s}\right)K^{(-)3}{}_\sigma(Z,s)+v_{\sigma 3}(Z)K^{(-)3}{}_3(Z,s)=0, \tag{II.32}$$

$$-i\left(\frac{\partial}{\partial Z}+\frac{\partial}{\partial s}\right)K^{(-)3}{}_3(Z,s)+\sum_{\gamma=1}^{2}v_{3\gamma}(Z)K^{(-)3}{}_\gamma(Z,s)=0,$$

for $Z<s$, with boundary conditions

$$K^{(-)3}{}_j(Z,-\infty)=0, \tag{II.33}$$

$$K^{(-)3}{}_\sigma(Z,Z)=-iv_{\sigma 3}(Z)/2. \tag{II.34}$$

Furthermore, one finds from an asymptotic expansion, at $|\zeta|\to\infty$, of the integral equations

$$\phi^{(-)\gamma}{}_j(Z,\zeta)=\delta^\gamma{}_j\exp\left(i\frac{\zeta}{\alpha}Z\right)+\int_{-\infty}^{Z}dy\sum_{l,m}g^{(-)}{}_{jl}(Z,y,\zeta)v_{lm}(y)\phi^{(-)\gamma}{}_m(y,\zeta), \tag{II.35}$$

where

$$g^{(-)}{}_{jl}(Z,y,\zeta)=-i\delta_{jl}\exp[i\zeta d_l\alpha^{-1}(Z-y)], \tag{II.36}$$

and $d_1=d_2=-d_3=1$, that

$$K^{(-)2}{}_1(Z,Z)=-\tfrac{1}{2}\int_{-\infty}^{Z}ds\,v_{13}(s)v_{32}(s), \tag{II.37}$$

and

$$K^{(-)1}{}_2(Z,Z)=-\tfrac{1}{2}\int_{-\infty}^{Z}ds\,v_{23}(s)v_{31}(s). \tag{II.38}$$

It follows from the nonlinear equation of v_{12} and v_{21}, i.e., the equation for q_3 in Eqs. (II.1) in the characteristic coordinates, that for $\tau\le 0$,

$$v_{12}(Z)=\beta_{13}k-2i\epsilon_1\epsilon_2\epsilon_3\beta_{12}\beta_{23}^{-1}K^{(-)2}{}_1(Z,Z), \tag{II.39}$$

$$v_{21}(Z)=\epsilon_1\epsilon_3\beta_{13}k+2i\epsilon_1\epsilon_2\epsilon_3\beta_{12}\beta_{23}^{-1}K^{(-)1}{}_2(Z,Z). \tag{II.40}$$

Thus, $K^{(-)i}{}_j(Z,Z)$ are the quantities we seek to solve from an inverse scattering theory.

The Gel'fand–Levitan equations for $K^{(-)i}{}_j(Z,s)$ can be derived by standard procedure,[14] namely, from the Fourier transform of the contour integrals

$$I^\gamma{}_j(Z,\zeta)=\frac{1}{2\pi i}\sum_{\delta=1}^{2}\int_C d\zeta'\,(\zeta'-\zeta)^{-1}\times\bar\Omega_{\gamma\delta}(\zeta')\phi^{(+)\delta}{}_j(Z,\zeta')\exp\left(-i\frac{\zeta'}{\alpha}Z\right), \tag{II.41}$$

$$I^3{}_j(Z,\zeta)=\frac{1}{2\pi i}\int_{\tilde C}d\zeta'\,(\zeta'-\zeta)^{-1}b_{33}^{-1}(\zeta')\times\phi^{(+)3}{}_j(Z,\zeta')\exp\left(i\frac{\zeta'}{\alpha}Z\right), \tag{II.42}$$

where

$$\sum_{\sigma=1}^{2}\bar\Omega_{\gamma\sigma}b_{\sigma\delta}=\delta_{\gamma\delta}, \tag{II.43}$$

C is a contour starting from $\zeta'=-\infty+i\epsilon$ ending at $\zeta'=\infty+i\epsilon$ and passing above all zeros of $a_{33}(\zeta')$, and $\tilde C$ is a similar contour passing below all zeros of $b_{22}(\zeta')$. We shall omit the details, and write down the equations.[15] For $y<Z$,

$$K^{(-)\gamma}{}_j(Z,y)-\int_{-\infty}^{Z}K^{(-)3}{}_j(Z,s)B^{(-)}{}_{\gamma 3}(s+y)\,ds=B^{(-)}{}_{\gamma 3}(Z+y)\delta_{3j}, \tag{II.44}$$

$$K^{(-)3}{}_j(Z,y)+\sum_{\sigma=1}^{2}\int_{-\infty}^{Z}ds\,K^{(-)\sigma}{}_j(Z,s)B^{(-)}{}_{3\sigma}(s+y)=(\delta_{j3}-1)B^{(-)}{}_{3j}(Z+y), \tag{II.45}$$

where

$$B^{(-)}{}_{\gamma 3}(x)=(2\pi i\alpha)^{-1}\int_{-\infty}^{-\infty}\frac{a_{\gamma 3}(\zeta)}{a_{33}(\zeta)}\exp\left(-i\frac{\zeta}{\alpha}x\right)d\zeta-\sum_{i=1}^{N}C_{\gamma 3}(i)[\dot a_{33}(\zeta_i)\alpha]^{-1}\exp\left(-i\frac{\zeta_i}{\alpha}x\right), \tag{II.46}$$

and

$$B^{(-)}{}_{3\gamma}(x)=(2\pi i\alpha)^{-1}\int_{-\infty}^{\infty}\frac{b_{3\gamma}(\zeta)}{b_{33}(\zeta)}\exp\left(i\frac{\zeta}{\alpha}x\right)d\zeta+\sum_{i=1}^{M}\tilde C_{3\gamma}(i)[\dot b_{33}(\tilde\zeta_i)\alpha]^{-1}\exp\left(i\frac{\tilde\zeta_i}{\alpha}x\right), \tag{II.47}$$

$\dot b_{33}$ and $\dot a_{33}$ denotes derivatives of b_{33} and a_{33} with respect to ζ.

In the case of interest here,

$$B^{(-)}{}_{3\gamma}(x)=\epsilon_3\epsilon_\gamma B^{(-)*}{}_{\gamma 3}(x). \tag{II.48}$$

From Eqs. (II.46) and (II.47), we see that the scattering datum needed are $b_{3\gamma}(\zeta)/b_{33}(\zeta)$, $a_{\gamma 3}(\zeta)/a_{33}(\zeta)$ and the bound state constants

$$\tilde p_{3\gamma}(i)=\tilde C_{3\gamma}(i)[\dot b_{33}(\tilde\zeta_i)\alpha]^{-1} \tag{II.49}$$

and

$$p_{\gamma 3}(i)=C_3(i)[\dot a_{33}(\zeta_i)\alpha]^{-1}. \tag{II.50}$$

The τ dependence of these quantities may be easily obtained from Eq. (II.11) and the asymptotic values of v_{ij} at $Z\to\pm\infty$. In general, if

$$\lim_{Z\to\pm\infty}v_{\gamma 3}(Z,\tau)=\lim_{Z\to\pm\infty}v_{3\gamma}(Z,\tau)=0, \tag{II.51}$$

and

$$\lim_{Z\to-\infty}\beta_{13}^{-1}v_{12}(Z,\tau)=f^{(-)}{}_{12}(\tau), \tag{II.52}$$

then

$$i\frac{\partial}{\partial\tau}\begin{bmatrix}b_{31}(\zeta,\tau)\\ b_{32}(\zeta,\tau)\end{bmatrix}=\begin{bmatrix}-\beta_{13}(C_2+C_3) & 2C_3f^{(-)}{}_{12}(\tau)\\ 2\epsilon_1\epsilon_2C_3f^{(-)*}{}_{12}(\tau) & -\beta_{23}\zeta(C_1+C_3)\end{bmatrix}\begin{bmatrix}b_{31}(\zeta,\tau)\\ b_{32}(\zeta,\tau)\end{bmatrix}, \tag{II.53}$$

and

$$i\frac{\partial}{\partial\tau}b_{33}(\zeta,\tau)=0. \tag{II.54}$$

When $f^{(-)}{}_{12}(\tau)$ takes the simple form of Eq. (II.12), we

have, for $\tau \leq 0$,

$$\begin{bmatrix} b_{31}(\zeta,\tau) \\ b_{32}(\zeta,\tau) \end{bmatrix} = A^{(+)}(\zeta,0) \begin{bmatrix} 1 \\ \xi-\eta \end{bmatrix} \exp(i\Omega^{(+)}\tau) + A^{(-)}(\zeta,0) \begin{bmatrix} -\epsilon_1\epsilon_2(\xi-\eta) \\ 1 \end{bmatrix} \exp(i\Omega^{(-)}\tau), \quad \text{(II. 55)}$$

where

$$\xi = \beta_{12}\zeta/2k, \quad \text{(II. 56)}$$

$$\eta = (\beta_{12}^2\zeta^2 + 4\epsilon_1\epsilon_2 k^2)^{1/2}(2k)^{-1}, \quad \text{(II. 57)}$$

$$\Omega^{(\pm)} = \zeta(C^2{}_3 - C_1C_2) \pm 2kC_3\eta, \quad \text{(II. 58)}$$

and

$$A^{(+)}(\zeta,0) = \tfrac{1}{2}[1+\epsilon_1\epsilon_2\xi(\xi-\eta)]^{-1} \times [b_{31}(\zeta,0) + \epsilon_1\epsilon_2(\xi-\eta)b_{32}(\zeta,0)], \quad \text{(II. 59)}$$

$$A^{(-)}(\zeta,0) = \tfrac{1}{2}[1+\epsilon_1\epsilon_2\xi(\xi-\eta)]^{-1} \times [b_{32}(\zeta,0) - (\xi-\eta)b_{31}(\zeta,0)]. \quad \text{(II. 60)}$$

For $\tau > 0$, we have simply

$$b_{3\gamma}(\zeta,\tau) = b_{3\gamma}(\zeta,0)\exp[i\zeta\beta_{\gamma 3}(C_{3-\gamma}+C_3)\tau]. \quad \text{(II. 61)}$$

The τ-dependences of $\tilde{\tilde{C}}_3(\zeta_i^*,\tau)$ are the same as $b_3(\zeta_i^*,\tau)$, i.e., Eq. (II. 55) with ζ replaced by ζ_i^* and $b_{3\gamma}$ replaced by $\tilde{C}_{3\gamma}$. The τ-dependences of $C_{\gamma 3}(\zeta_i,\tau)$ and $a_{\gamma 3}(\zeta,\tau)$ are deduced from Eqs. (II. 25) and (II. 26). Trivially, from Eq. (II. 54), $b_{33}(\zeta)$ and $a_{33}(\zeta)$ are independent of τ.

There also exists a set of Gel'fand—Levitan equations for $K^{(+)i}{}_j(Z,y)$, where $K^{(+)i}{}_j(Z,y)$ are defined by

$$\phi^{(+)i}{}_j(Z,\zeta) = \delta^i{}_j\exp(id_i\zeta Z/\alpha) + \int_Z^\infty K^{(+)i}{}_j(Z,s)\exp(id_i\zeta s/\alpha)\,ds, \quad \text{(II. 62)}$$

where

$$d_1 = d_2 = -d_3 = 1. \quad \text{(II. 63)}$$

These require a different set of scattering datum. The equivalence of the two sets of Gel'fand—Levitan equations will be demonstrated in the Appendix.

III. SOLITONS

The simplest solutions are the solitons, corresponding to $a_{\gamma 3}(\zeta) = a_{3\gamma}(\zeta) = 0$. In this case, Eqs. (II. 44)—(II. 45) reduce to a system of linear algebraic equations. Take the case when $\epsilon_1 = \epsilon_2 = \epsilon_3$, and let there be only one bound state at $\zeta_1 = \omega_1 + i\nu_1$, and ζ_1^*.

For $\tau \leq 0$, the bound state constants, $\tilde{p}_{31}$, $\tilde{p}_{32}$ evolve according to

$$\begin{bmatrix} \tilde{p}_{31}(\zeta_1^*,\tau) \\ \tilde{p}_{32}(\zeta_1^*,\tau) \end{bmatrix} = a^{(+)}(\zeta_1^*,0) \begin{bmatrix} 1 \\ \xi(\zeta_1^*) - \eta(\zeta_1^*) \end{bmatrix} \times \exp[i\Omega^{(+)}(\zeta_1^*)\tau] + a^{(-)}(\zeta_1^*,0) \times \begin{bmatrix} \eta(\zeta_1^*) - \xi(\zeta_1^*) \\ 1 \end{bmatrix} \exp[i\Omega^{(-)}(\zeta_1^*)\tau], \quad \text{(III. 1)}$$

where $\Omega^{(\pm)}$, ξ, η are defined by Eqs. (II. 56)—(II. 58).

For $\tau \geq 0$,

$$\tilde{p}_{3\gamma}(\zeta_1^*,\tau) = \tilde{p}_{3\gamma}(\zeta,0)\exp[i\zeta_1^*\beta_3(C_{3-\gamma}+C_3)\tau]. \quad \text{(III. 2)}$$

Let us set

$$K^{(-)\gamma}{}_\sigma(Z,y,\tau) = K^{(-)\gamma}{}_\sigma(Z,\tau)\exp[-i\zeta_1 y/\alpha], \quad \text{(III. 3)}$$

and

$$K^{(-)3}{}_\sigma(Z,y,\tau) = K^{(-)3}{}_\sigma(Z,\tau)\exp[i\zeta_1^* y/\alpha]. \quad \text{(III. 4)}$$

Solving Eqs. (II. 44) and (II. 45), and using Eqs. (II. 34) and (II. 3), we have

$$v_{\sigma 3}(Z,\tau) = -2i\tilde{p}_{3\sigma}(\zeta_1^*,\tau)\exp(i2\zeta_1^* Z/\alpha)\Delta^{-1}, \quad \text{(III. 5)}$$

$$v_{12}(Z,\tau) = \beta_{12}k + 2i\beta_{12}(\beta_{13}\beta_{23}^2\nu_1)^{-1}\tilde{p}_{31}(\zeta_1^*,\tau) \times \tilde{p}_{32}^*(\zeta_1^*,\tau)\exp[(2\nu_1 + i\zeta_1^*)Z/\alpha]\Delta^{-1}, \quad \text{(III. 6)}$$

where

$$\Delta = 1 + (\beta_{13}\beta_{23}\nu_1)^{-2}\exp(4\nu_1 Z/\alpha)[\,|\tilde{p}_{31}(\zeta_1^*,\tau)|^2 + |\tilde{p}_{32}(\zeta_1^*,\tau)|^2]. \quad \text{(III. 7)}$$

To analyze the solution, one should transform from Z and τ back to space—time coordinates x and t. In the region $x \leq C_3 t$ ($\tau \leq 0$), the solution is given by Eqs. (III. 1) and (III. 5)—(III. 7), and is the same as the two-soliton solution in Ref. 11 with v_{12} of the shock type, while $v_{\sigma 3}$ are spatially bounded. We take the factor

$$\eta = [(\beta_{12}\zeta/2k)^2 + 1]^{1/2}, \quad \text{(III. 8)}$$

to have a cut on the imaginary axis from $-i2k\beta_{12}^{-1}$ to $i2k\beta_{12}^{-1}$, and stay on the Riemann sheet $\mathrm{sign}(\mathrm{Im}\eta) = \mathrm{sign}(\mathrm{Im}\zeta)$. The solution in the region $x \leq C_3 t$ has two branches corresponding to $a^{(+)}$ and $a^{(-)}$ in Eq. (III. 1),[16] with different group velocities $C_{12}^{(+)}$ and $C_{12}^{(-)}$ respectively. Let

$$k\eta(\zeta_1^*) = (\lambda_1^{(r)} - i\lambda_1^{(i)}), \quad \text{(III. 9)}$$

where

$$\lambda_1^{(r)} = [\tfrac{1}{2}a + \tfrac{1}{2}(a^2+b)^{1/2}]^{1/2}(\mathrm{sign}\omega_1), \quad \text{(III. 10)}$$

$$\lambda_1^{(i)} = [-\tfrac{1}{2}a + \tfrac{1}{2}(a^2+b)^{1/2}]^{1/2}, \quad \text{(III. 11)}$$

$$a = \beta_{12}^2(\omega_1^2 - \nu_1^2)/4 + k^2, \quad \text{(III. 12)}$$

$$b = \beta_{12}^4(\omega_1\nu_1)^2/4. \quad \text{(III. 13)}$$

Then, according to Ref. 11,

$$C_{12}^{(\pm)} = \frac{[\nu_1(C_1C_3 + C_2C_3 - 2C_1C_2) \pm 2C_3\lambda_1^{(i)}]}{[\nu_1(2C_3 - C_1 - C_2) \pm 2\lambda_1^{(i)}]}. \quad \text{(III. 14)}$$

For a fixed ν_1, $C_{12}^{(\pm)}$ are monotonic functions of ω_1 in any quadrant, taking on extremum values at $\zeta_1 \to \infty$. It can be shown that

$$C_2 \geq C_{12}^{(+)} \geq (C_3C_1 + C_3C_2 - 2C_1C_2)/(2C_3 - C_1 - C_2), \quad \text{(III. 15)}$$

$$C_1 \leq C_{12}^{(-)} \leq (C_3C_1 + C_3C_2 - 2C_1C_2)/(2C_3 - C_1 - C_2). \quad \text{(III. 16)}$$

As $t \to -\infty$, in the region $x \leq C_3 t$, $v_{12} \to k_{13}$ and $v_{23} = v_{13} = 0$. As $t \to +\infty$, $x \leq C_3 t$, v_{ij} approaches the $a^{(-)}$ branch (slow) shock type soliton with the $a^{(+)}$ branch vanishing. The magnitude of $|v_{12}(\infty, r=0)| = \beta_{13}k$, but the phase is now $\theta_{12}^{(-)}$ where

$$\tan\theta_{12}^{(-)} = \nu_1\beta_{12}(2\lambda_1^{(r)} - \omega_1\beta_{12})[k^2(1 + |\xi(\zeta_1^*) - \eta(\zeta_1^*)|^2) + \nu_1\beta_{12}(2\lambda_1^{(i)} - \nu_1\beta_{12})]^{-1}. \quad \text{(III. 17)}$$

At finite Z and $\tau=0$, $|v_{12}|$ is, in general, not equal to $\beta_{13}k$. At $\tau=0$, let

$$\tilde{p}_{3\sigma}(\zeta_1^*,0)=\exp[-2i\zeta_1^* Z_\sigma/\alpha], \tag{III.18}$$

and define $Z_{\sigma 3}$ by

$$\exp(-4\nu_1 Z_{\sigma 3}/\alpha)=(\alpha/2\nu_1)^2\left[\sum_{\gamma=1}^{2}\exp(-4\nu_1 Z_\gamma/\alpha)\right], \tag{III.19}$$

then

$$v_{\sigma 3}(Z,0)=2i\exp\left(2i\frac{\omega_1}{\alpha}(Z-Z_\sigma)-\frac{2\nu_1}{\alpha}(Z_\sigma-Z_{\sigma 3})\right)\times\cosh\left(\frac{2\nu_1}{\alpha}(Z-Z_{\sigma 3})\right). \tag{III.20}$$

Thus, at $t=-\infty$, the long wave train $v_{12}=\beta_{13}k$ is undisturbed. The pulses of $v_{\sigma 3}$ gradually enter this constant background from the right. After approximately time t_0,

$$t_0\approx\max(Z_{\sigma 3})+\alpha/2\nu_1, \tag{III.21}$$

most of the $v_{\sigma 3}$ pulses have passed into the background. In the meantime, inside the background, the pulses split into a fast ($a^{(+)}$) and a slow ($a^{(-)}$) component both with group velocities less than C_2 so that they are propagating towards the left relative to the front of the background. The fast branch gradually vanishes until at $t\to-\infty$, there is only the slow component. The short pulses $v_{\sigma 3}$ modify the background within the widths of the pulses. After they are mostly inside the background, the amplitudes of the background in front and behind them are essentially the same; the only difference is the relative phase which will eventually approach the value $\theta_{12}^{(-)}$ given by Eq. (III.17).

As mentioned in the last section, what happens in the region $\tau>0$ is not too relevant to the self-induced transparency effect except that definite functional forms of $v_{\sigma 3}^{(Z)}$ are formed at $\tau=0$. In this case, at $t\to-\infty$, the solution is a pure v_{13}-soliton. The presence of infinitesimal amounts of v_{23}, v_{12} causes the v_{13}-soliton to be gradually transformed into v_{23}, v_{12}. Depending on when the $v_{\sigma 3}$ pulses hit the constant background, some amount of v_{12} will be present at $t\to+\infty$, $\tau>0$. One can find other solutions at $\tau>0$. One can find other solutions at $\tau>0$ which produce the same $v_{\sigma 3}$ at $\tau=0$. It is interesting, however, that solitons of $v_{\sigma 3}$ at $\tau>0$ produce solitons inside the background ($\tau<0$).

The solutions for the case when there are two bound states at ζ_1 and ζ_2 (ζ_1^* and ζ_2^*) can be found similarly. The expressions are rather complicated. However, one may deduce some general asymptotic behaviors. We assume that the group velocities given by Eq. (III.14) are such that $C_{12}^{(+)}(\zeta_2)>C_{12}^{(+)}(\zeta_1)$, then $C_{12}^{(-)}(\zeta_2)<C_{12}^{(-)}(\zeta_1)$. In this case, no matter whether the ζ_2-solitons or the ζ_1-solitons enter the background first, eventually, only the slow branches of each will remain, and the ζ_1-soliton will be to the right of the ζ_2-soliton. There is an asymptotic phase difference of the background between the front and back of the solitons. It is the sum of the phase change due to the slow branch of each bound state, i.e.,

$$\theta_{12}=\sum_{i=1}^{2}\theta_{12}^{(-)}(\zeta_i). \tag{III.22}$$

It is also interesting to note that according to Eq. (III.14), there is a locus in the ζ plane for each given $C_{12}^{(-)}$. All solitons along this locus will eventually move with the same speed.

IV. CONTINUUM

We have seen that the solitons inside a background behaves quite differently. The reason, as one might well suspect, is that the presence of the background causes induced dispersion. It is, therefore, not surprising that the continuum spectrum of the short pulses will gradually decay as any dispersive system. The energy is absorbed by the background, so that, the shape of the background may change after passage of the short pulses with a continuum.

To simplify the discussion, we assume that short pulses are of compact support and contain only a continuum contribution of small amplitudes. In this case, the first approximation to the solutions of Eqs. (II.44) and (II.45) is

$$-\frac{i}{2}v_{\sigma 3}(Z,\tau)=K^{(-)3}{}_\sigma(Z,Z,\tau)\approx-B^{(-)}{}_{3\sigma}(2Z,\tau), \tag{IV.1}$$

$$K^{(-)2}{}_1(Z,Z,\tau)\approx-\int_{-\infty}^{Z}B^{(-)}{}_{31}(Z+s,\tau)\times B^{(-)}{}_{23}(s+Z,\tau)\,ds. \tag{IV.2}$$

By Eqs. (II.54) and (II.55), for $\tau\le 0$,

$$\begin{bmatrix}B^{(-)}{}_{31}(2Z,\tau)\\ B^{(-)}{}_{32}(2Z,\tau)\end{bmatrix}=(2\pi i\alpha)^{-1}\int_{-\infty}^{\infty}d\zeta\begin{bmatrix}1\\ (\xi-\eta)\end{bmatrix}A^{(+)}(\zeta,0)\times\exp[if^{(+)}(\zeta,x,t)]+\begin{bmatrix}\eta-\xi\\ 1\end{bmatrix}A^{(-)}(\zeta,0)\exp\,if^{(-)}(\zeta,x,t)], \tag{IV.3}$$

where x and t are space–time coordinates, and

$$f^{(\pm)}(\zeta,x,t)=\left(\frac{\zeta}{\alpha}+\frac{\Omega^{(\pm)}}{2}\right)\frac{x}{C_3}+\left(\frac{\zeta}{\alpha}-\frac{\Omega^{(\pm)}}{2}\right)t. \tag{IV.4}$$

For large t, the major contributions to the integrals in Eqs. (IV.3) come from the stationary points of $f^{(\pm)}(\zeta,x,t)$. The stationary points of $f^{(+)}$ and $f^{(-)}$ turn out to be the same. They are at

$$\zeta_{1,2}=\pm k\beta_{12}^{-1}(\beta_{13}\beta_{23})^{-1/2}\left[\beta_{23}\left(\frac{x}{t}-C_1\right)+\beta_{13}\left(\frac{x}{t}-C_2\right)\right]\left[\left(C_2-\frac{x}{t}\right)\left(\frac{x}{t}-C_1\right)\right]^{-1/2}, \tag{IV.5}$$

where the + signs pertains to ζ_1, and the − sign to ζ_2. Since the short pulses are of compact support, and the phase velocities

$$W^{(\pm)}(\zeta)\equiv\frac{\partial f^{(\pm)}}{\partial x}\left(\frac{\partial f^{(\pm)}}{\partial t}\right)^{-1}, \tag{IV.6}$$

satisfy

$$C_1\le W^{(\pm)}(\zeta)\le C_2, \tag{IV.7}$$

where equality holds only at $|\zeta|\to\infty$, we may suppose that for some $x_-<0$, $v_{3\sigma}(x,t)$ is zero outside the region

$$C_1t+x_-<x<C_2t. \tag{IV.8}$$

Therefore, in looking at the asymptotic behavior, we may let

$$x = ut + x_0, \tag{IV.9}$$

where $C_2 > u > C_1$, $0 > x_0 > x_-$,

$$\zeta_{1,2} = \pm k\beta_{12}^{-1}(\beta_{13}\beta_{23})^{-1/2}[\beta_{23}(u - C_1) + \beta_{13}(u - C_2) + (\beta_{13} + \beta_{23})x_0 t^{-1}]\left[\left(C_2 - u - \frac{x_0}{t}\right)\left(u - C_1 + \frac{x_0}{t}\right)\right]^{-1/2}. \tag{IV.10}$$

For t large enough, $\zeta_{1,2}$ are on the real axis. A standard steepest descent calculation shows that

$$\begin{bmatrix} B_{31}^{(-)}(2Z,\tau) \\ B_{32}^{(-)}(2Z,\tau) \end{bmatrix} \sim (2\pi i\alpha)^{-1}\left[\sum_{i=1}^{2}\begin{pmatrix} 1 \\ \xi(\zeta_i) - \eta(\zeta_i) \end{pmatrix} A^{(+)}(\zeta_i, 0)\pi^{1/2} \times (2|f^{(+)\prime\prime}(\zeta_i)|)^{-1/2}\exp[i(f^{(+)}(\zeta_i) + \theta_i^{(+)})] + \sum_{i=1}^{2}\begin{pmatrix} \eta(\zeta_i) - \xi(\zeta_i) \\ 1 \end{pmatrix} A^{(-)}(\zeta_i, 0)\pi^{1/2} \times (2|f^{(-)\prime\prime}(\zeta_i)|)^{-1/2} \times \exp\{i[f^{(-)}(\zeta_i) + \theta_i^{(-)}]\}\right], \tag{IV.11}$$

where

$$\theta_1^{(+)} = -\theta_2^{(+)} = -\theta_1^{(-)} = \theta_2^{(-)} = -\frac{\pi}{4}, \tag{IV.12}$$

and

$$f^{(\pm)\prime\prime}(\zeta_i) = \pm\beta_{12}^2 t[4k\eta^3(\zeta_i)]^{-1}(u - C_3 + x_0/t). \tag{IV.13}$$

Hence, at sufficient large t, $B_{3\sigma}^{(-)}$, and consequently $v_{\sigma 3}$, decays as $t^{-1/2}$. It should be noted that as u approaches C_1 or C_2 arbitrarily closely, $|\zeta_i|$ becomes arbitrarily large and $|f^{(\pm)\prime\prime}(\zeta_i)|$ becomes arbitrarily small. The path of steepest descent is really not so steep, but $A^{(\pm)}(\zeta_i, 0)$ are also small; if $A^{(\pm)}(\zeta_i, 0) \sim 0(1/\zeta_i)$ as $|\zeta_i| \to \infty$, it can be deduced that $v_{\sigma 3} \sim O(t^{-1/2})$ even as u approaches these limiting cases arbitrarily close.

It is thus clear that the short pulses spread and decay as $t^{-1/2}$ for large times. For $v_{21}(x,t)$, it is convenient to consider $K^{(-)2}{}_1(Z,Z,\tau)$ in the coordinate system τ, t. Then, for $K^{(-)2}{}_1(\tau,t) = K^{(-)2}{}_1(Z,Z,\tau)$, we have, from Eq. (IV.2),

$$K^{(-)2}{}_1(\tau,t) \approx \frac{1}{4\pi\alpha}\int_{-\infty}^{\infty} d\zeta \frac{b_{32}^*}{b_{33}}(\zeta,\tau)\frac{b_{31}}{b_{33}}(\zeta,\tau) - i(2\pi)^{-2}\alpha^{-1} \times \int_{-\infty}^{\infty} d\zeta \frac{b_{32}^*}{b_{33}}(\zeta,\tau)\int_{-\infty}^{\infty} d\zeta' \exp[2i\alpha^{-1}(\zeta' - \zeta)(\tau + t)] \times \frac{b_{31}}{b_{33}}(\zeta',\tau)\frac{P}{\zeta' - \zeta}, \tag{IV.14}$$

where $P/(\zeta' - \zeta)$ denotes the principal value of the quantity $(\zeta' - \zeta)^{-1}$, and $\tau \leq 0$. For large t, this has a contribution depending only on τ,

$$K^{(-)2}{}_1(\tau,t) \sim \frac{1}{2\pi\alpha}\int_{-\infty}^{\infty} d\zeta \frac{b_{32}^*(\zeta,\tau)}{b_{33}(\zeta)}\frac{b_{31}(\zeta,\tau)}{b_{33}(\zeta)} + O(t^{-1/2}). \tag{IV.15}$$

Thus energy is absorbed by the background.

V. DISCUSSION

From the previous sections, we hope to have demonstrated that the three-wave self-induced transparency effect can be conveniently studied in the characteristic coordinates of the background wave. By studying the specific case in which $v_{12}(q_3)$ is the background, and $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1$ ($\omega_2 = \omega_3 + \omega_1$), we have shown that inside a background, three-wave resonant interaction is dispersive. The solitons are dependent on the amplitudes and widths (i.e., on the eigenvalue ζ_1). Each soliton has two branches, fast and slow, i.e., it is actually a soliton pair. In the case studied, which is characteristic of propagation in the background of the fastest (C_3) of the three waves, the slow branch is stable, and the fast branch eventually disappears. As other dispersive systems, the continuum asymptotically decay algebraically in time (as $t^{-1/2}$). Energy is fed into the background.

There are, of course, other cases of interest, which we have not studied in any great detail. First, ϵ's may take different signs. The main effect is on the bound states. Still considering the case with v_{12} as background, we assume that the waves are kept regular by some means for $\tau > 0$ (before entering the background). The bound states must be such that at $\tau = 0$, the solutions are nonsingular. It is instructive to look at the single soliton-pair solution. Formally, the solution looks like Eqs. (III.5)—(III.7) for $\tau < 0$, with the following modifications:

(i) The quantity $\eta(\zeta_1^*)$ is now given by Eq. (II.57) with the factor $\epsilon_1\epsilon_2$;

(ii) every term $\tilde{p}_{3\sigma}^*(\zeta_1^*,\tau)$ is replaced by $\epsilon_3\epsilon_\sigma\tilde{p}_{3\sigma}^*(\zeta_1^*,\tau)$. The only places where singularities occur are at the zeros of the expression Δ,

$$\Delta = 1 + (\beta_{13}\beta_{23}\nu_1)^{-2}\exp(4\nu_1 Z/\alpha)[\epsilon_1\epsilon_3|\tilde{p}_{31}(\zeta_1^*,\tau)|^2 + \epsilon_2\epsilon_3|\tilde{p}_{32}(\zeta_1^*,\tau)|^2], \tag{V.1}$$

where $\tilde{p}_{31}(\zeta_1^*,\tau)$ and $\tilde{p}_{32}(\zeta_1^*,\tau)$ are given by Eqs. (III.1). In the case $-\epsilon_1 = -\epsilon_2 = \epsilon_3$ ($\omega_1 = \omega_2 + \omega_3$), no regular solitons are possible, so there are no bound states. In the case $-\epsilon_1 = \epsilon_2 = \epsilon_3$ ($\omega_3 = \omega_1 + \omega_2$), the requirement on regularity at $\tau = 0$ implies that the term inside the square bracket of Eq. (V.1) is positive at $\tau = 0$, i.e.,

$$|1 + [\eta(\zeta_1^*) - \xi(\zeta_1^*)]s| < |[\xi(\zeta_1^*) - \eta(\zeta_1^*)] + s|,$$

where $s = a^{(-)}(\zeta_1^*, 0)/a^{(+)}(\zeta_1^*, 0)$. We require that the relation (V.2) holds for all arg(s). Let the function $\eta(\zeta)$ have a cut on the real axis $2k/\beta_{12} > \zeta > -2k/\beta_{12}$, and $\mathrm{Im}\eta(\zeta_1^*) \leq 0$. One may then induce that for $\tau < 0$, the term in the square bracket of Eq. (V.1) remains positive, so that no singularity will develop. In the case $\epsilon_1 = -\epsilon_2 = \epsilon_3$ ($\omega_1 + \omega_2 + \omega_3 = 0$), the inequality sign in Eq. (V.2) is reversed. Then one may show that a singularity will develop for some $\tau < 0$. However, we must look for solutions which have regular initial conditions at $t = 0$. For this purpose, we investigate the velocities $\bar{C}^{(\pm)}{}_{12}$. When $\epsilon_1\epsilon_2 = -1$, the velocities of solitons formally have the same expression, $\bar{C}^{(\pm)}{}_{12}$, as Eq. (III.14) with $\lambda_1^{(i)}$ defined by

$$\lambda_1^{(r)} - i\lambda_1^{(i)} = k[\xi^2(\zeta_1^*) - 1]^{1/2}. \tag{V.3}$$

The limits of $\bar{C}^{(\pm)}{}_{12}$ are different from Eqs. (III.15) and (III.16). One finds

$$C_2 < \bar{C}^{(+)}{}_{12} < C_3. \tag{V.4}$$

There is a half-ellipse on the lower half-plane,

$$\beta^{-2}{}_{12}\nu^2{}_1 + (\beta_{13}+\beta_{23})^{-2}\omega^2{}_1 = (\beta_{13}\beta_{23}\beta^2{}_{12})^{-1}k^2, \tag{V.5}$$

inside which $C^{(-)}{}_{12} > C_3$, and outside which $\bar{C}^{(-)}{}_{12} < C_1$. According to what was said about the stability of the two branches, the fast branch dominates initially, and for a regular initial condition, this must be the $a^{(+)}$ branch for explosive instability, i.e., regular initial conditions are possible only outside the ellipse, Eq. (V.5).

Secondly, a different wave may act as the background. When v_{23} (or q_1) is the background, the short pulses enter the background from the left. The background stays in the interval $0 < \tau = (x - C_1 t)(2C_1)^{-1} < \infty$. This case presents no difficulty, and one will find that inside the background, the fast branch of a soliton is the stable branch at $t \to +\infty$. As in Ref. 11, the velocities of the two branches have the limits (for $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1$ only)

$$C_3 > C_{\text{fast}} > C, \quad C > C_{\text{slow}} > C_2, \tag{V.6}$$

where

$$C = (C_3\beta_{12} + C_2\beta_{13})/(\beta_{12}+\beta_{13}). \tag{V.7}$$

When v_{13} (or q_2) acts as the background, the situation is different. Short pulses can enter from both the right (for q_1) and the left (for q_3) of the background. Then it seems that the phases at both ends, and the amplitude, of the background will change with time. The solitons found from the Bäcklund transformations[11] are such that the phases at both ends are fixed, which means that these waves are already in the background for a long time. We feel that these solitons are probably approximated only by very special initial conditions when the background has finite length, and are probably unstable in the sense that a slight change in initial conditions can drastically change the end result. From what was known about interaction of spatially bounded packets, we speculate that the q_2 background is rather unstable, and will change shape under passage of either a soliton or continuum.

ACKNOWLEDGMENT

The author would like to thank Professor K. M. Case for helpful discussions and comments.

APPENDIX

We have used the Gel'fand–Levitan equations for $K^{(-)i}{}_j(Z,y)$ for inverse scattering. An equivalent set of equations for inverse scattering are those for $K^{(+)i}{}_j(Z,y)$ defined in Eq. (II.61). These equations are

$$K^{(+)\gamma}{}_j(Z,y) - \int_Z^\infty ds\, K^{(+)3}{}_j(Z,s)B^{(+)}{}_{\gamma 3}(s+y) = B^{(+)}{}_{\gamma 3}(Z+y)\delta_{3j}, \tag{A1}$$

$$K^{(+)3}{}_j(Z,y) + \sum_{\sigma=1}^{2}\int_Z^\infty ds\, K^{(+)\sigma}{}_j(Z,s)B^{(+)}{}_{3\sigma}(s+y) = B^{(+)}{}_{3j}(Z+y)(\delta_{j3}-1), \tag{A2}$$

where

$$K^{(+)\gamma}{}_3(Z,Z) = \frac{i}{2}v_{\gamma 3}(Z) = -\epsilon_\gamma\epsilon_3 K^{(+)3*}{}_\gamma(Z,Z), \tag{A3}$$

$$K^{(+)1}{}_2(Z,Z) = -\tfrac{1}{2}\int_Z^\infty ds\, v_{23}(s)v_{31}(s) = \epsilon_1\epsilon_2 K^{(+)2*}{}_1(Z,Z), \tag{A4}$$

and

$$B^{(+)}{}_{\gamma 3}(y) = (2\pi i\alpha)^{-1}\int_{-\infty}^{\infty} d\zeta'\frac{b_{\gamma 3}(\zeta')}{b_{33}(\zeta')}\exp(-i\zeta' y\alpha^{-1}) + \sum_i \tilde{C}_{\gamma 3}(i)[\dot{b}_{33}(\zeta_i^*)\alpha]^{-1}\exp(-i\zeta_i^* y\alpha^{-1}), \tag{A5}$$

$$B^{(+)}{}_{3\gamma}(y) = \epsilon_3\epsilon_\gamma B^{(+)*}{}_{\gamma 3}(y). \tag{A6}$$

We are restricting ourselves to potentials with symmetry Eq. (II.9). The constants $\tilde{C}_{\gamma 3}(i)$ are related to $\tilde{C}_{3\gamma}(i)$ of Eq. (II.24) by

$$\tilde{C}_{13}(i) = -a_{12}(\zeta_i^*)\tilde{C}_{32}{}^{-1}(i) = a_{22}(\zeta_i^*)\tilde{C}_{31}{}^{-1}(i), \tag{A7}$$

$$\tilde{C}_{23}(i) = -a_{21}(\zeta_i^*)\tilde{C}_{31}{}^{-1}(i) = a_{11}(\zeta_i^*)\tilde{C}_{32}{}^{-1}(i). \tag{A8}$$

The two sets of scattering datum, $B^{(+)}{}_{\gamma 3}(y)$, $B^{(+)}{}_{3\gamma}(y)$, and $B^{(-)}{}_{\gamma 3}(y)$, $B^{(-)}{}_{3\gamma}(y)$ are presumably compatible so that knowing one set, we could find the other. We shall demonstrate this by showing that given $b_{3\gamma}(\omega)/b_{33}(\omega)$, $a_{\gamma 3}(\omega)/a_{33}(\omega) = \epsilon_\gamma\epsilon_3 b^*_{3\gamma}(\omega)/b^*_{33}(\omega)$, the positions of the bound states, ζ_i, which we assume to be simple, and the "polarization" vectors $\tilde{s}_{\sigma 3}(i)$, $s_{3\sigma}(i)$ of the bound states defined by

$$\tilde{s}_{\sigma 3}(i) = \tilde{C}_{\sigma 3}(i)/(\sum_\gamma \epsilon_\gamma\epsilon_3|\tilde{C}_{\sigma 3}(i)|^2), \tag{A9}$$

$$s_{3\sigma}(i) = C_{3\sigma}(i)/(\sum_\gamma \epsilon_\gamma\epsilon_3|C_{3\sigma}(i)|^2), \tag{A10}$$

it is possible to reconstruct the whole s matrix a_{ij} and b_{ij}. In the case $\epsilon_1\epsilon_2 = -1$, we must require that the denominators of Eqs. (A9) and (A10) are positive, which are the same as the regularity conditions of the solitons in Eq. (V.1). Essentially, one only needs the relation

$$\sum_j a_{ij}(\zeta)b_{jk}(\zeta) = \delta_{ik}, \tag{A11}$$

and the analytic properties of a_{33}, b_{33}, $a_{\sigma\gamma}$, and $b_{\sigma\gamma}$. First,

$$G(\omega) \equiv \sum_{\gamma=1}^{2}\frac{b_{3\gamma}(\omega)}{b_{33}(\omega)}\frac{a_{\gamma 3}(\omega)}{a_{33}(\omega)} + 1 = b_{33}{}^{-1}(\omega)a_{33}{}^{-1}(\omega). \tag{A12}$$

Thus, for $\zeta \geq 0$,

$$b_{33}(\zeta) = \prod_i\frac{(\zeta-\zeta_i^*)}{(\zeta-\zeta_i)}\exp\left[(2\pi i)^{-1}\int d\omega'\frac{G(\omega')}{\omega'-\zeta}\right], \tag{A13}$$

for $\zeta \leq 0$,

$$a_{33}(\zeta) = \prod_i\frac{(\zeta-\zeta_i)}{(\zeta-\zeta_i^*)}\exp\left[-(2\pi i)^{-1}\int d\omega'\frac{G(\omega)}{\omega'-\zeta}\right]. \tag{A14}$$

For the 2×2 matrices

$$(a')_{\sigma\gamma} = a_{\sigma\gamma}, \tag{A15}$$

$$(b')_{\sigma\gamma} = b_{\sigma\gamma}, \tag{A16}$$

we have

$$b'^{-1}(\omega)a'^{-1}(\omega) = g(\omega), \tag{A17}$$

where

$$g_{\sigma\gamma}(\omega)=\delta_{\sigma\gamma}+\frac{a_{\sigma 3}(\omega)}{a_{33}(\omega)}\,\frac{b_{3\sigma}(\omega)}{b_{33}(\omega)}. \tag{A18}$$

We define the matrix ${}^{(a)}P^{(b)}(i)$ by

$${}^{(a)}P^{(b)}{}_{\sigma\gamma}(i)\equiv \tilde{s}_{\sigma 3}(i)s_{3\gamma}(i). \tag{A19}$$

Then,

$${}^{(a)}P^{(b)}\,{}^{(a)}P^{(b)}={}^{(a)}P^{(b)}, \tag{A20}$$

and at ζ_i, ζ_i^*,

$$a'(\zeta_i^*)\,{}^{(a)}P^{(b)}(i)={}^{(a)}P^{(b)}(i)b'(\zeta_i)=0. \tag{A21}$$

The matrix

$$B'(\zeta)\equiv b'^{-1}(\zeta)\prod_i\left(\frac{\zeta-\zeta_i}{\zeta-\zeta_i^*}\,{}^{(a)}P^{(b)}(i)+I-{}^{(a)}P^{(b)}(i)\right), \tag{A22}$$

is analytic in the upper half ζ plane and

$$\lim_{|\zeta|\to\infty}B'(\zeta)\to I+O(\zeta^{-1}). \tag{A23}$$

The matrix

$$A'(\zeta)\equiv\prod_i\left(\frac{\zeta-\zeta_i^*}{\zeta-\zeta_i}\,{}^{(a)}P^{(b)}(i)+I-{}^{(a)}P^{(b)}(i)\right)a'^{-1}(\zeta), \tag{A24}$$

is analytic in the lower half ζ plane and

$$\lim_{|\zeta|\to\infty}A'(\zeta)\to I+O(\zeta^{-1}). \tag{A25}$$

On the real line $\zeta=\omega$,

$$B'(\omega)A'(\omega)=g(\omega). \tag{A26}$$

The solution of $B'(\zeta)$, $A'(\zeta)$, which are respectively analytic in the upper and lower ζ plane, and which satisfy Eq. (A26) on the real line is the standard homogeneous Hilbert problem studied by Gohberg and Krein,[17] and Newton and Jost.[18] The question of solvability and uniqueness depends crucially on the partial indices κ_1, κ_2 which are uniquely determined for a given nonsingular $g(\omega)$. Defining the real and imaginary part of $g(\omega)$ by

$$g_R(\omega)=\tfrac{1}{2}[g(\omega)+g^*(\omega)], \tag{A27}$$

$$g_I(\omega)=(2i)^{-1}[g(\omega)-g^*(\omega)], \tag{A28}$$

where g^* is the Hermitian conjugate of g, then as long as either $g_R(\omega)$ or $g_I(\omega)$ is definite, $\kappa_1=\kappa_2=0$. In this case, $B'(\zeta)$ and $A'(\zeta)$ are uniquely solvable. We assume that $g(\omega)$ of Eq. (A18) satisfy these conditions. In the case $\epsilon_1=\epsilon_2=\epsilon_3=1$, $g(\omega)$ is positive definite. For other cases, we require

$$\begin{aligned}&1-|\beta_{31}(\omega)|^2>0,\quad -\epsilon_1=\epsilon_2=\epsilon_3=1,\\ &1-|\beta_{32}(\omega)|^2>0,\quad \epsilon_1=-\epsilon_2=\epsilon_3=1,\\ &1-|\beta_{31}(\omega)|^2-|\beta_{32}(\omega)|^2>0,\quad -\epsilon_1=-\epsilon_2=\epsilon_3=1.\end{aligned} \tag{A29}$$

We must emphasize that these are only sufficient conditions. The matrices $a'(\zeta)$ and $b'(\zeta)$ can be obtained from

$$a'(\zeta)=A'^{-1}(\zeta)\prod_i\left(\frac{\zeta-\zeta_i^*}{\zeta-\zeta_i}\,{}^{(a)}P^{(b)}(i)+I-{}^{(a)}P^{(b)}(i)\right), \tag{A30}$$

$$b'(\zeta)=\prod_i\left(\frac{\zeta-\zeta_i}{\zeta-\zeta_i^*}\,{}^{(a)}P^{(b)}(i)+I-{}^{(a)}P^{(b)}(i)\right)B'^{-1}(\zeta). \tag{A31}$$

Finally,

$$a_{3\sigma}(\omega)=-\sum_{\gamma=1}^{2}\frac{b_{3\gamma}(\omega)}{b_{33}(\omega)}a_{\gamma\sigma}(\omega), \tag{A32}$$

and

$$b_{\sigma 3}(\omega)-\sum_{\gamma=1}^{2}b_{\sigma\gamma}(\omega)\frac{a_{\gamma 3}(\omega)}{a_{33}(\omega)}. \tag{A33}$$

[1]S. L. McCall and E. L. Hahn, Phys. Rev. **183**, 457 (1969).
[2]G. L. Lamb, Jr., Rev. Mod. Phys. **43**, 99 (1971).
[3]G. L. Lamb, Jr., Phys. Rev. A **9**, 422 (1974).
[4]M. Ablowitz, D. J. Kaup, and A. C. Newell, J. Math. Phys. **15**, 1852 (1974).
[5]J. A. Armstrong, S. JHA Sudhanshu, and N. S. Shiren, J. Quant. Elect. **QE-6**, 123 (1970).
[6]K. Nozaki, T. Tanuiti, J. Phys. Soc. Jpn. **34**, 796 (1973).
[7]Y. Ohsawa, K. Nozaki, J. Phys. Soc. Jpn. **36**, 591 (1974).
[8]V. E. Zakharov and S. V. Manakov, Pis'ma Zh. Eksp. Teor. Fiz. **18**, 413 (1973) [JETP Lett. **18**, 243 (1973)].
[9]V. E. Zakharov and S. V. Manakov, Zh. Eksp. Teor. Fiz. **69**, 1654 (1975) [Sov. Phys. JETP **42**, 842 (1976)].
[10]D. J. Kaup, Stud. Appl. Math. **55**, 9 (1976).
[11]K. M. Case and S. C. Chiu, Phys. Fluids **20**, 746 (1977).
[12]The v_{ij}'s in this paper differ from those in Ref. 11 by constant factors in β_{ij}.
[13]S. V. Manakov, Zh. Eksp. Teor. Fiz. **65**, 505 (1973) [Sov. Phys. JETP **38**, 248 (1974)].
[14]See, for example, M. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[15]The equivalent version of Eqs. (II.44) and (II.45), which are equations to determine $\phi^{(-)i}{}_j(Z,\zeta)$, instead of $K^{(-)i}{}_j(Z,s)$ here, was first derived in Ref. 13.
[16]The identification of the fast and slow branch in the lower half-plane is interchanged from that in Ref. 11. There, the quantity $\mathrm{Im}(2k\eta)$ (which was denoted as $\lambda^{(i)}{}_{12}$) always takes positive values for both $\nu<0$ and $\nu>0$, which corresponds to taking the values of $2k\eta_1$ at two different Riemann sheets. Except for the interchange of the (+) and (−) branches, the main conclusions all hold, provided one takes $\mathrm{Re}(2k\eta)$ consistently. This inconsistency however leads to an error in conclusions about explosive instability.
[17]I. C. Gohberg and M. G. Krein, Uspehi Mat. Nauk. (N. S.) **13**, No. 2 (80), 3—72 (1958) (Am. Math. Soc. Transl. Ser. 2, V. 14 1960, pp. 217—88.
[18]R. G. Newton and R. Jost, Nuovo Cimento **10**, Ser. 1, 590 (1955).

Perturbation theory for Coulomb scattering

J. Zorbas[a)]

Department of Mathematics, University of British Columbia, Vancouver, B. C., Canada V6T 1W5
(Received 28 December 1976)

Perturbation series for the "renormalized" complex-energy distorted plane waves (RCEW's) are defined and shown to converge to the series expansions for the physical distorted plane waves (PW's) for two-particle scattering via Coulomb-like potentials $V(x)=e_1e_2|x|^{-1}+V_s(x)$, where $V_s\in L^2$ has compact support. A perturbation series for the "renormalized" half-off-shell T matrix is defined and shown to converge to the series expansion for the pure Coulomb physical T matrix.

I. INTRODUCTION

In a recent paper[1] a "renormalized" off-energy-shell formalism was defined for scattering involving more than one charged fragment in either the incoming or outgoing channel. For the particular case of two-particle Coulomb-like scattering it was shown that the complex-energy distorted plane waves (CEW's) $\phi^{\mp\epsilon}(x,p)$ defined as follows

$$\phi^{\mp\epsilon}(x,p)=(\pm i\epsilon)\int dx'\,\phi(x',p)\,G(x,x';p^2/2m\mp i\epsilon), \tag{1.1}$$

$$\phi(x,p)=(2\pi)^{-3/2}\exp(ip\cdot x)$$

must be replaced by RCEW's

$$\hat{\varphi}^{\mp\epsilon}(x,p)=F^{*}_{\pm\epsilon}(|p|)\,\phi^{\mp\epsilon}(x,p), \tag{1.2}$$

where

$$F^{*}_{\pm\epsilon}(|p|)=\Gamma(1\pm i\gamma)^{-1}\exp[\pm i\gamma\log(\epsilon m/2p^2)],$$
$$\gamma=me_1e_2/|p| \tag{1.3}$$

with m the reduced mass and e_i the charge of the ith particle. In addition the half-off-shell "T matrix"

$$(p'|V_\lambda\Omega_-|p)=\int dx\,\overline{\phi(x,p)}\exp(-\lambda x)\,V(x)\,\phi^{+}(x,p), \tag{1.4}$$

where $\phi^{\mp}(x,p)$ denote the PW's must be replaced by the "renormalized" half-off-shell T matrix

$$\lim_{\lambda\to+0}T(p',p;\epsilon,\lambda),$$
$$T(p',p;\epsilon,\lambda)=\overline{F^{*}_{+\epsilon}(|p'|)}\,(p'|V_\lambda\Omega_-|p), \tag{1.5}$$

where the limit $\lambda\to+0$ is taken in the sense specified by (4.3).

The renormalized quantities (1.2) and (1.5) for the pure Coulomb potential have natural series expansions in powers of γ for each $\epsilon>0$. In this paper we apply the calculations of Hostler[2] and McDowell and Coleman[3] to verify the convergence of these "renormalized" perturbation series for (1.2) and (1.5) to the respective series expansions for the pure Coulomb PW's and T matrix for $|\gamma|<1$. Furthermore, the renormalized perturbation series corresponding to (1.2) for Coulomb-like potentials $V(x)=e_1e_2|x|^{-1}+V_s(x)$, where $V_s\in L^2$ has compact support is defined and shown to converge.

The derivation of the renormalized off-energy-shell formalism given in Ref. 1 is based on the self-adjointness of the Hamiltonian and the existence of the full Green's function and renormalized wave operators. In order to ensure these properties we assume in this paper that the Hamiltonian H has the form

$$H=H_0+V,\quad H_0=-(2m)^{-1}\nabla^2,$$
$$V(x)=e_1e_2|x|^{-1}+V_s(x), \tag{1.6}$$

where

$$V_s(x)=V^{(1)}(x)+V^{(2)}(x),$$
$$V^{(1)}\in L^1\cap L^2,\quad V^{(2)}\in L^\infty, \tag{1.7}$$
$$V^{(2)}(x)=O(|x|^{-2-\epsilon_0}),\quad \epsilon_0>0,\quad \text{as } |x|\to\infty.$$

The Hamiltonian H is self-adjoint on $D(H)=D(H_0)$. The requirement (1.7) ensures the existence and integrability of the full Green's function[4] [see Theorem (A1)], which guarantees the existence of the RCEW's for each $\epsilon>0$. In addition the renormalized wave operators $\Omega_\pm$, corresponding to H, exist.[5]

In Secs. II and III of this paper the convergence of the renormalized perturbation series in the limit $\epsilon\to+0$ for the pure Coulomb and Coulomb-like RCEW's respectively is shown. The renormalized perturbation series for the pure Coulomb T matrix is defined and shown to converge to the perturbation series for the physical T matrix in Sec. IV. A discussion of the Coulomb-like renormalized half-off-shell T matrix is given in Sec. V.

II. PERTURBATION THEORY FOR THE PURE COULOMB DISTORTED PLANE WAVES

In order to define the renormalized perturbation series corresponding to (1.2), we denote the formal iterative solutions of the off-shell Lippmann–Schwinger equations for the pure Coulomb potential [Eqs. (A5) with $\eta=0$] as follows,

$$\phi_c^{\mp\epsilon}(x,p)=\sum_{l=0}^{\infty}\gamma^l\phi_l^{\mp\epsilon}(x,p), \tag{2.1}$$

and the absolutely convergent expansion for $|\gamma|<\infty$ and $\epsilon>0$ of (1.3) as follows,

$$F^{*}_{\pm\epsilon}(|p|)=\sum_{l=0}^{\infty}\gamma^lF^{*}_{\pm\epsilon;l}(|p|). \tag{2.2}$$

The formal renormalized perturbation series for the pure Coulomb potential is defined as the product of (2.1) and (2.2), i.e.,

$$\hat{\phi}_c^{\mp\epsilon}(x,p)=\sum_{l=0}^{\infty}\gamma^l\Lambda_l(x,p,\mp\epsilon), \tag{2.3}$$

where

$$\Lambda_l(x,p,\mp\epsilon)=\sum_{\eta=0}^{l}F^{*}_{\pm\epsilon;\,l-\eta}(|p|)\,\phi_\eta^{\mp\epsilon}(x,p). \tag{2.4}$$

a)Supported by the National Research Council of Canada.

 0022-2488/78/1901-0177$1.00

The behavior of the renormalized perturbation series in the limit $\epsilon \to +0$ is investigated in the following theorem.

Theorem 2.1: For $|\gamma| < 1$ the renormalized perturbation series (2.3) converges absolutely and uniformly with respect to $\epsilon \in [0, \sigma]$, $\sigma > 0$. Furthermore, the limit $\epsilon \to +0$ can be taken term by term to yield the pure Coulomb distorted plane waves, i.e.,

$$\phi_c^{\mp}(x, p) = \lim_{\epsilon \to +0} \hat{\phi}_c^{\mp\epsilon}(x, p) = \sum_{l=0}^{\infty} \gamma^l \lim_{\epsilon \to +0} \Lambda_l(x, p, \mp\epsilon) \quad (2.5)$$

for each $(x, p) \in R^3 \times R^3 \setminus \{0\}$.

In order to verify the above theorem, we use Hostler's calculation of $\phi_c^{\mp\epsilon}(x, p)$ [2] which is given by[6]

$$\phi_c^{\mp\epsilon}(x, p) = \mp\epsilon 2k^2 m\, \phi(x, p) [\exp(-2\pi\nu) - 1]^{-1} \times \int_{+\infty;\, \arg(\zeta \pm 1) = 0}^{(1+)} d\zeta \left(\frac{\zeta + 1}{\zeta - 1}\right)^{-i\nu} \times \exp\left[\frac{ik|x|\zeta(k^2 - p^2) - ip \cdot x(k^2 - p^2)}{k^2\zeta^2 - p^2}\right] \left(\frac{2i\zeta}{(k^2\zeta^2 - p^2)^2} + \frac{k(\zeta^2 - 1)[|x|(k^2\zeta^2 + p^2) - 2x \cdot p\zeta k]}{(k^2\zeta^2 - p^2)^3}\right), \quad (2.6)$$

where $\nu = me_1e_2/k$, $k = (p^2 \mp 2mi\epsilon)^{1/2}$, $\mathrm{Im}k > 0$, and the integration starts at $\zeta = +\infty$ with $\arg(\zeta \pm 1) = 0$, runs along the real axis to the right of $\zeta = 1$, circles $\zeta = 1$, and returns along the real axis to $\zeta = +\infty$.

The anomalous ϵ dependence of (2.6) can be extracted by performing the transformations $\zeta = (t+1)/(1-t)$ and $t = \epsilon u$, which yields

$$\phi_c^{\mp\epsilon}(x, p) = (\mp 8k^2 m)\, \phi(x, p) [\exp(-2\pi\nu) - 1]^{-1} \epsilon^{i\nu} \times \int_{\epsilon^{-1};\, \arg u = 0}^{(0+)} du\, u^{i\nu} \exp\left[\frac{\pm 2mk|x|(1 - \epsilon^2u^2) \pm p \cdot x 2m(1 - \epsilon u)^2}{4p^2u \mp i2m(\epsilon u + 1)^2}\right] \times \left(\frac{i(1 - \epsilon^2u^2)}{[4p^2u \mp i2m(\epsilon u + 1)^2]^2} + \frac{2ku[|x|\{k^2(1 + \epsilon u)^2 + p^2(1 - \epsilon u)^2\} - 2x \cdot pk(1 - \epsilon^2u^2)]}{[4p^2u \mp i2m(\epsilon u + 1)^2]^3}\right). \quad (2.7)$$

The above concrete expressions for $\phi_c^{\mp\epsilon}(x, p)$ allow us to study the behavior of the series expansion (2.1) for small ϵ. To see this, let $\epsilon > 0$ to fixed. It will be shown in the proof of Lemma 2.2 that (2.7) has an absolutely convergent series expansion for $|\gamma| < 1$. Furthermore, by Proposition A.2 (2.7) is the unique solution of (A5), which for sufficiently small $|\gamma|$ has an iterative solution (2.1). Thus by the uniqueness of the power series expansion for each $\epsilon > 0$ the series expansion (2.1) is given by the series expansion of (2.7) which is absolutely convergent for $|\gamma| < 1$.

The RCEW's corresponding to the pure Coulomb potential are given by

$$\hat{\phi}_c^{\mp}(x, p) = \left(\frac{-ik^2\phi(x, p)[\exp(\pm 2\pi\gamma) - 1]\Gamma(1 \mp i\gamma)}{\pi m[\exp(-2\pi\nu) - 1]}\right) \times \exp\left(\pm i\gamma \log\frac{\epsilon m}{2p^2} + i\nu \log\epsilon \mp \pi\gamma\right) \times \left(\gamma^{-1} \int_{\epsilon^{-1};\, \arg u = 0}^{(0+)} du\, u^{i\nu} \exp\left[\frac{\pm k|x|(1 - \epsilon^2u^2) \mp p \cdot x(1 - \epsilon u)^2}{(2p^2/m)u \mp i(\epsilon u + 1)^2}\right] \times \left\{\frac{(1 - \epsilon^2u^2)}{[(2p^2/m)u \mp i(\epsilon u + 1)^2]^2} - \frac{iku[|x|\{k^2(1 + \epsilon u)^2 + p^2(1 - \epsilon u)^2\} - 2p \cdot xk(1 - \epsilon^2u^2)]}{m[(2p^2/m)u \mp i(\epsilon u + 1)^2]^3}\right\}\right). \quad (2.8)$$

Theorem 2.1 is a consequence of the following lemma together with the pointwise convergence of the pure Coulomb RCEW's, which is verified in Lemma 2.3. In this paper we let D and D' denote arbitrary compact subsets of R^3 and $R^3 \setminus \{0\}$, respectively.

Lemma 2.2: For all $(x, p, \epsilon) \in D \times D' \times [0, \sigma]$, $\sigma > 0$, there exist constants C_l, $l = 0, 1, 2, \cdots$, such that for $|\gamma| < 1$

$$|\Lambda_l(x, p, \mp\epsilon)| \le C_l, \quad \sum_{l=0}^{\infty} |\gamma|^l C_l \equiv C < \infty. \quad (2.9)$$

Furthermore, for each l and each $(x, p) \in R^3 \times R^3 \setminus \{0\}$

$$\lim_{\epsilon \to +0} \Lambda_l(x, p, \mp\epsilon) = \Lambda_l^{\mp}(x, p) \quad (2.10)$$

exist.

Proof: By the argument given after Theorem 2.1 the perturbation series (2.3) is given by the series expansion of (2.8). We show via the Weierstress test for uniform convergence that each factor in (2.8) has an absolutely and uniformly in $(x, p, \epsilon) \in D \times D' \times [0, \sigma]$ convergent series expansion for $|\gamma| < 1$ which verifies (2.9).

It is straightforward to see that the first factor on the right side of (2.8) has a series expansion which satisfies the Weierstress criteria for uniform convergence with respect to $(x, p, \epsilon) \in D \times D' \times [0, \sigma]$ for $|\gamma| < 1$.

In order that the second factor on the right side of (2.8) satisfies the Weierstrass criteria, we must show

$$\frac{||p|/k \pm 1|^l |\log\epsilon|^l}{l!} \le \hat{C}_l, \quad \sum_{l=0}^{\infty} |\gamma|^l \hat{C}_l < \infty, \quad |\gamma| < \infty, \quad (2.11)$$

where $\hat{C}_l$, $l = 0, 1, 2, \cdots$, are independent of $(p, \epsilon) \in D' \times [0, \sigma]$. The following inequality,

$$\left|\frac{|p|}{k} \pm 1\right|^2 = \left\{1 - \frac{|p|}{(p^4 + 4m^2\epsilon^2)^{1/4}} \cos\left[\tfrac{1}{2}\sin^{-1}\left(\frac{2m\epsilon}{(p^4 + 4m^2\epsilon^2)^{1/2}}\right)\right]\right\}^2 + \frac{p^2}{(p^4 + 4m^2\epsilon^2)^{1/2}} \sin^2\left[\tfrac{1}{2}\sin^{-1}\left(\frac{2m\epsilon}{(p^4 + 4m^2\epsilon^2)^{1/2}}\right)\right] = \left(|p| \int_0^{2m\epsilon} du \frac{d}{du}\left\{-\frac{1}{(p^4 + u^2)^{1/4}} \times \cos\left[\tfrac{1}{2}\sin^{-1}\left(\frac{u}{(p^4 + 4m^2\epsilon^2)^{1/2}}\right)\right]\right\}\right)^2 + \frac{p^2}{(p^4 + 4m^2\epsilon^2)^{1/2}} \times \left(\int_0^{2m\epsilon} du \frac{d}{du}\left\{\sin\left[\tfrac{1}{2}\sin^{-1}\left(\frac{u}{(p^4 + 4m^2\epsilon^2)^{1/2}}\right)\right]\right\}\right)^2 \le \left(\frac{m\epsilon}{|p|^2} + \frac{m\epsilon}{|p|^2}\right)^2 + \left(\frac{m\epsilon}{|p|^2}\right)^2,$$

together with $|\epsilon \log\epsilon| < 1 + \sigma^2$, verifies (2.11).

The absolute and uniform in $(x, p, \epsilon) \in D \times D' \times [0, \sigma]$ convergence of the series expansion corresponding to the last factor in (2.8) is a consequence of the following inequalities:

$$\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{(0+)} |du|\,|H_l(x,p,\mp\epsilon,u)| \le M_l, \tag{2.12}$$

where the constants M_l satisfy

$$\sum_{l=1}^{\infty}\frac{|\gamma|^{l-1}}{l!}M_l<\infty, \quad |\gamma|<1, \tag{2.13}$$

with $H_l(x,p,\mp\epsilon,u)$ given by

$$H_l(x,p,\mp\epsilon,u) = (\log u)^l \exp\left[\frac{\pm k|x|(1-\epsilon^2u^2)\mp p\cdot x(1-\epsilon u)^2}{(2p^2/m)u\mp i(\epsilon u+1)^2}\right] \times\left\{\frac{(1-\epsilon^2u^2)}{[(2p^2/m)u\mp i(\epsilon u+1)^2]^2} - \frac{iku[|x|\{k^2(1+\epsilon u)^2+p^2(1-\epsilon u)^2\}-2x\cdot pk(1-\epsilon^2u^2)]}{m[(2p^2/m)u\mp i(\epsilon u+1)^2]^3}\right\}. \tag{2.14}$$

The inequalities (2.12) and (2.13) can be verified by considering the separate contributions corresponding to the right side of the following inequality:

$$\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{(0+)} |du|\,|H_l(x,p,\mp\epsilon,u)| \le \left(\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{\rho}+\int_c+\int_{\rho}^{\epsilon^{-1};\,\mathrm{arc}u=2\pi}\right)|du|\,|H_l(x,p,\mp\epsilon,u)|, \tag{2.15}$$

where c is the contour $u=\rho e^{i\theta}$, $0\le\theta\le 2\pi$, and $\rho>0$ is an appropriately chosen constant.

The first term on the right side of (2.15) is bounded by (assuming $\epsilon<1$, the case $\epsilon\ge 1$ is immediate)

$$\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{\rho}|du|\,|H_l(x,p,\mp\epsilon,u)| \le \hat{C}_1+\hat{C}_2\int_1^{\epsilon^{-1}}du(\log u)^l/u^2 + \hat{C}_3\epsilon\int_1^{\epsilon^{-1}}du(\log u)^l/u+\hat{C}_4\epsilon^2\int_1^{\epsilon^{-1}}du(\log u)^l, \tag{2.16}$$

where $\hat{C}_i$, $i=1,2,3,4$, are constants independent of $(x,p,\epsilon)\in D\times D'\times[0,\sigma]$. Estimating the integrals in (2.16) yields

$$\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{\rho}|du|\,|H_l(x,p,\mp\epsilon,u)| \le \hat{C}_1+\hat{C}_2 l!+(\hat{C}_3+\hat{C}_4)(l+1)^l\exp[-(l+1)],$$

which verifies (2.12) and (2.13) for the contribution corresponding to $\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{\rho}$. A similar argument as given above verifies (2.12) and (2.13) for the contribution $\int_{\rho}^{\epsilon^{-1};\,\mathrm{arc}u=2\pi}$. The contribution corresponding to $\int_c$ will satisfy (2.12) and (2.13) if we can show there exists a constant $\hat{C}_5>0$ such that

$$|(2p^2/m)\rho\exp(i\theta)\mp i[\epsilon\rho\exp(i\theta)+1]^2| \ge \hat{C}_5 \tag{2.17}$$

for all $(x,p,\epsilon,\theta)\in D\times D'\times[0,\sigma]\times[0,2\pi]$. In order to verify (2.17), we require the following inequality:

$$\begin{aligned}|(2p^2/m)\rho\exp(i\theta)\mp i[\epsilon\rho\exp(i\theta)+1]^2|^2 &= (2p^2/m)^2\rho^2\mp 2(2p^2/m)\rho\sin\theta+1\\ &\pm 2(2p^2/m)\epsilon^2\rho^3\sin\theta+\epsilon^4\rho^4+4\epsilon^3\rho^3\cos\theta+6\epsilon^2\rho^2\\ &-4\epsilon^2\rho^2\sin^2\theta+4\epsilon\rho\cos\theta\\ &\ge[(2p^2/m)\rho-1]^2-2(2p^2/m)\epsilon^2\rho^3-4\epsilon^3\rho^3-4\epsilon\rho.\end{aligned} \tag{2.18}$$

Since there exists a finite constant K such that $|p|<K$, it follows that $(2p^2/m)\rho<\frac{1}{2}$ if $\rho<m/4K^2$. Furthermore, since $\epsilon\in[0,\sigma]$, it follows

$$4\epsilon^3\rho^3+\epsilon^2\rho^2+4\epsilon\rho<4\sigma^3\rho^3+\sigma^2\rho^2+4\sigma\rho.$$

Thus, choosing $\rho<\min\{1/24\sigma,\ m/4K^2\}$, (2.17) follows from (2.18).

The proof of (2.10) follows from the observation that the limit $\epsilon\to+0$ of each term of the series expansion corresponding to each factor in (2.8) exists.

Lemma 2.3: The pure Coulomb RCEW's satisfy:

(i) $\lim_{\epsilon\to+0}\hat{\phi}_c^{\mp\epsilon}(x,p)=\phi_c^{\mp}(x,p)$ for each $(x,p)\in R^3\times R^3\setminus\{0\}$. (2.19)

(ii) There exists a constant Q such that for all $(x,p,\epsilon)\in D\times D'\times[0,\lambda]$, for some $\lambda>0$,

$$|\hat{\phi}_c^{\mp\epsilon}(x,p)|\le Q. \tag{2.20}$$

Outline of proof: It is straightforward to see that the first two factors in (2.8) can be bound uniformly in $(x,p,\epsilon)\in D\times D'\times[0,\sigma]$, $\sigma>0$, and, furthermore, that the limit $\epsilon\to+0$ of these factors exists. Thus, if the last factor in (2.8) can be bound by a constant for all $(x,p,\epsilon)\in D\times D'\times[0,\lambda]$ for some $\lambda>0$ and has a limit $\epsilon\to+0$, it follows that (2.20) is valid and the limit $\epsilon\to+0$ of $\phi_c^{\mp\epsilon}(x,p)$ exists.

The boundedness of the last factor in (2.8) can be proven by an analogous argument as given in the proof of (2.12). For example the contribution corresponding to $\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{\rho}$ satisfies the following inequality:

$$\begin{aligned}&\left|\gamma^{-1}\int_{\epsilon^{-1};\,\mathrm{arc}u=0}^{\rho}du\,u^{i\nu}\exp\left[\frac{\pm k|x|(1-\epsilon^2u^2)\mp p\cdot x(1-\epsilon u)^2}{(2p^2/m)u\mp i(\epsilon u+1)^2}\right]\right.\\ &\left.\left\{\frac{(1-\epsilon^2u^2)}{[(2p^2/m)u\mp i(\epsilon u+1)^2]^2}-\frac{iku[|x|\{k^2(1+\epsilon u)^2+p^2(1-\epsilon u)^2\}-2p\cdot xk(1-\epsilon^2u^2)]}{m[(2p^2/m)u\mp i(\epsilon u+1)^2]^3}\right\}\right|\\ &\le\hat{C}_1+\hat{C}_2\int_1^{\epsilon^{-1}}du\frac{u^{|\gamma|m\epsilon|p|^2}}{u^2}+\hat{C}_3\epsilon\int_1^{\epsilon^{-1}}du\frac{u^{|\gamma|m\epsilon/|p|^2}}{u}\\ &+\hat{C}_4\epsilon^2\int_1^{\epsilon^{-1}}du\,u^{|\gamma|m\epsilon/|p|^2},\end{aligned} \tag{2.21}$$

where C_i, $i=1,2,3,4$, are constants independent of $(x,p,\epsilon)\in D\times D'\times[0,\sigma]$, $\sigma>0$. Since $|p|>K'$ for some positive constant K' depending on D', it follows that $|\gamma m\epsilon/p^2|\le 1-\alpha$, $\alpha>0$, for $\epsilon\le\lambda\equiv(1-\alpha)K'^3/m^2|e_1e_2|$. Thus restricting $\epsilon\in[0,\lambda]$ guarantees the existence of a constant which bounds the right side of (2.21) for all $(x,p,\epsilon)\in D\times D'\times[0,\lambda]$. A similar argument proves the uniform boundedness of the contributions $\int_c$, $\int_{\rho}^{\epsilon^{-1};\,\mathrm{arc}u=2\pi}$, which verifies part (ii) of the lemma.

It is straightforward to show via similar estimates as given in (2.21) that the terms involving an ϵ, ϵ^2, ϵ^3, or ϵ^4 factor in the last factor in (2.8) converge to zero. Furthermore, by the Lebesgue dominated convergenve theorem the term in the last factor in (2.8) not involving a power of ϵ as a factor converges to the following expression:

$$\gamma^{-1}\int_{+\infty;\,\arg u=0}^{(0+)} du\,u^{\mp i\gamma}\exp\left[\frac{-|p|\,|x|\mp p\cdot x}{(2p^2/m)u\mp i}\right]$$
$$\times\left\{\frac{1}{[(2p^2/m)u\mp i]^2}\pm\frac{2ip^2u[|p|\,|x|\pm p\cdot x]}{m[(2p^2/m)u\mp i]^3}\right\}.$$

Thus $\lim_{\epsilon\to+0}\hat\phi_c^{\mp\epsilon}(x,p)$ exist and are given by

$$\lim_{\epsilon\to+0}\hat\phi_c^{\mp\epsilon}(x,p)=\frac{-ip^2\phi(x,p)\Gamma(1\mp i\gamma)}{\pi m}\exp\left(\pm i\gamma\log\frac{m}{2p^2}\mp\pi\gamma\right)$$
$$\times\gamma^{-1}\int_{+\infty;\,\arg u=0}^{(0+)} du\,u^{\mp i\gamma}\exp\left[\frac{-|p|\,|x|\mp p\cdot x}{(2p^2/m)u\mp i}\right]$$
$$\times\left\{\frac{1}{[(2p^2/m)u\mp i]^2}\pm\frac{2ip^2u[|p|\,|x|\pm p\cdot x]}{m[(2p^2/m)u\mp i]^3}\right\}. \tag{2.22}$$

A rather lengthy manipulation shows that the right side of (2.22) equals the pure Coulomb physical distorted plane waves $\phi_c^{\mp}(x,p)$ which completes the proof of the lemma.

III. PERTURBATION THEORY FOR THE COULOMB-LIKE DISTORTED PLANE WAVES

The renormalized perturbation series corresponding to (1.2) for general Coulomb-like potentials is defined in analogy with the pure Coulomb case. We denote the formal iteration of (A5) as follows:

$$\phi^{\mp\epsilon}(x,p)=\sum_{l=0}^{\infty}\phi_l^{\mp\epsilon}(x,p), \tag{3.1}$$

where $\phi_l^{\mp\epsilon}(x,p)$ are given by (A9). The formal renormalized perturbation series is defined as the product of (3.1) and (2.2) as follows:

$$\hat\phi^{\mp\epsilon}(x,p)=\sum_{l=0}^{\infty}\hat\Lambda_l(x,p,\mp\epsilon), \tag{3.2}$$

where

$$\hat\Lambda_l(x,p,\mp\epsilon)=\gamma^0\eta^l\Lambda_{0,l}(x,p,\mp\epsilon)+\gamma\eta^{l-1}\Lambda_{1,l-1}(x,p,\mp\epsilon)$$
$$+\cdots+\gamma^{l-1}\eta\Lambda_{l-1,1}(x,p,\mp\epsilon)+\gamma^l\eta^0\Lambda_{l,0}(x,p,\mp\epsilon), \tag{3.3}$$

with $\Lambda_{i,j}(x,p,\mp\epsilon)$ the coefficient of the term $\gamma^i\eta^j$ in the product of (3.1) and (2.2).

The convergence of the perturbation series (3.2) in the limit $\epsilon\to+0$ is shown in the following theorem.

Theorem 3.1: Assume $V(x)=\gamma(\beta|x|)^{-1}+\eta V_s(x)$, $\beta=m/|p|$, where $V_s\in L^2$ has compact support contained in the compact set D. Then for $|\gamma|<1$ and $|\eta|<\eta_1$, $\eta_1=[G_c\sup_{x\in D}|\,|V_s|x-\cdot|^{-1}|\,|_1]^{-1}$, where G_c is defined in Lemma 3.2, the series (3.2) is absolutely and uniformly in $\epsilon\in[0,\sigma]$, $\sigma>0$, convergent. Furthermore,

$$\phi^{\mp}(x,p)=\lim_{\epsilon\to+0}\hat\phi^{\mp\epsilon}(x,p)=\sum_{l=0}^{\infty}\lim_{\epsilon\to+0}\hat\Lambda_l(x,p,\mp\epsilon) \tag{3.4}$$

for each $(x,p)\in D\times D'$, where $\phi^{\mp}(x,p)$ satisfy

$$\langle\chi_D\phi\,|\,\Omega_{\pm}\chi_{D'}\psi\rangle=\int_D dx\,\overline{\phi(x)}\int_{D'}dp\,\phi^{\mp}(x,p)\,\hat\psi(p) \tag{3.5}$$

with $\phi\in L^1\cap L^2$ and $\hat\psi\in C_0$.

In order to prove the above theorem we require the following lemmas whose proofs are given at the end of this section.

Lemma 3.2: The pure Coulomb Green's functions have absolutely convergent series expansions for $|\gamma|<1$ and each $(x,x')\in R^3\times R^3$, $x\neq x'$, i.e.,

$$G_c\left(x,x';\frac{p^2}{2m}\mp i\epsilon\right)=|x-x'|^{-1}\sum_{l=0}^{\infty}\gamma^l G_c^{(l)}\left(x,x';\frac{p^2}{2m}\mp i\epsilon\right). \tag{3.6}$$

Furthermore, for all $(x,x',p,\epsilon)\in D\times D\times D'\times[0,\sigma]$ there exist constants $G_c^{(l)}$, $l=0,1,2,\cdots$, such that

$$|G_c^{(l)}(x,x';p^2/2m\mp i\epsilon)|\le G_c^{(l)},$$
$$\sum_{l=0}^{\infty}|\gamma|^l G_c^{(l)}\equiv G_c<\infty,\quad |\gamma|<1. \tag{3.7}$$

The following lemma concerns the iterations of the equations:

$$\hat\phi^{\mp\epsilon}(x,p)$$
$$=\hat\phi_c^{\mp\epsilon}(x,p)-\eta\int dy\,G_c(x,y;p^2/2m\mp i\epsilon)\,V_s(y)\,\hat\phi^{\mp\epsilon}(y,p), \tag{3.8}$$

which we denote as follows:

$$\hat\phi^{\mp\epsilon}(x,p)=\sum_{l=0}^{\infty}\eta^l K_l(x,p,\mp\epsilon), \tag{3.9}$$

where

$$K_0(x,p,\mp\epsilon)=\hat\phi_c^{\mp\epsilon}(x,p),$$
$$K_l(x,p,\mp\epsilon)=-\int dy\,G_c(x,y;p^2/2m\mp i\epsilon)V_s(y)K_{l-1}(y,p,\mp\epsilon). \tag{3.10}$$

Lemma 3.3: Assume $V_s\in L^2$ has compact support contained in D. The iterations $K_l(x,p,\mp\epsilon)$ of (3.8) satisfy for each l:

(i) $K_l(x,p,\mp\epsilon)$ have absolutely and uniformly in $\epsilon\in[0,\sigma]$, $\sigma>0$, series expansions for $|\gamma|<1$, i.e.,

$$K_l(x,p,\mp\epsilon)=\sum_{j=0}^{\infty}\gamma^j K_{j,l}(x,p,\mp\epsilon). \tag{3.11}$$

(ii) There exist constants $K_{j,l}$, $j=0,1,2,\cdots$, such that for all $(x,p,\epsilon)\in D\times D'\times[0,\sigma]$

$$|K_{j,l}(x,p,\mp\epsilon)|\le K_{j,l}. \tag{3.12}$$

Furthermore, for $|\gamma|<1$

$$\sum_{j=0}^{\infty}|\gamma|^j K_{j,l}=C[G_c\sup_{x\in D}|\,|V_s|x-\cdot|^{-1}|\,|_1]^l, \tag{3.13}$$

where C and G_c are the constants appearing in Lemmas 2.2 and 3.2, respectively.

(iii) For each j and each $(x,p)\in R^3\times R^3\setminus\{0\}$ the limit $\epsilon\to+0$ of $K_{j,l}(x,p,\mp\epsilon)$ exists, i.e.,

$$\lim_{\epsilon\to+0}K_{j,l}(x,p,\mp\epsilon)=K_{j,l}^{(\mp)}(x,p).$$

(iv) The functions $K_{j,l}^{(\mp)}(x,p)$ are continuous in $p\in D'$.

Proof of Theorem 3.1: From (3.9) and Lemma 3.3 it follows that $\phi^{\mp\epsilon}(x,p)$ have the following double series expansions for $|\gamma|<1$ and $|\eta|<\eta_1$:

$$\phi^{\mp\epsilon}(x,p)=\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\eta^l\gamma^j K_{j,l}(x,p,\mp\epsilon), \tag{3.14}$$

where there exist constants $K_{j,l}$ such that

$$\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}|\eta|^l|\gamma|^j|K_{j,l}(x,p,\mp\epsilon)|\le\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}|\eta|^l|\gamma|^j K_{j,l}<\infty \tag{3.15}$$

for all $(x,p,\epsilon)\in D\times D'\times[0,\sigma]$.

For each fixed $\epsilon>0$ the series expansion (3.2) is

absolutely convergent for sufficiently small $|\gamma|$ and $|\eta|$ (Proposition A2) and thus by the uniqueness of the double power series expansion $K_{j,l}(x, p, \mp\epsilon) = \Lambda_{j,l}(x, p, \mp\epsilon)$ for each (j,l). Thus due to the inequalities (3.15) the series expansion (3.2) is absolutely and uniformly in $\epsilon \in [0,\sigma]$ convergent for $|\gamma| < 1$, $|\eta| < \eta_1$.

The equality (3.4) follows from the uniform convergence of the series (3.2) together with part (iii) of Lemma 3.3.

The equality (3.5) follows from the representation

$$\langle \chi_D \phi \mid \Omega_\pm \chi_{D'} \psi \rangle = \lim_{\epsilon \to +0} \int_D dx\, \overline{\phi(x)} \int_{D'} dp\, \hat{\phi}^{\mp\epsilon}(x, p)\, \hat{\psi}(p),$$

$\phi \in L^1 \cap L^2$, $\hat{\psi} \in C_0$, which is a consequence of Theorem 4.1, Ref. 1.

Proof of Lemma 3.2: The following integral representations of the Whittaker functions are valid[7]:

$$W_{\kappa;\mu/2}(z) = \frac{\exp(-z/2) z^{(1-\mu)/2}}{\Gamma((1+\mu)/2-\kappa)} \times \int_0^{+\infty} \exp(-t) t^{(\mu-1)/2+\kappa} (z+t)^{(\mu-1)/2+\kappa}\, dt, \tag{3.16}$$

where $|\arg z| < \pi$ and $\mathrm{Re}[(1+\mu)/2 - \kappa] > 0$

$$M_{\kappa;\mu/2}(z) = \frac{\exp(-z/2) z^{(1+\mu)/2}}{\Gamma((1+\mu)/2+\kappa)\Gamma((1+\mu)/2-\kappa)} \times \int_0^1 \exp(zu)\, u^{(\mu-1)/2-\kappa} (1-u)^{(\mu-1)/2+\kappa}\, du, \tag{3.17}$$

where $|\arg z| < \pi$ and $\mathrm{Re}[(1+\mu)/2 \pm \kappa] > 0$.

The integral representations (3.16) and (3.17) together with the definition of the pure Coulomb Green's function (A1) yield

$$\begin{aligned}
&|x-x'|\, G_c(x,x'; p^2/2m \mp i\epsilon) \\
&= \frac{m \exp[ik(\alpha_1+\alpha_2)/2]}{2\pi\Gamma(1-i\nu)\Gamma(1+i\nu)} \Bigg\{ \int_0^{+\infty} dt\, \exp(-t) t^{i\nu} (-ik\alpha_2 + t)^{-i\nu} \\
&\times \Bigg[(1 + ik\alpha_1/2) \int_0^1 du \exp(-ik\alpha_1 u) u^{i\nu} (1-u)^{-i\nu} - ik\alpha_1 \\
&\times \int_0^1 du\, \exp(-ik\alpha_1 u) u^{i\nu+1} (1-u)^{-i\nu} \Bigg] - ik\alpha_1 \\
&\times \int_0^1 du \exp(-ik\alpha_1 u) u^{i\nu} (1-\mu)^{-i\nu} \Bigg[\tfrac{1}{2} \int_0^{+\infty} dt \exp(-t) \\
&\times t^{i\nu} (-ik\alpha_2 + t)^{-i\nu} + i\nu \int_0^{+\infty} dt \exp(-t) t^{i\nu} (-ik\alpha_2 + 1)^{-i\nu-1} \Bigg] \Bigg\}.
\end{aligned} \tag{3.18}$$

In order to verify the lemma, we must show that each factor of each term on the right side of (3.18) has a series expansion for $|\gamma| < 1$ which satisfies the Weierstrass criteria for uniform convergence with respect to $(x, x', p, \epsilon) \in D \times D \times D' \times [0,\sigma]$, from which it follows that the series expansion (3.6) satisfies (3.7).

Since the proof for each term in (3.18) is similar, we only give the proof for the last term in (3.18) which is given by

$$\frac{-mk\alpha_1 \exp[ik(\alpha_1+\alpha_2)/2] i\iota}{2\pi k\alpha_2 \Gamma(1-i\nu)\Gamma(1+i\nu)} \times \left[\int_0^1 du \exp(-ik\alpha_1 u) u^{i\nu} (1-u)^{-i\nu} \right] \times \left[ik\alpha_2 \int_0^\infty dt \exp(-t) t^{i\nu} (-ik\alpha_2 + t)^{-i\nu-1} \right]. \tag{3.19}$$

It is straightforward to show that the first factor in (3.19) has a series expansion which satisfies the Weierstrass criteria for $|\gamma| < \infty$.

From the inequalities

$$\int_0^1 du\, |\exp(-ik\alpha_1 u)|\, |\log(1/u - 1)|^l \le 2 \exp(A) l!,$$

where

$$A = \sup_{(x,x',p,\epsilon) \in D \times D \times D' \times [0,\sigma]} |k\alpha_2|$$

it follows for $|\gamma| < 1$

$$\int_0^1 du \exp(-ik\alpha_1 u) u^{i\nu} (1-u)^{-i\nu} = \sum_{l=0}^\infty \frac{\gamma^l (-i|p|/k)^l}{l!} \int_0^1 du \exp(-ik\alpha_1 u) [\log(1/u-1)]^l,$$

where

$$\left| \frac{(-i|p|/k)^l}{l!} \int_0^1 du \exp(-ik\alpha_1 u) [\log(1/u-1)]^l \right| \le 2\exp(A).$$

Thus for $|\gamma| < 1$ the second factor in (3.19) has a series expansion which converges uniformly with respect to $(x, x', p, \epsilon) \in D \times D \times D' \times [0,\sigma]$.

In order to verify the uniform convergence of the series expansion for the last factor in (3.19), we require the following inequalities where $k = |k| \exp(i\psi)$:

$$\begin{aligned}
&\int_0^{+\infty} dt \exp(-t) \left| \frac{-ik\alpha_2}{-ik\alpha_2 + t} \right| \left| \log\left(\frac{-ik\alpha_2}{t} + 1 \right) \right|^l \\
&\le \int_0^{+\infty} dt \exp(-t) \left\{ \log^2\left(\frac{|k|^2\alpha_2^2}{t^2} + \frac{2|k|\alpha_2 \sin\psi}{t} + 1 \right) + \arctan^2\left(\frac{-ik\alpha_2}{t} + 1 \right) \right\}^{l/2} \\
&\le \int_0^{+\infty} dt \exp(-t) \left\{ \log\left(\frac{A}{t} + 1 \right) + 2\pi \right\}^l \\
&\le \int_0^A dt \left\{ \log\left(\frac{A}{t} + 1 \right) + 2\pi \right\}^l + \int_A^\infty dt \exp(-t) \{\log(2) + 2\pi\}^l \\
&\le 2A \exp(2\pi) l! + \{\log(2) + 2\pi\}^l.
\end{aligned}$$

Thus for $|\gamma| < 1$

$$\int_0^{+\infty} dt \exp(-t)(-ik\alpha_2) t^{i\nu} (-ik\alpha_2 + t)^{-i\nu-1} = \sum_{l=0}^\infty \gamma^l M_l(x, x', p, \mp\epsilon),$$

where for each l there exist constants $\hat{M}_l$ such that

$$|M_l(x, x', p, \mp\epsilon)| \le \hat{M}_l, \quad \sum_{l=0}^\infty |\gamma|^l \hat{M}_l < \infty.$$

Proof of Lemma 3.3: In Sec. II it was shown that $K_0(x, p, \mp\epsilon)$ have series expansions for $|\gamma| < 1$ which satisfy

$$|K_0(x, p, \mp\epsilon)| \le \sum_{j=0}^\infty |\gamma|^j |\Lambda_j(x, p, \mp\epsilon)| \le \sum_{l=0}^\infty |\gamma|^j C_j = C$$

for all $(x, p, \epsilon) \in D \times D' \times [0, \sigma]$. Furthermore the limit $\epsilon \to +0$ of $\Lambda_j(x, p, \mp\epsilon)$ exists for each j and the functions $\Lambda_j^{(\mp)}(x, p)$ are continuous in $p \in D'$. Thus Lemma (3.3) is valid for $l = 0$.

Assume $K_{q-1}(x, p, \mp\epsilon)$ has the following series expansion

$$K_{q-1}(x, p, \mp\epsilon) = \sum_{j=0}^{\infty} \gamma^j K_{j,q-1}(x, p, \mp\epsilon), \tag{3.20}$$

where there exist constants $K_{j,q-1}$, $j = 0, 1, 2, \cdots$, such that for $|\gamma| < 1$

$$\begin{aligned} &|K_{j,q-1}(x, p, \mp\epsilon)| \leq K_{j,q-1}, \\ &\sum_{j=0}^{\infty} |\gamma|^j K_{j,q-1} = C\left(G_c \sup_{x \in D} \| |V_s|x - \cdot|^{-1}| \|_1\right)^{q-1} \end{aligned} \tag{3.21}$$

for all $(x, p, \epsilon) \in D \times D' \times [0, \sigma]$. Furthermore, assume the limit $\epsilon \to +0$ of $K_{j,q-1}(x, p, \mp\epsilon)$, $j = 0, 1, 2, \cdots$, exists and $K_{j,q-1}^{(\mp)}(x, p)$ are continuous in $p \in D'$.

From (3.10), Lemma 3.2, (3.20), and (3.21) we have

$$K_q(x, p, \mp\epsilon) = \sum_{j=0}^{\infty} \gamma^j K_{j,q}(x, p, \mp\epsilon),$$

where for $j = 0, 1, 2, \cdots$

$$\begin{aligned} K_{j,q}(x, p, \mp\epsilon) = -\sum_{k=0}^{j} \int dy \frac{G_c^{(k)}(x, y; p^2/2m \mp i\epsilon)}{|x - y|} \\ \times V_s(y) K_{(j-k),q-1}(y, p, \mp\epsilon). \end{aligned} \tag{3.22}$$

Furthermore,

$$\begin{aligned} &|K_{j,q}(x, p, \mp\epsilon)| \\ &\leq \sum_{k=0}^{j} G_c^{(k)} K_{(j-k),q-1} \sup_{x \in D} \| |V_s|x - \cdot|^{-1}| \|_1 \equiv K_{j,q} \end{aligned}$$

for all $(x, p, \epsilon) \in D \times D' \times [0, \sigma]$ where, for $|\gamma| < 1$, $K_{j,q}$ satisfy

$$\sum_{j=0}^{\infty} |\gamma|^j K_{j,q} = C\left(G_c \sup_{x \in D} \| |V_s|x - \cdot|^{-1}| \|_1\right)^q,$$

which verifies the first two parts of the lemma for $K_q(x, p, \mp\epsilon)$.

The existence of the limit $\epsilon \to +0$ of $K_{j,q}(x, p, \mp\epsilon)$ for each j follows from the existence of the limit $\epsilon \to +0$ of $G_c^{(l)}(x, y; p^2/2m \mp i\epsilon)$, $l = 0, 1, 2, \cdots$, and $K_{l,q-1}(y, p, \mp\epsilon)$, $l = 0, 1, 2, \cdots$, together with (3.22). Similarly the continuity of $K_{j,q}^{(\mp)}(x, p)$ in $p \in D'$ follows from the continuity of $\lim_{\epsilon \to +0} G_c^{(l)}(x, y; p^2/2m \mp i\epsilon)$ and $K_{l,q-1}^{(\mp)}(y, p)$ in $p \in D'$ for $l = 0, 1, 2, \cdots$ together with (3.22).

Thus the lemma is valid for $l = q$, and by induction it is valid for all l.

IV. PERTURBATION THEORY FOR THE HALF-OFF-SHELL T MATRIX

The perturbation series for the pure Coulomb renormalized half-off-shell T matrix is defined as the expansion of (1.5) in powers of γ and is denoted as follows:

$$T(p', p, \epsilon, \lambda) = \sum_{k=0}^{\infty} \gamma^{k+1} T_k(p', p, \epsilon, \lambda). \tag{4.1}$$

This expression can be obtained from the perturbation series for the physical distorted plane waves via the definition (1.5) of the renormalized half-off-shell T matrix.

In this section we show that the nth partial sum of (4.1) converges to the nth partial sum $\sum_{k=0}^{n} \gamma^{k+1} \times T_k(p', |p'|, \theta, \phi)$ for the on-energy-shell T matrix in the following sense

$$\begin{aligned} &\lim_{\epsilon \to +0} \lim_{\lambda \to +0} \pi^{-1} \int_{\Delta' \times \Delta} dp' \, dp \, \overline{\hat{\phi}(p')} \, \hat{\psi}(p) \sum_{k=0}^{n} \gamma^{k+1} T_k(p', p, \epsilon, \lambda) \\ &\quad \times \frac{\epsilon}{(p'^2/2m - p^2/2m)^2 + \epsilon^2} \\ &= \int_{\Delta'} dp' \int_0^{\eta_1 - \alpha} d\theta \sin\theta \int_0^{2\pi} d\phi \, \overline{\hat{\phi}(p')} \, \hat{\psi}(|p'|, \theta, \phi)(|p'|m) \\ &\quad \times \sum_{k=0}^{n} \gamma^{k+1} T_k(p', |p'|, \theta, \phi), \end{aligned} \tag{4.2}$$

where for $0 < \alpha < \eta_1 < \pi/2$

$$\Delta' = \{p' = (|p'|, \theta', \phi') \mid |p'| \in [0, +\infty), \ \theta' \in [\eta_1, \pi - \eta_1], \ \phi \in [0, 2\pi]\}$$

and

$$\Delta = \{p = (|p|, \theta, \phi) \mid |p| \in [0, +\infty), \ \theta \in [0, \eta_1 - \alpha], \ \phi \in [0, 2\pi]\}$$

and $\hat{\phi}, \hat{\psi} \in C_0^{\infty}(R^3 \setminus \{0\})$. The restriction $\Delta' \cap \Delta = \phi$ enables us to neglect the difficulties associated with the singular behavior of the Coulomb T matrix in the forward direction. The sense of convergence specified in (4.2) is motivated by the following representation[1] for the restricted S operator, $S(\Delta', \Delta)$, which defines the sense in which the limits $\lambda \to +0$ and $\epsilon \to +0$ are to be taken

$$\begin{aligned} \langle \phi | S(\Delta', \Delta) \psi \rangle &= \lim_{\epsilon \to +0} \lim_{\lambda \to +0} \pi^{-1} \int_{\Delta' \times \Delta} dp' \, dp \, \overline{\hat{\phi}(p')} \, \hat{\psi}(p) \\ &\quad \times T(p', p, \epsilon, \lambda) \frac{\epsilon}{(p'^2/2m - p^2/2m)^2 + \epsilon^2} \end{aligned} \tag{4.3}$$

where

$$S(\Delta', \Delta) \equiv (-2\pi i)^{-1} \Omega_+^{\Delta'} {}^* \Omega_-^{\Delta}, \quad \Omega_{\pm}^{\Delta} = \Omega_{\pm} P^{\Delta}$$

with $P^{\Delta} = \chi_{\Delta}(p)$ the characteristic function of Δ. The relation (4.2) allows us to calculate the approximate on-energy-shell physical Coulomb T matrix to any power of γ for the momentum restricted to $(|p'|, \theta, \phi) \in \Delta$ and $(|p'|, \theta', \phi') \in \Delta'$.

For the pure Coulomb potential McDowell and Coleman have calculated (1.4) (Eq. 5.4.60, Ref. 5), which yields the following expression for the renormalized half-off-shell T matrix:

$$\begin{aligned} T(p', p, \epsilon, \lambda) &= \frac{|p|\gamma}{2\pi^2 m} \exp\left(\frac{-im_1 e_2}{|p'|} \log\frac{\epsilon m}{2p'^2} - \frac{\pi\gamma}{2}\right) \Gamma(1 + i\gamma) \\ &\quad \times \Gamma\left(1 - \frac{ime_1 e_2}{|p'|}\right)^{-1} \frac{[p'^2 - (|p| + i\lambda)^2]^{i\gamma}}{[(p' - p)^2 + \lambda^2]^{1 + i\gamma}}. \end{aligned} \tag{4.4}$$

From (4.4) we obtain (4.1) for $|\gamma| < 1$ with

$$\begin{aligned} &T_k(p', p, \epsilon, \lambda) \\ &= \frac{|p|}{2\pi^2 m} [(p' - p)^2 + \lambda^2]^{-1} \sum_{l=0}^{k} \frac{\Gamma_l(|p'|, |p|)}{(k - l)!} \left(\frac{-i|p|}{|p'|} \log\frac{\epsilon m}{2p'^2} \right. \\ &\quad \left. - \frac{\pi}{2} + i \log[p'^2 - (|p| + i\lambda)^2] - i \log[(p' - p)^2 + \lambda^2]\right)^{(k-l)}, \end{aligned}$$

where $\Gamma_l(|p'|, |p|)$ are defined by

$$\Gamma(1 + i\gamma) \Gamma(1 - i\gamma|p|/|p'|)^{-1} = \sum_{l=0}^{\infty} \gamma^l \Gamma_l(|p'|, |p|).$$

It is straightforward to verify that the limit $\lambda \to +0$ appearing in (4.2) can be performed, i.e.,

$$\lim_{\lambda\to+0} \int_{\Delta'\times\Delta} dp'\,dp\,\overline{\hat\phi(p')}\,\hat\psi(p)\sum_{k=0}^{n}\gamma^{k+1}\,T_k(p',p,\epsilon,\lambda)\frac{\epsilon}{(p'^2/2m-p^2/2m)^2+\epsilon^2}$$
$$= \int_{\Delta'\times\Delta} dp'\,dp\,\overline{\hat\phi(p')}\,\hat\psi(p)\sum_{k=0}^{n}\gamma^{k+1}T_k(p',p,\epsilon,0)\ \frac{\epsilon}{(p'^2/2m-p^2/2m)^2+\epsilon^2}, \tag{4.5}$$

where

$$T_k(p',p,\epsilon,0)=\frac{|p|}{2\pi^2 m(p'-p)^2}\sum_{l=0}^{k}\frac{\Gamma_l(|p'|,|p|)}{(k-l)!}\Big[\Big(\frac{-i|p|}{|p'|}\log\frac{\epsilon m}{2p'^2}+i\log|p'^2-p^2|-i\log[(p'-p)^2]-\tfrac{1}{2}\pi\Big)^{(k-l)}$$
$$\times\chi_{[0,|p'|)}(|p|)+\Big(\frac{-i|p|}{|p'|}\log\frac{\epsilon m}{2p'^2}+i\log|p'^2-p^2|-i\log[(p'-p)^2]+\tfrac{1}{2}\pi\Big)^{(k-l)}\chi_{[|p'|,+\infty)}(|p|)\Big].$$

In order to perform the limit $\epsilon \to +0$, we make the substitution $p'^2=p^2+2m\epsilon u$ in the $|p'|$-integral on the right side of (4.5), which yields

$$\sum_{k=0}^{n}\sum_{j=0}^{k}\int_{\Delta} dp\,\hat\psi(p)\frac{\gamma^{k+1}}{(k-j)!}\frac{|p|}{2\pi^2}\int_{\eta_1}^{\pi-\eta_1}d\theta'\sin\theta'\int_0^{2\pi}d\phi'\Big\{\int_0^{+\infty}du\frac{(p^2+2m\epsilon u)^{1/2}}{u^2+1}\Gamma_j((p^2+2m\epsilon u)^{1/2},|p|)$$
$$\Big(-\frac{i|p|}{(p^2+2m\epsilon u)^{1/2}}\log\frac{\epsilon m}{2(p^2+2m\epsilon u)}+i\log 2m\epsilon u-i\log\{2p^2+2m\epsilon u-2|p|(p^2+2m\epsilon u)^{1/2}\cos\hat\theta\}-\frac{\pi}{2}\Big)^{(k-j)}\{2p^2+2m\epsilon u$$
$$-2|p|(p^2+2m\epsilon u)^{1/2}\cos\hat\theta\}^{-1}\hat\phi((p^2+2m\epsilon u)^{1/2},\theta',\phi')+\int_{-p^2/2m\epsilon}^{0}du\frac{(p^2+2m\epsilon u)^{1/2}}{u^2+1}\Gamma_j((p^2+2m\epsilon u)^{1/2},|p|)$$
$$\Big(-\frac{i|p|}{(p^2+2m\epsilon u)^{1/2}}\log\frac{\epsilon m}{2(p^2+2m\epsilon u)}+i\log 2m\epsilon|u|-i\log\{2p^2+2m\epsilon u-2|p|(p^2+2m\epsilon u)^{1/2}\cos\hat\theta\}+\frac{\pi}{2}\Big)^{(k-j)}\{2p^2+2m\epsilon u$$
$$-2|p|(p^2+m\epsilon u)^{1/2}\cos\hat\theta\}^{-1}\hat\phi((p^2+2m\epsilon u)^{1/2},\theta',\phi'),$$

where $\cos\hat\theta=(p\cdot p')/|p|\,|p'|$. The verification that the limit $\epsilon\to+0$ can be interchanged with the p, θ', and ϕ' integrals is straightforward and allows us to write

$$\lim_{\epsilon\to+0}\pi^{-1}\int_{\Delta'\times\Delta} dp'\,dp\,\overline{\hat\phi(p')}\,\hat\psi(p)\sum_{k=0}^{n}\gamma^{k+1}T_k(p',p,\epsilon,0)\frac{\epsilon}{(p'^2/2m-p^2/2m)^2+\epsilon^2}$$
$$=\int_{\Delta'}dp'\int_0^{\eta_1-\alpha}d\theta\sin\theta\int_0^{2\pi}d\phi\,\overline{\hat\phi(p')}\,\hat\psi(|p'|,\theta,\phi)\sum_{k=0}^{n}\frac{\gamma^{k+1}}{8\pi^3}\Big(\sin\frac{\hat\theta}{2}\Big)^{-2}\int_0^{+\infty}du(u^2+1)^{-1}\sum_{j=0}^{k}\Big\{\Big[i\log u-i\log\sin^2\frac{\hat\theta}{2}-\frac{\pi}{2}\Big]^{(k-j)}$$
$$+\Big[i\log u-i\log\sin^2\Big(\frac{\hat\theta}{2}\Big)+\frac{\pi}{2}\Big]^{(k-j)}\Big\}\frac{\Gamma_j(|p'|,|p'|)}{(k-j)!}. \tag{4.6}$$

The Coulomb T matrix is given by

$$T(p',p)_{|p'|=|p|}=\frac{e_1e_2}{8\pi^2p^2[\sin^2(\hat\theta/2)]^{1+i\gamma}}\frac{\Gamma(1+i\gamma)}{\Gamma(1-i\gamma)},$$

which can be rewritten as

$$T(p',p)_{|p'|=|p|}=\frac{e_1e_2}{(2\pi)^3p^2\sin^2(\hat\theta/2)}\int_0^{+\infty}du(u^2+1)^{-1}\frac{\Gamma(1+i\gamma)}{\Gamma(1-i\gamma)}\Big(\exp\Big\{\gamma\Big[i\log u-i\log\sin^2\Big(\frac{\hat\theta}{2}\Big)+\frac{\pi}{2}\Big]\Big\}$$
$$+\exp\Big\{\gamma\Big[i\log u-i\log\sin^2\frac{\hat\theta}{2}-\frac{\pi}{2}\Big]\Big\}\Big). \tag{4.7}$$

For $|\gamma|<1$, (4.7) has a series expansion in powers of γ whose nth term is related to the integrand of (4.6) by the factor $|p'|m$, which verifies (4.2).

V. THE COULOMB-LIKE HALF-OFF-SHELL T-MATRIX

In this section we discuss the renormalized perturbation series for the general Coulomb-like renormalized half-off-shell T matrix.

The renormalized Coulomb-like half-off-shell T matrix has the form (for $\lambda>0$)

$$T(p',p,\epsilon,\lambda)=\overline{F^*_{+\epsilon}(|p'|)}(p'|V_{c;\lambda}\Omega_-|p)$$
$$+F^*_{+\epsilon}(|p'|)(p'|V_{s;\lambda}\Omega_-|p), \tag{5.1}$$

where $V_{c;\lambda}(x)=\exp(-\lambda|x|)\gamma(\beta|x|)^{-1}$, $V_{s;\lambda}(x)=\exp(-\lambda|x|)\eta V_s(x)$. Assuming Theorem 3.1 and Lemma 3.3 are valid, we have for $|\gamma|<1$ and $|\eta|<\eta_1$

$$\phi^*(x,p)=\sum_{l=0}^{\infty}\lim_{\epsilon\to+0}\hat\Lambda_l(x,p,+\epsilon), \tag{5.2}$$

where each term in the above series is bounded by a constant independent of $(x,p)\in D\times D'$.

Using (5.2), we obtain

$$(p'|V_{s;\lambda}\Omega_-|p)=\sum_{l=0}^{\infty}\eta\int dx\,\phi(x,p')\exp(-\lambda|x|)V_s(x)\times\left(\lim_{\epsilon\to+0}\hat{\Lambda}_l(x,p,+\epsilon)\right),$$

where each term of the above series is a continuous function of $p'\in R^3$, $p\in D'$, and $\lambda>0$. Thus in the on-energy-shell limit, $\epsilon\to+0$, the contribution to the renormalized perturbation series corresponding to

$$\overline{F^*_{+\epsilon}(|p'|)}(p'|V_{s;\lambda}\Omega_-|p) \tag{5.3}$$

must contain $(\log\epsilon)^l$, $l=1,2,3,\cdots$, divergencies.

The above argument shows that the validity of the renormalized perturbation series for (1.5) requires that the divergencies in the series expansion of (5.3) cancel divergencies in the series expansion for the first term on the right side of (5.1). Whether such a cancellation takes place is an open question.

The following theorem shows that the last term in (5.1) does not contribute to the on-energy-shell T matrix.

Theorem 5.1: Assume $V_s\in L^1$ and there exists a constant Q such that for all $(x,p)\in D\times D'$

$$|\phi^+(x,p)|\le Q \tag{5.4}$$

and $\phi^+(x,p)$ are continuous functions of $p\in D'$. Then

$$\lim_{\epsilon\to+0}\lim_{\lambda\to+0}\pi^{-1}\int_{\Delta'\times\Delta}dp'\,dp\,\overline{\hat{\phi}(p')}\,\hat{\psi}(p)\,T(p',p,\epsilon,\lambda)\times\frac{\epsilon}{(p'^2/2m-p^2/2m)^2+\epsilon^2}$$
$$=\lim_{\epsilon\to+0}\lim_{\lambda\to+0}\pi^{-1}\int_{\Delta'\times\Delta}dp'\,dp\,\overline{\hat{\phi}(p')}\,\hat{\psi}(p)\,\overline{F^*_{+\epsilon}(|p'|)}\times(p'|V_{c;\lambda}\Omega_-|p)\frac{\epsilon}{(p'^2/2m-p^2/2m)+\epsilon^2}, \tag{5.5}$$

where $\hat{\phi},\hat{\psi}\in C_0^\infty(R^3\backslash\{0\})$.

Proof: We must show

$$\lim_{\epsilon\to+0}\lim_{\lambda\to+0}A(\epsilon,\lambda)=0,$$
$$A(\epsilon,\lambda)\equiv\pi^{-1}\int_{\Delta'\times\Delta}dp'\,dp\,\overline{\hat{\phi}(p')}\,\hat{\psi}(p)\,\overline{F^*_{+\epsilon}(|p'|)}(p'|V_{s;\lambda}\Omega_-|p)\times\overline{(p'^2/2m}\,\frac{\epsilon}{p^2/2m)^2+\epsilon^2} \tag{5.6}$$

for $\phi,\psi\in C_0^\infty(R^3\backslash\{0\})$. Setting $p^2=p'^2-2m\epsilon u$, we obtain

$$A(\epsilon,\lambda)=\frac{m}{\pi}\int_{\Delta'}dp'\,\overline{\hat{\phi}(p')}\,\overline{F^*_{+\epsilon}(|p'|)}\int_0^{\eta_1-\alpha}d\theta\sin\theta\int_0^{2\pi}d\phi\times\int_{-\infty}^{p'^2/2m\epsilon}du(p'^2-2m\epsilon u)^{1/2}(u^2+1)^{-1}$$
$$\times\hat{\psi}((p'^2-2m\epsilon u)^{1/2},\theta,\phi)\times(p'|V_{s;\lambda}\Omega_-|(p'^2-2m\epsilon u)^{1/2},\theta,\phi).$$

In order to verify (5.6), we require the following inequality:

$$|A(\epsilon,\lambda)|\le\frac{m}{\pi}\int_{\Delta'}dp'\,|\hat{\phi}(p')|\left|\Gamma\left(1-\frac{ime_1e_2}{|p'|}\right)^{-1}\right|\int_0^{\eta_1-\alpha}d\theta\sin\theta$$
$$\times\int_0^{2\pi}d\phi\left|\int_{-\infty}^{p'^2/2m\epsilon}du(u^2+1)^{-1}(p'^2-2m\epsilon u)^{1/2}\right.$$
$$\times\hat{\psi}((p'^2-2m\epsilon u)^{1/2},\theta,\phi)(p'|V_{s;\lambda}\Omega_-|(p'^2-2m\epsilon u)^{1/2},\theta,\phi)$$
$$-\pi|p'|\hat{\psi}(|p'|,\theta,\phi)(p'|V_{s;\lambda}\Omega_-||p'|,\theta,\phi)|$$
$$+m\left|\int_{\Delta'}dp'\,\overline{\hat{\phi}(p')}\,\overline{F^*_{+\epsilon}(|p'|)}\right||p'|\int_0^{\eta_1-\alpha}d\theta\sin\theta$$
$$\times\left.\int_0^{2\pi}d\phi\,\hat{\psi}(|p'|,\theta,\phi)(p'|V_{s;\lambda}\Omega_-||p'|,\theta,\phi)\right|. \tag{5.7}$$

Due to assumption (5.4) there exists a constant $\hat{Q}$ such that

$$|(p'|V_{s;\lambda}\Omega_-|(p'^2-2m\epsilon u)^{1/2},\theta,\phi)|\le\hat{Q}. \tag{5.8}$$

Furthermore, due to the continuity of $\phi^+(x,p')$ in $p'\in D'$ we have

$$\lim_{\epsilon\to+0}\lim_{\lambda\to+0}|(p'|V_{s;\lambda}\Omega_-|(p'^2-2m\epsilon u)^{1/2},\theta,\phi)-(p'|V_s\Omega_-||p'|,\theta,\phi)|=0.$$

Thus by the Lebesgue dominated convergence theorem the limit $\lambda\to+0$ and $\epsilon\to+0$ of the first term on the right side of (5.7) is zero. Furthermore, by (5.8)

$$|p'|\hat{\phi}(p')\Gamma(1-ime_1e_2/|p'|)^{-1}\int_0^{\eta_1-\alpha}d\theta\sin\theta\times\int_0^{2\pi}d\phi\,\hat{\psi}(|p'|,\theta,\phi)(p'|V_s\Omega_-||p'|,\theta,\phi)\in L^1(R^3).$$

Thus by a similar argument as used to prove the Riemann–Lebesgue lemma the last term in (5.7) converges to zero as $\epsilon\to+0$, which proves the theorem.

Remark: It must be emphasized that, while the last term in (5.1) does not contribute to the on-energy-shell T matrix, the series expansion of this term may cancel divergencies in the series expansion of the first term on the right side of (5.1).

VI. DISCUSSION

In this paper we have shown that the recently formulated renormalized off-energy-shell scattering formalism[1] leads to natural perturbation expansions for the Coulomb distorted plane waves and T matrix. A two-step prescription has been presented for calculating the on-energy-shell Coulomb T matrix. The approximate physical distorted plane wave is calculated first via the renormalized perturbation series for the RCEW. The approximate physical distorted plane wave is then used to generate the perturbation series for the renormalized half-off-shell T matrix, which yields the approximate physical T matrix.

The results of this paper suggest that the stationary renormalization terms appearing in the definition of the renormalized off-energy-shell formalism[1] cancel the divergencies encountered in perturbation theory based on the short-range off-energy-shell equations. The validity of a three-particle renormalized perturbation theory based on the RCEW's and renormalized half-off-shell T matrix is of particular importance for Coulomb scattering. The case of three particles interacting via pure Coulomb potentials is of practical importance since

closed form expressions for the kernels of the Faddeev equations can be derived. Thus a prescription for canceling the divergencies in the iterations of the off-energy-shell Faddeev equations would yield a method for calculating the physical three-particle T matrices.

In a recent paper by van Haeringen[8] a stationary Coulomb scattering formalism based on Coulombian asymptotic states was derived together with a discussion of the energy-shell limit of the renormalized off-energy-shell T matrix [Eq. (51), Ref. 8]. The validity of a perturbation approach for the renormalized off-energy-shell T matrix has not been considered in this paper.

APPENDIX: THE COULOMB-LIKE CEW'S

In this appendix we summarize various properties of the Coulomb-like Green's function and CEW's which are required in this paper.

The pure Coulomb Green's function has been explicitly calculated[9] and is given by

$$G_c\left(x,x';\frac{p^2}{2m}\mp i\epsilon\right) = \frac{m\Gamma(1+i\nu)}{2\pi|x-x'|}\{W_{-i\nu;1/2}(-ik\alpha_2) \times \dot{M}_{-i\nu;1/2}(-ik\alpha_1) - M_{-i\nu;1/2}(-ik\alpha_1)\dot{W}_{-i\nu;1/2}(-ik\alpha_2)\}, \tag{A1}$$

where $\alpha_2 = |x| + |x'| + |x-x'|$ and $\alpha_1 = |x| + |x'| - |x-x'|$ and the dot denotes differentiation with respect to the argument of the Whittaker functions.

The following theorem summarizes various properties of the Coulomb-like Green's function which are proven in Ref. 4.

Theorem A.1: Assume V_s satisfies (1.7). Then for each $\epsilon > 0$ and $p \in R^3\backslash\{0\}$:

(i) The full Green's function $G(x,x';p^2/2m\mp i\epsilon)$ exists.

(ii) There exist finite constants $K_1(p,\epsilon)$ and $K_2(p,\epsilon)$, depending on $\epsilon > 0$ and $p \in R^3\backslash\{0\}$, such that

$$\int dy\,|G(x,y;p^2/2m\mp i\epsilon)| \le K_1(p,\epsilon), \tag{A2}$$

$$\int dy\,|G(x,y;p^2/2m\mp i\epsilon)V_s(y)| \le K_2(p,\epsilon) \tag{A3}$$

for all $x \in R^3$.

(iii) $G(x,x';p^2/2m\mp i\epsilon)$ satisfy the following equations:

$$\begin{aligned} G(x,x';p^2/2m\mp i\epsilon) &= G_0(x,x';p^2/2m\mp i\epsilon) - \int dy\, G_0(x,y;p^2/2m\mp i\epsilon)V(y)\,G(y,x';p^2/2m\mp i\epsilon) \\ &= G_0(x,x';p^2/2m\mp i\epsilon) - \int dy\, G(x,y;p^2/2m\mp i\epsilon) \times V(y)\,G_0(y,x';p^2/2m\mp i\epsilon) \\ &= G_c(x,x';p^2/2m\mp i\epsilon) - \int dy\, G_c(x,y;p^2/2m\mp i\epsilon) \times V_s(y)\,G(y,x';p^2/2m\mp i\epsilon) \\ &= G_c(x,x';p^2/2m\mp i\epsilon) - \int dy\, G(x,y;p^2/2m\mp i\epsilon) \times V_s(y)\,G_c(y,x';p^2/2m\mp i\epsilon) \end{aligned} \tag{A4}$$

for almost all $(x,x') \in R^3\times R^3$.

Proposition A.2: Assume $V(x) = \gamma(\beta|x|)^{-1} + \eta V_s(x)$, $\beta = m/|p|$, where V_s satisfies (1.7). Then for each fixed $\epsilon > 0$, $p \in R^3\backslash\{0\}$, $\phi^{\mp\epsilon}(x,p)$ defined by (1.1) satisfy:

(i) $\phi^{\mp\epsilon}(\cdot,p) \in L^\infty(R^3)$.

(ii) $\phi^{\mp\epsilon}(x,p)$ are the unique solutions in L^∞ of each of the following equations:

$$\phi^{\mp\epsilon}(x,p) = \phi(x,p) - \int dy\, G_0(x,y;p^2/2m\mp i\epsilon)V(y)\phi^{\mp\epsilon}(y,p), \tag{A5}$$

$$\phi^{\mp\epsilon}(x,p) = \phi_c^{\mp\epsilon}(x,p) - \eta\int dy\, G_c(x,y;p^2/2m\mp i\epsilon)V_s(y) \times \phi^{\mp\epsilon}(y,p). \tag{A6}$$

(iii) $$\begin{aligned} \phi^{\mp\epsilon}(x,p) &= \phi(x,p) - \int dy\, G(x,y;p^2/2m\mp i\epsilon) \times V(y)\phi(y,p), \\ &= \phi_c^{\mp\epsilon}(x,p) - \eta\int dy\, G(x,y;p^2/2m\mp i\epsilon) \times V_s(y)\,\phi_c^{\mp\epsilon}(y,p). \end{aligned}$$

(iv) There exist positive finite constants $\hat\gamma$ and $\hat\eta$, depending on $\epsilon > 0$ and $p \in R^3\backslash\{0\}$, such that the iterations of (A5) are absolutely convergent and satisfy (A5) for $|\gamma| < \hat\gamma$ and $|\eta| < \hat\eta$.

Proof: Parts (i) and (iii) and the validity of the off-energy-shell equations of part (ii) follow immediately from Theorem A.1.

We now verify that the solutions of (A5) are unique in $L^\infty(R^3)$. Thus we must show if $f \in L^\infty$ satisfies

$$f(x) = -\int dy\, G(x,y;p^2/2m\mp i\epsilon)V(y)f(y), \tag{A7}$$

then $f \equiv 0$.

Due to (1.7) there exists a constant $\hat C_1$ such that $|V^{(2)}(x)| \le \hat C_1|x|^{-1}$. The following inequalities,

$$\begin{aligned} &\int dx\,|x|^{-2}\Big[\int dy\,\exp[-\mathrm{Im}k\,|x-y|]|x-y|^{-1}|V(y)|\Big]^2 \\ &\quad\le \int dx\,|x|^{-2}\Big[\hat C_2\{\exp[-\mathrm{Im}k\,|x|]-1\}|x|^{-1} \\ &\qquad + \hat C_3\int dy\,\exp[-\mathrm{Im}k\,|x-y|]|x-y|^{-1}\,|V^{(1)}(y)|\Big]^2 < \infty, \end{aligned}$$

where $\hat C_2$ and $\hat C_3$ are constants, show that $Vf \in L^2$. It follows $f = (H_0-\zeta)^{-1}Vf \in D(H_0)$, $\zeta = p^2/2m\mp i\epsilon$, i.e., $(H-\zeta)f = 0$. However, H is self-adjoint and $\epsilon > 0$; thus $f \equiv 0$.

A similar argument verifies the uniqueness of solutions to Eqs. (A6).

Denote the formal iterations of (A5) as follows:

$$\phi^{\mp\epsilon}(x,p) = \sum_{l=0}^{\infty}\phi_l^{\mp\epsilon}(x,p), \tag{A8}$$

where

$$\begin{aligned} \phi_0^{\mp\epsilon}(x,p) &= \phi(x,p), \\ \phi_l^{\mp\epsilon}(x,p) &= -\int dy\, G_0(x,y;p^2/2m\mp i\epsilon)V(y)\phi_{l-1}^{\mp\epsilon}(y,p). \end{aligned} \tag{A9}$$

Due to (1.7) we have

$$V(x) = \gamma V_c^{(1)}(x) + \gamma V_c^{(2)}(x) + \eta V_s^{(1)}(x) + \eta V_s^{(2)}(x),$$

where $V_c^{(1)}$, $V_s^{(1)} \in L^1 \cap L^2$ and $V_c^{(2)}$, $V_s^{(2)} \in L^\infty$. Thus we obtain for each l

$$|\phi_l^{\mp\epsilon}(x,p)| \leq [\{\gamma||V_c^{(2)}||_\infty + \eta||V_s^{(2)}||_\infty\}||G_0(\cdot)||_1 + \{\gamma||V_c^{(1)}||_2 + \eta||V_s^{(1)}||_2\}||G_0(\cdot)||_2]^l.$$

Clearly for each fixed $\epsilon > 0$ and $p \in R^3 \setminus \{0\}$ there exist $\hat{\gamma}$ and $\hat{\eta}$ such that

$$\{\gamma||V_c^{(2)}||_\infty + \eta||V_s^{(2)}||_\infty\}||G_0(\cdot)||_1 + \{\gamma||V_c^{(1)}||_2 + \eta||V_s^{(1)}||_2\}||G_0(\cdot)||_2 < 1$$

for $|\gamma| < \hat{\gamma}$ and $|\eta| < \hat{\eta}$, which proves part (iv) of the proposition.

[1] J. Zorbas, J. Math. Phys. **18**, 1112 (1977).
[2] L. Hostler, J. Math. Phys. **5**, 1235 (1964).
[3] M.R.C. McDowell and J.P. Coleman, *Introduction to the Theory of Ion–Atom Collisions* (North-Holland, Amsterdam, 1970).
[4] J. Zorbas, "Two and three particle Coulomb Green's function," unpublished (1976).
[5] J.D. Dollard, J. Math. Phys. **5**, 729 (1964).
[6] In this paper we set $\hbar = 1$. In order to obtain (2.6) from Hostler's expression [Eqs. (5) and (13) of Ref. 2], we note that $G(x,x';p^2/2m \mp i\epsilon)$ corresponding to $(H-\zeta)^{-1}$, $\zeta = p^2/2m \mp i\epsilon$, are related to the Green's function $\hat{G}(x,x';p^2/2m \mp i\epsilon)$ determined by Hostler [Eq. (4), Ref. 2] as follows: $G(x,x';p^2/2m \mp i\epsilon) = -(2m)\hat{G}(x,x';p^2/2m \mp i\epsilon)|_{Ze^2/4\pi = -e_1e_2}$.
[7] H. Buchholz, *The Confluent Hypergeometric Function*, Springer Tracts in Natural Philosophy, Vol. 15 (Springer-Verlag, Berlin, 1969).
[8] H. van Haeringen, J. Math. Phys. **17**, 995 (1976).
[9] L. Hostler, J. Math. Phys. **5**, 591 (1964).

Forms of all spacetime metrics which admit [(11) (11)] Killing tensors with nonconstant eigenvalues[a)]

I. Hauser and R. J. Malhiot

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 8 June 1977)

Forms are obtained for all of the spacetime metrics which admit or are conformal to those which admit Killing tensors whose Segre characteristics are [(11) (11)] and whose eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ are not constants. No *a priori* assumptions are made concerning separability, isometries, invertibility, Petrov type, or the matter tensor. Some of our metrical forms, viz., those for which the Hamilton–Jacobi equation is separable or partially separable after multiplication by an integrating factor, are already known; in particular, they include *all* of Carter's Hamilton–Jacobi separable spacetimes. Also, some of our metrical forms are new and may be applicable to finding interesting nonvacuum metrics which admit [(11) (11)] Killing tensors, but which do not necessarily have any Killing vectors. Though our results include a class of metrics for which $\lambda^{(1)}$ and/or $\lambda^{(2)}$ are constant, they do not include all such metrics; in an appendix, we prove that the set of all metrics which admit [(11) (11)] Killing tensors with constant $\lambda^{(1)}$ and $\lambda^{(2)}$ is identical *except for conformal factors* with the set of *all* metrics which admit [(11) (11)] conformal Killing tensors.

1. INTRODUCTION

A. Carter's Hamilton–Jacobi separable spacetimes

The objective of this paper is a nontrivial generalization of Carter's[1] family of Hamilton–Jacobi (H–J) separable spacetimes.[2,3] It will serve that objective to start with a review of the general structure of these spacetimes and of some key points which have not been adequately stressed in the literature.

Carter introduced his family of spacetimes in 1967–68. The idea had come from his observations[4] on the separability of the Hamilton–Jacobi equation for the geodesic orbits of the Kerr metric. Like the Kerr metric, each of Carter's H–J separable spacetimes admits a coordinate system x^α such that the Hamiltonian for geodesic orbits takes the form[1,5,6]

$$\tfrac{1}{2}g^{\alpha\beta}p_\alpha p_\beta = (\lambda^{(2)}-\lambda^{(1)})^{-1}(H_1+H_2), \tag{1}$$

where

$$2H_1 = F^{(1)}p_1^2 + [G^{(1)}]^{-1}(p_3 - J^{(1)}p_4)^2,$$

$$2H_2 = F^{(2)}p_2^2 + [G^{(2)}]^{-1}(p_4 - J^{(2)}p_3)^2.$$

The fields $\lambda^{(i)}$, $F^{(i)}$, $G^{(i)}$, $J^{(i)}$ depend at most on only one coordinate x^i $(i=1,2)$. This and the structure of the Hamiltonian H imply that

$$H,\quad A=\delta_3^\alpha p_\alpha,\quad B=\delta_4^\alpha p_\alpha,$$

$$K \equiv \tfrac{1}{2}K^{\alpha\beta}p_\alpha p_\beta = (\lambda^{(2)}-\lambda^{(1)})^{-1}(\lambda^{(2)}H_1+\lambda^{(1)}H_2) \tag{2}$$

constitute a set of linear and quadratic forms whose six mutual Poisson brackets all vanish identically; the *set is irreducible* in the sense that K is not expressible as a linear combination with constant coefficients of the quadratic forms A^2, B^2, AB, H.

The above description of the Carter family in terms of first integrals of geodesic motion has its equivalent in terms of symmetries of the metric. Specifically, each Carter spacetime admits a pair of Killing vectors A^α, B^α [δ_3^α, δ_4^α, in the coordinate system for which Eq. (1) holds] and a Killing tensor[7] $K^{\alpha\beta}$ such that

(1) the Lie derivatives of B^α and $K^{\alpha\beta}$ with respect to A^α are zero;

(2) the Abelian isometry group G_2 which is generated by A^α, B^α has nonnull surfaces of transitivity and is invertible;

(3) the Segre characteristic of $K^{\alpha\beta}$ is [(11)(11)];

(4) $K^{\alpha\beta}$ is not *reducible by* G_2, which means that it is not equal to any linear combination of $A^\alpha A^\beta$, $B^\alpha B^\beta$, $A^{(\alpha}B^{\beta)}$, $g^{\alpha\beta}$, with constant coefficients.

It is essential to keep in mind some key points concerning the known solutions[1,8–10] which are in Carter's H–J separable family and which, therefore, have the symmetries described above. Recall that any Killing tensor whose Segre characteristic is [(11)(11)] has two linearly independent real null eigenvectors k_α, m_α which are shear free and geodesic.[2,11] Therefore, by the Goldberg–Sachs theorem, every vacuum solution and every electrovac whose Maxwell field has k_α, m_α as eigenvectors must be conformally flat or type D.[12] Also, if a given spacetime is in Carter's family and if the coordinate system of Eq. (1) is chosen, the Schrödinger equation is separable if and only if the component R_{12} of the Ricci tensor vanishes.[13]

All of the vacuum metrics and all electrovac augmentations of these metrics are, in fact, known and were computed by Carter, with quite a few new results.[1,8] The Kerr–NUT is the most general type D vacuum metric[14] in the H–J separable family. Also, there is Wahlquist's[10] type D rigidly rotating perfect fluid solution which has the Kerr–NUT as its vacuum subcase and which is manifestly in the Schrödinger separable subset of the Carter family.

The glaring exceptions[15] to the rule that type D vacuums[14] belong in the Carter family are the C metric and its generalization, the C–NUT which was discovered by Kinnersley[9] and which is the most general type D vacuum metric, from which all others can be obtained by limiting processes. However, as Plebański[16] and

a)Work supported in part by the National Science Foundation under Grant No PHY 75-08750.

 0022-2488/78/1901-0187$1.00

Demiański[16] have made clear by the elegant form of their seven parameter electrovac augmentation of the C—NUT, *this metric and its electrovac augmentation are each conformal to a Schrödinger separable member of the Carter H—J separable family. So, it is not an exception provided we extend the Carter family by including a conformal factor in Eq.* (1). *For the electrovac augmentation of* C—NUT, *this conformal factor is* $1+|\lambda^{(1)}\lambda^{(2)}|^{1/2}$.

One possible direction of further research on the H—J separable family is to look into nonvacuum solutions for which $R_{12}\neq 0$ and which are, therefore, not Schrödinger separable. Our interest overlaps that one but is not the same.

B. Objective

The property of Carter's H—J separable family which distinguishes it from the broader set of all spacetimes which admit a two-parameter Abelian group G_2 is the existence of a [(11)(11)] KT (Killing tensor). It has occurred to us, as it has unquestionably occurred to others, that here is a place to study symmetry breaking by dropping the assumption that any isometries exist and retaining only the existence of a [(11)(11)] KT.

In view of the facts that the vacuum solutions and their electrovac augmentations are all known[1,8] and all possess a G_2, there is nothing new to be found in that ball park. Nor is there any guarantee that a physically interesting metric will be found in our contemplated generalization of the H—J separable family. We have no evidence for that. However, the search for physically plausible matter tensors which permit the hidden symmetry characterized by a [(11)(11)] KT is still an open field, and even a negative conclusion to that search will be of interest provided it comes at the end of a fairly thorough hunt. We want to set up tools for that hunt in this paper.

In a previous paper,[2] the authors found a general metrical form for those spacetimes which admit [(11)(11)] Killing tensors whose eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ satisfy the condition that $d\lambda^{(1)}\wedge d\lambda^{(2)}$ not be zero or null. They obtained a result which is exactly like Eq. (1) except that $F^{(i)}$, $G^{(i)}$, and $J^{(i)}$ can depend on x^{i+1} as well as on x^i $(i=1,2)$. The dependences are not arbitrary, but they were not determined in that paper since its objectives were different.

There remain the nontrivial tasks of extending our results to the case where $d\lambda^{(1)}\wedge d\lambda^{(2)}$ is a null 2-form, of clarifying some important parts of our previous derivation, and of obtaining explicit results for the dependences on x^3 and x^4. However, though we want to generalize the H—J separable family by dropping the isometry premise, we must stress that we are not going to embrace *all* spacetimes which admit [(11)(11)] Killing tensors. For, as we prove in the Appendix, *the set of all metrics which admit* [(11)(11)] *Killing tensors with constant* $\lambda^{(1)}$ *and* $\lambda^{(2)}$ *is identical except for conformal factors with the set of all metrics which admit* [(11)(11)] *conformal Killing tensors* {which, of course, automatically includes *all* [(11)(11)] Killing tensors}. This is an appreciably broader family than the one which we are prepared to determine.

Our objective is the more limited one of finding the metrics of all spacetimes which admit or are conformal to those which admit nonsingular [(11)(11)] *Killing tensors*; by a *nonsingular* [(11)(11)] KT, we mean one whose eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ are neither of them identically constant. No *a priori* assumptions are made concerning separability, the existence of isometries, invertibility,[17] Petrov—Pirani type, or the matter tensor.

This objective should not be confused with a different extension of Carter's H—J separable family which has been considered by Dietz,[3] on the basis of Woodhouse's[3] separability theorems. Dietz constructed the entire family of metric tensors for which the Hamilton—Jacobi equation for geodesic orbits becomes separable or partially separable after it is multiplied by an integrating factor. The result is a family of metrics which overlaps ours but is not the same, and the overlap does not include all of our metrical forms; in particular, it does not include those of our metrical forms which arouse our greatest interest and curiosity. The details will be given in Sec. 4.

Our results fall into four subfamilies which are each characterized by different dependences of the metric components on the *post-ignorable coordinates* x^3 and x^4. For each of these subfamilies, there are three distinct cases which are labeled by a trichotomic variable ϵ which equals 1, −1, or 0 depending on whether the 2-surfaces of constant x^1 and x^2 are timelike, spacelike, or null. As in the H—J separable family of Carter, the metric is a functional of eight fields $\lambda^{(i)}$, $F^{(i)}$, $G^{(i)}$, $J^{(i)}$ some of which may be redundant or identically zero in a given subfamily. As before, $\lambda^{(i)}$ depends at most on x^i but $F^{(i)}$, $G^{(i)}$, $J^{(i)}$ can now depend on both x^i and x^{i+1} $(i=1,2)$. In one of the four subfamilies, which are detailed in Sec. 3, the dependence on x^{i+1} is specific and involves two parameters; this is the subfamily which includes the Kerr metric. In each of the other subfamilies, some of the metrical fields depend arbitrarily on x^{i+1}, and some are independent of x^{i+1}.

The general form of our results independent of the particular subfamily and not including any details on x^3, x^4 dependence is given by[18]

$$\begin{aligned}
&\tfrac{1}{2}g^{\alpha\beta}p_\alpha p_\beta = e^{-2\sigma}(\lambda^{(2)}-\lambda^{(1)})^{-1}(H_1+H_2),\\
&2H_1 = F^{(1)}p_1^2+[G^{(1)}]^{-1}(p_3-J^{(1)}p_4)^2,\\
&2H_2 = \epsilon\{F^{(2)}p_2^2-[G^{(2)}]^{-1}(p_4-J^{(2)}p_3)^2\}\\
&\qquad +2(1-\epsilon^2)(F^{(2)}/G^{(2)})^{1/2}p_2(p_4-J^{(2)}p_3).
\end{aligned}\tag{3}$$

If the conformal factor $\exp(-2\sigma)$ is unity, then there is a [(11)(11)] KT given by Eqs. (2) and (3).

When $\epsilon=0$, note that $F^{(2)}/G^{(2)}$ in H_2 actually contains only one independent field; we leave it in the form of a ratio, because that is how it occurs in our calculations. When $\epsilon=\pm 1$, and the conformal factor $\exp(-2\sigma)$ is unity, and all metrical fields are independent of x^3 and x^4, then Eqs. (3) reduce to Carter's H—J separable form as given by Eqs. (1).[19]

To avoid any misunderstanding, we repeat that all Carter H—J separable metrics[1] are contained in our

family. That includes H–J separable metrics whose [(11)(11)] Killing tensors are singular. To obtain the singular cases, let $\exp(-2\sigma)=[\lambda^{(2)\prime}-\lambda^{(1)\prime}]^{-1}[\lambda^{(2)}-\lambda^{(1)}]$, where $\lambda^{(1)\prime}$ or/and $\lambda^{(2)\prime}$ are constants, or simply let the conformal factor be unity and replace $\lambda^{(1)}$ or/and $\lambda^{(2)}$ by constants. Examples of H–J separable metrics which admit singular [(11)(11)] Killing tensors are the Schwarzschild metric and all nondiverging type D vacuums.[1,8]

The existence of a coordinate system which gives us the canonical structure of Eq. (3) is proven in Sec. 2. In Sec. 3, we solve the affine connection equations to determine exactly what can be said about the coordinate dependences of the metrical components. In Sec. 4, we summarize our results, give a convenient form of the line element, and relate this form to the charged Kerr metric in Boyer–Lindquist coordinates.[4]

We open Sec. 2 with some derivations which were introduced in a previous paper[2] by the authors. The treatment here differs, however, in some important respects. First, we now include the case where the 2-surfaces of constant $\lambda^{(1)}$ and $\lambda^{(2)}$ are null, which we did not do previously and which involves some special treatment. Second, we previously used the eigenvalues $\lambda^{(i)}$ as coordinates, because we were not interested in the singular cases at that time. We now use generally applicable coordinates x^1, x^2, and we replace our former notations ϕ^3, ϕ^4 by x^3, x^4. Third, our derivation here is an appreciable improvement of our previous one and has been boiled down to essentials as far as we can tell.

2. CHOICE OF NULL TETRAD AND COORDINATES

We first introduce a few notations.

k, m, t, t^* denote any null tetrad consisting of real 1-forms k, m and complex 1-forms t, t^* such that t^* is the c.c. of t, and $k\circ m=t\circ t^*=1$. We define real connection forms P,Q and complex connection forms v,w by the equations

$$P\equiv dx^\alpha(\nabla_\alpha k_\beta)m^\beta,\quad iQ\equiv dx^\alpha(\nabla_\alpha t_\beta)t^{*\beta},$$
$$v\equiv dx^\alpha(\nabla_\alpha k_\beta)t^\beta,\quad w\equiv dx^\alpha(\nabla_\alpha m_\beta)t^{*\beta}. \tag{4}$$

Then, the covariant derivatives of the null tetrad members are given by

$$\nabla_\alpha k_\beta=P_\alpha k_\beta+v^*{}_\alpha t_\beta+v_\alpha t^*{}_\beta,$$
$$\nabla_\alpha m_\beta=-P_\alpha m_\beta+w_\alpha t_\beta+w^*{}_\alpha t^*{}_\beta, \tag{5}$$
$$\nabla_\alpha t_\beta=-w^*{}_\alpha k_\beta-v_\alpha m_\beta+iQ_\alpha t_\beta,$$

and the exterior derivatives are given by

$$dk=Pk+v^*t+vt^*,$$
$$dm=-Pm+wt+w^*t^*, \tag{6}$$
$$dt=-w^*k-vm+iQt.$$

Note that we follow the convention of suppressing the symbol $\wedge$ in exterior products of forms; e.g., $d\wedge k=dk$ and $P\wedge k=Pk$.

The script variables which we use for coordinate tetrad components of tensors are

$\alpha,\beta,\cdots$ with values 1,2,3,4,

i,j with values 1,2,

r,s with values 3,4.

For null tetrad components, we use the script variables[20]

$a,b,\cdots$ with values k, m, t, n, where $n=t^*$.

The components v_k, v_t, $\mathrm{Re}v_n$, $\mathrm{Im}v_n$ are the geodesy, shear, divergence, and twist of the null congruence defined by k; the corresponding optical parameters for m are w_m, w_n, $\mathrm{Re}w_t$, $\mathrm{Im}w_t$. As a note of caution, we always use the convention $v^*{}_a=(v_a)^*$.

The first step in the derivation is the selection of a null tetrad. Since the Segre characteristic of $K_{\alpha\beta}$ is assumed to be [(11)(11)], the KT has eigenvectors k_α, m_α, t_α, $t^*{}_\alpha$ and eigenvalues $\lambda^{(1)}$, $\lambda^{(2)}$ such that

$$K_{\alpha\beta}=\lambda^{(1)}(k_\alpha m_\beta+m_\alpha k_\beta)+\lambda^{(2)}(t_\alpha t^*{}_\beta+t^*{}_\alpha t_\beta). \tag{7}$$

The defining equation for a KT is

$$\nabla_{(\gamma}K_{\alpha\beta)}=0. \tag{8}$$

From Eqs. (5), we derive the following conditions which are collectively equivalent to Eqs. (7) and (8):

$$v_k=v_t=w_m=w_n=0, \tag{9}$$
$$d\lambda^{(1)}=\rho^2[(w_k+v_m^*)t+(w_k^*+v_m)t^*],$$
$$d\lambda^{(2)}=\rho^2[(w_t+w^*{}_t)k+(v_n+v_n^*)m], \tag{10}$$

where we let

$$\rho^2=\lambda^{(2)}-\lambda^{(1)}. \tag{11}$$

Note that $d\lambda^{(1)}\wedge d\lambda^{(2)}=0$ if and only if $d\lambda^{(1)}=0$ or $d\lambda^{(2)}=0$. If $d\lambda^{(1)}$ or $d\lambda^{(2)}$ is identically zero, the KT can always be modified by adding a suitable constant multiple of the metric tensor so that $\lambda^{(1)}$ or $\lambda^{(2)}$ becomes identically zero. For this reason, we say that the KT is *singular* if $d\lambda^{(1)}\wedge d\lambda^{(2)}=0$. We consider only nonsingular $K_{\alpha\beta}$ until the end of this section.

Observe from Eq. (10) that $d\lambda^{(1)}$ is spacelike ($d\lambda^{(1)}\circ d\lambda^{(1)}>0$), whereas $d\lambda^{(2)}$ is spacelike, timelike, or null. The sign of $d\lambda^{(2)}\circ d\lambda^{(2)}$ will be denoted by

$$\epsilon=1,\ -1,\ \text{or}\ 0.$$

We will see that a fairly unified derivation can be given of all of the metrical forms regardless of the value of ϵ.

The null tetrad has a residual arbitrariness since the canonical form of Eq. (7) is invariant under the continuous group of transformations

$$k\to e^{\chi}k,\quad m\to e^{-\chi}m,\quad t\to e^{-i\varphi}t$$
$$\chi,\varphi\ \text{any real scalar fields}, \tag{12a}$$

and the inversion

$$k\to m,\quad m\to k,\quad t\to t^*. \tag{12b}$$

We now remove almost all of this arbitrariness. Note that, under the above transformations,

$$v_m\to e^{-i\varphi}v_m,\quad w_k\to e^{i\varphi}w_k,$$
$$v_n\to e^{\chi}v_n,\quad w_t\to e^{-\chi}w_t, \tag{13a}$$

and

$$v_m \to w_k, \quad w_k \to v_m, \quad v_n \to w_t, \quad w_t \to v_n. \tag{13b}$$

We use the inversion (13b) only if $\epsilon = 0$. If $\epsilon = 0$, either $\mathrm{Re}v_n = 0$ or $\mathrm{Re}w_t = 0$; we use (13b) to make $\mathrm{Re}w_t = 0$ if it is not already true. Then, for all ϵ, we use the transformation (13a) to make

$$\gamma_1 \equiv \tfrac{1}{2}(v_m + w_k^*) \text{ real}, \tag{14a}$$

$$\tfrac{1}{2}(w_t + w_t{}^*) = \epsilon\gamma_2, \quad \text{where } \gamma_2 \equiv \tfrac{1}{2}(v_n + v_n^*), \tag{14b}$$

$$\gamma_2 \equiv \tfrac{1}{2}(v_n + v_n^*) = f(\lambda^{(2)})/\rho^3 \quad \text{if } \epsilon = 0. \tag{14c}$$

By way of explanation, Eq. (14a) is obtained for any ϵ by selecting an appropriate φ in Eqs. (13a). If $\epsilon = \pm 1$, Eq. (14b) is obtained by selecting an appropriate χ in Eqs. (13a). If $\epsilon = 0$, Eq. (14b) holds trivially, because $\mathrm{Re}w_t = 0$; therefore, we can use the boost transformation to make $\mathrm{Re}v_n$ whatever we please, and Eq. (14c) is what we please. The function $f(\lambda^{(2)})$ of $\lambda^{(2)}$ could have been made unity, but we prefer to leave it open for future flexibility.

For given $f(\lambda^{(2)})$, the null tetrad is now uniquely determined except for a discrete group, and Eqs. (10) become

$$d\lambda^{(1)} = 2\rho^2\gamma_1(l + l^*), \quad d\lambda^{(2)} = 2\rho^2\gamma_2(m + \epsilon k). \tag{15}$$

In addition to γ_1 and γ_2, the following combinations of connection form components will frequently occur in our calculations[21]:

$$\begin{aligned} &\alpha_1 \equiv \tfrac{1}{2}(Q_l + Q_l{}^*), \quad i\beta_1 \equiv \tfrac{1}{2}(Q_l - Q_l{}^*), \\ &\alpha_2 \equiv \tfrac{1}{2}(\epsilon P_k + P_m), \quad \beta_2 \equiv \tfrac{1}{2}(P_k - \epsilon P_m), \\ &i\delta_1 \equiv \tfrac{1}{2}(v_m - w_k{}^*), \quad i\delta_2 \equiv \tfrac{1}{2}(v_n - v_n{}^*). \end{aligned} \tag{16}$$

As we will see, δ_1 is real; so, all eight scalar fields α_i, β_i, γ_i, δ_i are real.

The next step is to apply the integrability conditions $d^2\lambda^{(i)} = 0$ to Eqs. (15), with the aid of Eqs. (6). Only the following algebraic implications of these straightforward computations are needed in the sequel:

$$w_k = v_m = \gamma_1 + i\delta_1, \quad w_l = \epsilon v_n = \epsilon(\gamma_2 + i\delta_2), \tag{17a}$$

$$P_l = i\delta_1, \quad Q_k = -\delta_2, \quad Q_m = \epsilon\delta_2, \quad P_m = \epsilon^2 P_m. \tag{17b}$$

We are now ready to select a chart.

Consider the tetrad[21]

$$\begin{aligned} &\omega^1 \equiv l + l^*, \quad \omega^2 \equiv m + \epsilon k, \\ &\omega^3 \equiv -i(l - l^*), \quad \omega^4 \equiv k - \epsilon m. \end{aligned} \tag{18}$$

We claim that there locally exist scalar fields x^i, x^r, $E^{(i)}$, and $A^r{}_s$ ($i = 1, 2$ and $r, s = 3, 4$) such that

$$\omega^i = (\rho/E^{(i)})dx^i, \quad \omega^r = \rho A^r{}_s dx^s, \tag{19}$$

where summation convention is not used for i, but is used for r, s. The fields x^α constitute our chart.

To prove the existence of this chart, use Eqs. (11) and (15) to show that

$$\rho^{-1}d\rho = \gamma_2\omega^2 - \gamma_1\omega^1. \tag{20}$$

Then, use Eqs. (6), (9), (16), (17), (20) to compute the 2-forms

$$\begin{aligned} &d(\rho^{-1}\omega^1) = \alpha_1(\rho^{-1}\omega^3\omega^1), \\ &d(\rho^{-1}\omega^2) = -2(1+\epsilon^2)^{-1}\alpha_2(\rho^{-1}\omega^4\omega^2), \\ &d(\rho^{-1}\omega^3) = (\gamma_1 - \beta_1)(\rho^{-1}\omega^1\omega^3) + 4(1+\epsilon^2)^{-1}\delta_1(\rho^{-1}\omega^2\omega^4), \\ &d(\rho^{-1}\omega^4) = (1+\epsilon^2)\delta_2(\rho^{-1}\omega^1\omega^3) \\ &\qquad + \left(\frac{2\beta_2}{1+\epsilon^2} - \gamma_2\right)(\rho^{-1}\omega^2\omega^4), \end{aligned} \tag{21}$$

whereupon the theorem of Frobenius implies the local existence of x^i, x^r, $E^{(i)}$, $A^r{}_s$ such that Eqs. (19) hold.

The full set of implications of the affine connection is derived by substituting Eqs. (19) back into Eqs. (21). The results are conveniently grouped into three types. First, there are the following expressions for the connection form components, in which we let

$$\det A \equiv A^3{}_3 A^4{}_4 - A^3{}_4 A^4{}_3,$$

$$R = 5 - i, \quad S = 2 + i.$$

$$\left(\frac{2}{1+\epsilon^2}\right)^{i-1}\alpha_i = \frac{1}{\rho \det A}\left(A^R{}_3\frac{\partial \ln E^{(i)}}{\partial x^4} - A^R{}_4\frac{\partial \ln E^{(i)}}{\partial x^3}\right), \tag{22a}$$

$$\left(\frac{2}{1+\epsilon^2}\right)^{i-1}\beta_i = \gamma_i + \frac{E^{(i)}}{\rho \det A}\left(A^R{}_3\frac{\partial A^S{}_4}{\partial x^i} - A^R{}_4\frac{\partial A^S{}_3}{\partial x^i}\right), \tag{22b}$$

$$\left(\frac{2}{1+\epsilon^2}\right)^{2i-3}\delta_{3-i} = \frac{(-1)^i E^{(i)}}{2\rho \det A}\left(A^R{}_3\frac{\partial A^R{}_4}{\partial x^i} - A^R{}_4\frac{\partial A^R{}_3}{\partial x^i}\right), \tag{22c}$$

$$\gamma_1 dx^1 = \tfrac{1}{2}\rho^{-3}E^{(1)}d\lambda^{(1)}, \quad \gamma_2 dx^2 = \tfrac{1}{2}\rho^{-3}E^{(2)}d\lambda^{(2)}. \tag{22d}$$

Second, there are the following equations which place direct constraints on the x^i-dependences of the metric tensor:

$$\partial E^{(i)}/\partial x^j = 0 \quad \text{if } i \neq j, \tag{23a}$$

$$A^r{}_3\frac{\partial A^s{}_4}{\partial x^i} - A^r{}_4\frac{\partial A^s{}_3}{\partial x^i} = 0 \quad \text{if } i = r - 2, \text{ and for all } r, s. \tag{23b}$$

Finally, there are the following equations which place direct constraints on the x^r dependences of the metric tensor and which hold for $r = 3, 4$:

$$A^r{}_3\frac{\partial E^{(i)}}{\partial x^4} - A^r{}_4\frac{\partial E^{(i)}}{\partial x^3} = 0 \quad \text{if } i = r - 2, \tag{24a}$$

$$\partial E^{(2)}/\partial x^r = 0 \quad \text{when } \epsilon = 0, \tag{24b}$$

$$\frac{\partial A^r{}_4}{\partial x^3} - \frac{\partial A^r{}_3}{\partial x^4} = 0. \tag{24c}$$

By way of explanation, Eq. (24b) is derived with the aid of Eqs. (16) and (17b), which imply $\alpha_2 = 0$ when $\epsilon = 0$.

Let us see where we stand with respect to our choice of coordinates. Suppose p_0 is any given point in the domain of $K_{\alpha\beta}$. We have established the existence of a null tetrad and a chart which cover p_0 such that Eqs. (19), (23), and (24) hold. This chart is not unique; it is subject to any coordinate transformation of the form,

$$x^{i'} = f^i(x^i), \quad x^{r'} = f^r(x^3, x^4).$$

Consider, first, the question of choosing the coordinates x^1, x^2. Observe that Eqs. (22d) imply that $\lambda^{(i)}$ depends only on the coordinate x^i. Therefore, we can always select $x^i = \lambda^{(i)}$, as the authors did previously.[2] However, we are now convinced that it is important

to maintain flexibility by not imposing a premature choice of the coordinates x^i. This flexibility was employed by Carter in his treatment of the Hamilton–Jacobi separable spacetimes.[1] *One of its advantages is that it enables us to include a host of cases involving singular* [(11)(11)] *Killing tensors by admitting metric tensors such that*

$$\lambda^{(1)} \text{ or } \lambda^{(2)} \equiv \text{a constant.}$$

All of our results remain viable when $\lambda^{(1)}$ or $\lambda^{(2)}$ *is identically equal to a constant.* As an example, consider the Kerr metric for which $\lambda^{(1)} = -(a\cos\theta)^2$ and $\lambda^{(2)} = (r)^2$; the singular limit of Schwarzschild occurs when $a = 0$.

We next discuss where we stand with respect to our choice of the coordinates x^3, x^4. Let c^1, c^2 be the values of x^1, x^2 at the point p_0. We can always restrict our chart so that all points in its domain are connected by coordinate lines to the 2-surface $x^1 = c^1$, $x^2 = c^2$; then, (c^1, c^2, x^3, x^4) is in the range of the chart whenever x^1, x^2 exist such that (x^1, x^2, x^3, x^4) is in its range. With this in mind, we prove that Eq. (24c) enables us to select x^3, x^4 so that

$$A^r{}_s(c^1, c^2, x^3, x^4) = \delta^r{}_s \text{ for all values of } x^3, x^4. \tag{25}$$

(We follow the old fashioned practice of designating a point by one of its coordinate representations, with the hope that the context rescues us from ambiguity.) The proof proceeds by introducing scalar fields $a^r{}_s$ defined over the domain of our chart and with values

$$a^r{}_s(x^1, x^2, x^3, x^4) \equiv A^r{}_s(c^1, c^2, x^3, x^4). \tag{26}$$

Equation (24c) implies the existence of scalar fields $x^{3'}, x^{4'}$ defined over a neighborhood of p_0 such that

$$dx^{r'} = a^r{}_s dx^s,$$

which implies $\omega^r = \rho A^r{}_s dx^{s'}$ where $A^{r'}{}_s \equiv (A a^{-1})^r{}_s$. From Eq. (26), A' is the unit matrix on the 2-surface $x^1 = c^1$, $x^2 = c^2$. QED

We do have some concern that Eq. (25) may impose a poor choice of x^3 and x^4 in some cases. However, Eq. (25) will amply justify itself by the ease which it lends to solving Eqs. (24a), (24c). Once we have the solution, it will be seen that the condition (25) can be dispensed with.

For the general class of metrics which admit non-singular [(11)(11)] Killing tensors or their singular limits, Eqs. (23) and (24) are the only constraints on the fields $E^{(i)}$ and $A^r{}_s$. The solving of Eqs. (23b) and (24a), (24c) is the next objective.

3. THE METRICAL FORMS

The general solution of Eq. (23b) presents no unusual difficulty and is given below without derivation:

$$\begin{aligned} &A^3{}_3 = \Delta^{-1}H^{(1)}, \quad A^3{}_4 = A^3{}_3 J^{(2)}, \\ &A^4{}_3 = A^4{}_4 J^{(1)}, \quad A^4{}_4 = \Delta^{-1}H^{(2)}, \\ &\Delta \equiv 1 - J^{(1)}J^{(2)} > 0, \end{aligned} \tag{27}$$

where the four fields $H^{(i)}, J^{(i)}$ satisfy

$$\frac{\partial H^{(i)}}{\partial x^j} = \frac{\partial J^{(i)}}{\partial x^j} = 0 \quad \text{if } i \neq j. \tag{28}$$

In terms of these four fields, Eq. (25) becomes

$$H^{(i)}(c^i, x^3, x^4) = 1, \quad J^{(i)}(c^i, x^3, x^4) = 0, \quad \text{for all } x^3, x^4. \tag{29}$$

We next consider the differential–functional Eqs. (24a, 24c) with $A^r{}_s$ subject to Eqs. (27)–(29). These now become

$$\frac{\partial E^{(2)}}{\partial x^3} - J^{(1)}\frac{\partial E^{(2)}}{\partial x^4} = J^{(2)}\frac{\partial E^{(1)}}{\partial x^3} - \frac{\partial E^{(1)}}{\partial x^4} = 0, \tag{30a}$$

$$\frac{\partial(\Delta^{-1}H^{(1)})}{\partial x^4} - \frac{\partial(\Delta^{-1}H^{(1)}J^{(2)})}{\partial x^3} = 0, \tag{30b}$$

$$\frac{\partial(\Delta^{-1}H^{(2)}J^{(1)})}{\partial x^4} - \frac{\partial(\Delta^{-1}H^{(2)})}{\partial x^3} = 0. \tag{30c}$$

We proceed to use Eqs. (29) and the fact that $\Delta = 1$ if $J^{(1)}J^{(2)} = 0$. For any values of x^2, x^3, x^4, set $x^1 = c^1$ in Eqs. (30); the implications are

$$\frac{\partial E^{(2)}}{\partial x^3} = \frac{\partial J^{(2)}}{\partial x^3} = \frac{\partial H^{(2)}}{\partial x^3} = 0.$$

Likewise, if for any values of x^1, x^3, x^4, we set $x^2 = c^2$, we get

$$\frac{\partial E^{(1)}}{\partial x^4} = \frac{\partial J^{(1)}}{\partial x^4} = \frac{\partial H^{(1)}}{\partial x^4} = 0.$$

Therefore, each of the six fields depends at most on two coordinates as follows:

$$E^{(i)}(x^i, x^{2+i}), \quad H^{(i)}(x^i, x^{2+i}), \quad J^{(i)}(x^i, x^{2+i}). \tag{31}$$

Upon substituting Eqs. (31) back into Eqs. (30), we get

$$J^{(1)}\frac{\partial E^{(2)}}{\partial x^4} = J^{(2)}\frac{\partial E^{(1)}}{\partial x^3} = 0, \tag{32a}$$

$$(1 - J^{(1)}J^{(2)})J^{(2)}\frac{\partial H^{(1)}}{\partial x^3} + \left[(J^{(2)})^2\frac{\partial J^{(1)}}{\partial x^3} - J^{(1)}\frac{\partial J^{(2)}}{\partial x^4}\right]H^{(1)} = 0, \tag{32b}$$

$$(1 - J^{(1)}J^{(2)})J^{(1)}\frac{\partial H^{(2)}}{\partial x^4} + \left[(J^{(1)})^2\frac{\partial J^{(2)}}{\partial x^4} - J^{(2)}\frac{\partial J^{(1)}}{\partial x^3}\right]H^{(2)} = 0. \tag{32c}$$

To proceed any further, it is necessary to distinguish between four cases which depend on whether or not δ_i $(i = 1, 2)$ is identically zero in a neighborhood of p_0. Also, within each of these cases, $E^{(i)}$ is independent of x^{i+2} if and only if α_i vanishes identically. To help us understand these matters, we need the expressions for the connection form components α_i and δ_i in terms of the fields $E^{(i)}, H^{(i)}, J^{(i)}$. From Eqs. (22a), (22c), (24b), (27), and (31),

$$\alpha_1 = \frac{-1}{\rho H^{(1)}}\frac{\partial \ln E^{(1)}}{\partial x^3}, \quad \alpha_2 = \frac{1}{\rho H^{(2)}}\frac{\partial \ln E^{(2)}}{\partial x^4}, \tag{33a}$$

$$\alpha_2 = 0 \text{ when } \epsilon = 0, \tag{33b}$$

$$\delta_1 = \frac{(1+\epsilon^2)E^{(2)}H^{(1)}}{4\rho\Delta H^{(2)}}\frac{\partial J^{(2)}}{\partial x^2}, \qquad \delta_2 = \frac{E^{(1)}H^{(2)}}{(1+\epsilon^2)\rho\Delta H^{(1)}}\frac{\partial J^{(1)}}{\partial x^1}. \tag{33c}$$

Equations (33a), (33c) hold for all ϵ.

Suppose, for example, δ_1 is identically zero in a neighborhood (connected) of p_0. We restrict our chart to the intersection of this neighborhood and the original domain of the chart. Since $J^{(2)}(c^2,x^4)=0$, Eq. (33c) implies that $J^{(2)}$ is also identically zero, and Eq. (32a) imposes no constraint on the x^3 dependence of $E^{(1)}$, i.e., no constraint on α_1. Equation (32b) is identically satisfied, and Eq. (32c) reduces to a simple form.

On the other hand, suppose there exists no neighborhood of p_0 in which δ_1 is identically zero. Then $J^{(2)}$ cannot vanish identically in any neighborhood of p_0, and there exists a sequence of points p_n with coordinates x_n^α such that

$$\begin{aligned} &x_n^\alpha \to c \text{ as } n\to\infty \quad (c^\alpha = \text{coord. of } p_0),\\ &J^{(2)}(x_n^2,x_n^4)\neq 0 \text{ for all } n. \end{aligned} \tag{34a}$$

Also, $E^{(1)}$ is independent of x^3, and $\alpha_1=0$ at all points of the neighborhood. The mode of solving Eq. (32b), (32c) will be illustrated later.

Similar remarks hold for δ_2, *with the only crucial difference being that Eqs.* (33a), (33b) *imply* $E^{(2)}$ *is always independent of* x^4 *when* $\epsilon=0$ (this is true even if $\delta_2=0$). Corresponding to Eq. (34a), if δ_2 is not identically zero in any neighborhood of p_0, there exists a sequence of points q_n with coordinates y_n^α such that

$$\begin{aligned} &y_n^\alpha \to c \text{ as } n\to\infty,\\ &J^{(1)}(y_n^1,y_n^3)\neq 0 \text{ for all } n. \end{aligned} \tag{34b}$$

These sequences will be used later.

Without more ado, we give the final results for all four cases below. To apply these results to the contravariant metric in Eqs. (3) of Sec. 1, note that

$$F^{(i)}:=2[E^{(i)}]^2, \quad G^{(i)}:=\tfrac{1}{2}[H^{(i)}]^2. \tag{35}$$

The derivations for cases (0,1), (1,0), (0,0) are obvious from the discussion which we just completed, and the derivation for case (1,1) will be given immediately after the list of results.

(1,1) *RESULTS when* $\delta_1\delta_2\neq 0$:

$$\begin{aligned} &E^{(i)}=e^{(i)}(x^i), && (i=1,2),\\ &H^{(i)}=h^{(i)}(x^i)[K^{(i)}]^{-1} && (i=1,2),\\ &J^{(i)}=j^{(i)}(x^i)[K^{(i)}]^{-1}\exp(\sigma_i x^{i+2}) && (i=1,2),\\ &K^{(1)}=1-(\sigma_2/\sigma_1)[\exp(\sigma_1 x^3)-1]j^{(1)}(x^1),\\ &K^{(2)}=1-(\sigma_1/\sigma_2)[\exp(\sigma_2 x^4)-1]j^{(2)}(x^2). \end{aligned} \tag{36a}$$

σ_1,σ_2 are any real constants;

$e^{(i)},h^{(i)},j^{(i)}$ are arbitrary smooth functions of x^i except that $dj^{(i)}/dx^i$ is not identically zero.

(0,1) *RESULTS when* $\delta_1=0$, $\delta_2\neq 0$:

$$\begin{aligned} &E^{(1)}=E^{(1)}(x^1,x^3), \quad E^{(2)}=e^{(2)}(x^2),\\ &H^{(1)}=H^{(1)}(x^1,x^3), \quad H^{(2)}=h^{(2)}(x^2),\\ &J^{(1)}=J^{(1)}(x^1,x^3), \quad J^{(2)}\equiv 0. \end{aligned} \tag{36b}$$

All are arbitrary smooth functions of respective arguments except that $\partial J^{(1)}/\partial x^1$ is not identically zero.

(1,0) *RESULTS when* $\delta_1\neq 0$, $\delta_2=0$:

Like (0,1) except for interchange of scripts, 1 with 2, and 3 with 4, and *except for the additional constraint expressed by*

$$E^{(2)}=E^{(2)}(x^2,\epsilon^2 x^4), \tag{36c}$$

(0,0) *RESULTS when* $\delta_1=0$, $\delta_2=0$:

$$\begin{aligned} &E^{(1)}=E^{(1)}(x^1,x^3), \quad E^{(2)}=E^{(2)}(x^2,\epsilon^2 x^4),\\ &H^{(1)}=H^{(1)}(x^1,x^3), \quad H^{(2)}=H^{(2)}(x^2,x^4),\\ &J^{(1)}\equiv 0, \quad J^{(2)}\equiv 0. \end{aligned} \tag{36d}$$

All are arbitrary smooth functions of respective arguments.

Now, let us go through the derivation of the results for case (1,1), $\delta_1\delta_2\neq 0$. Both of the sequences (34a) and (34b) exist in this case. In Eq. (32b), while keeping (x^1,x^3) arbitrary and fixed, set $(x^2,x^4)=(x_n^2,x_n^4)$, divide through by $J^{(2)}$, and let $n\to\infty$. In Eq. (32c), while keeping (x^2,x^4) arbitrary and fixed, set $(x^1,x^3)=(y_n^1,y_n^3)$, divide through by $J^{(1)}$, and let $n\to\infty$. It can be seen that the limits

$$\sigma_1 \equiv \lim_{n\to\infty}\left[\frac{\partial \ln J^{(1)}}{\partial x^3}\right]_n, \quad \sigma_2 \equiv \lim_{n\to\infty}\left[\frac{\partial \ln J^{(2)}}{\partial x^4}\right]_n,$$

exist and are independent of the choice of our sequences; moreover, we obtain the differential equations

$$\frac{\partial H^{(1)}}{\partial x^3}-\sigma_2 J^{(1)}H^{(1)}=0, \quad \frac{\partial H^{(2)}}{\partial x^4}-\sigma_1 J^{(2)}H^{(2)}=0. \tag{37}$$

We substitute the above Eqs. (37) back into Eqs. (32b), (32c), and we get

$$\sigma_2(1-J^{(1)}J^{(2)})J^{(1)}J^{(2)}+(J^{(2)})^2\frac{\partial J^{(1)}}{\partial x^3}-J^{(1)}\frac{\partial J^{(2)}}{\partial x^4}=0, \tag{38a}$$

$$\sigma_1(1-J^{(1)}J^{(2)})J^{(1)}J^{(2)}+(J^{(1)})^2\frac{\partial J^{(2)}}{\partial x^4}-J^{(2)}\frac{\partial J^{(1)}}{\partial x^3}=0. \tag{38b}$$

Next, in Eq. (38a), while keeping (x^2,x^4) arbitrary and fixed, set $(x^1,x^3)=(y_n^1,y_n^3)$, divide through by $J^{(1)}$, and let $n\to\infty$. In Eq. (38b), keep (x^1,x^3) arbitrary and fixed, set $(x^2,x^4)=(x_n^2,x_n^4)$, divide by $J^{(2)}$, and let $n\to\infty$. We obtain

$$\begin{aligned} &\sigma_2 J^{(2)}+\sigma_1(J^{(2)})^2-\frac{\partial J^{(2)}}{\partial x^4}=0,\\ &\sigma_1 J^{(1)}+\sigma_2(J^{(1)})^2-\frac{\partial J^{(1)}}{\partial x^3}=0. \end{aligned} \tag{39}$$

If we now substitute Eqs. (39) back into Eqs. (38a), (38b), we find that the latter are identically satisfied by any solution of Eqs. (39). So, our cycle of substitutions is completed, and we proceed to solve Eqs. (39) and (37) for $J^{(i)}$ and $H^{(i)}$. The rest is straightforward.

In spite of the fact that the choice of a chart for which Eqs. (25) hold played a useful role in simplifying our derivation, it is easily verified that our results in Eqs. (36) satisfy Eqs. (32) without resort to Eqs. (25). So, Eqs. (25) are of no further concern to us, and the functions appearing in our results are quite arbitrary except for the usual demands of smoothness and signature and except for the condition that $\partial J^{(i)}/\partial x^i$ is not identically zero whenever $J^{(i)}$ is not identically zero.[22]

4. SUMMARY

We now summarize our results by giving the general form of the line element for any spacetime which admits a nonsingular [(11)(11)] KT or which is conformal to one which does so. From Eqs. (18), (19), and (27), the line element for all $\epsilon = \pm 1, 0$ is

$$ds^2 = \tfrac{1}{2}e^{2\sigma}\rho^2\left\{\left[\frac{dx^1}{E^{(1)}}\right]^2 + \left[\frac{H^{(1)}(dx^3 + J^{(2)}dx^4)}{1 - J^{(1)}J^{(2)}}\right]^2 + \epsilon\left[\frac{dx^2}{E^{(2)}}\right]^2 - \epsilon\left[\frac{H^{(2)}(J^{(1)}dx^3 + dx^4)}{1 - J^{(1)}J^{(2)}}\right]^2 + 4(1-\epsilon^2)\left[\frac{H^{(2)}dx^2(J^{(1)}dx^3 + dx^4)}{E^{(2)}(1 - J^{(1)}J^{(2)})}\right]\right. . \quad (40)$$

Above, $\exp 2\sigma$ is the conformal factor; $\rho^2 = \lambda^{(2)} - \lambda^{(1)}$, where $\lambda^{(1)}$ and $\lambda^{(2)}$ are arbitrary smooth functions of x^1 and x^2, respectively. $E^{(i)}, H^{(i)}, J^{(i)}$ are given by Eqs. (36a)—(36d) for four distinct cases. The contravariant metric corresponding to Eqs. (40) can easily be read off from Eq. (3), where we have used the notations $F^{(i)}$ and $G^{(i)}$ as defined in Eqs. (35). When $\exp 2\sigma = 1$, there is a [(11)(11)] KT which is given by Eqs. (2) and (3); otherwise, all we can say in general is that there is a conformal KT given by $H_2 - H_1$.

As we stated in Sec. 1, there is an overlap between our metrical forms and those which admit separability or partial separability for the H—J equation after it is multiplied by an integrating factor U. The relations between our metrical forms and the H—J separable or partially separable ones are most easily obtained by comparing our contravariant metrics with the H—J separable or partially separable contravariant spacetime metrics as listed in a convenient table given by Collinson[3] or as derived from the general expressions given by Woodhouse[3] and by Dietz.[3] We omit the details of this comparison and simply summarize the results for the various cases (1,1), (0,1), (1,0), (0,0) given by Eqs. (36a)—(36d).

The case (1,1) *when* $\sigma_1 \neq 0$ *and/or* $\sigma_2 \neq 0$ [*and for all* $\epsilon = 1, -1, 0$] *represents completely new metrical forms* which do not fall under any conventional separable or partially separable classification. When $\sigma_1 = \sigma_2 = 0$ and $\epsilon = 1, -1, 0$, however, case (1,1) is separable if

$$\rho^2 \exp 2\sigma = U = U_1(x^1) + U_2(x^2),$$

where $U_1(x^1)$ and $U_2(x^2)$ are any smooth functions of x^1 and x^2, respectively. (So, separability exists in this case if $\exp 2\sigma = 1$.) The H—J equation then has completely separated solutions of the form

$$S = S_1(x^1) + S_2(x^2) + k_3x^3 + k_4x^4,$$

where k_3 and k_4 are constants.

The other three cases (0,1), (1,0), and (0,0) each involves at least partially separable H—J equations regardless of the value of ϵ. Case (0,1) has H—J solutions of the form[23]

$$S = S_1(x^1, x^3) + S_2(x^2) + k_4x^4$$

provided that

$$e^{2\sigma}\rho^2 = U = U_1(x^1, x^3) + U_2(x^2),$$

where U_1 and U_2 depend arbitrarily on their respective arguments. Note that $\partial/\partial x^4$ is a Killing vector for case (0,1).

For case (1,0), $\partial/\partial x^3$ is a Killing vector, and the H—J equations have solutions of the form[23]

$$S = S_1(x^1) + S_2(x^2, x^4) + k_3x^3$$

provided that

$$e^{2\sigma}\rho^2 = U_1(x^1) + U_2(x^2, x^4).$$

Finally, for case (0,0), the H—J equation has solutions of the form[23]

$$S = S_1(x^1, x^3) + S_2(x^2, x^4)$$

provided that

$$e^{2\sigma}\rho^2 = U_1(x^1, x^3) + U_2(x^2, x^4).$$

In all of the above cases, U_1 and U_2 are arbitrary smooth functions of their respective arguments.

The most interesting of our metrical forms is, perhaps, the one corresponding to case (1,1) with $\epsilon = 1$,[24] since the Kerr metric and its various augmentations have this form. For the charged Kerr metric expressed in Boyer—Lindquist coordinates,[4]

$$x^1 = \cos\theta, \quad x^2 = r, \quad x^3 = \varphi, \quad x^4 = t - a\varphi,$$

we have $\exp 2\sigma = 1$, and

$$\lambda^{(1)} = -a^2\cos^2\theta, \quad \lambda^{(2)} = r^2,$$

$$F^{(1)} = \sin^2\theta, \quad F^{(2)} = \Delta = r^2 - 2mr + a^2 + e^2,$$

$$G^{(1)} = \sin^2\theta, \quad G^{(2)} = r^4\Delta,$$

$$J^{(1)} = a\cos^2\theta, \quad J^{(2)} = -ar^{-2}.$$

One of the points which we intend to investigate in some detail is whether there can possibly exist any meaningful matter tensor for case (1,1) when $\sigma_1 \neq 0$ and/or $\sigma_2 \neq 0$. In this investigation, great caution must be preserved with regard to the interpretation of the coordinates x^3 and x^4; i.e., when $\sigma_1 \neq 0$ and/or $\sigma_2 \neq 0$, Eq. (36a) makes it abundantly clear that we cannot simply regard x^3 and x^4 as some kind of "post-ignorable" φ and $t - a\varphi$ *even if the metric reduces to the Kerr metric* (in Boyer—Lindquist coordinates) *in the limit* $\sigma_1 \to 0$ *and* $\sigma_2 \to 0$.

ACKNOWLEDGMENTS

We thank C. D. Collinson for a preprint of his work on H—J separable spacetimes. Also, we thank Fred Ernst for his kind encouragement.

APPENDIX

Let $K^{\alpha\beta}$ be any symmetric contravariant tensor field on any given differentiable manifold M of dimension four, and let $g_{\alpha\beta}$ and

$$g'_{\alpha\beta} = e^{2\sigma}g_{\alpha\beta} \quad (A1)$$

denote any conformally related Lorentzian metrics for M. Suppose, furthermore, $K^{\alpha\beta}$ is a conformal Killing tensor in the spacetime $(M, g_{\alpha\beta})$; i.e., there exists a vector field ξ_α in $(M, g_{\alpha\beta})$ such that

$$\nabla_{(\alpha}K_{\beta\gamma)} = \xi_{(\alpha}g_{\beta\gamma)}, \tag{A2}$$

where ∇_α denotes covariant differentiation in $(M, g_{\alpha\beta})$, and where $K_{\alpha\beta} = g_{\alpha\gamma}g_{\beta\delta}K^{\gamma\delta}$. Then, we recall that $K^{\alpha\beta}$ is also a conformal Killing tensor in $(M, g'_{\alpha\beta})$. Moreover, it is easily shown that the conformal transformation (A1) induces the following transformation of the vector field ξ^α:

$$\xi^\alpha \to \xi^{\alpha'} = \xi^\alpha + 2K^{\alpha\beta}\nabla_\beta\sigma. \tag{A3}$$

Without loss of generality, we can make $K^{\alpha\beta}$ traceless by replacing it with $K^{\alpha\beta} - \frac{1}{4}g^{\alpha\beta}K^\gamma_\gamma$; it will be understood that $K^\gamma_\gamma = 0$ in the sequel.

We will now prove the following theorem:

Suppose the Segre characteristic of $K^{\alpha\beta}$ is [(11)(11)], and $K^{\alpha\beta}$ is a traceless conformal Killing tensor in $(M, g_{\alpha\beta})$. Then, there exists a conformal mapping (A1) such that $K^{\alpha\beta}$ is a [(11)(11)] Killing tensor in $(M, g'_{\alpha\beta})$ and such that the eigenvalues of this Killing tensor are constants.

The proof proceeds by noting that

$$K_\alpha{}^\beta K_\beta{}^\gamma = \lambda^2\delta_\alpha{}^\gamma \tag{A4}$$

where λ and $-\lambda$ are the eigenvalues of the [(11)(11)] traceless conformal KT. Therefore,

$$\nabla_\gamma(K_\alpha{}^\beta K_\beta{}^\gamma) = \nabla_\alpha(\lambda^2). \tag{A5}$$

By a straightforward calculation which we do not reproduce here, we use Eqs. (A2), (A4), and (A5) to prove that

$$\xi^\alpha = K^{\alpha\beta}\nabla_\beta \ln\lambda. \tag{A6}$$

From Eqs. (A3) and (A6), we see that $\xi^{\alpha'} = 0$ if we choose $\sigma = -\frac{1}{2}\ln\lambda$. However, $\xi^{\alpha'} = 0$ if and only if $K^{\alpha\beta}$ is a Killing tensor in $(M, g'_{\alpha\beta})$. Moreover, by applying Eq. (A6) to the spacetime $(M, g'_{\alpha\beta})$, we see that $\xi^{\alpha'} = 0$ if and only if $\lambda' = \text{const}$. So, the proof of our theorem is completed.

Now, every Killing tensor is also a conformal Killing tensor. Therefore, the above theorem has the following corollary:

The set of all spacetimes which admit [(11)(11)] conformal Killing tensors is identical with the set of all spacetimes which admit or which are conformal to those which admit [(11)(11)] Killing tensors whose eigenvalues are constants.

[1]B. Carter, Commun. Math. Phys. **10**, 280 (1968).

[2]Preliminary work by the authors on the subject of the present paper is given by I. Hauser and R. J. Malhiot, J. Math. Phys. **17**, 1306 (1976).

[3]Another generalization can be based on the separability theorems of N. M. J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975). A generalization of this kind is given by W. Dietz, J. Phys. A **9**, 519 (1976). A summary of the vacuum solutions corresponding to the metrical forms of Dietz has been worked out by C. D. Collinson and J. Fugère (preprint). Our generalization differs from that of Dietz, though there is some overlap which is detailed in Sec. 4.

[4]B. Carter, Phys. Rev. **174**, 1559 (1968).

[5]Our $x^1, x^2, x^3, x^4, \lambda^{(1)}, \lambda^{(2)}, F^{(1)}, F^{(2)}, G^{(1)}, G^{(2)}, J^{(1)}, J^{(2)}$ are respectively equal to Carter's $\lambda^{-1}, \lambda^1, \phi^2, \phi^{-2}, -U_{-1}, U_1, \Delta_{-1}, \Delta_1, \Delta_{-1}(Z^{-1}{}_{-1})^{-2}, \Delta_1(Z^1{}_1)^{-2}, -Z^1{}_{-1}(Z^{-1}{}_{-1})^{-1}, -Z^{-1}{}_1(Z^1{}_1)^{-1}$, as they appear in Eq. (80) of Ref. 1.

[6]The notations H_1, H_2, and K are adapted from W. Dietz, Ref. 3.

[7]M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).

[8]B. Carter, Phys. Lett. A **26**, 399 (1968).

[9]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).

[10]H. D. Wahlquist, Phys. Rev. **172**, 1291 (1968).

[11]C. D. Collinson, Int. J. Theor. Phys. **15**, 311 (1975). Also, W. Dietz (preprint). The same conclusion holds for [(11)(11)] conformal Killing tensors; the derivation is essentially the same.

[12]We thank R. Geroch for a helpful discussion of this point.

[13]R. J. Malhiot and I. Hauser, Bull. Am. Phys. Soc. **20**, 543 (1975). An extension of this theorem to n-dimensional Riemannian spaces is given by W. Dietz in Ref. 3.

[14]For a discussion of type D vacuum solutions in the context of a review of exact solutions, see W. Kinnersley, in *General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975), pp. 109—35.

[15]L. P. Hughston and P. Sommers, Commun. Math. Phys. **32**, 147 (1973); **33**, 129 (1973).

[16]J. F. Plebański, "A Class of Solutions of Einstein-Maxwell Equations," Inst. of Theor. Phys., Warsaw Univ., IFT-17-73 (1973); J. F. Plebański and M. Demiański, "Accelerating Solutions of Einstein-Maxwell Equations," Inst. of Theor. Phys., Warsaw Univ., IFT-17-73 (1973); J. F. Plebański, in *Gravitational Radiation and Gravitational Collapse*, edited by C. DeWitt-Morette (Reidel, Dorecht, 1974), pp. 188—90; J. F. Plebański and M. Demiański, Ann. Phys. (N.Y.) **98**, 98 (1976). See Ref. 14, pp. 128—29 for additional references.

[17]The concept of invertibility when an isometry group exists is defined and discussed by B. Carter, J. Math. Phys. **10**, 70 (1969). An extension of this concept which may be applicable when isometries do not exist has been communicated to us by R. Debever.

[18]Our signature is +2; i.e., in Eq. (3), the coordinate ranges are subject to the constraints that $(\lambda^{(2)} - \lambda^{(1)})^{-1}F^{(1)}$ and $(\lambda^{(2)} - \lambda^{(1)})G^{(1)}$ be positive and that $F^{(2)}$ and $G^{(2)}$ have the same sign.

[19]In Eq. (1), the factor $\epsilon = \pm 1$ is absorbed into $F^{(2)}$ and $G^{(2)}$.

[20]The use of the subscript n in place of t^* is for the printer's convenience; we used t^* in Ref. 2, and the result was a typographical nightmare.

[21]The definitions of α_2 in Eqs. (16) and of ω^2 in Eqs. (18) differ from Ref. 2 by a sign when $\epsilon = -1$. The change facilitates our unified treatment for $\epsilon = \pm 1, 0$. Also, we now use the notations $\omega^1, \omega^2, \omega^3, \omega^4$ in place of $\omega_x, \omega_y, \omega_+, \omega_-$ respectively. (Incidentally, our notations $\alpha_i, \beta_i, \gamma_i$ should not be confused with the use of these symbols in Ref. 1; there is no simple relation.)

[22]This constraint on $\partial J^{(i)}/\partial x^i$ can be dropped, but then the various cases (1,1), (1,0), (0,1), (0,0) will no longer be disjoint. Also it is possible to have a spacetime with different cases in different regions.

[23]Obvious modifications occur when additional Killing vectors exist. For example, for case (0,1), if $\partial/\partial x^3$ is also a Killing vector, then $S_1 \to S_1(x^1) + k_3x^3$, and $U_1 = U_1(x^1)$ is independent of x^3.

[24]A nonrelativistic analog which we discussed in Ref. 2 is closest to case (1,1) with $\epsilon = 1$.

On the relationship between conservation laws and invariance groups of nonlinear field equations in Hamilton's canonical form [a)]

Sukeyuki Kumei [b)]

Department of Physics, University of the Pacific, Stockton, California 95204
(Received 28 January 1977)

It is shown that whenever fields governed by the equations $\partial/\partial t p_\alpha = -\delta H/\delta q_\alpha$, $\partial/\partial t q_\alpha = \delta H/\delta p_\alpha$ allow a conservation law of the form $\partial\rho/\partial t + \mathrm{div}\mathbf{J} = 0$, there exists a corresponding Lie–Bäcklund infinitesimal contact transformation which leaves the Hamiltonian equations invariant. A condition that an invariant Lie–Bäcklund infinitesimal contact transformation gives rise to a conservation law is established. Each such transformation, which may involve derivatives of arbitrary order, yields a one-parameter local Lie group of invariance transformations. The results are established with the aid of a Lie bracket formalism for Hamiltonian fields. They account for a number of recently discovered conservation laws associated with nonlinear time evolution equations.

INTRODUCTION

In previous papers,[1,2] we have studied invariance properties of various nonlinear time evolution equations by applying the theory of groups of Lie—Bäcklund tangent transformations[3] (not to be confused with the Bäcklund transformations of recent literature[4]) and we have shown that each of the well-known series of conservation laws assoicated with the sine-Gordon equation, the Korteweg—de Vries equation, and the nonlinear Schrödinger equation is related to a different one-parameter group which leaves the corresponding differential equation invariant.

The group generators obtained in these papers depend upon derivatives of arbitrary order, so that they are not of the type considered in Lie's general theory of continuous groups of transformations. The question naturally arises: To what extent can the previous results be generalized?

In the present paper, we study invariance properties of Hamilton's equations governing the time evolution of multicomponent fields $p_\alpha(x)$, $q_\alpha(x)$,

$$\dot{p}_\alpha = -\delta H/\delta q_\alpha, \quad \dot{q}_\alpha = \delta H/\delta p_\alpha, \quad \alpha = 1, 2, \ldots, N, \qquad (1)$$

where $x = (x^0, x^1, x^2, x^3)$ and $\dot{p}_\alpha = \partial_{x^0} p_\alpha$, $\dot{q}_\alpha = \partial_{x^0} q_\alpha$. We assume that an energy density H associated with H can depend on coordinates x (including x^0), p_α, and q_α, and their spatial derivatives of arbitrary order.[5] The main interest of the study is: to examine the relationship between invariance groups admitted by Eq. (1) and conservation laws obeyed by the fields. We will prove that: *The existence of N independent conservation laws associated with the fields of Eq.* (1) *necessarily requires the existence of N one-parameter groups which leave Eq.* (1) *invariant.* The precise result will be stated here as a theorem. The notations in the theorem are the following: $\mathcal{A}$ and $\mathcal{J}^i$ are quantities associated with the fields and are functions of x, p_α, and q_α, and of their spatial derivatives of arbitrary order; D_i represents a differentiation with respect to x^i, and the quantity $\delta A/\delta f$ ($f = q_\alpha$ or p_α) is defined by

$$\frac{\delta A}{\delta f} = \mathcal{A}_f - D_i \mathcal{A}_{f,i} + D_i D_j \mathcal{A}_{f,ij} + \cdots + (-D_i)\cdots(-D_j)\mathcal{A}_{f,i\ldots j} + \cdots,$$

with

$$\mathcal{A}_{f,i\ldots j} = \partial_{f,i\ldots j}\mathcal{A} \quad \text{and} \quad f_{,i\cdots j} = \partial_{x^i}\cdots\partial_{x^j} f.$$

Theorem: If, when p_α and q_α are solutions of the Hamiltonian equations (1), the functions $\mathcal{A}(x, p_\alpha, q_\alpha, \cdots)$ and $\mathcal{J}^i(x, p_\alpha, q_\alpha, \cdots)$ obey the conservation law $D_0\mathcal{A} + \sum_{i=1}^3 D_i \mathcal{J}^i = 0$, then the prolongation of the operator $\mathbf{A} = (\delta A/\delta p_\alpha)\partial_{q_\alpha} - (\delta A/\delta q_\alpha)\partial_{p_\alpha}$ is a generator of an invariance group of the Hamiltonian equations. Conversely, for any operator of the form $\mathbf{A}' = (\delta A'/\delta p_\alpha)\partial_{q_\alpha} - (\delta A'/\delta q_\alpha)\partial_{p_\alpha}$ whose prolongation becomes a generator of an invariance group of the Hamiltonian equations, there exists a flux $\mathcal{J}^i$ which together with a density $\mathcal{A}'$ forms a conservation law $D_0\mathcal{A} + \sum_{i=1}^3 D_i \mathcal{J}^i = 0$.

The corresponding result for Hamiltonian systems with finite degrees of freedom governed by the equations $\dot{q}_\alpha = \partial H/\partial p_\alpha$, $\dot{p}_\alpha = -\partial H/\partial q_\alpha$ has been obtained by Peterson.[6]

We will prove the theorem by using a Lie bracket formalism, instead of a Poisson bracket formalism, for Eq. (1). To establish the Lie bracket formulation, one needs to associate appropriate operators with physical quantities of the system. Such a formalism is known for Hamiltonian systems with finite degrees of freedom.[7] In the following we will develop a similar formalism for the field equations (1) by applying the theory of Lie—Bäcklund tangent transformations. The formalism turns out to be very appropriate in studying the connection of invariance groups of Eq. (1) to conservation laws. In this approach no reference is made to invariance properties of an action integral $\int L\, dx$: We deal directly with invariance properties of differential equations.

All the results in the following sections remain valid for a general case of n spatial variables.

[a)]This work has been supported by a Research Corporation grant.

[b)]Present address: 1513-69 Sekido, Tama-shi, Tokyo 192-02, Japan.

 0022-2488/78/1901-0195$1.00

I. LIE BRACKET FORMALISM

We consider groups of Lie–Bäcklund tangent transformations generated by the operators[8]

$$U = F_\alpha \partial_{q_\alpha} - G_\alpha \partial_{p_\alpha} + (D_i F_\alpha)\partial_{q_{\alpha,i}} - (D_i G_\alpha)\partial_{p_{\alpha,i}} + \cdots + \{(-D_i)\cdots(-D_j)F_\alpha\}\partial_{q_{\alpha,i\cdots j}} - \{(-D_i)\cdots(-D_j)G_\alpha\}\partial_{p_{\alpha,i\cdots j}} + \cdots, \tag{2}$$

where D_i represents a total derivative operator

$$D_i = \partial_{x^i} + p_{\alpha,i}\partial_{p_\alpha} + q_{\alpha,i}\partial_{q_\alpha} + p_{\alpha,ij}\partial_{p_{\alpha,j}} + q_{\alpha,ij}\partial_{q_{\alpha,j}} + \cdots, \tag{3}$$

and $p_{\alpha,i\cdots j}$, $q_{\alpha,i\cdots j}$ represent coordinates associated with derivatives $\partial_{x^i}\cdots\partial_{x^j}p_\alpha(x)$, $\partial_{x^i}\cdots\partial_{x^j}q_\alpha(x)$. Throughout the paper we adopt a summation convention for repeated indices: a greek index runs from 1 to N and a Roman from 0 to 3. In contrast to conventional contact transformations, we allow F and G to be functions of x and $p_\alpha(x)$, $q_\alpha(x)$, and any of their derivatives of *arbitrary* order. In the study of Eq. (1) which is a time evolution type we can assume without a loss of generality that the F_α and G_α are not functions of time derivatives of $p_\alpha(x)$ and $q_\alpha(x)$. This will be assumed in the following for all the operators of the form (2). To avoid a complex expression we write the operator (2), which we call a Lie–Bäcklund operator, as

$$U = F_\alpha \partial_{q_\alpha} - G_\alpha \partial_{p_\alpha}. \tag{4}$$

We must always consider this to be the infinite series given by (2). We denote a set of operators of the form (2) by Λ. It is known that the U have the properties

(a) If U^1, $U^2 \in \Lambda$, then $U^3 = [U^1, U^2] \in \Lambda$ with $F^3_\alpha = U^1 F^2_\alpha - U^2 F^1_\alpha$ and $G^3_\alpha = U^1 G^2_\alpha - U^2 G^1_\alpha$.[2,9]

(b) If U^1, U^2, $U^2 \in \Lambda$, the Jacobi identity holds[9]: $[[U^1,U^2],U^3] + [[U^2,U^3],U^1] + [[U^3,U^1],U^2] = 0$.

(c) Members of Λ commute with the total derivative operator D_i: $[U, D_i] = 0$.[2,9,10]

This last property will be used frequently in the following without comment. We define the time derivative of the U, which we denote by U_{x^0}, by

$$U_{x^0} = (\partial_{x^0}F_\alpha)\partial_{q_\alpha} - (\partial_{x^0}G_\alpha)\partial_{p_\alpha}. \tag{5}$$

Again, this is a simplified expression; the full expression is obtained by replacing F_α and G_α in (2) by $\partial_{x^0}F_\alpha$ and $\partial_{x^0}G_\alpha$.

Now, let us consider a variational problem of a functional

$$J[p,q,x^0] = \int_G \mathcal{J}(x,p,q)\,dx', \quad dx' = dx^1dx^2dx^3. \tag{6}$$

The density $\mathcal{J}(x,p,q)$ depends on x and p_α, q_α and their derivatives $p_{\alpha,i\cdots j}$, $q_{\alpha,i\cdots j}$ of arbitrary order except ones involving time derivatives. For the variation $p_\alpha(x) \to p_\alpha(x) + \epsilon\psi_\alpha(x)$ we have

$$\delta J = \epsilon \int_G \left(\frac{\delta J}{\delta p_\alpha}\right)\psi_\alpha\,dx' + \text{surface integral} \tag{7}$$

with

$$\frac{\delta J}{\delta p_\alpha} = \mathcal{J}_{p_\alpha} - D_i\mathcal{J}_{p_{\alpha,i}} + D_iD_j\mathcal{J}_{p_{\alpha,ij}} + (-D_i)\cdots(-D_j)\mathcal{J}_{p_{\alpha,i\cdots j}} + \cdots. \tag{8}$$

Similarly, for a variation $q_\alpha(x) \to q_\alpha(x) + \epsilon\phi_\alpha(x)$, we have

$$\frac{\delta J}{\delta q_\alpha} = \mathcal{J}_{q_\alpha} - D_i\mathcal{J}_{q_{\alpha,i}} + D_iD_j\mathcal{J}_{q_{\alpha,ij}} + (-D_i)\cdots(-D_j)\mathcal{J}_{q_{\alpha,i\cdots j}} + \cdots. \tag{9}$$

We adopt (8) and (9) as the defining equations of $\delta J/\delta p_\alpha$ and $\delta J/\delta q_\alpha$. We call $\mathcal{J}$ a density of J. With the functional J we associate an operator $\mathbf{J}$ which is obtained from (2) by substituting $\delta J/\delta p_\alpha$ and $\delta J/\delta q_\alpha$ for F_α and G_α: In simplified notation

$$\mathbf{J} = \left(\frac{\delta J}{\delta p_\alpha}\right)\partial_{q_\alpha} - \left(\frac{\delta J}{\delta q_\alpha}\right)\partial_{p_\alpha}. \tag{10}$$

We designate the operators of this particular form by boldface letters. Then, with the energy functional H, the following Lie–Bäcklund Hamiltonian operator will be associated:

$$\mathbf{H} = \left(\frac{\delta H}{\delta p_\alpha}\right)\partial_{q_\alpha} - \left(\frac{\delta H}{\delta q_\alpha}\right)\partial_{p_\alpha}. \tag{11}$$

The operator corresponding to a functional $\int \partial_{x^0}\mathcal{J}\,dx'$ is found to be equivalent to

$$\mathbf{J}_{x^0} = \left[\partial_{x^0}\left(\frac{\delta J}{\delta p_\alpha}\right)\right]\partial_{q_\alpha} - \left[\partial_{x^0}\left(\frac{\delta J}{\delta q_\alpha}\right)\right]\partial_{p_\alpha}. \tag{12}$$

Let us denote the set of all the operators of the form (10) by Ω. We can prove that Ω closes under the commutation operation defined in (a) above:

Proposition: If two operators $\mathbf{A}$ and $\mathbf{B}$ belong to Ω, the commutator $\mathbf{C} = [\mathbf{B}, \mathbf{A}]$ also belongs to Ω, and its density $\mathcal{C}$ is given by any one of the following:

$$\mathcal{C}_1 = \left(\frac{\delta B}{\delta p_\alpha}\right)\left(\frac{\delta A}{\delta q_\alpha}\right) - \left(\frac{\delta A}{\delta p_\alpha}\right)\left(\frac{\delta B}{\delta q_\alpha}\right),$$
$$\mathcal{C}_2 = \mathbf{B}\mathcal{A}, \quad \mathcal{C}_3 = -\mathbf{A}\mathcal{B}. \tag{13}$$

The proof will be given in the Appendix. Following the usual definition of a Poisson bracket for fields, we have $C = \int \mathcal{C}_1\,dx' = \{B,A\}$. Thus, we might state this as: The commutator of the operators associated with the functionals A and B is equal to the operator assoicated with the functional $\{A,B\}$. We note that the canonical commutation relations among p_α and q_α are not carried over to the operator formalism: The operators corresponding to p_α and q_α are $\mathbf{P}_\alpha = \partial_{q_\alpha}$ and $\mathbf{Q}_\alpha = -\partial_{p_\alpha}$ and they all commute.

II. INVARIANCE GROUPS OF HAMILTON'S EQUATIONS AND CONSERVATION LAWS

We now turn our attention to the theorem stated earlier. The well-known equation which describes the time evolution of a functional $A = \int \mathcal{A}\,dx'$ is

$$\frac{d}{dx^0}A = \{H,A\} + \int \partial_{x^0}\mathcal{A}\,dx'. \tag{14}$$

We associate an operator $\mathbf{K} = [\mathbf{H},\mathbf{A}] + \mathbf{A}_{x^0}$ with the quantity on the right-hand side. In view of (12) and (13), it is obvious that:

The density $\mathcal{K}$ corresponding to the operator $\mathbf{K} = [\mathbf{H},\mathbf{A}] + \mathbf{A}_{x^0}$ is any of the following:

$$\mathcal{K}_1 = \left(\frac{\delta H}{\delta p_\alpha}\right)\left(\frac{\delta A}{\delta q_\alpha}\right) - \left(\frac{\delta A}{\delta p_\alpha}\right)\left(\frac{\delta H}{\delta q_\alpha}\right) + \partial_{x^0}\mathcal{A},$$

or (**)

$$K_2 = \mathbf{H}\mathcal{A} + \partial_{x^0}\mathcal{A}, \quad K_3 = -\mathbf{A}\mathcal{H} + \partial_{x^0}\mathcal{A}.$$

In the following, we prove the theorem by showing basically the following equivalences:

A is a generator of an invariance group of Eq. (1).

A satisfies $[\overset{\updownarrow}{\mathbf{H}}, \mathbf{A}] + \mathbf{A}_{x^0} = 0$

$\mathcal{A}$ satisfies $\overset{\updownarrow}{D_0}\mathcal{A} + \sum_{i=1}^{3} D_i \mathcal{J}^i = 0.$

According to the theory of groups of differential equations,[11] the operator U of (2) becomes a generator of an invariance group of Eqs. (1) if and only if U satisfies the equations

$$U\left(\dot{p}_\alpha + \frac{\delta H}{\delta q_\alpha}\right)\Big|_W = 0, \quad U\left(\dot{q}_\alpha - \frac{\delta H}{\delta p_\alpha}\right)\Big|_W = 0. \tag{15}$$

Here, the symbol $(\cdots)|_W$ means: Evaluate the quantities under conditions (1) and the conditions implied by them. We note that there exist generators which do not take the special form given by (10). We start from the following properties of a generator of an invariance group of Eq. (1):

Lemma 1: The Lie-Bäcklund operator U defined by (2) satisfies the equations

$$([\mathbf{H}, U] + U_{x^0})p_\alpha|_W = 0, \quad ([\mathbf{H}, U] + U_{x^0})q_\alpha|_W = 0, \tag{16}$$

if and only if U is a generator of an invariance group of Hamilton's equation (1).

Proof: In view of the definition of H, under the condition $(\cdots)|_W$ we have an identity $D_0 = \partial_{x^0} + \mathbf{H}$. Using this relation, we obtain $([\mathbf{H}, U] + U_{x^0})p_\alpha|_W = \{-\mathbf{H}G_\alpha + U(\delta H/\delta q_\alpha) - \partial_{x^0}G_\alpha\}|_W = \{U(\delta H/\delta q_\alpha) - D_0 G_\alpha\}|_W = U(\delta H/\delta q_\alpha + \mathring{p}_\alpha)|_W$. Similarly, $([\mathbf{H}, U] + U_{x^0})q_\alpha|_W = U(-\delta H/\delta p_\alpha + \mathring{q}_\alpha)|_W$. These relations obviously prove the statement.

In the following analysis, it is often helpful to consider an initial value problem of Eq. (1). We say that functions $f_\alpha(x')$ and $g_\alpha(x')$, $x' = (x^1, x^2, x^3)$, are admissible if the initial value problem $p_\alpha|_{x^0=t} = f_\alpha$, $q_\alpha|_{x^0=t} = g_\alpha$ has a solution. A set of all such admissible functions will be denoted by I. The following lemma states that Eq. (16) holds without the condition $|_W$.

Lemma 2: If U is a generator of an invariance group of Hamilton's equations (1), the operator $[\mathbf{H}, U] + U_{x^0}$ vanishes identically for arbitrary functions $f_\alpha(x')$ and $g_\alpha(x')$ which belong to I.

Proof: We have, by definition, $[\mathbf{H}, U] + U_{x^0} = M_\alpha \partial_q - N_\alpha \partial_p$, where $M_\alpha(x,p,q) = \mathbf{H}F_\alpha - U(\delta H/\delta p_\alpha) + \partial_{x^0}F_\alpha$, $N_\alpha(x,p,q) = \mathbf{H}G_\alpha - U(\delta H/\delta q_\alpha) + \partial_{x^0}G$. By Lemma 1, if p_α, q_α are solutions of Hamilton's equation, then $M_\alpha = N_\alpha = 0$. We let $x^0 \to t =$ initial time. At t, both M_α and N_α are well defined (note that F_α and G_α do not depend on any x^0 derivatives of p_α and q_α), hence, $M_\alpha = N_\alpha = 0$ at t. Suppose that inital conditions were $p_\alpha = f_\alpha(x')$, $q_\alpha = g_\alpha(x')$. Then, $M_\alpha(x,f,g) = N_\alpha(x,f,g) = 0$ with $x = (t, x^1, x^2, x^3)$. Because t is a parameter of arbitrary value, we may replace x by (x^0, x^1, x^2, x^3) to obtain the desired result.

Remark: If f_α, g_α or any of their derivatives were not defined at some point, the function M_α and N_α, hence the operator $[\mathbf{H}, U] + U_{x^0}$, would not be defined at the point. We note that the relation $[\mathbf{H}, U] + U_{x^0} \equiv 0$ holds even pointwise: For any given values of x, p_α, q_α, $p_{\alpha,i}$, $q_{\alpha,i}$, $\cdots$, the operator vanishes. This should be true as long as there exists a solution which takes the designated values at the given point x. It is also clear that we can allow f_α and g_α to be functions of x instead of x' because x^0, if it appears in f_α and g_α, acts simply as a parameter and has no consequence for the proof given here.

Now, we combine the results obtained above to prove the theorem stated at the beginning:

Proof of theorem: In the following, we assume, that the index i runs from 1 to 3. First we show $D_0\mathcal{A} + D_i\mathcal{J}^i = 0 \to \mathbf{A}$ is a generator. Under condition (1), we have $\mathbf{H}\mathcal{A} + \partial_{x^0}\mathcal{A} = D_0\mathcal{A}$, hence, by the hypothesis

$$\mathbf{H}\mathcal{A} + \partial_{x^0}\mathcal{A} = -D_i\mathcal{J}^i. \tag{†}$$

Because we may assume that neither $\mathcal{A}$ nor $\mathcal{J}^i$ contains time derivatives of p_α and q_α, this relation must hold at initial time $x^0 = t$ where arbitrary initial values may be imposed on p_α and q_α. Consequently, the equation (†) holds not only for solutions p_α and q_α but also for arbitrary functions $f_\alpha(x')$ and $g_\alpha(x')$. Thus, noticing that the left-hand side of the equation (†) is the K_2 of (**), we have $K_2 = -D_i\mathcal{J}^i$. This implies that $\delta K/\delta p_\alpha$ and $\delta K/\delta q_\alpha$ vanish identically, and, as a result, $[\mathbf{H}, \mathbf{A}] + \mathbf{A}_{x^0} \equiv 0$ by (**). In view of Lemmas 1 and 2, we see that this is the necessary and sufficient condition for **A** to be a generator of an invariance group of Eq. (1). Conversely, if **A** is a generator of an invariance group, in view of Lemma 1 we obtain two equations $\delta K/\delta p_\alpha = 0$ and $\delta K/\delta q_\alpha = 0$. According to Lemma 2, these vanish identically. This implies that the density K in (**) must have a divergent form; for instance, $K_2 = (\mathbf{H} + \partial_{x^0})\mathcal{A} = -D_i\mathcal{J}^i$ with $\mathcal{J}^i = \mathcal{J}^i(x,p,q)$. Now if we let p_α and q_α be solutions of Eq. (1), the quantity in the middle of this equation becomes equal to $D_0\mathcal{A}$, and the equation leads to the desired result $D_0\mathcal{A} + D_i\mathcal{J}^i = 0$.

III. INTEGRABILITY OF GENERATORS TO CONSERVED DENSITIES

We have proved that with every conservation law obeyed by the Hamiltonian fields one invariance group is always associated. In the present formulation, the converse of this is true only if the coefficients of the generator U take the special form $F_\alpha = \delta A/\delta p_\alpha$, $G_\alpha = \delta A/\delta q_\alpha$. Because there exists a systematic algorithm for finding generators of invariance groups, it is important to know whether the generators found are integrable to conserved densities. For simplicity, we adopt the following notation:

$$\begin{aligned} &f_\alpha(x) = q_\alpha(x), \quad f_{\alpha+N}(x) = p_\alpha(x) \quad \text{with } \alpha = 1,2,\ldots,N, \\ &S_\alpha(f_\lambda) = F_\alpha, \quad S_{\alpha+N}(f_\lambda) = G_\alpha \quad \text{with } \alpha = 1,2,\ldots,N. \end{aligned} \tag{17}$$

In this notation, our problem is to tell whether a given set of S_α have the property $S_\alpha = \delta A/\delta f_\alpha$ for some functional $A[f_\lambda] = \int \mathcal{A}(x, f_\lambda)\,dx'$. As a general property of a functional, we have

$$\left\{\frac{d}{d\epsilon_\alpha}\frac{d}{d\epsilon_\beta}A[f_\lambda + \epsilon_\lambda\phi_\lambda]\right\}_{\epsilon=0} = \left\{\frac{d}{d\epsilon_\beta}\frac{d}{d\epsilon_\alpha}A[f_\lambda + \epsilon_\lambda\phi_\lambda]\right\}_{\epsilon=0},$$

where $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_{2N})$. If S_α has the desired property,

then, because of the definition $\{(d/d\epsilon_\nu)A[f_\lambda+\epsilon_\lambda\phi_\lambda]\}_{\epsilon=0}$ $=\int(\delta A/\delta f_\nu)\phi_\nu\,dx'$ (ν fixed), this relation is written for fixed α and β as,

$$\left\{\frac{d}{d\epsilon_\alpha}\int S_\beta(f_\lambda+\epsilon_\lambda\phi_\lambda)\phi_\beta\,dx'\right\}_{\epsilon=0} =\left\{\frac{d}{d\epsilon_\beta}\int S_\alpha(f_\lambda+\epsilon_\lambda\phi_\lambda)\phi_\alpha\,dx'\right\}_{\epsilon=0}. \tag{18}$$

This is the integrability condition of the set S_α to a conserved density $\mathcal{A}$. It is not difficult to obtain from this a condition which does not involve integration: Using the fact that the functions ϕ_α and ϕ_β are arbitrary, we can reduce (18) to

$$\sum_{\langle ij\rangle} D^-_{i\cdots j}(\phi\partial_{f_{\beta,i\cdots j}}S_\alpha)=\sum_{\langle ij\rangle}\phi_{,i\cdots j}\partial_{f_{\alpha,i\cdots j}}S_\beta, \tag{19}$$

$\phi(x)=$ arbitrary function

where $D^-_{i\cdots j}=(-D_i)\cdots(-D_j)$, $\phi_{,i\cdots j}=\partial_{x^i}\cdots\partial_{x^j}\,\varphi(x)$, and the notation $\sum_{\langle ij\rangle}a_{i\cdots j}=a+a_i+a_{ij}+\cdots$. All the Roman indices run from 1 to 3. Because ϕ is arbitrary, the coefficients of each $\phi_{,i\cdots j}$ on both sides of (20) must match.

Example: sine-Gordon equation

To illustrate the results obtained above, we study the sine-Gordon equation, using $t=x^0$, $x=x^1$,

$$u_{tt}-u_{xx}+\sin u=0. \tag{20}$$

A canonical form $q_t=\delta H/\delta p$, $p_t=-\delta H/\delta q$ for this equation is obtained by letting $q=u$, $H=\frac{1}{2}p^2+\frac{1}{2}q_x^2-\cos q$:

$$q_t=p,\quad p_t=-q_{xx}+\sin q. \tag{21}$$

In the previous paper,[1] we have shown the equation $u_{xt}=\sin u$ admits an infinite number of invariance groups, and it is straightforward to adapt these results to Eq. (21); four of the generators of invariance groups of Eq. (21) are

$$U_1=q_x\partial_q+p_x\partial_p,\quad U_2=p\partial_q-(-q_{xx}+\sin q)\partial_p.$$

$$U_3=(4q_{xxx}-3q_x\cos q+\tfrac{1}{2}q_x^3+\tfrac{3}{2}q_xp^2)\partial_q-(-4p_{xxx}+3p_x\cos q-\tfrac{3}{2}q_x^2p_x-\tfrac{3}{2}p_xp^2-3q_xq_{xx}p)\partial_p,$$

$$U_4=(4p_{xx}-p\cos q+\tfrac{3}{2}q_x^2p+\tfrac{1}{2}p^3)\partial_q-(-4q_{xxx}+5q_{xx}\cos q-\tfrac{5}{2}q_x^2\sin q-\sin q\cos q-\tfrac{3}{2}q_x^2q_{xx}+\tfrac{1}{2}p^2\sin q-3q_xp_xp-\tfrac{3}{2}p^2q_{xx})\partial_p.$$

Using a theorem given in the previous paper,[8] we see that U_1 and U_2 are equivalent to the space and time translation operators ∂_x and ∂_t. To find conserved densities from these operators we must check condition (19). All of them satisfy the equation, and the conserved density $\mathcal{A}$ associated with each of the generators is found to be:

$$\mathcal{A}_1=pq_x=\text{momentum density},$$

$$\mathcal{A}_2=\tfrac{1}{2}p^2+\tfrac{1}{2}q_x^2-\cos q=\text{energy density},$$

$$\mathcal{A}_3=4pq_{xxx}-3pq_x\cos q+\tfrac{1}{2}q_x^3p+\tfrac{1}{2}q_xp^3,$$

$$\mathcal{A}_4=-2p_x^2-2q_{xx}^2-\tfrac{1}{2}p^2\cos q+\tfrac{3}{4}q_x^2p^2+\tfrac{1}{8}p^4-\tfrac{5}{2}q_x^2\cos q+\tfrac{1}{2}\cos^2 q+\tfrac{1}{8}q_x^4.$$

These conserved densities are related to those obtained by Lamb,[12] and their group theoretic aspects have been studied by the author[1] and by Steudel.[13]

In the previous paper,[2] we also have shown that a series of conservation laws admitted by the nonlinear Shrödinger equation are related to invariance groups of the equation, where we have made use of a special property of the equation. The present results provide a unified view to the previous results.

CONCLUSION

In this paper, we have developed a new group theoretic way of looking at conservation laws associated with field equations in Hamilton's canonical form, and we have proved that the existence of N independent conservation laws necessarily implies the existence of at least N local one-parameter Lie groups which leave the field equations invariant. The condition that a given invariance group is integrable to a conserved density also has been given. Because there exists a well established algorithm for finding generators of invariance groups of differential equations, and because many Euler–Lagrange equations can be put into Hamilton's canonical form, the present results should be useful in finding conservation laws for a variety of systems.[14]

Clearly, the present approach to conservation laws via a Lie bracket formalism is quite different from conventional approachs which make use of Noether's theorem; Noether's theorem as originally derived is too restrictive to give rise to conservation laws such as those dealt with here. However, as this work was being completed, the author learned in a personal communication from N. H. Ibragimov that he has been able to generalized Noether's theorem and with the aid of Lie–Bäcklund contact transformations he has obtained results similar in part to those obtained here.[10]

ACKNOWLEDGMENTS

I am grateful to Professor Carl Wulfman for his continuous interest and encouragement in the present work, and greatly appreciate his helpful discussions and valuable comments on the manuscript. I also would like to express my thanks to the Research Corporation for supporting the research and to Professor N. H. Ibragimov for a personal communication.

APPENDIX: PROOF OF THE PROPOSITION

To simplify expressions, we use notations $D_{i\cdots}D_j=D_{i-j}$ and $(-D_i)\cdots(-D_j)=D^-_{i-j}$; $p_{\alpha,i\cdots j}=p_{\alpha i-j}$ and $q_{\alpha,i\cdots j}=q_{\alpha i-j}$. We represent a sum of the form $f+f_i+f_{ij}+\cdots$ by one term f_{i-j}. For instance, Eq. (2) and Eq. (8) become $U=(D_{i-j}F_\alpha)\partial_{q_{\alpha i-j}}-(D_{i-j}G_\alpha)\partial_{p_{\alpha i-j}}$ and $\delta J/\delta p_\alpha=D^-_{i-j}\mathcal{J}_{p_{\alpha i-j}}$. We first prove the following relations:

$$U\left(\frac{\delta M}{\delta q_\beta}\right)=D^-_{i-j}\left[F_\alpha\left(\frac{\delta M}{\delta q_\alpha}\right)_{q_{\beta i-j}}-G_\alpha\left(\frac{\delta M}{\delta p_\alpha}\right)_{q_{\beta i-j}}\right] \tag{A1}$$

$$U\left(\frac{\delta M}{\delta p_\beta}\right)=D^-_{i-j}\left[F_\alpha\left(\frac{\delta M}{\delta q_\alpha}\right)_{p_{\beta i-j}}-G_\alpha\left(\frac{\delta M}{\delta p_\alpha}\right)_{p_{\beta i-j}}\right]. \tag{A2}$$

This relationship is entirely independent from Eq. (1).

We prove the first relation; the second follows similarly. To prove (A1), we assume that functions p_α and q_α decay sufficiently fast as $[(x^1)^2+(x^2)^2+(x^3)^2]^{1/2}\to\infty$ so that all the surface integrals which appear in the process vanish. Let us consider an integral $\int v(x')U(\delta M/\delta q_\beta)\,dx'$ where v is some arbitrary function except that it does not diverge at infinity. If we write $\partial_{x^i}\cdots\partial_{x^j}v=v_{i\text{-}j}$, then integrating over the whole space

$$
\begin{aligned}
&\int vU\left(\frac{\delta M}{\delta q_\beta}\right)dx' \\
&\quad=\int vUD^-_{i\text{-}j}/\!\!/_{q_{\beta i\text{-}j}}dx'=\int vD^-_{i\text{-}j}U/\!\!/_{q_{\beta i\text{-}j}}dx' \\
&\quad=\int v_{i\text{-}j}U/\!\!/_{q_{\beta i\text{-}j}}dx' \\
&\quad=\int v_{i\text{-}j}[(D_{i'\text{-}j'}F_\alpha)/\!\!/_{q_{\beta i\text{-}j}q_{\alpha i'\text{-}j'}}-(D_{i'\text{-}j'}G_\alpha)/\!\!/_{q_{\beta i\text{-}j}p_{\alpha i'\text{-}j'}}]dx' \\
&\quad=\int[F_\alpha D^-_{i'\text{-}j'}(v_{i\text{-}j}/\!\!/_{q_{\beta i\text{-}j}q_{\alpha i'\text{-}j'}}) \\
&\qquad-G_\alpha D^-_{i'\text{-}j'}(v_{i\text{-}j}/\!\!/_{q_{\beta i\text{-}j}p_{\alpha i'\text{-}j'}})]dx' \\
&\quad=\int(F_\alpha v_{i\text{-}j}\partial_{q_{\beta i\text{-}j}}D^-_{i'\text{-}j'}/\!\!/_{q_{\alpha i'\text{-}j'}} \\
&\qquad-G_\alpha v_{i\text{-}j}\partial_{q_{\beta i\text{-}j}}D^-_{i'\text{-}j'}/\!\!/_{p_{\alpha i'\text{-}j'}})\,dx' \\
&\qquad\qquad\text{(using } [D_i,\ v_{i\text{-}j}\partial_{q_{\beta i\text{-}j}}]=0\text{)} \\
&\quad=\int\left[F_\alpha v_{i\text{-}j}\partial_{q_{\beta i\text{-}j}}\left(\frac{\delta M}{\delta q_\alpha}\right)-G_\alpha v_{i\text{-}j}\partial_{q_{\beta i\text{-}j}}\left(\frac{\delta M}{\delta p_\alpha}\right)\right]dx' \\
&\quad=\int vD^-_{i\text{-}j}\left[F_\alpha\left(\frac{\delta M}{\delta q_\alpha}\right)_{q_{\beta i\text{-}j}}-G_\alpha\left(\frac{\delta M}{\delta p_\alpha}\right)_{q_{\beta i\text{-}j}}\right]dx'.
\end{aligned}
$$

Now, we have the right hand quantity of Eq. (A1) in the last integrand. Both in the starting and in this final form of the integral the function v appears as a factor. Because v is arbitrary, this equation necessarily implies (A1). Next, by definition, $[\mathbf{B},\mathbf{A}]=C^{q\alpha}\partial_{q_\alpha}-C^{p\alpha}\partial_{p_\alpha}$ with

$$
C^{q\alpha}=\mathbf{B}\left(\frac{\delta A}{\delta p_\alpha}\right)-\mathbf{A}\left(\frac{\delta B}{\delta p_\alpha}\right),
$$

$$
C^{p\alpha}=\mathbf{B}\left(\frac{\delta A}{\delta q_\alpha}\right)-\mathbf{A}\left(\frac{\delta B}{\delta q_\alpha}\right).
$$

In view of the equalities (A2), we obtain

$$
\begin{aligned}
C^{q\alpha}&=D^-_{i\text{-}j}\left[-\left(\frac{\delta B}{\delta q_\beta}\right)\left(\frac{\delta A}{\delta p_\beta}\right)_{p_{\alpha i\text{-}j}}+\left(\frac{\delta B}{\delta p_\beta}\right)\left(\frac{\delta A}{\delta q_\beta}\right)_{p_{\alpha i\text{-}j}}\right. \\
&\quad\left.+\left(\frac{\delta A}{\delta q_\beta}\right)\left(\frac{\delta B}{\delta p_\beta}\right)_{p_{\alpha i\text{-}j}}-\left(\frac{\delta A}{\delta p_\beta}\right)\left(\frac{\delta B}{\delta q_\beta}\right)_{p_{\alpha i\text{-}j}}\right] \\
&=D^-_{i\text{-}j}\left[\left(\frac{\delta B}{\delta p_\beta}\right)\left(\frac{\delta A}{\delta q_\beta}\right)-\left(\frac{\delta A}{\delta p_\beta}\right)\left(\frac{\delta B}{\delta q_\beta}\right)\right]_{p_{\alpha i\text{-}j}} \\
&\equiv D^-_{i\text{-}j}\,[C_1]_{p_{\alpha i\text{-}j}}.
\end{aligned}
$$

Similarly, we obtain $C^{p\alpha}=D^-_{i\text{-}j}[C_1]_{q_{\alpha i\text{-}j}}$. Thus, we have proved the assertion for C_1. To prove it for C_2 and C_3, we simply note that they are related to C_1 by $C_2=C_1+D_if^i$ and $C_3=C_1+D_ig^i$ where f^i and g^i are functions of x, p_α, q_α, and of their derivatives and the index i runs from 1 to 3. The fact that functional derivatives of the functional $\int \mathrm{div}h(x,p,q)\,dx'$ always vanish leads to the desired results.

[1]S. Kumei, J. Math. Phys. **16**, 2461 (1975).

[2]S. Kumei, J. Math. Phys. **18**, 256 (1977).

[3]N.H. Ibragimov and R.L. Anderson, Soviet Math. Dokl. **17**, 437 (1976); N.H. Ibragimov and R.L. Anderson, "Lie–Bäcklund Tangent Transformations," to appear in J. Math. Anal. Appl. As these authors have clearly shown, the transformations which have been considered in previous papers [R.L. Anderson, S. Kumei, and C.E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972), and Ref. 1 and 2 above] form groups of infinite order contact transformations; for instance, if we take our present problem as an example, the operator U defined by Eq. (2) is a tangent vector and generates a transformation e^{aU} which is a local one-parameter group of contact transformations defined in a vector space of infinite dimension with coordinates $(p_\alpha, q_\alpha, p_{\alpha,i}, q_{\alpha,i}, p_{\alpha,ij}, q_{\alpha,ij}, \cdots)$. In this space, no finite dimensional subspace exists which closes under the transformation e^{aU} except for special cases where the transformation becomes an ordinary point transformation or a first order contact transformation. The basic idea of the method of calculating group generators is the same as the one due to Lie; for instance, see G.W. Bluman and J.D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).

[4]For instance, see R.M. Miura, Ed., *Bäcklund Transformations, The Inverse Scattering Methods, Solutions, and Their Applications* (Springer, New York, 1976).

[5]A method of casting Euler–Lagrange equations into Hamilton's canonical form is well known for the case where Lagrangian densities involve no derivatives of fields whose orders are higher than one. The case where Lagrangian densities depend on higher derivatives has been studied by T.S. Chang; Proc. Cambridge Philos. Soc. **42**, 132 (1945); **44**, 76 (1948).

[6]D.R. Peterson, M.S. thesis, University of the Pacific, 1976 (unpublished).

[7]R. Abraham and J.E. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1967).

[8]These operators, as they appear, do not generate transformations in independent variables x. However, such transformations are contained in them in disguise as stated in the previous paper [Lemma 2 in Ref. 2 above], and we are not excluding any of such transformations. See the example in Sec. III for instance.

[9]H.H. Johnson, Proc. Am. Math. Soc. **15**, 433, 675 (1964).

[10]N.H. Ibragimov, Dokl. Akad. Nauk SSSR **230**, 26 (1976).

[11]For instance, see L.V. Ovsjanikov, "Group properties of differential equations," Ozdat. Sibirsk. Otdel. Akad. Nauk SSSR, Novosivirsk, 1962 (in Russian).

[12]G.W. Lamb, Jr., Phys. Lett. A **32**, 251 (1970).

[13]H. Steudel, Ann. der Phys. **32**, 205 (1975).

[14]*Note added in proof*: Recently, it has been shown that many of the time evolution equations which are solvable by an inverse scattering method are written in Hamilton's canonical form; Y. Kodama, Prog. Theor. Phys. **54**, 669 (1975); H. Flaschka and A.C. Newell, in *Dynamical Systems: Theory and Applications*, edited by J. Moser (Springer, New York, 1975).

Symmetry and separation of variables for the Hamilton–Jacobi equation $W_t^2 - W_x^2 - W_y^2 = 0$

C. P. Boyer

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México 20, D. F., Mexico

E. G. Kalnins

Mathematics Department, University of Waikato, Hamilton, New Zealand

W. Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota
(Received 3 December 1976)

We present a detailed group theoretical study of the problem of separation of variables for the characteristic equation of the wave equation in one time and two space dimensions. Using the well-known Lie algebra isomorphism between canonical vector fields under the Lie bracket operation and functions (modulo constants) under Poisson brackets, we associate, with each R-separable coordinate system of the equation, an orbit of commuting constants of the motion which are quadratic members of the universal enveloping algebra of the symmetry algebra $o(3,2)$. In this, the first of two papers, we essentially restrict ourselves to those orbits where one of the constants of the motion can be split off, giving rise to a reduced equation with a nontrivial symmetry algebra. Our analysis includes several of the better known two-body problems, including the harmonic oscillator and Kepler problems, as special cases.

INTRODUCTION

This is the first of two papers in which we study the problem of separation of variables for solutions of

$$W_t^2 - W_x^2 - W_y^2 = 0, \tag{0.1}$$

the characteristic equation of the wave equation

$$\Psi_{tt} - \Psi_{xx} - \Psi_{yy} = 0. \tag{0.2}$$

As is well known[1,2] the symmetry algebra of (0.1) defined in terms of operators acting on t, x, y is o(3, 2). Furthermore, there is an isomorphism between the symmetry algebras of (0.1) and (0.2).

In Refs. 3–5 it was shown that every R-separable coordinate system for (0.2) is characterized by a pair of commuting second order symmetric differential operators in the enveloping algebra of o(3, 2). Furthermore, coordinates whose operators lie on the same orbit under the adjoint action of O(3, 2) can be considered as equivalent. The separated special function solutions are eigenfunctions of the commuting symmetry operators and this relationship is a powerful tool for the derivation of special function identities.

In Ref. 2 it was shown that the symmetry algebra of the Hamilton–Jacobi equation for the free particle

$$S_t + S_x^2 = 0 \tag{0.3}$$

(acting in t, x, S space) was also isomorphic to o(3, 2). The problem of additive R-separation of variables for (0.3) was studied in Ref. 6. There all solutions of the form

$$S = R(u, v) + U(u) + V(v) \tag{0.4}$$

were classified where u, v is a new coordinate system, either $R \equiv 0$ or $R \not\equiv 0$ and is not expressible as a sum of a function of u alone and a function of v alone, and U, V are arbitrary solutions of first order ordinary differential equations. It was shown that the separable coordinates agree exactly with the separable coordinates for the free-particle Schrödinger equation

$$i\Psi_t + \Psi_{xx} = 0 \tag{0.5}$$

as derived in Ref. 7. There coordinates are all associated with the Schrödinger subalgebra of o(3, 2). Furthermore, the other elements of o(3, 2) lead to symmetry adapted solutions which do not separate additively as in (0.4) but in some more complicated fashion. In exact analogy with the linear results in Ref. 7 it was also shown that the Hamilton–Jacobi equations for the harmonic oscillator, repulsive oscillator, and linear potential are equivalent to (0.3). Finally, it was pointed out that all these equations are equivalent to (0.1) in the sense that (0.1) is the equation of the graph of each of the considered Hamilton–Jacobi equations.

Here, we look for additive R-separable solutions of (0.1) in the form

$$W = R(u_k) + \sum_{j=1}^{3} U_j(u_j). \tag{0.6}$$

We show by example that the separable systems are the same as those derived in Refs. 3–5 for (0.2) and, properly interpreted the Lie algebraic characterizations of the systems are the same. The proper interpretation of the symmetry algebra of (0.1) is that it is an algebra of functions in the six-dimensional phase space. $\{t, x, y; p_0 = W_t, p_1 = W_x, p_2 = W_y\}$, linear in the p_j, where the commutator is the Poisson bracket. The second order symmetries are formed by taking linear combinations of products of these functions. For the linear case, separable solutions are eigenfunctions of commuting second order differential operators. For the Hamilton–Jacobi case, separable solutions are those for which the corresponding second order functions, commuting under the Poisson bracket operation,

 0022-2488/78/1901-0200$1.00

take constant values. The orbit analysis for the classification of separable systems is identical to that in the linear case.

We follow closely the procedure of Ref. 3 and concentrate here on those systems where one coordinate can be split off by diagonalizing a first order symmetry, leaving a reduced equation in two variables. The additively separable coordinates for the reduced equations (Hamilton–Jacobi equations for the free particle, harmonic oscillator, repulsive oscillator, linear potential, the equation of geometrical optics, etc.) correspond to proper subalgebras of o(3,2). However, utilizing the full symmetry algebra, we find many solutions of these reduced equations which separate nonadditively. On one hand each reduced equation is a special case of (0.1), but also (0.1) is the equation of the graph of the reduced equation. Thus each reduced equation is equivalent to (0.1) and admits the symmetry algebra o(3,2).

We show that the passage from a Hamilton–Jacobi equation to the associated Hamiltonian system provides us with the analogy of a momentum space model in the linear theory. We also indicate how the results of Ref. 4 concerning cyclidic R-separable coordinates for (0.2), in which it is impossible to split off one variable at a time, carry over directly to (0.1). In the second paper we will provide complete proofs concerning the identity of separable coordinates for these two equations. Finally, due to the fact that Lie algebra computations are much easier (though equivalent) for (0.1) than for (0.2), we have been able to find and correct some computational errors in Refs. 3 and 8.

Although this paper concerns only the nonlinear equation (0.1), it should be obvious to the reader that our Lie algebraic procedure can be applied with little change to more general Hamilton–Jacobi equations. Indeed there has been a recent revival of interest in separation of variables[9,10] for general Hamilton–Jacobi equations owning to its usefulness as a solution technique for the Einstein and Einstein–Maxwell equations.[11] (For the classical literature see Ref. 12, and the book by Hagihara,[13] where many applications to celestial mechanics are given.) Of the recent literature dealing with separation of variables for the Hamilton–Jacobi and related second order differential equations in general Riemannian (and pseudo-Riemannian) spaces we mention the works of Havas,[9] Dietz,[10] and Woodhouse.[10] Havas[9] has given the general form of the metric tensor for coordinates which admit complete or partial separation. He also gives the general form of linear and quadratic integrals of motion. Dietz[10] and Woodhouse[10] consider a much more restrictive definition of separation of variables; they additively split off a single variable at a time. In this way one cannot obtain the more general type separable coordinates.[14] None of the above authors allows nontrivial R separation or considers nonadditive separation such as appears here. Furthermore, this and our subsequent paper are the only ones to associate with separable systems orbits of second order members of the enveloping algebra of the symmetry algebra as the correspondings integrals of the motions, thus allowing us to give explicit lists of separable coordinates classified in equivalence classes.

It is our hope that this treatment of a most interesting example will aid in the establishment of more general results for Hamilton–Jacobi equations.

1. BASIC PRINCIPLES

We begin with the equation

$$W_{x^0}^2 - W_{x^1}^2 - W_{x^2}^2 = 0, \quad W_{x^\mu} = \partial_{x^\mu} W(x) \tag{1.1}$$

for the characteristics of the wave equation. As is well known[1,2] the space–time symmetry algebra of (1.1) is o(3,2). That is, the set of Lie derivatives

$$\mathbf{L} = \sum_{\mu=0}^{2} a_\mu(x)\partial_{x^\mu}$$

such that $\mathbf{L}W$ is a solution of (1.1) whenever W is a solution forms the Lie algebra o(3,2) under the operations of addition of Lie derivatives and commutator bracket. A basis for o(3,2) is provided by the elements

$$\begin{aligned} &\mathbf{M}_{\mu\nu} = x_\mu\partial_{x^\nu} - x_\nu\partial_{x^\mu}, \quad \mu \neq \nu, \\ &\mathbf{P}_\mu = \partial_{x^\mu}, \quad \mathbf{D} = x\cdot\partial_x \equiv x^\mu\partial_{x^\mu}, \\ &\mathbf{K}_\mu = 2x_\mu x\cdot\partial_x - x^2\partial_{x^\mu}, \quad 0 \leq \mu, \nu \leq 2, \end{aligned} \tag{1.2}$$

where $x_0 = x^0$, $x_j = -x^j$ for $j = 1,2$, $x^2 = x^\mu x_\mu$ and the Einstein summation convention for repeated indices is adopted. The commutation relations are

$$\begin{aligned} &[\mathbf{M}_{\mu\nu}, \mathbf{M}_{\sigma\rho}] = g_{\mu\rho}\mathbf{M}_{\nu\sigma} + g_{\nu\sigma}\mathbf{M}_{\mu\rho} - g_{\mu\sigma}\mathbf{M}_{\nu\rho} - g_{\nu\rho}\mathbf{M}_{\mu\sigma}, \\ &[\mathbf{M}_{\mu\nu}, \mathbf{P}_\sigma] = g_{\nu\sigma}\mathbf{P}_\mu - g_{\mu\sigma}\mathbf{P}_\nu, \\ &[\mathbf{P}_\mu, \mathbf{P}_\nu] = [\mathbf{K}_\mu, \mathbf{K}_\nu] = [\mathbf{M}_{\mu\nu}, \mathbf{D}] = 0, \\ &[\mathbf{D}, \mathbf{P}_\mu] = -\mathbf{P}_\mu, \quad [\mathbf{D}, \mathbf{K}_\mu] = \mathbf{K}_\mu, \\ &[\mathbf{M}_{\mu\nu}, \mathbf{K}_\sigma] = g_{\nu\sigma}\mathbf{K}_\mu - g_{\mu\sigma}\mathbf{K}_\nu, \\ &[\mathbf{K}_\mu, \mathbf{P}_\nu] = -2(\mathbf{M}_{\mu\nu} + g_{\mu\nu}\mathbf{D}), \end{aligned} \tag{1.3}$$

where $g_{00} = -g_{jj} = 1$, $g_{\mu\nu} = 0$ for $\mu \neq \nu$.

These operators can be exponentiated to yield a local Lie transformation group[2] of symmetries of (1.1). Indeed the operator $\mathbf{M}_{\mu\nu}$, $\mathbf{P}_\mu$ generate the Poincaré symmetry group

$$W(x) \to W(\Lambda^{-1}(x-a)), \quad a = (a_0, a_1, a_2), \quad \Lambda \in \mathrm{SO}(1,2), \tag{1.4}$$

the dilatation operator generates the symmetry

$$(\exp\lambda\mathbf{D})W(x) = X(e^\lambda x) \tag{1.5}$$

and the $\mathbf{K}_\mu$ generate the special conformal transformations

$$\exp(a^\mu K_\mu)W(x_\nu) = W\left(\frac{x_\nu - a_\nu x^2}{1 - 2a\cdot x + a^2x^2}\right). \tag{1.6}$$

We shall also consider the inversion, space reflection, and time reflection symmetries of (1.1),

$$\begin{aligned} &\mathbf{I}W(x) = W(-x/x^2), \quad \mathbf{R}W(x) = W(x_0, x_1, -x_2), \\ &\mathbf{T}W(x) = W(-x_0, x_1, x_2), \end{aligned}$$

which are not generated by the Lie derivatives (1.2).

In Ref. 6 the full infinite-dimensional symmetry algebra of an arbitrary first order partial differential equation was computed and shown that this algebra splits into symmetries which are contact transforma-

tions and an ideal of characteristics. Now there is a well-known Lie algebra isomorphism between canonical vector fields on phase space with the usual Lie brackets and functions on phase space (modulo constants) with the Poisson brackets.[1] Explicitly, given phase space with coordinates (x^μ, p_μ) we have the canonical 2-form $\omega = dp_\mu \wedge dx^\mu$. Then with each vector field X on phase space which leaves ω invariant, we can associate a function $F(x^\mu, p_\nu)$ such that

$$X \lrcorner \omega = dF, \tag{1.7}$$

where $\lrcorner$ denotes the inner product between vector fields and forms. To the Lie brackets for vector fields there correspond the Poisson brackets

$$\{F(x,p), G(x,p)\} = \Sigma\left(\frac{\partial G}{\partial x^\mu}\frac{\partial F}{\partial p_\mu} - \frac{\partial F}{\partial x^\mu}\frac{\partial G}{\partial p_\mu}\right) \tag{1.8}$$

for functions. Explicitly for the Lie derivatives (1.2) we have, using (1.7),

$$\begin{aligned} &M_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu, \quad P_\mu = p_\mu, \quad D = x^\mu p_\mu, \\ &K_\mu = 2x_\mu(x^\nu p_\nu) - x^2 p_\mu. \end{aligned} \tag{1.9}$$

One can easily check that the basis functions (1.9) satisfy relations (1.3) under the Poisson bracket operation. From the point of view of separation of variables of (1.1), the Lie algebraic characterization (1.9) in terms of functions on phase space is superior to that of Lie derivatives.

By taking all possible products of operators (1.2) we can generate an enveloping algebra[15] of so(3, 2). Furthermore, we can identify the subspace $\mathcal{S}_k'$ of homogeneous symmetric kth-order elements in the enveloping algebra with the space $\mathcal{S}_k$ of kth-order polynomials in the basis functions (1.9). That is, the two subspaces are isomorphic as vector spaces and the adjoint action of so(3, 2) on $\mathcal{S}_k'$ induced by the commutator $[\cdot, \cdot]$ agrees with the adjoint action on $\mathcal{S}_k$ induced by the Poisson bracket. In particular, $\mathcal{S}_2'$ is spanned by elements of the form $[\mathbf{L}_1, \mathbf{L}_2]_+ \equiv \mathbf{L}_1\mathbf{L}_2 + \mathbf{L}_2\mathbf{L}_1$, where the $\mathbf{L}_j$ are Lie derivatives belonging to the symmetry algebra. Let $\mathbf{L}_1, \mathbf{L}_2$ be the corresponding functions in the Lie algebra (1.9). Then the correspondence

$$[\mathbf{L}_1, \mathbf{L}_2]_+ \to 2L_1L_2, \tag{1.10}$$

extended by linearity provides the stated isomorphism between $\mathcal{S}_2'$ and $\mathcal{S}_2$.

As is well known there is an intimate relationship between a first-order partial differential equation

$$H(x^\mu, p_\mu) = 0, \quad p_\mu = \frac{\partial W}{\partial x^\mu}, \quad 0 \le \mu \le n, \tag{1.11}$$

and the Hamiltonian system of ordinary differential equations[16]

$$\frac{\partial p_\mu}{\partial \tau} = -H_{x^\mu}, \quad \frac{dx^\mu}{d\tau} = H_{p_\mu}, \quad 0 \le \mu \le n. \tag{1.12}$$

Indeed, consider the n-dimensional surface $x_\mu = x_\mu(t_1, \ldots, t_n)$ and prescribe initial data on this surface:

$$W = W(t_1, \ldots, t_n), \quad p_\mu = p_\mu(t_1, \ldots, t_n), \tag{1.13}$$

subject to the requirements

$$\frac{\partial W}{\partial t_j} = p_\mu \frac{\partial x^\mu}{\partial t_j}, \quad j = 1, \ldots, n,$$

$$H(x^\mu(t, \ldots, t_n), p_\mu(t_1, \ldots, t_n)) = 0.$$

Then, provided

$$\det\begin{bmatrix} H_{p_0} \cdots H_{p_n} \\ \frac{\partial x_0}{\partial t_1} \cdots \frac{\partial x_n}{\partial t_1} \\ \cdots\cdots\cdots \\ \frac{\partial x_0}{\partial t_n} \cdots \frac{\partial x_n}{\partial t_n} \end{bmatrix} \neq 0$$

on the surface, the solutions of (1.12) with initial data (1.13) generate a local solution of (1.11). The function W can be obtained either from the equation

$$\frac{dW}{d\tau} = p_\mu H_{p_\mu} \tag{1.14}$$

or the defining relations $p_\nu = W_{x^\nu}$.

Conversely, let

$$W = f(x^\mu, a_1, \ldots, a_n) + a_0 \tag{1.15}$$

be a complete integral of (1.11), i.e., W is a solution of (1.11) for each choice of the $n+1$ real constants a_μ and the $n \times (n+1)$ matrix

$$(\partial_{a_j}\partial_{x^\mu} f) = (f_{a_j x^\mu})$$

has rank n. Then (1.15) and relations

$$\begin{aligned} &f_{a_j}(x^\mu, a_\mu) = \lambda_j, \quad j = 1, \ldots, n \\ &p_\nu = f_{x^\nu}(x^\mu, a_\mu), \quad \nu = 0, \ldots, n \end{aligned} \tag{1.16}$$

with $a_0, \ldots, a_n$, $\lambda_1, \ldots, \lambda_n$ fixed, define a solution of the characteristic system (1.12).

It is also well known that the canonical transformation generated by (1.12) preserves Poisson brackets.[1] Thus, if

$$F_j^\tau(x^\mu, p_\nu) = F_j(x^\mu(\tau), p_\nu(\tau)), \quad j = 1, 2,$$

where

$$x^\mu(\tau) = x^\mu(\tau, x^{\mu'}, p_{\nu'}), \quad p_\nu(\tau) = p_\nu(\tau, x^{\mu'}, p_{\nu'}), \tag{1.17}$$

are solutions of (1.12) such that $x^\mu(0) = x^\mu$, $p_\nu(0) = p_\nu$, then

$$\{F_1^\tau, F_2^\tau\} = \{F_1, F_2\}. \tag{1.18}$$

Furthermore,

$$\frac{d}{d\tau}F^\tau = F_{x^\mu}H_{p_\mu} - F_{p_\mu}H_{x^\mu} = \{H^\tau, F^\tau\}$$

so that $F^\tau \equiv F$ if F commutes with H.

Applying this theory to Eq. (1.1), we find

$$H = p_0^2 - p_1^2 - p_2^2, \tag{1.19}$$

so that the associated Hamiltonian system is

$$\frac{dp_\mu}{d\tau} = 0, \quad \frac{dx^\mu}{d\tau} = \tfrac{1}{2}g^{\mu\nu}p_\nu, \quad 0 \le \mu, \nu \le 2. \tag{1.20}$$

Thus we can obtain a solution of (1.1) by prescribing initial data for W, x^μ, and p_μ on a two-dimensional

surface in x-space and solving Eqs. (1.20). For some of our computations we shall choose this surface and data in the special form

$$x^0=0,\quad p_0=(p_1^2+p_2^2)^{1/2},\quad x^1=t_1,\quad x^2=t_2,$$
$$p_1=p_1(t_1,t_2),\quad p_2=p_2(t_1,t_2). \tag{1.21}$$

Note that the basis functions (1.8) restricted to this surface become

$$\mathcal{P}_0=(p_1^2+p_2^2)^{1/2}=p_0,\quad \mathcal{P}_j=p_j,\quad \mathcal{M}_{ij}=x_ip_j-x_jp_i,$$
$$\mathcal{M}_{0j}=x^jp_0,\quad \mathcal{K}_0=[(x_1)^2+(x_2)^2]p_0,\quad \mathcal{D}=x^ip_i, \tag{1.22}$$
$$\mathcal{K}_j=2x_j(x^ip_i)+[(x_1)^2+(x_2)^2]p_j,\quad i,j=1,2.$$

Model (1.22) and its relationship to (1.1) via integration of the Hamiltonian system (1.20) is an analogy of the Fourier transform model for the solution space of the wave equation (0.2) which was treated in Ref. 3.

We now introduce another basis for the symmetry algebra which makes explicit the isomorphism with the usual matrix realization of o(3,2). The matrix algebra o(3,2) is usually defined as the ten-dimensional Lie algebra of 5×5 matrices A such that $AG+GA^t=0$, where 0 is the zero matrix and

$$G=\begin{pmatrix}1&&&&\\&1&&0&\\&&1&&\\&0&&-1&\\&&&&-1\end{pmatrix}. \tag{1.23}$$

Let $\mathcal{E}_{ij}$ be the 5×5 matrix with a 1 in row i, column j and zeros elsewhere. Then the matrices

$$\Gamma_{ab}=\mathcal{E}_{ab}-\mathcal{E}_{ba}=-\Gamma_{ba},\quad a\neq b,$$
$$\Gamma_{aB}=\mathcal{E}_{aB}+\mathcal{E}_{Ba}=-\Gamma_{Ba},\quad 1\leq a,b\leq3, \tag{1.24}$$
$$\Gamma_{45}=\mathcal{E}_{54}-\mathcal{E}_{45}=-\Gamma_{54},\quad B=4,5$$

form a basis for o(3,2) with commutation relations

$$[\Gamma_{\alpha\beta},\Gamma_{\gamma\zeta}]=G_{\beta\gamma}\Gamma_{\alpha\zeta}+G_{\alpha\zeta}\Gamma_{\beta\gamma}+G_{\gamma\alpha}\Gamma_{\zeta\beta}+G_{\zeta\beta}\Gamma_{\gamma\alpha}. \tag{1.25}$$

This basis is related to our basis (1.2) by the identifications

$$\mathbf{P}_0=\Gamma_{14}+\Gamma_{45},\quad \mathbf{P}_1=\Gamma_{25}+\Gamma_{12},\quad \mathbf{P}_2=\Gamma_{35}+\Gamma_{13},$$
$$\mathbf{K}_0=\Gamma_{45}-\Gamma_{14},\quad \mathbf{K}_1=\Gamma_{25}-\Gamma_{12},\quad \mathbf{K}_2=\Gamma_{35}-\Gamma_{13}, \tag{1.26}$$
$$\mathbf{M}_{12}=\Gamma_{32},\quad \mathbf{M}_{01}=\Gamma_{24},\quad \mathbf{M}_{02}=\Gamma_{34},\quad \mathbf{D}=\Gamma_{51}.$$

Here and hereafter we denote by $\mathbf{\Gamma}(\Gamma)$ the vector field (function) which corresponds to the matrix Γ.

Returning to our Poisson bracket model (1.9), we note that if we impose the relation

$$P_0^2-P_1^2-P_2^2=0 \tag{1.27}$$

to obtain solutions of (1.1), we introduce linear dependencies among the elements of $\mathcal{S}_2$. Although there are formally 35 independent terms L_iL_j where the L_i run over a basis for o(3,2), among the explicit functions (1.9) subject to (1.27) there are 20 independent relations obeyed by the L_iL_j. Hence, if $\mathcal{Q}_2$ is the subspace of $\mathcal{S}_2$ which is mapped to zero under this identification, then $\mathcal{Q}_2$ is actually an ideal under the adjoint action of O(3,2) and the factor space $\mathcal{J}_2\equiv\mathcal{S}_2/\mathcal{Q}_2$ is 15-dimensional. In particular, on $\mathcal{J}_2$ we have the relations

$$K_0^2-K_1^2-K_2^2=0,\quad \Gamma_{12}^2+\Gamma_{13}^2+\Gamma_{23}^2=\Gamma_{45}^2,$$
$$M_{12}^2-M_{01}^2-M_{02}^2=-D^2,\quad \Gamma_{45}^2-\Gamma_{41}^2-\Gamma_{51}^2=\Gamma_{23}^2. \tag{1.28}$$

[Of course, these relations also hold in the model (1.22).]

We shall now classify all separable solutions of (1.1) which are analogous to the separable solutions of the wave equation (1.23), studied in Refs. 3–5. In these references each (R-) separable solution of (1.23) was characterized by the fact that the corresponding separated solutions were common eigenfunctions of a commuting pair S_1, S_2 of second-order differential symmetry operators from the enveloping algebra of so(3,2). Analogously we will characterize separated solutions W of (1.1) by the conditions $F_1(x^\mu,p_\nu)=\lambda$, $F_2(x^\mu,p_\nu)=\mu$, where $F_1, F_2\in\mathcal{J}_2$ and $\{F_1,F_2\}=0$. Here $p_\nu=W_{x^\nu}$ and λ,μ are the separation constants. For all cases treated in this paper we shall see that coordinates which yield separation for (1.23), corresponding to commuting operators S_1, S_2, also yield separation for (1.1) corresponding to commuting functions $F_1, F_2\in\mathcal{J}_2$. Here, S_j and F_j are related by (1.10). We also mention, although we will not make use of it, that corresponding to each commuting pair (F_1,F_2) there is via (1.7) a commuting pair of vector fields (X_1,X_2) on phase space and hence a two-parameter local Abelian group of symmetries.

In analogy with Ref. 3 we shall begin by studying separable systems which can be characterized by commuting functions F_1, F_2 such that $F_1=A^2$ for some $A\in$ so(3,2) with $\{A,F_2\}=0$—and where A has a nontrivial centralizer.

In Ref. 3 the corresponding systems were called "semi-subgroup" coordinates. However, as has been pointed out by Winternitz, this name is not appropriate because many coordinate systems which correspond to the restriction of so(3,2) to a subalgebra are not "semi-subgroup" systems. Thus, we shall now call these systems "semi-split" coordinates.

In the following seven sections we shall study systems of the form A^2, F_2 by first splitting off the coordinate associated with A to obtain a reduced equation (†) such that F_2 belongs to the space of second-order elements in the enveloping algebra of the symmetry algebra (centralizer of A) of (†). We shall then investigate the coordinate systems in which (†) separates. In the last section we shall give an example of a completely nonsplitting type coordinate. In the study of separable coordinates it is common to associate with each coordinate system $x^\mu=x^\mu(y_\nu)$ a metric tensor $g_{\mu\nu}$ and a quadratic differential form

$$ds^2=dx_0^2-dx_1^2-dx_2^2=g_{\mu\nu}dy^\mu dy^\nu \tag{1.29}$$

then the Hamilton–Jacobi equation (1.1) takes the form

$$g^{\mu\nu}W_{y^\mu}W_{y^\nu}=0 \tag{1.30}$$

in the new coordinate system, where $g^{\mu\nu}$ is the inverse tensor to $g_{\mu\nu}$ given by

$$g^{\mu\nu}=\frac{(-1)^{\mu+\nu}\,\mathrm{cof}\,g_{\mu\nu}}{\det g_{\mu\nu}}. \tag{1.31}$$

We also mention for convenience that for orthogonal coordinates (1.29) takes the form

$$ds^2 = \sum_i h_i^2 (d\xi^i)^2 \tag{1.32}$$

and the relation between the canonical moment in p_μ in Cartesian coordinates and arbitrary curvalinear coordinates ξ^i is

$$p^\mu = \sum_i h_i^{-2} \frac{\partial x^\mu}{\partial \xi^i} p_{\xi^i}. \tag{1.33}$$

The above relations will be used throughout our computations without further mention.

2. THE SPHERE

For our case we consider the function $A = \Gamma_{45} = \frac{1}{2}(P_0 + K_0)$. Setting $\Gamma_{45} = \lambda$, λ const, we see from (1.29) that (1.1) reduces to

$$\Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{23}^2 = \lambda^2. \tag{2.1}$$

Since the centralizer of Γ_{45} in o(3,2) is $\{\Gamma_{45}\} \oplus o(3)$, where o(3) is the subalgebra with basis $\Gamma_{12}, \Gamma_{13}, \Gamma_{23}$, it follows that o(3) is a symmetry algebra for (2.1). We call it the *reduced* symmetry algebra.

Equation (2.1) can be viewed as the result of separating off one variable ψ in W. Indeed, we choose new coordinates such that $\Gamma_{45} = -\partial_\psi$. Standard Lie theory gives

$$x^0 = \frac{\sin\psi}{y_1 - \cos\psi}, \quad x^1 = \frac{y_2}{y_1 - \cos\psi}, \quad x^2 = \frac{y_3}{y_1 - \cos\psi}, \quad y_1^2 + y_2^2 + y_3^2 = 1. \tag{2.2}$$

Thus, choosing any parametrization $y_j(\sigma, \alpha)$ of the unit sphere S_2, we obtain a new set of coordinates for space time. In these coordinates we have $\Gamma_{23} = y_2\partial_{y_3} - y_3\partial_{y_2}$,

$$\Gamma_{12} = y_1\partial_{y_2} - y_2\partial_{y_1}, \quad \Gamma_{13} = y_1\partial_{y_3} - y_3\partial_{y_1}.$$

The equation $\Gamma_{45} = \lambda$ or, what is the same thing, $\Gamma_{45} W = \lambda$ implies

$$W = -\lambda\psi + S(\sigma, \alpha). \tag{2.3}$$

Substituting this expression into (1.1), we obtain the reduced Eq. (2.1) for S.

As is well known[17] the space of second order symmetry operators in o(3), modulo the invariant $\Gamma_{12}^2 + \Gamma_{13}^2 + \Gamma_{23}^2$, splits into exactly two orbit types under the adjoint action of o(3). A representative on each orbit type is given by the assignment

(1) $\Gamma_{45}^2, \Gamma_{23}^2$,

(2) $\Gamma_{45}^2, \Gamma_{12}^2 + a^2\Gamma_{13}^2, \quad 1 > a > 0$.

For the orbit of type (1) we introduce spherical coordinates on S_2

$$y_1 = \cos\sigma, \quad y_2 = \sin\sigma\cos\alpha, \quad y_3 = \sin\sigma\sin\alpha. \tag{2.4a}$$

Then (2.1) becomes

$$\csc^2\sigma p_\alpha^2 + p_\sigma^2 = \lambda^2, \quad p_\alpha = \frac{\partial S}{\partial \alpha}, \quad p_\sigma = \frac{\partial S}{\partial \sigma}, \tag{2.4b}$$

and the requirement $\Gamma_{23} = p_\alpha = m$ yields the separated solution

$$S = m\alpha + \int (m^2 \csc^2\sigma - \lambda^2)^{1/2} d\sigma,$$
$$= m\alpha + \frac{im}{2} \ln\left(\frac{\Delta + \cos\sigma}{\Delta - \cos\sigma}\right) - i\lambda \ln\left(\frac{\lambda}{m}\cos\sigma + \Delta\right) + c,$$
$$\Delta = [1 - (\lambda^2/m^2)\sin^2\sigma]^{1/2}. \tag{2.4c}$$

For the orbit of type (2) we introduce elliptic coordinates on S_2:

$$y_1 = k'^{-1}\,\mathrm{dn}(\alpha, k)\,\mathrm{dn}(\sigma, a), \quad y_2 = ikk'^{-1}\,\mathrm{cn}(\alpha, k)\,\mathrm{cn}(\sigma, k),$$
$$y_3 = k\,\mathrm{sn}(\alpha, k)\,\mathrm{sn}(\sigma, k), \tag{2.5a}$$

where $k' = (1 - k^2)^{1/2}$ and $\mathrm{dn}(\alpha), \mathrm{cn}(\alpha), \mathrm{sn}(\alpha)$ are Jacobi elliptic functions.[18] Then (2.1) becomes

$$p_\alpha^2 - p_\sigma^2 = -\lambda^2 k^2(\mathrm{sn}^2\alpha - \mathrm{sn}^2\sigma)$$

and the condition

$$\Gamma_{12}^2 + k^2\Gamma_{13}^2 = (\mathrm{sn}^2\alpha - \mathrm{sn}^2\sigma)^{-1}(\mathrm{sn}^2\alpha p_\alpha^2 - \mathrm{sn}^2\sigma p_\alpha^2) = \mu \tag{2.5b}$$

leads to the separated solution

$$S = \int (-\lambda^2 a^2 \mathrm{sn}^2\alpha + \mu)^{1/2} d\alpha + \int (-\lambda^2 a^2 \mathrm{sn}^2\sigma + \mu)^{1/2} d\sigma. \tag{2.5c}$$

There is a close relationship between our own study of Γ_{45} and the Hamilton–Jacobi equation for the Kepler problem with closed orbits in two-dimensional space. Indeed, on the surface (1.21) the condition $\Gamma_{45} = \lambda$ for a solution of (1.1) becomes

$$(x^1)^2 + (x^2)^2 - 2\lambda/p_0 = -1, \quad p_0 = (p_1^2 + p_2^2)^{1/2}. \tag{2.6}$$

Performing the canonical transformation $p_j \to x^j$, $x^j \to -p_j$, which preserves Poisson brackets, we transform (2.6) to the Hamilton–Jacobi equation for the Kepler problem with energy normalized to -1 (bound orbits), viz.,

$$p_1^2 + p_2^2 - 2\lambda/r = -1, \quad r = (x_1^2 + x_2^2)^{1/2}. \tag{2.7}$$

Moreover, the o(3) symmetry algebra for (2.6) generated by $\Gamma_{12}, \Gamma_{13}, \Gamma_{23}$ is mapped to an o(3) symmetry algebra for (2.7). If $\psi(\xi_1, \xi_2)$ is a solution of (2.7) (with $x^1 = \xi_1$, $x^2 = \xi_2$), then by prescribing the initial data $x^j = \psi_{\xi_j}$, $p_j = \xi_j$ on the surface (1.21) and integrating along characteristics, we find a solution of (1.1) with $\Gamma_{45} = \lambda$. Conversely, if $W(x^\mu)$ is a solution of (1.1) with $\Gamma_{45} = \lambda$, then a function $\psi(\xi_1, \xi_2)$ such that

$$x^j = \psi_{\xi_j}, \quad W_{x^j}(0, x^l) = \xi_j, \quad j, l = 1, 2 \tag{2.8}$$

with $\det(W_{x^j x^l}) \neq 0$ is a solution of (2.7). This relationship is a classical analogy of Fock's treatment of the quantum mechanical hydrogen atom,[19] and underlies the group theoretical approach to the Kepler problem.[20]

We have obtained the reduced equation (2.1) from (1.1) by additively separating off dependence on the variable ψ. However, (1.1) can also be viewed as the equation for the graph of (2.1). Indeed, set $\lambda = 1$ for simplicity and let $S(\sigma, \alpha)$ be a solution of (2.1). A graph of S is a function $W(\sigma, \alpha, S)$ such that $W(\sigma, \alpha, S(\sigma, \alpha)) = 0$. Since $W_\sigma + W_S S_\sigma = 0$, $W_\alpha + W_S S_\alpha = 0$ it follows from (2.2)–(2.4) that W satisfies Eq. (1.1) with ψ replaced by S. In this sense Eqs. (1.1) and (2.1) are equivalent.

Since (1.1) admits the symmetry algebra o(3,2), it induces an action of o(3,2) as a symmetry algebra[2,6] of (2.1). Now, however, o(3,2) acts not only on σ, α but also on S. We have used only the o(3) subgroup of o(3,2) to explain the two systems in which (2.1) admits an additive separation of variables. (We believe that there are only two such systems and will settle this point in a future publication.) However, we can use commuting pairs of second-order elements in the enveloping algebra of o(3,2) to distinguish many other symmetry adapted solutions of (2.1). [For example, some other solutions may correspond to a product separation in (2.1).] Indeed, every separable solution of (1.1) corresponds via the graph to some symmetry adapted solution of (2.1). Our restriction to the subalgebra o(3) merely picks out those solutions which are additively separable for (2.1).

3. THE EQUATION OF GEOMETRICAL OPTICS

In this section we consider the equation obtained from (1.1) by partial separation via the operator $\mathbf{P}_0$, i.e., we treat the usual equation of geometrical optics obtained from (1.1) by putting $P_0^2 = \lambda^2$, viz.,

$$W = \lambda x^0 + S(x^1, x^2), \tag{3.1a}$$

$$S_{x^1}^2 + S_{x^2}^2 = P_1^2 + P_2^2 = \lambda^2. \tag{3.1b}$$

It is easy to check that the centralizer of $\mathbf{P}_0$ in o(3,2) is $\{\mathbf{P}_0\} \oplus e(2)$, where e(2) is generated by $\{\mathbf{M}_{12}, \mathbf{P}_1, \mathbf{P}_2\}$. However, in this case P_0 has a normalizer which is bigger than its centralizer, and for the purpose of separation of variables it is convenient to classify orbits using the full normalizer group $D \otimes E(2)$, where D is the one-parameter group of dilatations generated by D. For (3.1b) there are four separable orthogonal coordinate systems corresponding precisely to the four orbit types[21] of the quadratic members of the universal enveloping algebra e(2) (modulo $P_1^2 + P_2^2$) under the adjoint action of $D \otimes E(2)$. The list of pairs of orbit representatives is

(3) P_0^2, P_1^2, Cartesian,

(4) P_0^2, M_{12}^2, polar,

(5) P_0^2, $M_{12}P_2$, parabolic,

(6) P_0^2, $M_{12}^2 + P_2^2$, elliptic.

The coordinate systems and corresponding solutions (3.1a) are now given for each of the above cases:

(3) *Cartesian*: The coordinates are the usual Cartesian coordinates

$$x^0 = t, \quad x^1 = x, \quad x^2 = y \tag{3.2a}$$

with the solution

$$S = \mu x + (\lambda^2 - \mu)^{1/2} y, \tag{3.2b}$$

whose separation constant is

$$P_1^2 = p_x^2 = \mu^2. \tag{3.2c}$$

(4) *Polar*: The coordinates are

$$x^0 = t, \quad x^1 = r\cos\theta, \quad x^2 = r\sin\theta \tag{3.3a}$$

and the well-known solution

$$S = \mu\theta + \int dr(\lambda^2 - \mu^2/r^2)^{1/2} \tag{3.3b}$$

with the constant of the motion

$$M_{12}^2 = p_\theta^2 = \mu^2. \tag{3.3c}$$

(5) *Parabolic*: The coordinates are

$$x^0 = t, \quad x^1 = (\xi^2 - \eta^2)/2, \quad x^2 = \xi\eta \tag{3.4a}$$

with $-\infty < \xi < \infty$, $0 \leq \eta < \infty$. The solution is

$$S = \int (\lambda^2\xi^2 - \mu)^{1/2} d\xi + \int (\lambda^2\eta^2 + \mu)^{1/2} d\eta \tag{3.4b}$$

with separation constant

$$2M_{12}P_2 = (\xi^2 + \eta^2)^{-1}(\xi^2 p_\eta^2 - \eta^2 p_\xi^2) = \mu.$$

(6) *Elliptic*: The coordinates are

$$x^0 = t, \quad x^1 = \cosh\rho\cos\sigma, \quad x^2 = \sinh\rho\sin\sigma, \tag{3.5a}$$

with $-\infty < \rho < \infty$, $0 \leq \sigma < 2\pi$. The solution is

$$S = \int (\lambda^2\cosh^2\rho - \mu)^{1/2} d\rho + \int (\mu - \lambda^2\cos^2\sigma)^{1/2} d\sigma \tag{3.5b}$$

with constant of the motion

$$M_{12}^2 + P_1^2 = (\cosh^2\rho - \cos^2\sigma)^{-1}(\cos^2\sigma p_\rho^2 + \cosh^2\rho p_\sigma^2) = \mu. \tag{3.5c}$$

Just as in the previous section, we can also interpret (1.1) as the equation for the graph of a solution of (3.1b). In this sense (3.1b) admits the full symmetry algebra and any separable system for (1.1) gives rise to a symmetry adapted solution of (3.1b).

4. THE FREE RELATIVISTIC PARTICLE

By separating off one space variable, say x^2, via the operator $\mathbf{P}_2$, Eq. (1.1) reduces to the equation for a free relativistic particle in one space and one time dimension. Explicitly, putting $P_2^2 = \lambda^2$, we find

$$W = \lambda x^2 + S(x^1, x^0), \tag{4.1a}$$

$$(S_{x^0})^2 - (S_{x^1})^2 = P_0^2 - P_1^2 = \lambda^2. \tag{4.1b}$$

Again a straightforward calculation shows that the centralizer of $\mathbf{P}_2$ in o(3,2) is $\{\mathbf{P}_2\} \oplus e(1,1)$. A basis for the subalgebra e(1,1) is given by $\{\mathbf{M}_{01}, \mathbf{P}_0, \mathbf{P}_1\}$. The orbit analysis of the quadratic members of the universal enveloping algebra of e(1,1) (modulo $\mathbf{P}_0^2 - \mathbf{P}_1^2$) under the adjoint action of the normalizer group $D \otimes E(1,1)$ extended by certain discrete transformations was given in Ref. 22. There it was also seen that there is a nonuniqueness for the orbit corresponding to the separation of the Klein–Gordon equation in the usual Cartesian coordinates. This nonuniqueness was resolved by considering nonorthogonal coordinates. Here, however, only orthogonal coordinates are considered. Furthermore, there is one orbit for which the Klein–Gordon equation (Laplace operator) does not admit a separation of variables. The question as to whether there is also no separable coordinate system for (3.6b) corresponding to this orbit representative $[\mathbf{M}_{01}(\mathbf{P}_0 + \mathbf{P}_1)]$ will be answered in Paper II.

The list of orbits representatives is

(3′) P_2^2, P_0P_1, Cartesian,

(7) P_2^2, M_{01}^2, polar,

(8) P_2^2, $2M_{01}P_1$, parabolic—type 1,

(9) P_2^2, $M_{01}^2-(P_0+P_1)^2$, hyperbolic—type 3,

(10) P_2^2, $M_{01}^2+(P_0+P_1)^2$, hyperbolic—type 2,

(11) P_2^2, $2M_{01}(P_0+P_1)+(P_0-P_1)^2$, parabolic—type 2,

(12) P_2^2, $M_{01}^2-P_0P_2$, hyperbolic—type 1,

(13) P_2^2, $M_{01}^2+P_1^2$, elliptic—type 1,

(14) P_2^2, $M_{01}^2-P_1^2$, elliptic—type 2.

From the point of view of separable coordinates (3′) is equivalent to (3). The remainder of the coordinates and solutions are:

(7) *Polar*: The coordinates for the region $(x^0)^2-(x^1)^2>0$ are

$$x^0=\pm r\cosh\eta,\quad x^1=r\sinh\eta, \tag{4.2a}$$

$0\le r<\infty$, $-\infty<\eta<\infty$. The solutions are

$$S=\pm\mu\eta+\int dr(\lambda^2+\mu^2/r^2)^{1/2}. \tag{4.2b}$$

For the region $(x^0)^2-(x^1)^2<0$, interchange x^0 and x^1 and find (4.2b) with $\lambda^2\to-\lambda^2$. The separation constant is

$$M_{01}^2=P_1^2=\mu^2. \tag{4.2c}$$

(8) *Parabolic—type* 1: The coordinates are

$$x^0=\pm\tfrac{1}{2}(\xi^2+\eta^2),\quad x^1=\xi\eta, \tag{4.3a}$$

with $-\infty<\xi<\infty$, $0\le\eta<\infty$. This parametrizes the region $(x^0)^2-(x^1)^2>0$. The solutions are

$$S=\int d\xi(\mu+\lambda^2\xi^2)^{1/2}+\int d\eta(\mu+\lambda^2\eta^2)^{1/2} \tag{4.3b}$$

with separation constant

$$\pm 2M_{01}P_1=(\xi^2-\eta^2)^{-1}(\xi^2p_\eta^2-\eta^2p_\xi^2)=\mu. \tag{4.3c}$$

The spacelike region $(x^0)^2-(x^1)^2<0$ can be parametrized by interchanging x^0 and x^1; however, the orbit representative then changes (see Ref. 22 for further discussion).

(9) *Hyperbolic—type* 3: The coordinates are

$$x^0=\cosh(\eta-\zeta)+\exp(\eta+\zeta),\quad x^1=\cosh(\eta-\zeta)-\exp(\eta+\zeta) \tag{4.4a}$$

with $-\infty<\eta,\zeta<\infty$. The solutions are

$$S=\int d\eta(\mu+2\alpha\lambda^2e^{2\eta})^{1/2}+\int d\zeta(\mu+2\alpha\lambda^2e^{2\zeta})^{1/2} \tag{4.4b}$$

and separation constant

$$M_{01}^2-(P_0+P_1)^2=(e^{2\eta}-e^{2\zeta})^{-1}(e^{2\eta}p_\xi^2-e^{2\zeta}p_\eta^2)=\mu. \tag{4.4c}$$

The coordinates (4.4a) parametrize the region $x^0+x^1>2$, $x^0-x^1>0$. In this case we can introduce similar parametrizations for all region except the strip $|x^0+x^1|<2$. For this strip there appear to be no separable coordinates of this type for (4.1b).

(10) *Hyperbolic—type* 2: The coordinates are

$$x^0=\sinh(\eta-\zeta)+\exp(\eta+\zeta)\quad x^1=\sinh(\eta-\zeta)-\exp(\eta+\zeta), \tag{4.5a}$$

where $-\infty<\eta,\zeta<\infty$. These coordinates parametrize the region $x^0-x^1>0$. By interchanging x^0 and x^1 we can parametrize the region $x^0-x^1<0$. The solutions are

$$S=\int d\eta(\mu+2\lambda^2e^{2\eta})^{1/2}+\int d\zeta(\mu-2\lambda^2e^{2\zeta})^{1/2} \tag{4.5b}$$

with constant of the motion

$$M_{01}^2+(P_0+P_1)^2=(e^{2\eta}+e^{2\zeta})^{-1}(e^{2\zeta}p_\eta^2+e^{2\eta}p_\xi^2)=\mu. \tag{4.5c}$$

(11) *Parabolic—type* 2: The coordinates are

$$x^0=\tfrac{1}{2}(\eta-\xi)^2+(\eta+\xi),\quad x^1=\tfrac{1}{2}(\eta-\xi)^2-(\eta+\xi) \tag{4.6a}$$

with $-\infty<\xi<\infty$ and $0\le\eta<\infty$. These coordinates cover the half-plane $x^0+x^1>0$. Similarly we can parametrize the remaining half-plane. The solutions are

$$S=\int d\eta(\mu+4\lambda\eta)^{1/2}+\int(\mu+4\lambda\xi)^{1/2}d\xi \tag{4.6b}$$

with separation constant

$$2M_{01}(P_0+P_1)+(P_0-P_1)^2=(\eta-\xi)^{-1}(\eta p_\xi^2-\xi p_\eta^2)=\mu. \tag{4.6c}$$

(12) *Hyperbolic—type* 1:

$$x^0=\tfrac{1}{2}[\cosh\tfrac{1}{2}(\eta-\zeta)+\sinh\tfrac{1}{2}(\eta+\zeta)],\qquad x^1=\tfrac{1}{2}[\cosh\tfrac{1}{2}(\eta-\zeta)-\sinh\tfrac{1}{2}(\eta+\zeta)], \tag{4.7a}$$

where $0\le\eta<\infty$, $-\infty<\zeta<\infty$ and (4.7a) parametrizes the half-plane $x^0+x^1\ge1$. By taking $(x^0,x^1)\to(x^0,\mp x^1)$ we can parametrize the remaining portion. The solutions are

$$S=\tfrac{1}{2}\int d\eta(\mu+\tfrac{1}{2}\lambda^2\sinh\eta)^{1/2}+\tfrac{1}{2}\int d\zeta(\mu+\tfrac{1}{2}\lambda^2\sinh\zeta)^{1/2} \tag{4.7b}$$

with separation constant

$$M_{01}^2-4P_0P_1=4(\sinh\eta-\sinh\zeta)^{-1}(\sinh\eta p_\xi^2-\sinh\zeta p_\eta^2)=\mu. \tag{4.7c}$$

(13) *Elliptic—type* 1: The coordinates are

$$x^0=\sinh\rho\cosh\sigma,\quad x^1=\cosh\rho\sinh\sigma, \tag{4.8a}$$

where the full plane is parametrized with $-\infty<\rho,\sigma<\infty$. The solutions are

$$S=\int d\rho(\mu+\lambda^2\cosh^2\rho)^{1/2}+\int d\sigma(\mu-\lambda^2\sinh^2\sigma)^{1/2} \tag{4.8b}$$

with separation constant

$$M_{01}^2+P_1^2=(\cosh^2\rho+\sinh^2\sigma)^{-1}(\sinh^2\sigma p_\rho^2+\cosh^2\rho p_\sigma^2)=\mu. \tag{4.8c}$$

(14) *Elliptic—type* 2: For the last case we have different coordinates for different regions of Minkowski space, viz.,

$$x^0=\cosh\rho\cosh\sigma,\quad x^1=\sinh\rho\sinh\sigma, \tag{4.9a}$$

with $-\infty<\rho<\infty$, $0\le\sigma<\infty$ for the region $x^0\ge1$. If we let $(x^0,x^1)\to(-x^0,x^1)$, we can treat the region $x^0\le-1$. However, for the region $-1\le x^0\le1$ we have

$$x^0=\cos\phi\cos\psi,\quad x^1=\sin\phi\sin\psi \tag{4.9b}$$

with $0\le\phi<2\pi$, $0\le\beta<\pi$. With (4.9a) and (4.9b) we still miss the region $-1<x^0<1$, $1<|x^1|$. This region can be handled by interchanging x^0 and x^1; hence, we can cover the full plane with elliptic—type 2 coordinates.

The solution corresponding to (4.9a) is

$$S=\int d\rho(\mu+\lambda^2\sinh^2\rho)^{1/2}+\int d\sigma(\mu+\lambda^2\sinh^2\sigma)^{1/2} \tag{4.9c}$$

while for (4.9b)

$$S=\int d\phi(\lambda^2\sin^2\phi-\mu)^{1/2}+\int d\psi(\lambda^2\sin^2\psi-\mu)^{1/2} \tag{4.9d}$$

with constants of the motion in the corresponding regions

$$M_{01}^2-P_1^2=(\cosh^2\rho-\cosh^2\sigma)^{-1}(\sinh^2\rho p_\sigma^2-\sinh^2\sigma p_\rho^2)=\mu, \tag{4.9e}$$

$$M_{01}^2-P_1^2=(\cos^2\psi-\cos^2\phi)^{-1}(\sin^2\psi p_\phi^2-\sin^2\phi p_\psi^2)=\mu.$$

5. THE HYPERBOLOIDS (DOUBLE AND SINGLE SHEETED)

In this case we consider the function D given in (1.9). Putting $D=\lambda$ and using (1.28), we see that (1.1) reduces to

$$M_{01}^2+M_{02}^2-M_{12}^2=\lambda^2. \tag{5.1}$$

Now the centralizer of D in o(3,2) is $\{D\}\oplus$ o(2,1), where o(2,1) is the subalgebra of o(3,2) with basis M_{01}, M_{02}, M_{12}; hence, o(2,1) is the reduced symmetry algebra for (5.1). Introducing the real variable $0<\rho=(x\cdot x)^{1/2}$, $x\cdot x>0$, for which $D=\rho P_\rho$, we obtain

$$x_0=\rho y_0,\quad x_1=\rho y_1,\quad x_2=\rho y_2,$$
$$y_0^2-y_1^2-y_2^2=1. \tag{5.2}$$

Thus we have separated off one variable in such a way that we are left with a double-sheeted hyperboloid. We will hereafter restrict ourselves to the upper sheet ($y_0>0$). Furthermore, for the region $x\cdot x<0$ with $\rho=(-x\cdot x)^{1/2}>0$, we obtain the single-sheeted hyperboloid, viz.,

$$x_0=\rho y_0,\quad x_1=\rho y_1,\quad x_2=\rho y_2,$$
$$y_0^2-y_1^2-y_2^2=-1. \tag{5.3}$$

We note that, on the light-cone $x\cdot x=0$, (5.1) reduces to an ordinary differential equation and the problem of separation does not exist.

Now using the coordinates (5.2) or (5.3) it is straightforward to show that $\mathbf{M}_{\mu\nu}=y_\mu\partial_\nu-y_\nu\partial_\mu$; thus by diagonalizing $\mathbf{D}$, i.e., $\mathbf{D}W=\lambda$ we have

$$W=\lambda\ln\rho+S(y^\mu) \tag{5.4}$$

and (1.1) reduces to (4.1) where the M's in (5.1) are interpreted now as $\mathbf{M}_{\mu\nu}S$, i.e., $M_{\mu\nu}\to\mathbf{M}_{\mu\nu}$.

It was shown by Winternitz, Lukač, and Smorodinskiĭ[23] that the space of second order members of the universal enveloping algebra of o(2,1) modulo its center splits into nine orbits under the adjoint action of O(2,1). Furthermore, the connection with the separation of variables for the Laplace–Beltrami operator was established. Here, we list the pairs of commuting functions which separate variables in (1.1) keeping with the notation used previously[3,23]

(15) D^2, M_{12}^2, spherical ,

(16) D^2, M_{02}^2, equidistant ,

(7′) D^2, $(M_{12}-M_{02})^2$, horocycle ,

(17) D^2, $M_{12}^2+a^2M_{01}^2$, elliptic, $0<a<1$,

(18) D^2, $M_{01}^2-a^2M_{12}^2$, hyperbolic, $0<a<1$,

(19) D^2, $2aM_{01}^2-M_{02}M_{12}$, semihyperbolic, $0<a<\infty$,

(20) D^2, $(M_{01}+M_{12})^2+aM_{02}^2$, elliptic–parabolic, $0<a<\infty$,

(21) D^2, $(M_{01}+M_{12})^2-aM_{02}^2$, hyperbolic–parabolic, $0<a<\infty$,

(22) D^2, $\frac{1}{4}M_{02}(M_{01}-M_{12})$, semicircular–parabolic.

We now give the solutions for the above orbit representatives on the upper sheet of the double-sheeted hyperboloid. The coordinates on the single-sheeted hyperboloid can be obtained by $y^\mu\to iy^\mu$ with the appropriate change of parametrization; however, different parametrizations are sometimes needed for different regions. There is also no guarantee that Eq. (5.1) is separable in all regions (see Ref. 24). Each of the separated solutions is found by solving (4.1) (i.e., $D^2=\lambda$) and taking one of the above orbit representatives as constant of the motion.

(15) *Spherical*: The coordinates are

$$y^0=\cosh\eta,\ y^1=\sinh\eta\cos\phi,\quad y^2=\sinh\eta\sin\phi, \tag{5.5a}$$

with $0<\eta<\infty$ and $0\le\phi<2\pi$ and the separated solution is

$$S=\mu\phi+\int(\lambda^2-\mu^2/\sinh^2\eta)^{1/2}d\eta \tag{5.5b}$$

with separation constant

$$M_{12}^2=p_\phi^2=\mu^2. \tag{5.5c}$$

(16) *Equidistant*: The coordinates are

$$y^0=\cosh\rho\cosh\sigma,\quad y^1=\sinh\rho,\quad y^2=\cosh\rho\sinh\sigma \tag{5.6a}$$

with $-\infty<\rho,\sigma<\infty$, and separated solution

$$S=\mu\sigma+\int(\alpha^2-\mu^2/\cosh^2\rho)^{1/2}d\rho \tag{5.6b}$$

with separation constant

$$M_{02}^2=p_\sigma^2=\mu^2. \tag{5.6c}$$

(7′) is the same as (7).

(17) *Elliptic*: The coordinates are

$$y^0=(k')^{-1}\,\mathrm{dn}u\,\mathrm{dn}v,\quad y^1=k(k')^{-1}\,\mathrm{cn}u\,\mathrm{cn}v,$$
$$y^2=-ik\,\mathrm{sn}u\,\mathrm{sn}v \tag{5.7a}$$

with $u\in(0,4K)$ and $v\in(0,iK')$. The separated solution of (4.1) is

$$S=\int du(\mu-\lambda^2k^2\,\mathrm{cn}^2u)^{1/2}+\int dv(\mu-\lambda^2k^2\,\mathrm{cn}^2v)^{1/2} \tag{5.7b}$$

and separation constant is

$$k'^2M_{12}^2+k^2M_{01}^2=(\mathrm{cn}^2u-\mathrm{cn}^2v)^{-1}(\mathrm{cn}^2up_v^2-\mathrm{cn}^2vp_u^2)=\mu. \tag{5.7c}$$

(18) *Hyperbolic*: The coordinates are

$$y^0=ik(k')^{-1}\,\mathrm{cn}u\,\mathrm{cn}v,\quad y^1=ik\,\mathrm{sn}u\,\mathrm{sn}v,$$
$$y^2=i(k')^{-1}\,\mathrm{dn}u\,\mathrm{dn}v \tag{5.8a}$$

with $u \in (iK', iK+2K)$ and $v \in (-iK', iK)$. The separated solutions are

$$S=(1/k^2)\int du(\lambda^2k^2\,\mathrm{sn}^2u-\mu)^{1/2} + (1/k^2)\int dv(\lambda^2k^2\,\mathrm{sn}^2v-\mu)^{1/2} \tag{5.8b}$$

with constant of the motion

$$M_{01}^2-k^2M_{12}^2=k^4(\mathrm{sn}^2u-\mathrm{sn}^2v)^{-1}(\mathrm{sn}^2vp_u^2-\mathrm{sn}^2up_v^2)=\mu. \tag{5.8c}$$

(19) *Semihyperbolic*: The coordinates are

$$y_1^2=\frac{-\nu\eta}{\alpha^2+\beta^2},$$

$$y_2^2=\frac{\nu\eta}{2(\alpha^2+\beta^2)}-\frac{1}{2}+\frac{1}{2\beta}\left[\frac{((\nu-\alpha)^2+\beta^2)((\eta-\alpha)^2+\beta^2)}{\alpha^2+\beta^2}\right]^{1/2}, \tag{5.9a}$$

$$y_0^2=\frac{-\nu\eta}{2(\alpha^2+\beta^2)}+\frac{1}{2}+\frac{1}{2\beta}\left[\frac{((\nu-\alpha)^2+\beta^2)((\eta-\alpha)^2+\beta^2)}{\alpha^2+\beta^2}\right]^{1/2}$$

with $0\le\eta<\infty$, $-\infty<\nu<0$, and α,β real. The separated solutions are

$$S=\int d\nu(\mu+\tfrac{1}{4}\lambda^2\nu)^{1/2}\nu^{-1/2}[(\nu-\alpha)^2+\beta^2]^{-1/2} + \int d\eta(\mu+\tfrac{1}{4}\lambda^2\eta)^{1/2}\eta^{-1/2}[(\eta-\alpha)^2+\beta^2]^{1/2} \tag{5.9b}$$

with separation constant

$$\alpha M_{01}^2-2\beta M_{12}M_{02} =\eta\nu(\nu-\eta)^{-1}[((\eta-\alpha)^2+\beta^2)p_\eta^2-((\nu-\alpha)^2+\beta^2)p_\nu^2]=\mu. \tag{5.9c}$$

(20) *Elliptic–parabolic*: For simplicity we consider the nondegenerate point $a=1$; the coordinates are

$$y^0=\frac{1}{2}\left(\frac{\cosh^2\rho+\cos^2\theta}{\cosh\rho\cos\theta}\right),\quad y^2=\frac{1}{2}\left(\frac{\sin^2\theta-\cosh^2\rho}{\cosh\rho\cos\theta}\right),$$

$$y^1=-\tanh\rho\tan\theta \tag{5.10a}$$

with $-\infty<\rho<\infty$ and $-\pi/2<\theta\le\pi/2$. The solutions are

$$S=\int d\rho(\mu+\lambda^2/\cosh^2\rho)^{1/2}+\int d\theta(-\mu-\lambda^2/\cos^2\theta)^{1/2} \tag{5.10b}$$

with constant of the motion

$$M_{02}^2+(M_{01}+M_{12})^2 =(\cosh^2\rho-\cos^2\theta)^{-1}(\cosh^2\rho p_\rho^2+\cos^2\theta p_\theta^2)=\mu. \tag{5.10c}$$

(21) *Hyperbolic–parabolic*: Again for $a=1$ the coordinates are

$$y^0=\frac{1}{2}\left(\frac{\cosh^2\rho+\cos^2\theta}{\sinh\rho\sin\theta}\right),\quad y^2=\frac{1}{2}\left(\frac{\sin^2\theta-\sinh^2\rho}{\sinh\rho\sin\theta}\right),$$

$$y^1=\coth\rho\cot\theta \tag{5.11a}$$

with $0\le\rho<\infty$ and $0\le\theta<\pi$. The separated solutions are

$$S=\int d\rho(\mu-\lambda^2/\sinh^2\rho)^{1/2}+\int d\theta(-\mu-\lambda^2/\sin^2\theta)^{1/2} \tag{5.11b}$$

with separation constant

$$-(M_{01}+M_{12})^2+M_{02}^2 =(\sinh^2\rho+\sin^2\theta)^{-1}(\sinh^2\rho p_\rho^2-\sin^2\theta p_\theta^2)=\mu. \tag{5.11c}$$

(22) *Semicircular–parabolic*: The coordinates are

$$y^0=\frac{(\xi^2+\eta^2)^2+4}{8\xi\eta},\quad y^1=\frac{\eta^2-\xi^2}{2\xi\eta},\quad y^2=\frac{(\xi^2+\eta^2)^2-4}{8\xi\eta} \tag{5.12a}$$

with $0\le\xi,\eta<\infty$. The solutions are

$$S=\int d\eta(\mu+\lambda^2/\eta^2)^{1/2}+\int d\xi(\lambda^2/\xi^2-\mu)^{1/2} \tag{5.12b}$$

with the constant of the motion

$$\tfrac{1}{4}M_{02}(M_{01}-M_{12})=(\xi^2+\eta^2)^{-1}(\eta^2p_\eta^2-\xi^2p_\xi^2)=\mu. \tag{5.12c}$$

As with the sphere in Sec. 2 there is a close relationship between our model on the hyperboloid and the Hamilton–Jacobi equation for the Kepler problem with unbounded orbits (positive energies). Indeed the operator $\mathbf{D}=\Gamma_{15}$ is conformally equivalent to $\Gamma_{14}=\frac{1}{2}(\mathbf{P}_0-\mathbf{K}_0)$. Explicitly, $\mathrm{Ad}\exp(\frac{1}{2}\pi\Gamma_{45})\Gamma_{15}=\Gamma_{14}$. Thus $\Gamma_{14}=\lambda$, and on the surface (1.21) we have

$$x_1^2+x_2^2+2\lambda/p_0=1,\quad p_0=(p_1^2+p_2^2)^{1/2}. \tag{5.13}$$

Again implementing the canonical transformation $p_j\to x^j$, $x^j\to -p_j$, we obtain

$$p_1^2+p_2^2+2\lambda/r=1\quad r=(x_1^2+x_2^2)^{1/2}, \tag{5.14}$$

i.e., the positive energy Kepler problem. It is clear that under the above canonical transformation the reduced symmetry algebra o(2,1) is preserved.

Again, we emphasize that Eq. (1.1) can also be interpreted as the equation for the graph of a solution $S(\eta,\phi)$ of (5.1), here parametrized by the spherical coordinates (5.5a). Thus o(3,2) is the full symmetry group of (5.1) and every separated solution of (1.1) gives rise via the graph to a symmetry adapted solution of (5.1).

6. THE NONRELATIVISTIC FREE PARTICLE

We now look at the only partial separation of (1.1) which involves nonorthogonal coordinates. Since this case was already treated in detail in Ref. 6, we will be brief here. Considering the reduced equation corresponding to the operator $\mathbf{P}_0+\mathbf{P}_1$, we set $P_0+P_1=\lambda$; then (1.1) reduces to

$$\lambda S_t-S_y^2=0, \tag{6.1a}$$

where

$$t=(x^0-x^1),\quad y=x^2 \tag{6.1b}$$

and

$$W=\lambda(x^0+x^1)+S(t,y). \tag{6.1c}$$

Clearly (6.1) is equivalent to the equation studied in Ref. 6 (take $t\to-\lambda^{-1}t$ and $y\to x$). Its reduced symmetry algebra is the Schrödinger algebra s_1 generated by $\{P_0-P_1,\ K_0+K_1,\ D+M_{01},\ P_2,\ M_{02}-M_{12},\ P_0+P_1\}$. Notice that in this case we no longer have a Lie algebra direct sum of the operator corresponding to the partial separation (here $\epsilon=P_0+P_1$) and its centralizer. However, since ϵ is in the center of s_1, we can consider the factor algebra s_1/ϵ. Because the partial separation in this case involves nonorthogonal coordinates, the R-separable coordinates of the reduced equation (6.1a) are nonorthogonal and are characterized by orbits in the factor

algebra s_1/ϵ. The list of representatives is:

(3″) P_0+P_1, P_2, free particle,

(23) P_0+P_1, $P_0-P_1-\frac{1}{4}(K_0+K_1)$, attractive oscillator,

(24) P_0+P_1, $P_0-P_1\pm(M_{12}-M_{02})$, free fall (linear potential),

(25) P_0+P_1, $D+M_{01}$, repulsive oscillator.

In Ref. 6 orbits in s_1 were classified by equivalence under the full conformal group. It is easy to see from those results that the orbits in s_1/ϵ under the conformal group are precisely those listed above. Moreover, because all of the above are members of the Lie algebra, one can construct constants of the motion via functions on phase space as done here or equivalently construct relative invariants of vector fields as done in Ref. 6. One can easily check by using the Lie algebra isomorphism (1.7) that the two methods are indeed equivalent, keeping in mind that orbits of relative invariants in s_1 considered as vector fields correspond to orbits in s_1/ϵ considered as functions to be set equal to constants.

Again as shown in Refs. 2 and 6, o(3,2) is the full symmetry algebra of (6.1a) corresponding to the fact that Eq. (1.1) can be interpreted as the equation for the graph of a solution of (6.1a). Thus all separated solutions of (1.1) give rise to symmetry adapted solutions of (6.1a). Indeed all those corresponding to first order operators have been given, up to equivalence, in Ref. 6.

7. A NONLINEAR EPD EQUATION

Now we look for coordinate systems yielding separation of variables in (1.1) such that $A=\Gamma_{32}=m$, m constant. Setting $x^0=t$, $x=r\cos\phi$, $x^2=r\sin\phi$, we have $\Gamma_{32}=-p_\phi$ so

$$W=-m\phi+S(t,r), \tag{7.1a}$$

where

$$S_t^2-S_r^2-m^2/r^2=0 \tag{7.1b}$$

or, from (1.29),

$$\Gamma_{45}^2-\Gamma_{14}^2-\Gamma_{15}^2=\Gamma_{32}^2=m^2. \tag{7.2}$$

Since the centralizer of Γ_{32} in o(3,2) is $\{\Gamma_{32}\}\oplus$ o(2,1), where o(2,1) is the subalgebra with basis Γ_{45}, Γ_{14}, Γ_{15}, we see that o(2,1) is a symmetry algebra for the reduced equation (6.1). Here,

$$\begin{aligned}\Gamma_{45}&=\tfrac{1}{2}(1+t^2+r^2)p_t+trp_r,\\ \Gamma_{14}&=\tfrac{1}{2}(1-t^2-r^2)p_t-trp_r,\\ \Gamma_{15}&=-tp_t-rp_r,\quad p_t=S_t,\ p_r=S_r.\end{aligned} \tag{7.3}$$

It is well known[23] that the space of second order symmetry operators in o(2,1), modulo the invariant $\Gamma_{45}^2-\Gamma_{14}^2-\Gamma_{15}^2$, splits into exactly nine orbits types under the adjoint action of o(2,1). A representative of each orbit type is given by the assignment

(1′) Γ_{32}^2, Γ_{45}^2 ,

(4′) Γ_{32}^2, $(\Gamma_{45}+\Gamma_{14})^2$,

(15′) Γ_{32}^2, Γ_{15}^2 ,

(26) Γ_{32}^2, $\Gamma_{14}^2+\Gamma_{14}\Gamma_{45}-a\Gamma_{15}^2$, $a>-\frac{1}{2}$,

(27) Γ_{32}^2, $\Gamma_{45}^2+\Gamma_{14}\Gamma_{45}+a\Gamma_{15}^2$, $a>-\frac{1}{2}$,

(28) Γ_{32}^2, $a\Gamma_{15}^2+\Gamma_{14}\Gamma_{45}$,

(29) Γ_{32}^2, $\Gamma_{15}^2+a^2\Gamma_{45}^2$, $0<a<1$,

(30) Γ_{32}^2, $\Gamma_{15}^2-a^2\Gamma_{45}^2$, $0<a<1$,

(31) Γ_{32}^2, $(\Gamma_{14}+\Gamma_{45})\Gamma_{15}$.

We shall show explicitly that (1.1), hence (7.1), admits an additive separation of variables corresponding to each of these orbits. The separable coordinate systems are exactly those studied in Ref. 8.

Orbits (1′), (4′), and (15′) have been treated above.

(26) We consider for simplicity the nondegenerate point $a=0$ and

$$t=\cos\theta\cosh\alpha,\quad r=\sin\theta\sinh\alpha. \tag{7.4a}$$

Then (7.1) yields the separated solutions

$$\begin{aligned}S(\theta,\alpha)&=\int(2\mu^2-m^2\cot^2\theta)^{1/2}d\theta\\&\quad+\int(2\mu^2-m^2\coth^2\alpha)^{1/2}d\alpha\end{aligned} \tag{7.4b}$$

with constant of the motion

$$\begin{aligned}&\Gamma_{14}^2+\Gamma_{14}\Gamma_{45}\\&=\tfrac{1}{2}(\sin^2\alpha-\sin^2\theta)^{-1}(\cos^2\theta\sin^2\alpha p_\alpha^2-\sin^2\theta\cos^2\alpha p_\theta^2)\\&=-\mu^2.\end{aligned} \tag{7.4c}$$

The coordinates θ,α are valid only for $|t|+r\leq 1$. As shown in Refs. 4 and 8 there are similar separable parametrizations for $|t|\geq r+1$ and $|t|\leq r-1$, but not all regions of the $r-t$ plane with $r>0$ are covered with parametrizations which permit separation of variables.

(27) With $a=0$ the separable coordinates are

$$t=\cosh\theta\sinh\alpha,\quad r=\sinh\theta\cosh\alpha \tag{7.5a}$$

and the solutions have a form similar to (7.4b).

Solutions corresponding to orbits (28)–(30) are rather similar. For (28) the separable coordinates are

$$\begin{aligned}&t=2(kk')^{1/2}(k+ik')\,\mathrm{sn}(\theta,l)\,\mathrm{sn}(\alpha,l)/R,\\&r=2(kk')^{1/2}/R,\\&R=(k-ik')\,\mathrm{dn}(\theta,l)\,\mathrm{dn}(\alpha,l)+(k+ik')\,\mathrm{cn}(\theta,l)\,\mathrm{cn}(\alpha,l),\\&4a=k'/k-k/k',\quad k'=(1-k^2)^{1/2},\\&l=(k+ik')/(k-ik'),\end{aligned} \tag{7.6}$$

for (29) the coordinates are

$$\begin{aligned}&t=\mathrm{dn}(\theta,a)\,\mathrm{dn}(\alpha,a)/a'R,\quad r=1/R,\\&R=a\,\mathrm{sn}(\theta,a)\,\mathrm{sn}(\alpha,a)+a\,\mathrm{cn}(\theta,a)\,\mathrm{cn}(\alpha,a)/a',\\&a'=(1-a^2)^{1/2},\end{aligned} \tag{7.7}$$

and for (30) they are

$$\begin{aligned}&t=k\,\mathrm{sn}(\theta,k)\,\mathrm{sn}(\alpha,k)/R,\quad r=1/R,\\&R=(k')^{-1}\,\mathrm{dn}(\theta,k)\,\mathrm{dn}(\alpha,k)+(k/k')\,\mathrm{cn}(\theta,k)\,\mathrm{cn}(\alpha,k),\\&a=k',\quad k'=(1-k^2)^{1/2}.\end{aligned} \tag{7.8}$$

(For a discussion of the ranges of the variables θ,α, see Ref. 8.)

As an example of the form of the solutions we insert coordinates (7.6) in (7.1) to obtain

$$p_\theta^2 - p_\alpha^2 = m^2k^2(\mathrm{sn}^2\alpha - \mathrm{sn}^2\theta) \tag{7.9a}$$

with separated solutions

$$S = \int (\mu^2 + m^2\,\mathrm{dn}^2\theta)^{1/2}\,d\theta + \int (\mu^2 + m^2k^2\,\mathrm{dn}^2\alpha)^{1/2}\,d\alpha. \tag{7.9b}$$

Here, $\Gamma_{15}^2 - (k')^2\Gamma_{45}^2 = \mu^2$. (Some errors in the corresponding list of elliptic coordinates for the EPD equation, contained in Ref. 8, have been corrected here.)

(31) For this orbit we set

$$t = \pm(\theta^2 + \alpha^2), \quad r = 2\theta\alpha, \quad t \geq |r|, \tag{7.10a}$$

in which case (7.1) becomes

$$p_\theta^2 - p_\alpha^2 - m^2(1/\alpha^2 - 1/\theta^2) = 0. \tag{7.10b}$$

The condition

$$-(\Gamma_{14} + \Gamma_{45})\Gamma_{15} = \tfrac{1}{4}(\theta^2 - \alpha^2)^{-1}(\theta^2p_\theta^2 - \alpha^2p_\alpha^2) = \mu^2 \tag{7.10c}$$

yields the separated solution

$$S(\theta, \alpha) = \int (4\mu^2 - m^2/\theta^2)^{1/2}\,d\theta + \int (4\mu^2 - m^2/\alpha^2)^{1/2}\,d\alpha. \tag{7.10d}$$

(Contrary to the statement in Ref. 8, variables do not separate for $r \geq |t|$.)

In analogy with the previous reduced equations it is easy to show that (1.1) is the equation of the graph of (7.1). Thus, o(3,2) is the full symmetry algebra of (7.1).

8. THE SYMMETRY $\Gamma_{23} - \Gamma_{45}$

We next separate a variable from (1.1) by requiring $L = \frac{1}{2}(\Gamma_{23} - \Gamma_{45}) = K$. In terms of the coordinates (2.2), (2.4) with $\beta = \alpha + \psi$, $\phi = \alpha - \psi$ we have $L = p_\beta$, so

$$W = K\beta + S(\phi, \sigma),$$

where

$$\cot^2\sigma S_\phi^2 + 2K(\csc^2\sigma + 1)S_\phi + S_\sigma^2 + K^2\cot^2\sigma = 0. \tag{8.1}$$

The centralizer of L in o(3,2) is $\{L\} \oplus$ o(2,1), where o(2,1) is the subalgebra with basis A, B, C such that

$$A = \tfrac{1}{2}(\Gamma_{23} + \Gamma_{45}), \quad B = \tfrac{1}{2}(\Gamma_{24} + \Gamma_{35}), \quad C = \tfrac{1}{2}(\Gamma_{25} - \Gamma_{34}), \tag{8.2}$$

$$[A, B] = C, \quad [C, A] = B, \quad [C, B] = A.$$

Thus o(2,1) is a symmetry algebra for the reduced equation (8.1). Here

$$A = S_\phi, \quad -B = \sin\phi S_z + \coth z\cos\phi S_\phi + K\cos\phi/\sinh z$$
$$C = -\cos\phi S_z + \coth z\sin\phi S_\phi + K\sin\phi/\sinh z, \tag{8.3}$$
$$\sin\sigma = \tanh(z/2),$$

and in terms of these symmetries equation (8.1) reads

$$A^2 - B^2 - C^2 = 0. \tag{8.4}$$

Note: The simple computation leading to this identity shows that the corresponding identity for the wave equation as given in Refs. 3 and 5 is in error. The correct result for the wave equation is

$$\mathbf{A}^2 - \mathbf{B}^2 - \mathbf{C}^2 = \tfrac{3}{16}. \tag{8.5}$$

Indeed, the eigenvalues of $-i\mathbf{L}$ are $\lambda_s = \frac{1}{2}(s + \frac{1}{2})$, $s = 0, 1, 2, \cdots$. The eigenspace V_s corresponding to eigenvalue λ_s is irreducible under o(2,1) and transforms according to the unitary representation D_l^+ where $l = -[(-1)^s + 2]/4$.

As usual we try to associate separable coordinates for (8.4) with the nine orbits of second order symmetries in the enveloping algebra for o(2,1). It is guaranteed that there are separable coordinates corresponding to the three orbits which correspond to squares of first order symmetries:

(1′) L^2, A^2,

(32) L^2, C^2,

(33) L^2, $(A - B)^2$.

In particular, (1′) is equivalent to (1). To obtain the remaining systems, we note that for $K = 0$ the operators and coordinates (8.3) agree with system (15) on the hyperboloid, i.e., coordinates (5.5a). Since the separable coordinates for (8.4) must be independent of K it follows that a separable system for (8.4) must be one of the systems (7′), (15)–(22). However, one of the latter systems need not necessarily yield separation for (8.4).

We are guaranteed success for systems (32) and (33). For (32) we set $\cosh z = \cosh\xi\cosh\eta$, $\tan\phi = \tanh\xi/\sinh\eta$ to obtain

$$\frac{1}{\cosh^2\xi}S_\eta^2 + S_\xi^2 + \frac{1}{(1 - \cosh^2\xi\cosh^2\eta)} \times (2K\cosh\xi\sinh\eta\cosh\eta S_\xi - 2K\sinh\xi\cosh^2\eta S_\eta + K^2) = 0. \tag{8.6a}$$

The condition

$$C = -S_\eta + K\sinh\xi/(\cosh^2\xi\cosh^2\eta - 1) = \mu \tag{8.6b}$$

yields the R-separated solution

$$S = -K\tan^{-1}(\sinh\xi\coth\eta) - \mu\eta + \int [(K^2 - \mu^2 - 2\mu K\sinh\xi)^{1/2}/\cosh\xi]\,d\xi. \tag{8.6c}$$

For (33) we set $\cosh z = \frac{1}{2}[e^{-\xi} + (\eta^2 + 1)e^{\xi}]$, $\tan\phi = -2\eta e^{\xi}/[e^{-\xi} + (\eta^2 - 1)e^{\xi}]$ and use the condition

$$A - B = S_\eta - 2K[e^{-\xi} + (\eta^2 - 1)e^{\xi}]/\{[e^{-\xi} + (\eta^2 + 1)e^{\xi}]^2 - 4\} = \mu \tag{8.7a}$$

to obtain the R-separated solution

$$S = -K\tan^{-1}\left[\frac{2\eta}{e^{\xi}(1 + \eta^2) - e^{-\xi}}\right] + m\eta + \int [-(2Ke^{-\xi}m + m^2e^{-2\xi})]^{1/2}\,d\xi. \tag{8.7b}$$

We have carefully studied the coordinates corresponding to system (17) and have found that they do not lead to R separation of variables for (8.4). It appears that only the subgroups systems (15′), (32) and (33) yield additive variable separation for this equation, although we have not explicitly checked this for all systems (15)–(22).

Just as for the other reduced equations it is easy to show that (1.1) is the equation of the graph of (8.4). Thus o(3,2) is the full symmetry algebra of (8.4) and all of the additively separated solutions of (1.1) lead to symmetry adapted solutions of (8.4).

9. NONSPLIT COORDINATES

In analogy with Ref. 3 for the wave equation, we have listed, with the exception of some degenerate nonorthogonal systems,[5] all separable systems for (1.1) in which it is possible to additively split off one variable. In Ref. 4 a classification of all orthogonal R-separable coordinates systems for the wave equation was given for which the coordinate surfaces were families of confocal cyclides. 53 such systems were found, and, except for degenerate cases, it was shown that the variables intertwine in such a complicated fashion that it is necessary to separate them simultaneously, i.e., it is not possible to split off a single variable. Each such system was shown to be characterized by a commuting pair of second order symmetric operators in the enveloping algebra of o(3,2).

The results of Ref. 4 can be applied directly to obtain orthogonal separable coordinates for (1.1) simply by interpreting the Lie algebra of differential operators o(3,2) as a Lie algebra of functions under the Poisson bracket.

For example, the system [311] (i) of Ref. 4 leads to coordinates

$$x^0 = -\tfrac{1}{2}(\cos^2\alpha + \cos^2\beta + \cos^2\gamma), \quad x^1 = \sin\alpha \sin\beta \sin\gamma,$$
$$x^2 = \cos\alpha \cos\beta \cos\gamma. \tag{9.1}$$

In these coordinates (1.1) reduces to

$$(\sin^2\beta - \sin^2\gamma)p_\alpha^2 + (\sin^2\gamma - \sin^2\alpha)p_\beta^2 + (\sin^2\alpha - \sin^2\beta)p_\gamma^2 = 0. \tag{9.2}$$

It is not possible to additively separate one of these variables from the other two. However, use of the defining symmetry elements

$$2(P_0 - P_1)M_{02} + P_0^2 + P_1^2 = \frac{\sin^2\alpha p_\alpha^2}{(\sin^2\alpha - \sin^2\beta)(\sin^2\alpha - \sin^2\gamma)} + \frac{\sin^2\beta p_\beta^2}{(\sin^2\beta - \sin^2\alpha)(\sin^2\beta - \sin^2\gamma)} + \frac{\sin^2\gamma p_\gamma^2}{(\sin^2\gamma - \sin^2\alpha)(\sin^2\gamma - \sin^2\beta)} = \mu \tag{9.3a}$$

$$2P_2M_{02} - M_{12}^2 + P_2^2 = \frac{\sin^2\beta \sin^2\alpha p_\alpha^2}{(\sin^2\alpha - \sin^2\beta)(\sin^2\alpha - \sin^2\gamma)} + \frac{\sin^2\alpha \sin^2\gamma p_\beta^2}{(\sin^2\beta - \sin^2\alpha)(\sin^2\beta - \sin^2\gamma)} + \frac{\sin^2\alpha \sin^2\beta p_\gamma^2}{(\sin^2\gamma - \sin^2\alpha)(\sin^2\gamma - \sin^2\beta)} = \nu \tag{9.3b}$$

leads to the separated solution

$$W = \int (\mu \sin^2\alpha + \nu)^{1/2} d\alpha + \int (\mu \sin^2\beta + \nu)^{1/2} d\beta + \int (\mu \sin^2\gamma + \nu)^{1/2} d\gamma. \tag{9.4}$$

In a similar fashion each of the orthogonal R-separable coordinate systems for the wave equation is additively separable for (1.1). In Paper II we shall examine the relationship between the wave equation and (1.1) more closely and provide proofs of own assertions concerning variable separation.

ACKNOWLEDGMENT

One of us (W.M., Jr.) would like to thank the Institute de Investigaciones en Matemáticas Aplicadas y en Sistemas for financial support and their hospitality during the period of this work.

[1] J.E. Campbell, *Introductory Treatise on Lie's Theory of Finite Continuous Transformation Groups* (Chelsea, New York, 1966), reprint of 1903 edition, p. 32.
[2] C.P. Boyer and M. Peñafiel N., Nuovo Cimento B **31**, 195 (1976).
[3] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **16**, 2507 (1975).
[4] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **17**, 331 (1976).
[5] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **17**, 356 (1976).
[6] C.P. Boyer and E.G. Kalnins, J. Math. Phys. **18**, 1032 (1977).
[7] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **15**, 1728 (1974).
[8] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **17**, 369 (1976).
[9] P. Havas, J. Math. Phys. **16**, 1461, 2476 (1975).
[10] W. Dietz, J. Phys. A **9**, 519 (1976); N.M.J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975).
[11] B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[12] J. Liouville, J. de Math. **11**, 345 (1846); P. Stackel, Habilitationschrift Halle, (1891); Math. Ann. 42, 548 (1893); **35**, 91 (1890); Ann. Mat. **25**, 55 (1897); T. Levi-Civita, Math. Ann. **59**, 383 (1904); F. Dall'Acqua, Math. Ann. **66**, 398 (1908); Rend. Circ. Matem. Palermo 33, 341 (1912); P. Burgatti, Rend. Acad. Lincei 20 (1), 108 (1911); H.P. Robertson, Math. Ann. **98**, 749 (1927); L.P. Eisenhart, Ann. Math. **35**, 284 (1934); M.S. Iarov-Iarovoi, J. Appl. Math. Mech. **27**, 1499 (1964).
[13] Y. Hagihara, *Celestial Mechanics*, *Vol.* 1 (MIT Press, Cambridge, Mass., 1970).
[14] N. Levison, B. Bogart, and R.M. Redheffer, J. Appl. Math. **7**, 241 (1949); P. Moon and D.E. Spencer, J. Franklin Inst. **255**, 302 (1953).
[15] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[16] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, *Vol. II* (Interscience, New York, 1962).
[17] J. Patera and P. Winternitz, J. Math. Phys. **14**, 1130 (1973).
[18] A. Erdelyi *et al.*, *Higher Transcendental Functions*, *Vol.* 2 (McGraw-Hill, New York, 1951).
[19] V. Fock, Z. Phys. **98**, 145 (1935).
[20] H. Bacry, Nuovo Cimento A **41**, 222 (1966); G. Gyorgyi, Nuovo Cimento A **53**, 717 (1968).
[21] P. Winternitz and I. Fris, Sov. J. Nucl. Phys. **1**, 636 (1965).
[22] E.G. Kalnins, SIAM J. Math. Anal. **6**, 340 (1975).
[23] P. Winternitz, I. Lukač, and Y.A. Smorodinskiĭ, Sov. J. Nucl. Phys. **7**, 139 (1968).
[24] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **18**, 1 (1977).

Symmetries in gauge theories[a)]

Peter G. Bergmann and Edward J. Flaherty, Jr.

Department of Physics, Syracuse University, Syracuse, New York 13210
(Received 26 April 1977)

A general definition of symmetry for solutions of the field equations of gauge theories is proposed, and some of its properties and consequences are discussed.

Gauge theories have enjoyed a surge of popularity in recent years, chiefly in the context of elementary particle physics. Gauge theories of the Yang–Mills type play a central role in unified models of the weak and electromagnetic interactions and in theories of the strong interactions. There has also been interest generated by the realization that general relativity is a gauge theory, in a well-defined sense, and this fact has provided a point of departure for several alternative theories of gravity. When the gauge group of a theory is non-Abelian, the field equations are necessarily nonlinear, and exact solutions are notoriously hard to come by. Experience with the field equations of general relativity has shown that searching for solutions with specific symmetries (e.g., isometries) often simplifies the field equations to the extent that exact solutions can be found, such as the class of all static, axially symmetric spacetimes. The corresponding concepts of static, and of spherically symmetric SU(2) Yang–Mills fields have also been formulated.[1,2] [Throughout this paper the term "symmetry" refers to properties of solutions (e.g., isometry), not of the field equations.] However, a general definition of symmetry or isometry for gauge theories is so far lacking in the literature; in this note we shall provide what we believe to be a reasonable and useful definition of symmetry for such theories. For special cases of this definition, see Refs. 1 and 2.

In order to introduce our notation and conventions, we shall present a very brief sketch, or review, of gauge theories. By a gauge theory we mean a relativistic field theory in which at least one of the physical fields is represented by the curvature matrix of a connection on a vector bundle over a (flat or curved) spacetime M. The "gauge group" of the theory is then the general linear group, or one of its subgroups, acting on the fibres of the bundle. A (local) "gauge transformation" corresponds to a change of frame field in the vector bundle over some open subset N of the base space M, together with the associated transformation of the connection matrix or "gauge potential." Let $e=(e_1\cdots e_n)$ be a frame field over N, i.e., if the dimension of the fibres is n, e consists of n linearly independent local cross sections of the bundle over N. Relative to this frame field, the connection is represented by an $n\times n$ matrix of 1-forms B. This defines a covariant derivative operator for cross sections of the bundle: If $V=e_i v^i$ is represented by the column vector $v=(v^1\cdots v^n)^{\mathrm{tr}}$, then we have

$$Dv=dv+Bv. \tag{1}$$

Under a change of frame field given by $e'=eU^{-1}$, we have

$$v'=Uv, \quad B'=UBU^{-1}-dUU^{-1}=(UB-dU)U^{-1}. \tag{2}$$

The physical field is represented by the curvature matrix F associated with B; with respect to the frame field e, it is given by an $n\times n$ matrix of 2-forms:

$$F=dB+B\wedge B, \tag{3}$$

where d denotes the exterior derivative of the entries (elements) of the matrix B. Under the change of frame $e'=eU^{-1}$, it is easy to show that

$$F'=UFU^{-1}. \tag{4}$$

Thus, the components of F transform as a "tensor" under U, whereas those of B do not.

The simplest gauge theory is electromagnetism, in which the fibres are one-dimensional complex vector spaces, and the gauge group is taken as the unitary group $U(1)\subset \mathrm{GL}(1,C)$ (or as the additive group R). In this case the connection matrix is simply a 1-form $B=A$, the ordinary vector potential of electromagnetism (but only locally defined for a non-trivial line bundle); and the curvature matrix is simply the (globally defined) 2-form $F=dA$, representing the electromagnetic field. The Lagrangian density for Maxwell–Lorentz theory is given by $(-g)^{1/2}F_{mn}F_{rs}g^{mr}g^{ns}$, where g_{mn} is the spacetime metric. Yang–Mills theories[3] are a straightforward generalization of electromagnetism, in the following sense: The fibres are allowed to be vector spaces (real or complex) of any dimension, the group of the bundle being any semisimple Lie group which acts linearly on the fibres; the Lagrangian density is usually chosen to be $\mathrm{tr}(-g)^{1/2}F_{mn}F_{rs}g^{mr}g^{ns}$. It is also ordinarily required that the physical fields F be restricted to being real in an appropriate sense: In the complex case with SU(n) as gauge group, this is accomplished by considering only frame fields which are unitary with respect to some Hermitian inner product on the bundle, and by requiring the connection matrix B to be anti-Hermitian with respect to each such frame; it then follows that the matrix F is also anti-Hermitian for such frames.

It can be argued that general relativity is the "most natural" gauge theory,[4] insofar as the vector bundle involved is the "most natural" bundle associated with a manifold, the tangent bundle; the connection is the Riemannian connection, also in some sense the "most natural" choice, and the group is the Lorentz group SO(3,1) with only orthonormal frame fields with respect to the Riemannian metric being considered. However, standard general relativity is not a Yang–Mills theory,

[a)]Supported in part by NSF Grant #MPS74-15246.

 0022-2488/78/1901-0212$1.00

because the Lagrangian density $(-g)^{1/2}R_{mnrs}g^{mr}g^{ns}$ is linear rather than quadratic in the curvature.

What we have described so far are "free-field" gauge theories. Of course, other fields may be introduced and interaction terms added to the Lagrangian. A good example is Einstein–Maxwell–Dirac theory, featuring gravity represented by the Riemannian connection on the tangent bundle, electromagnetism represented by a connection on a complex line bundle, and a Dirac matter field represented by a cross section of the line bundle.

We now proceed to the question of symmetries. Recall that the Lie derivative $T^{a\cdots}_{b\cdots}$ of a tensor field $T^{a\cdots}_{b\cdots}$ with respect to a vector field ξ^m is defined, in term of coordinates, as follows:

$$\pounds_\xi T^{a\cdots}_{b\cdots} = T^{a\cdots}_{b\cdots,m}\xi^m - T^{m\cdots}_{b\cdots}\xi^a_{,m} - \cdots + T^{a\cdots}_{m\cdots}\xi^m_{,b} + \cdots, \tag{5}$$

In intrinsic terms, the Lie derivative has the following meaning. The vector field ξ^m gives rise (under suitable assumptions such as compact support or restriction to local considerations) to a one-parameter family of diffeomorphisms $\delta(\epsilon)$ of the manifold M (or a subset thereof) onto itself. Each diffeomorphism gives rise to a bundle map $\delta^*(\epsilon): E \to E$ for each tensor bundle E defined over M. To every local cross section T of E there corresponds a cross section $T(\epsilon)$ of E for each diffeomorphism $\delta(\epsilon)$. The Lie derivative of T can then be defined as

$$\pounds_\xi T = \lim_{\epsilon\to 0}(1/\epsilon)[T - T(\epsilon)]. \tag{6}$$

Less formally, the vector field ξ^m gives rise to a set of transformations $x'^m = x^m + \epsilon\xi^m$, which can be thought of as *active* point transformations (point-to-point mappings) of the manifold M. Then with every tensor field $T^{a\cdots}_{b\cdots}$ there is associated a tensor field $T^{a\cdots}_{b\cdots}(\epsilon)(x') = \partial x'^a/\partial x^m \cdots \partial x^n/\partial x'^b \cdots T^{m\cdots}_{n\cdots}(x)$, and the Lie derivative at p of T is the limit of $1/\epsilon$ times the difference $T - T(\epsilon)$ evaluated at p.

Since the connection matrix B with respect to some frame field over $N \subseteq M$ is a matrix of 1-forms, it makes sense to consider the matrix whose entries are the Lie derivatives of the entries of B, i.e., of covariant vector fields. We shall denote this object by $\pounds_\xi B$, and refer to it as the ordinary Lie derivative of B with respect to ξ^m.

Now, again in informal terms, consider an infinitesimal change of frame field given by $U = I + \alpha S$, where α is infinitesimal. Consulting Eq. (2) and neglecting terms of order α^2, we find

$$B' = B + \alpha[S, B] - \alpha dS. \tag{7}$$

We define the Lie derivative of B with respect to the one parameter family of gauge transformations $U = I + \alpha S$ as

$$\underset{S}{\pounds} B \equiv \lim_{\alpha\to 0}(1/\alpha)(B' - B) = [S, B] - dS. \tag{8}$$

In analogy to the ordinary Lie derivative, the intrinsic interpretation is as follows. We think of the gauge transformation U not as a passive change of frame, but rather as an *active* bundle isomorphism, defined locally, by specifying a U for a local chart of the bundle. For each connection B defined for the chart over $N \subseteq M$, there is a one-parameter family of connections defined by Eq. (7), and the Lie derivative of B with respect to S is the limit of $1/\alpha$ times the difference $B(\alpha) - B$. Under a passive change of frame field $e' = eU^{-1}$, a straightforward computation reveals that

$$\pounds_{S'}B' = U(\pounds_S B)U^{-1}. \tag{9}$$

where $S' = USU^{-1}$.

We now have two types of Lie derivative defined for a connection on a vector bundle, and it is natural to combine these to obtain the concept of the Lie derivative of a connection with respect to the vector field ξ^m and the infinitesimal isomorphism S:

$$\pounds_{\xi,S}B = \pounds_\xi B + [S, B] - dS. \tag{10}$$

(Note that we could easily define an operator $\pounds_{\xi,S}$ on cross sections of the vector bundle itself:

$$\pounds_{\xi,S}v = \pounds_\xi v + Sv. \tag{11}$$

The Lie derivative of B would then have the following natural-looking property:

$$\pounds_{\xi,S}(Dv) - D(\pounds_{\xi,S}v) = (\pounds_{\xi,S}B)v. \tag{12}$$

However, the Lie derivative of a cross section would appear to have very little relevance for the type of applications we have in mind, since for any cross section v it would be possible to find many pairs ξ^m and S for which $\pounds_{\xi,S}v = 0$.)

Incidentally, considerations analogous to those above lead to the following definition for a generalized Lie derivative of a curvature matrix:

$$\pounds_{\xi,S}F = \pounds_\xi F + [S, F]. \tag{13}$$

The absence of a term involving dS comes from the fact that F transforms by Eq. (4) rather than Eq. (2) under a gauge transformation. A lengthy but straightforward computation confirms the following relationship:

$$\pounds_{\xi,S}F = d\pounds_{\xi,S}B + (\pounds_{\xi,S}B)\wedge B + B\wedge(\pounds_{\xi,S}B). \tag{14}$$

In deriving this result, use has been made of the fact that the action of the operator $\pounds_\xi$ on forms or matrices of forms can be expressed as

$$\pounds_\xi = d\iota_\xi + \iota_\xi d, \tag{15}$$

where d is the exterior derivative operator and ι_ξ is contraction on the first index with the vector ξ^m. Note that Eq. (14) insures that

$$\pounds_{\xi,S}F = 0 \text{ whenever } \pounds_{\xi,S}B = 0;$$

but the converse of this is not true in general.

We list here a pair of alternative expressions for

$$\pounds_{\xi,S}B \text{ and } \pounds_{\xi,S}F,$$

which are of some interest. First, again making use of Eq. (15), it can be shown that

$$\pounds_{\xi,S}B = \iota_\xi F + [\iota_\xi B + S, B] - d(\iota_\xi B + S). \tag{16}$$

Secondly, we have

$$\pounds_{\xi,S}F = (D\iota_\xi + \iota_\xi D)F + [\iota_\xi B + S, F], \tag{17}$$

where D is the "gauge covariant curl" operator, defined by

$$DM = dM + B \wedge M + (-1)^k M \wedge B, \tag{18}$$

where M is an $n \times n$ matrix of k-forms whose transformation law under a gauge change is given by $M' = UMU^{-1}$. (This is the operator which appears in the "Bianchi identities" $DF = 0$, which hold true for any curvature matrix F.)

Using Eqs. (16) and (17), we can derive the following properties of our Lie derivative operators:

$$\pounds_{\xi,(USU^{-1}+\iota_\xi dUU^{-1})} B' = U(\pounds_{\xi,S} B)U^{-1}, \tag{19}$$

$$\pounds_{\xi,(USU^{-1}+\iota_\xi dUU^{-1})} F' = U(\pounds_{\xi,S} F)U^{-1}, \tag{20}$$

Since, under the passive change of frame, the representation of the active isomorphism S changes according to $S' = USU^{-1}$, we must for consistency have

$$\pounds_{\xi,S'} B' = U(\pounds_{\xi,S} B)U^{-1} - \pounds_{0,\iota_\xi dUU^{-1}}(UBU^{-1} - dUU^{-1}) \tag{21}$$

and

$$\pounds_{\xi,S'} F' = U(\pounds_{\xi,S} F)U^{-1} - \pounds_{0,\iota_\xi dUU^{-1}}(UFU^{-1}), \tag{22}$$

under a change of frame. Because of this complicated transformation law,

$$\pounds_{\xi,S} B = 0 \text{ does not imply } \pounds_{\xi,S'} B' = 0,$$

and likewise for $\pounds_{\xi,S} F$. However, from Eqs. (19) and (20), we see that if $\pounds_{\xi,S} B = 0$

$$\pounds_{\xi,S} B = 0 \text{ in one frame,}$$

there is a *different* infinitesimal isomorphism $T' = USU^{-1} + \iota_\xi dUU^{-1}$ such that

$$\pounds_{\xi,T'} B' = 0 \text{ in the transformed frame,}$$

and likewise for $\pounds_{\xi,S} F = 0$.

We are now prepared to state the proposed definition of symmetry for a gauge theory. We say that a gauge potential B admits an (infinitesimal) symmetry, if there exists a vector field ξ^m and a matrix of scalars S for each allowable frame field, for which $\pounds_{\xi,S} B = 0$, with $S \to USU^{-1} + \iota_\xi dUU^{-1}$ for a change of frame; similarly, we say that a gauge field F admits an infinitesimal symmetry if $\pounds_{\xi,S} F = 0$ for some ξ^m and S. As mentioned above, this definition can be thought of as a generalization of (the infinitesimal versions of) certain special cases of symmetries which have appeared in the literature.

It is worthwhile to examine our definitions in the case of two familiar gauge theories, electromagnetism, and general relativity. In the case of electromagnetism, we are dealing with "1×1 matrices" so that expressions involving commutators of matrices vanish and our definitions of the generalized Lie derivatives reduce to

$$\pounds_{\xi,S} A = \pounds_\xi A - dS \tag{23}$$

and

$$\pounds_{\xi,S} F = \pounds_\xi F. \tag{24}$$

Note that in this very special Abelian case, $\pounds_{\xi,S} F$ is independent of S (since F itself is) and

$$\pounds_\xi F = 0 \text{ implies } \pounds_{\xi,S} A = 0 \text{ for some } S.$$

Neither of these statements is true in general in the non-Abelian case.

In the case of general relativity, there is already a natural definition of symmetry at hand, by virtue of the fact that the vector bundle involved is the tangent bundle of the spacetime. The natural definition for a symmetry of the gravitational field is the notion of an (infinitesimal) isometry, i.e., the existence of a Killing vector field ξ^m for which

$$\pounds_\xi g_{mn} = 0.$$

Our proposed definition of symmetry in this case is considerably more general than the notion of isometry, but the latter is contained as a special case: It is straightforward to show that

$$\pounds_\xi g_{mn} = 0 \text{ implies } \pounds_{\xi,0} \Gamma^r_{mn} = 0 \text{ and } \pounds_{\xi,0} R^r_{mns} = 0,$$

where the Christoffel symbol Γ^m_{nr} plays the role of the mn entry of the connection matrix with respect to a coordinate frame, and the Riemann tensor plays the role of the curvature matrix. Our definition of symmetry is equivalent to the requirement that there exists a chart (local coordinate system) in which the Christoffel symbols Γ^m_{nr} are independent of one of the coordinates, whereas in a true isometry this is also true of the metric tensor components g_{mn}. Obviously, the latter condition is sufficient, but not necessary for fulfilling the former. In fact, the condition

$$\pounds_{\xi,0} \Gamma = \pounds_\xi \Gamma = 0,$$

which states that the parallelism of the spacetime is invariant under the infinitesimal diffeomorphisms associated with ξ^m, is a well-known one; and the vector field ξ^m is said to generate an affine motion of the spacetime. Such a ξ^m is referred to as an affine Killing vector.

[1] Mineo Ikeda and Yoshihiko Miyachi, Progr. Theor. Phys. **27**, 474 (1962).
[2] Erick J. Weinberg and Alan H. Guth, "Non-Existence of Spherically Symmetric Monopoles with Multiple Magnetic Charge," preprint, Columbia University.
[3] C.N. Yang and R.L. Mills, Phys. Rev. **96**, 191 (1954); Tai Tsun Wu and Chen Ning Yang, Phys. Rev. D **12**, 3845 (1975).
[4] R. Utiyama, Phys. Rev. **101**, 1597 (1956).

Coherent electromagnetic waves in pair-correlated random distributions of aligned scatterers[a)]

Victor Twersky

Department of Mathematics, University of Illinois, Chicago, Illinois 60680
(Received 18 April 1977)

Recent results for the corresponding scalar problem are generalized to coherent electromagnetic waves in random distribution of pair-correlated obstacles (aligned or averaged over alignment). Proceeding essentially as before, we obtain dispersion equations by averaging the vector–dyadic functional equation relating the multiple and single scattered amplitudes of the obstacles. In general, for aligned nonradially symmetric scatterers, the resulting bulk indices of refraction specify anisotropic media; the anisotropy arises either from the scatterers' properties (physical parameters or shape, or both) or from their distribution, or from both. The illustrations include both isotropic and anisotropic cases, and the explicit results generalize earlier ones.

1. INTRODUCTION

In a recent paper,[1] we considered scattering of a scalar wave $\exp(i\mathbf{k}\cdot\mathbf{r})$ by a slab region of randomly distributed pair-correlated obstacles, and determined the average wave, and the associated bulk index of refraction (η) in terms of the scattering amplitude $g(\hat{\mathbf{r}},\hat{\mathbf{k}})$ for one isolated obstacle. Now, we extend the development to the analogous electromagnetic problems for incident $\hat{\delta}\exp(i\mathbf{k}\cdot\mathbf{r})$ by using vector forms[2] of the multiple scattered field and amplitude (**G**), and vector and tensor representations ($\mathbf{g},\tilde{\mathbf{g}}$) for the single obstacle. Relationships with earlier results[1–15] are mentioned in context.

Essentially as before,[1] we average the functional equation[2] $\mathbf{G}=\mathbf{G}[\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})]$ (with the ensemble average for two fixed scatterers replaced by that for one,[6,7] a procedure which may be interpreted by alternative approaches[8,9,11]) to obtain dispersion equations for η. In general, for aligned nonradially symmetric scatterers, η specifies an anisotropic medium; the anisotropy arises either from the scatterers' properties (physical parameters or shape, or both) or from their distribution, or from both.

For spherically symmetric pair correlations, we use vector spherical harmonics to consider spherically symmetric scatterers, and also to develop corresponding forms for ellipsoidal scatterers which we apply to electric plus magnetic dipoles. The earlier scalar development in terms of Mathieu functions for the two-dimensional problems of elliptic cylinders and elliptic correlations (with nonconfocal, nonsimilar, and nonparallel scatterer and exclusion surfaces) suffices for the electromagnetic cases, and as before we generalize the low-frequency results directly to the analogous problems for triaxial ellipsoids.

We begin with several definitions and representations, and use (1 : 34) for Eq. (34) of Ref. 1, etc., as well as essentially the same notation as before.[1,2]

[a)]Work supported in part by National Science Foundation Grants GP-33368X and MPS 75-07391.

2. DEFINITIONS AND REPRESENTATIONS

We write Maxwell's equations [after factoring $\exp(-i\omega t)$] in an anisotropic medium imbedded in free space as

$$\nabla\times\mathbf{H}=-ik\mathbf{D}=-ik\tilde{\epsilon}\cdot\mathbf{E},\quad \nabla\times\mathbf{E}=ik\mathbf{B}=ik\tilde{\mu}\cdot\mathbf{H},$$
$$k=|k|=2\pi/\lambda, \tag{1}$$

from which $\nabla\cdot\mathbf{B}=\nabla\cdot\mathbf{D}=0$. The relative parameters $\tilde{\epsilon}$ and $\tilde{\mu}$ are piecewise constant dyadics, which reduce to the identity dyadic $\tilde{\mathbf{I}}$ in free space, and to $\epsilon\tilde{\mathbf{I}}$, $\mu\tilde{\mathbf{I}}$ in an isotropic medium. At discontinuities (interfaces of different media), in terms of the unit normal $\hat{\mathbf{n}}$, we require continuity of $\hat{\mathbf{n}}\times\mathbf{E}$, $\hat{\mathbf{n}}\times\mathbf{H}$, $\hat{\mathbf{n}}\cdot\mathbf{D}$ and $\hat{\mathbf{n}}\cdot\mathbf{B}$. Equivalently, we work with

$$\Psi=\begin{Bmatrix}\mathbf{E}\\ \mathbf{H}\end{Bmatrix},\quad \tilde{\mathfrak{C}}=\begin{Bmatrix}\tilde{\epsilon}\\ \tilde{\mu}\end{Bmatrix},\quad \tilde{\mathfrak{B}}=\begin{Bmatrix}\tilde{\mu}^{-1}\\ \tilde{\epsilon}^{-1}\end{Bmatrix};$$
$$\Psi^{M}=\begin{Bmatrix}\tilde{\mathbf{H}}\\ -\tilde{\mathbf{E}}\end{Bmatrix}=\frac{\tilde{\mathfrak{B}}\cdot\nabla\times\Psi}{ik},\quad \frac{\nabla\times\Psi^{M}}{\tilde{\mathbf{I}}k}=-\tilde{\mathfrak{C}}\circ\Psi, \tag{2}$$

from which

$$\nabla\times(\tilde{\mathfrak{B}}\cdot\nabla\times\Psi)=k^2\tilde{\mathfrak{C}}\cdot\Psi,\quad \nabla\cdot(\tilde{\mathfrak{C}}\cdot\Psi)=\tilde{\mathfrak{C}}:\nabla\Psi=0 \tag{3}$$

with interface continuity of

$$\hat{\mathbf{n}}\times\Psi,\quad \hat{\mathbf{n}}\times\tilde{\mathfrak{B}}\cdot(\nabla\times\Psi),\quad \hat{\mathbf{n}}\cdot\tilde{\mathfrak{C}}\cdot\Psi,\quad \hat{\mathbf{n}}\cdot(\nabla\times\Psi). \tag{4}$$

We emphasize Ψ (which, for detailed applications, we identify with **E**, in general) but refer to its mate Ψ^{M} when convenient. For some purposes we also consider the polarizabilities (electric **P** and magnetic **M**) such that

$$(\tilde{\mathfrak{C}}-\tilde{\mathbf{I}})\cdot\Psi=\mathfrak{L}=\begin{Bmatrix}\mathbf{P}\\ \mathbf{M}\end{Bmatrix},$$
$$-(\tilde{\mathfrak{B}}-\tilde{\mathbf{I}})\cdot\nabla\times\Psi/ik=\mathfrak{L}^{M}=\begin{Bmatrix}\mathbf{M}\\ -\mathbf{P}\end{Bmatrix},$$
$$\nabla\cdot\Psi=-\nabla\cdot\mathfrak{L}. \tag{5}$$

Plane wave solutions of (3) correspond to

$$\Psi=\mathbf{A}\exp(i\mathbf{K}\cdot\mathbf{r}),\quad (\mathbf{K}\times\tilde{\mathfrak{B}}\times\mathbf{K}+k^2\tilde{\mathfrak{C}})\cdot\mathbf{A}\equiv\tilde{\mathbf{N}}\cdot\mathbf{A}=0,$$
$$\mathbf{K}=K\hat{\mathbf{K}},\quad K=k\eta,\quad \hat{\mathbf{K}}\cdot\tilde{\mathfrak{C}}\cdot\mathbf{A}=0 \tag{6}$$

with η as the complex relative index of refraction, and $\hat{\mathbf{K}}$ as the unit propagation vector. The eigenvalues of the

 0022-2488/78/1901-0215$1.00

determinantal equation $|\tilde{N}(\eta)| = 0$ (the dispersion equation) are the pair $(i = 1, 2)$ of indices $\eta_i(\hat{\mathbf{K}}_i)$ with corresponding unit eigenvectors determined by $\tilde{N}(\eta_i) \cdot \hat{\mathbf{A}}_i = 0$. From (2) and (5) with $\mathfrak{L} = \mathbf{L} \exp(i\mathbf{K} \cdot \mathbf{r})$, etc., we also have $\eta \hat{\mathbf{K}} \times \mathbf{A} = \mathbf{A}^M + \mathbf{L}^M$, $-\eta \hat{\mathbf{K}} \times \mathbf{A}^M = \mathbf{A} + \mathbf{L}$; consequently,

$$
\begin{aligned}
&-\eta^2 \hat{\mathbf{K}} \times (\hat{\mathbf{K}} \times \mathbf{A}) - \mathbf{A} = (\eta^2 - 1)\mathbf{A}_T - \mathbf{A}_K = \mathbf{L} - \eta \hat{\mathbf{K}} \times \mathbf{L}^M, \\
&\mathbf{A}_T = \mathbf{A} - \mathbf{A}_K = (\tilde{\mathbf{I}} - \hat{\mathbf{K}}\hat{\mathbf{K}}) \cdot \mathbf{A} = \mathbf{I}_T \cdot \mathbf{A}, \\
&(\eta^2 - 1)\mathbf{A}_T = \mathbf{L}_T - \eta \hat{\mathbf{K}} \times \mathbf{L}^M = (\eta^2 - 1)\tilde{\mathbf{I}}_T \cdot (\tilde{\mathfrak{C}} - \tilde{\mathbf{I}})^{-1} \cdot \mathbf{L}
\end{aligned}
\tag{7}
$$

with $\mathbf{I}_T(\hat{\mathbf{K}})$ as the planar identity transverse to $\hat{\mathbf{K}}$.

In an isotropic medium, $\tilde{\mathfrak{C}} = \mathfrak{C}\tilde{\mathbf{I}}$, $\tilde{\mathfrak{B}} = \mathfrak{B}\tilde{\mathbf{I}}$, (3) reduces to

$$
\nabla \times \nabla \times \Psi = \Psi k^2 \mathfrak{C}/\mathfrak{B} = \Psi K^2, \quad (\nabla^2 + K^2)\Psi = 0, \quad \nabla \cdot \Psi = 0. \tag{8}
$$

The plane wave solutions $\mathbf{A} \exp(i\mathbf{K} \cdot \mathbf{r})$ satisfy

$$
\begin{aligned}
&[-K^2 \mathfrak{B}\tilde{\mathbf{I}}_T(\hat{\mathbf{K}}) + k^2 \mathfrak{C}\tilde{\mathbf{I}}] \cdot \mathbf{A} = k^2 \mathfrak{C}\hat{\mathbf{K}}\hat{\mathbf{K}} \cdot \mathbf{A} = 0, \\
&\mathbf{A} = \tilde{\mathbf{I}}_T \cdot \mathbf{A} = A_T, \quad \eta^2 = \mathfrak{C}/\mathfrak{B} = \epsilon\mu,
\end{aligned}
$$

where η is independent of $\hat{\mathbf{K}}$ and the only constraint on $\mathbf{A}$ is that it be transverse to $\hat{\mathbf{K}}$.

The special case $\mathfrak{C} = \mathfrak{B} = \eta = 1$ represents free space, the isotropic imbedding medium for the scatterers. For a plane wave

$$
\boldsymbol{\phi} = \hat{\boldsymbol{\delta}}\phi(\mathbf{k}) = \hat{\boldsymbol{\delta}} \exp(i\mathbf{k} \cdot \mathbf{r}), \quad \hat{\boldsymbol{\delta}} \cdot \hat{\mathbf{k}} = 0, \tag{9}
$$

incident on a single obstacle, with center (the center of the smallest circumscribing sphere, a sphere of radius a) at the phase origin ($r = 0$), the external field satisfies

$$
\nabla \times \nabla \times \psi = k^2 \psi, \quad \psi = \psi(k) = \boldsymbol{\phi} + \mathbf{u}. \tag{10}
$$

The scattered wave $\mathbf{u}$ is the radiative function[2]

$$
\begin{aligned}
\mathbf{u}(\mathbf{r}) &= -c_o \int [(\tilde{\mathbf{h}} \times \hat{\mathbf{n}}) \cdot (\nabla \times \mathbf{u}) - (\nabla \times \tilde{\mathbf{h}}) \cdot (\hat{\mathbf{n}} \times \mathbf{u})]\, d\mathfrak{S}(\mathbf{r}') \\
&\equiv \{\tilde{\mathbf{h}}(k|\mathbf{r} - \mathbf{r}'|), \mathbf{u}(\mathbf{r}')\}, \\
\tilde{\mathbf{h}} &= \nabla \times \nabla \times \tilde{\mathbf{I}} h/k^2 = (\tilde{\mathbf{I}} + \nabla\nabla/k^2) h(k|\mathbf{r} - \mathbf{r}'|), \\
h(x) &= \exp(ix)/ix, \quad c_o = k/i4\pi,
\end{aligned}
\tag{11}
$$

with $\mathfrak{S}$ as the obstacle's surface, $\hat{\mathbf{n}}$ as the outward normal, and $\nabla = \nabla_{r'}$. In the scatterer's interior $\mathfrak{V}$ (characterized by the scalars $\mathfrak{C}'$, $\mathfrak{B}'$), the field is a nonsingular solution of

$$
\begin{aligned}
&\nabla \times \nabla \times \psi = K'^2 \psi, \quad K' = k\eta', \\
&\eta'^2 = \mathfrak{C}'/\mathfrak{B}' = \epsilon'\mu', \quad \psi(K') = \psi'.
\end{aligned}
\tag{12}
$$

On $\mathfrak{S}$ we require

$$
\hat{\mathbf{n}} \times \psi = \hat{\mathbf{n}} \times \psi', \quad \hat{\mathbf{n}} \times (\nabla \times \psi) = \mathfrak{B}' \hat{\mathbf{n}} \times (\nabla \times \psi'), \tag{13}
$$

as well as $\hat{\mathbf{n}} \cdot \psi = \mathfrak{C}' \hat{\mathbf{n}} \cdot \psi'$, $\hat{\mathbf{n}} \cdot (\nabla \times \psi) = \hat{\mathbf{n}} \cdot (\nabla \times \psi')$. Since $\{\tilde{\mathbf{h}}, \boldsymbol{\phi}\} = 0$ for $\mathbf{r}$ outside of $\mathfrak{V}$, we also have $\mathbf{u} = \{\tilde{\mathbf{h}}, \psi\}$, and from (13),

$$
\mathbf{u} = -c_o \int [\mathfrak{B}'(\tilde{\mathbf{h}} \times \hat{\mathbf{n}}) \cdot (\nabla \times \psi') - (\nabla \times \tilde{\mathbf{h}}) \cdot (\hat{\mathbf{n}} \times \psi')]\, d\mathfrak{S} \equiv [\tilde{\mathbf{h}}, \psi]. \tag{14}
$$

Then from Gauss' theorem and (12),

$$
\begin{aligned}
\mathbf{u} &= \psi - \boldsymbol{\phi} \\
&= c_o \int [(\mathfrak{C}' - 1) k^2 \tilde{\mathbf{h}} \cdot \psi' - (\mathfrak{B}' - 1)(\nabla \times \tilde{\mathbf{h}}) \cdot (\nabla \times \psi)]\, d\mathfrak{V}(\mathbf{r}') \\
&\equiv [\![\tilde{\mathbf{h}}, \psi]\!],
\end{aligned}
\tag{15}
$$

which also holds for $\psi' - \boldsymbol{\phi}$ at $\mathbf{r}$ in $\mathfrak{V}$ as a principal value with singularity excluded by $|\mathbf{r} - \mathbf{r}'| \to 0$.

For a perfect conductor, we take $\psi' = 0$ in $\mathfrak{V}$ and consider either $\hat{\mathbf{n}} \times (\boldsymbol{\phi} + \mathbf{u}) = 0$ or $\hat{\mathbf{n}} \times \nabla \times (\boldsymbol{\phi} + \mathbf{u}) = 0$ on $\mathfrak{S}$. For this case we use only the brace operation in the form $\mathbf{u} = \{\tilde{\mathbf{h}}, \boldsymbol{\phi} + \mathbf{u}\}'$ with the integrand containing the surviving term.

Asymptotically, for $kr \gg 1$ and $r \gg a \geq r'$,

$$
\begin{aligned}
&\tilde{\mathbf{h}}(k|\mathbf{r} - \mathbf{r}'|) \sim h(kr)\, \tilde{\boldsymbol{\phi}}(-\mathbf{k}_r), \quad \tilde{\boldsymbol{\phi}} \equiv (\tilde{\mathbf{I}} + \nabla\nabla/k^2)\, \phi, \\
&\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r) = \tilde{\mathbf{I}}_T(\hat{\mathbf{r}})\, \phi(-\mathbf{k}_r), \quad \mathbf{k}_r = k\hat{\mathbf{r}},
\end{aligned}
\tag{16}
$$

with $\tilde{\mathbf{I}}_T(\hat{\mathbf{r}}) = \tilde{\mathbf{I}} - \hat{\mathbf{r}}\hat{\mathbf{r}} = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\varphi}}\hat{\boldsymbol{\varphi}}$. Consequently,

$$
\mathbf{u} \sim h(kr)\, \mathbf{g}(\hat{\mathbf{r}}, \hat{\mathbf{k}} : \hat{\boldsymbol{\delta}}), \quad \hat{\mathbf{r}} \cdot \mathbf{g}(\hat{\mathbf{r}}) = 0, \quad \mathbf{g}(\hat{\mathbf{r}}) = \tilde{\mathbf{I}}_T(\hat{\mathbf{r}}) \cdot \mathbf{g}(\hat{\mathbf{r}}). \tag{17}
$$

Corresponding to the representations (11), (14), (15), we write the scattering amplitude $\mathbf{g}$ as

$$
\begin{aligned}
&\mathbf{g}\{\mathbf{k}_r, \mathbf{k} : \hat{\boldsymbol{\delta}}\} = \mathbf{g}\{\mathbf{k}_r\} = \{\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r), \mathbf{u}\}, \\
&\mathbf{g}[\mathbf{k}_r] = [\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r), \psi], \quad \mathbf{g}[\![\mathbf{k}_r]\!] = [\![\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r), \psi]\!].
\end{aligned}
\tag{18}
$$

In terms of $\mathbf{g}$ we have [at least for $r > a$ for all $\hat{\mathbf{r}}$, and for $r > (\hat{\mathbf{r}} \cdot \mathbf{r}')_{\max}$ for given $\hat{\mathbf{r}}$][2],

$$
\mathbf{u}(\mathbf{r}) = \int_c \exp(i\mathbf{k}_c \cdot \mathbf{r})\, \mathbf{g}(\hat{\mathbf{r}}_c), \quad \mathbf{k}_c = k\hat{\mathbf{r}}_c = k\hat{\mathbf{r}}(\theta_c, \varphi_c),
$$

$$
\int_c = \frac{1}{2\pi} \iint d\Omega(\theta_c, \varphi_c) \tag{19}
$$

with contours as for $h_0^{(1)}$.

Writing ψ_i for the field resulting from $\boldsymbol{\phi}_i = \hat{\boldsymbol{\delta}}_i \phi(\mathbf{k}_i)$, the reciprocity relation $\hat{\boldsymbol{\delta}}_1 \cdot \mathbf{g}(-\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2 : \hat{\boldsymbol{\delta}}_2) = \hat{\boldsymbol{\delta}}_2 \cdot \mathbf{g}(-\hat{\mathbf{k}}_2, \hat{\mathbf{k}}_1 : \hat{\boldsymbol{\delta}}_1)$ follows from $\{\psi_1, \psi_2\} = 0$. Decomposing $\{\psi^*, \psi\}$ yields the energy theorem in the form

$$
\begin{aligned}
&-\mathrm{Re}\, \hat{\boldsymbol{\delta}} \cdot \mathbf{g}(\hat{\mathbf{k}}, \hat{\mathbf{k}} : \hat{\boldsymbol{\delta}}) = \frac{\sigma_A + \sigma_S}{\sigma_O}, \quad \sigma_0 = \frac{4\pi}{k^2}, \quad \frac{\sigma_A}{\sigma_O} = \tfrac{1}{2}\{\psi^*, \psi\}, \\
&\frac{\sigma_S}{\sigma_O} = \tfrac{1}{2}\{\mathbf{u}^*, \mathbf{u}\} = \frac{1}{4\pi} \int \mathbf{g}^* \cdot \mathbf{g}\, d\Omega(\hat{\mathbf{r}}) = \mathfrak{M}|\mathbf{g}(\hat{\mathbf{r}}, \hat{\mathbf{k}} : \hat{\boldsymbol{\delta}})|^2
\end{aligned}
\tag{20}
$$

with σ_A and σ_S as the absorption and scattering cross sections and $\mathfrak{M}$ as the mean over direction of observation.

For a dyadic incident wave[2]

$$
\tilde{\boldsymbol{\phi}}(\mathbf{k}) = \tilde{\mathbf{I}}_T(\hat{\mathbf{k}})\, \phi(\mathbf{k}), \quad \tilde{\mathbf{I}}_T(\hat{\mathbf{k}}) = \tilde{\mathbf{I}} - \hat{\mathbf{k}}\hat{\mathbf{k}} = \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}, \quad \hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\delta}} \times \hat{\mathbf{k}}, \tag{21}
$$

the resulting $\tilde{\mathbf{u}}, \tilde{\psi}$ are linear combinations of the appropriate vectors, i.e., $\tilde{\mathbf{u}} = \mathbf{u}(\hat{\boldsymbol{\alpha}})\, \hat{\boldsymbol{\alpha}} + \mathbf{u}(\hat{\boldsymbol{\delta}})\, \hat{\boldsymbol{\delta}} = \{\mathbf{h}, \tilde{\mathbf{u}}\}$, etc. The corresponding dyadic scattering amplitude

$$
\begin{aligned}
&\tilde{\mathbf{g}}(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \mathbf{g}(\hat{\mathbf{r}}, \hat{\mathbf{k}} : \hat{\boldsymbol{\alpha}})\, \hat{\boldsymbol{\alpha}} + \mathbf{g}(\hat{\mathbf{r}}, \hat{\mathbf{k}} : \hat{\boldsymbol{\delta}})\, \hat{\boldsymbol{\delta}} = \tilde{\mathbf{I}}_T(\hat{\mathbf{r}}) \cdot \tilde{\mathbf{g}}(\hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \tilde{\mathbf{I}}_T(\hat{\mathbf{k}}), \\
&\hat{\mathbf{r}} \cdot \tilde{\mathbf{g}}(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \tilde{\mathbf{g}}(\hat{\mathbf{r}}, \hat{\mathbf{k}}) \cdot \hat{\mathbf{k}} = 0
\end{aligned}
\tag{22}
$$

satisfies $\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})=\tilde{\mathbf{g}}^c(-\hat{\mathbf{k}},-\hat{\mathbf{r}})$ with $\tilde{\mathbf{g}}^c$ as the transpose (Gibbs' conjugate). We may represent $\tilde{\mathbf{g}}$ in any of the forms in (18), e.g.,

$$\tilde{\mathbf{g}}[\![\mathbf{k}_r,\mathbf{k}]\!]=[\![\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r),\tilde{\Psi}(K';\mathbf{k})]\!]$$
$$=c_o\int_{\mathfrak{V}}[(\mathbb{C}'-1)k^2\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r)\cdot\tilde{\Psi}-(\mathfrak{B}'-1)(\nabla\times\tilde{\boldsymbol{\phi}})\circ(\nabla\times\tilde{\Psi})],$$
$$\int_{\mathfrak{V}}=\int d\mathfrak{V}(\mathbf{r}'). \tag{23}$$

For a fixed configuration of N obstacles with centers located by $\mathbf{r}_s$, we write the net field as[2,5-9]

$$\Psi=\phi+\sum_{s=1}^{N}\mathbf{U}_s(\mathbf{r}-\mathbf{r}_s),\quad \mathbf{U}_s(\mathbf{r}-\mathbf{r}_s)=\mathbf{U}_s(\mathbf{r}-\mathbf{r}_s;\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N), \tag{24}$$

with $\mathbf{U}_s\sim h(k|\mathbf{r}-\mathbf{r}_s|)\mathbf{G}_s$ for $|\mathbf{r}-\mathbf{r}_s|\sim\infty$, and $\mathbf{G}_s\sim\mathbf{g}_s\times\exp(i\mathbf{k}\cdot\mathbf{r}_s)$ as $|\mathbf{r}_t-\mathbf{r}_s|\sim\infty$ for fixed $\mathbf{r}_s$. Equivalently, with reference to scatterer t,

$$\Psi=\Psi_t=\Phi_t+\mathbf{U}_t,\quad \Phi_t=\phi+\sum_s'\mathbf{U}_s,\quad \sum_s'=\sum_{s\neq t}, \tag{25}$$

where Φ_t may be regarded as the net excitation. The functions Ψ_t, Φ_t, $\mathbf{U}_t$ satisfy the same relations (10)—(15) at scatterer t as ψ, ϕ, $\mathbf{u}$ for the single obstacle.

The corresponding multiple scattered amplitude $\mathbf{G}_t$ may be expressed in any of the forms in (18), e.g., as the volume integral over $\mathfrak{V}_t(\mathbf{r}')$ with $\mathbf{r}'$ as the local vector from $\mathbf{r}_t$,

$$\mathbf{G}_t(\hat{\mathbf{r}})=[\![(\mathfrak{T}-\hat{\mathbf{r}}\hat{\mathbf{r}})\exp(-i\mathbf{k}_r\circ\mathbf{r}'),\quad \Psi_t(\mathbf{r}_t+\mathbf{r}')]\!]$$
$$=[\![\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r),\Psi_t]\!] \tag{26}$$

or as $\mathbf{G}_t=\{\tilde{\boldsymbol{\phi}},\mathbf{U}_t\}$ over the surface $\mathfrak{S}_t$, etc. The analog of (19) is

$$\mathbf{U}_t(\mathbf{r}-\mathbf{r}_t)=\int_c\exp[i\mathbf{k}_c\circ(\mathbf{r}-\mathbf{r}_t)]\mathbf{G}_t(\hat{\mathbf{r}}_c). \tag{27}$$

The requirement that Ψ_t and $\psi_a=\phi_a+\mathbf{u}_a$ (for arbitrary $\hat{\mathbf{k}}_a$ and perpendicular $\hat{\boldsymbol{\delta}}_a$) satisfy the same conditions on $\mathfrak{S}_t$ and in $\mathfrak{V}_t$ corresponds to $\{\psi_a,\Psi_t\}_t=0$ over $\mathfrak{S}_t(\mathbf{r}')$; consequently $\hat{\boldsymbol{\delta}}_a\circ\mathbf{G}_t(-\hat{\mathbf{k}}_a)=\{\boldsymbol{\phi}_a,\mathbf{U}_t\}_t=\{\Phi_t,\mathbf{u}_a\}_t$. Similarly from $\{\tilde{\psi}_a,\Psi_t\}_t=0$, we have $\mathbf{G}_t(-\hat{\mathbf{k}}_a)=\{\tilde{\boldsymbol{\phi}}_a,\mathbf{U}_t\}_t=\{\Phi_t,\tilde{\mathbf{u}}_a\}_t^c$, which reduces to[2]

$$\mathbf{G}_t(\hat{\mathbf{r}})=\tilde{\mathbf{g}}_t(\hat{\mathbf{r}},\hat{\mathbf{k}})\cdot\hat{\boldsymbol{\delta}}\exp(i\mathbf{k}\cdot\mathbf{r}_t)$$
$$+\sum_s'\int_c\tilde{\mathbf{g}}_t(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)\cdot\mathbf{G}_s(\hat{\mathbf{r}}_c)\exp(i\mathbf{k}_c\cdot\mathbf{R}_{ts}),$$
$$\mathbf{R}_{ts}=\mathbf{r}_t-\mathbf{r}_s, \tag{28}$$

on expressing Φ_t of (25) in terms of (27), and using $\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})=\tilde{\mathbf{g}}^c(-\hat{\mathbf{k}},-\hat{\mathbf{r}})$.

In Secs. 3 and 4 we average (24) and (28) over an ensemble of configurations to derive representations (from Ψ) and dispersion equations (from G) for the bulk index η. As before,[1] the obstacles (with centers within the slab region $0\leq z\leq d$) may be assymetrical, and either similarly aligned, or averaged over alignment; if symmetrized by averaging, the distribution of alignments is uniform and uncorrelated with position or separation. For the ensemble of configurations, we specify the one-particle statistics by the average number (ρ) of scatterers in unit volume, and the two-particle statistics by $\rho f(\mathbf{R})$ with $f(\mathbf{R})$ as the distribution function for the separation ($\mathbf{R}_{ts}=\mathbf{r}_t-\mathbf{r}_s$) of pairs. The minimum separation of centers as a function of $\hat{\mathbf{R}}$ specifies the exclusion surface $\mathbf{R}=\mathbf{b}(\hat{\mathbf{R}})$; we require $f(\mathbf{R})=0$ for $R<b(\hat{\mathbf{R}})$, and $f(\mathbf{R})\sim1$ for $R\sim\infty$. If $\mathbf{b}=b\hat{\mathbf{R}}$ is a sphere with radius $b\geq2a$, then $f(R)$ is the usual radial distribution function. The general case we consider corresponds to differently aligned nonsimilar scatterer (**a**) and exclusion (**b**) surfaces.

3. THE AVERAGE WAVE

We write the average of Ψ over a statistically homogeneous ensemble of configurations of N identical, aligned obstacles with center $\mathbf{r}_s$ uniformly distributed in V as

$$\langle\Psi(\mathbf{r})\rangle=\phi+\rho\int_V\langle U_s(\mathbf{r}-\mathbf{r}_s;\mathbf{r}_s)\rangle_s\,d\mathbf{r}_s,$$
$$\rho=N/V,\quad \langle\mathbf{U}_s\rangle_s=\langle\Psi_s\rangle_s-\langle\Phi_s\rangle_s, \tag{29}$$

and use the radiative form $\langle\mathbf{U}_s(k)\rangle_s$ if $\mathbf{r}$ is outside of $\mathfrak{V}_s=\mathfrak{V}$, and $\langle\Psi_s(K')\rangle_s-\langle\Phi_s(k)\rangle_s$ if inside. We take V as the slab region $0\leq z\leq d$, and without loss of generality write the incident wave as

$$\phi=\hat{\mathbf{y}}\phi(\mathbf{k}),\quad \phi(\mathbf{k})=\exp(i\mathbf{k}\circ\mathbf{r}),$$
$$\mathbf{k}=k\hat{\mathbf{k}}(\alpha)=k(\hat{\mathbf{z}}\cos\alpha+\hat{\mathbf{x}}\sin\alpha)=\hat{\mathbf{z}}\gamma+\hat{\mathbf{x}}\tau,\quad 0\leq\alpha<\pi/2. \tag{30}$$

From Snell's law (preservation of phase parallel to planar boundaries), we have

$$\langle\Psi(\mathbf{r})\rangle=\langle\Psi(\mathbf{z})\rangle\exp(i\tau x),$$
$$\langle\Psi_s(\mathbf{r}_s+\mathbf{r}')\rangle_s=\langle\Psi_s(\mathbf{z}_s+\mathbf{r}')\rangle_s\exp(i\tau x_s), \tag{31}$$

and we may write the average multiple scattering amplitude $\langle\mathbf{G}_s\rangle_s=[\![\tilde{\boldsymbol{\phi}},\langle\Psi_s\rangle_s]\!]$ as

$$\langle\mathbf{G}_s(\hat{\mathbf{r}})\rangle_s=\mathbf{G}(\mathbf{r}_s;\hat{\mathbf{r}})=\exp(i\tau x_s)\,\mathbf{G}(z_s:\hat{\mathbf{r}}),$$
$$\mathbf{G}(z_s:\hat{\mathbf{r}})=[\![\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r),\langle\Psi_s(\mathbf{z}_s+\mathbf{r}')\rangle_s]\!]. \tag{32}$$

Using (27) for $\langle\mathbf{U}_s\rangle_s$ and substituting (32), we reduce the external form of (29) by integrating over x_s and y_s (to obtain δ functions). Thus for the transmitted field, at least for $z>d+a$,

$$\langle\Psi\rangle=\phi[\hat{\mathbf{y}}+c\int_0^d\exp(-i\tau z_s)\,\mathbf{G}(z_s:\hat{\mathbf{k}})\,dz_s]=\phi\mathfrak{T},$$
$$c=2\pi\rho/\gamma k. \tag{33}$$

Similarly, for the reflected field, at least for $z<-a$,

$$\langle\Psi\rangle-\phi=\phi' c\int_0^d\exp(i\gamma z_s)\,\mathbf{G}(z_s:\hat{\mathbf{k}}')\,dz_s=\phi'\,\mathfrak{R},$$
$$\phi'=\exp(i\mathbf{k}'\cdot\mathbf{r}),\quad \mathbf{k}'=\mathbf{k}(\pi-\alpha)=-\hat{\mathbf{z}}\gamma+\hat{\mathbf{x}}\tau. \tag{34}$$

The corresponding internal field, at least for $a<z<d-a$, consists in general of essentially four waves

$$\langle\Psi\rangle=\sum A_i\exp(i\mathbf{K}_i\cdot\mathbf{r})=\sum\Psi_i(\mathbf{r}),$$
$$\mathbf{K}_i=\hat{\mathbf{z}}\Gamma_i+\hat{\mathbf{x}}\tau=k\eta_i\hat{\mathbf{K}}_i=K_i\hat{\mathbf{K}}_i,\quad \mathrm{Im}\eta_i>0. \tag{35}$$

At least for $\mathbf{r}_s$ not within boundary layers (say, at least for $l<z_s<d-l'$), the functions $\langle\,\rangle_s$ consist of corresponding terms with translational property $\mathbf{f}_i(\mathbf{z}_s+\mathbf{r}')=\mathbf{f}_i(\mathbf{r}')\exp(i\Gamma_i z_s)$. For the simplest cases (e.g., spherically symmetric scatterers and statistics), the set of K's reduces to two which are images in $z=0$, with η independent of α.

Proceeding essentially as for (1:48), we write the average field at a fixed point $\mathbf{r}$ within the distribution as[10]

$$\langle\Psi(\mathbf{r})\rangle=\phi(\mathbf{r})+\rho\int_{V-\mathfrak{V}}\langle\mathbf{U}_s(k)\rangle_s\,d\mathbf{r}_s$$
$$+\rho\int_{\mathfrak{V}}[\langle\Psi_s(K')\rangle_s-\langle\Phi_s(k)\rangle_s]\,d\mathbf{r}_s \tag{36}$$

with arguments $\mathbf{r}-\mathbf{r}_s$ suppressed in the integrands. The second term represents the radiative contributions to $\mathbf{r}$ from scatterers whose centers ($\mathbf{r}_s$) are excluded from the region $\mathfrak{V}$ (the volume of one scatterer) around $\mathbf{r}$, and the last term represents the contributions of the scatterers with centers within $\mathfrak{V}$. Similarly, for the mate Ψ^M as in (2); equivalently, with $\tilde{\mathfrak{B}}$ now representing the bulk parameter associated with the coherent wave,

$$\tilde{\mathfrak{B}}\cdot\nabla\times\langle\Psi\rangle=\nabla\times\phi+\rho\int_{V-\mathfrak{V}}\nabla\times\langle U_s\rangle_s\,d\mathbf{r}_s+\rho\int_{\mathfrak{V}}[\mathfrak{B}'\nabla\times\langle\Psi_s\rangle_s-\nabla\times\langle\Phi_s\rangle_s]\,d\mathbf{r}_s, \tag{37}$$

with ∇ acting on $\mathbf{r}$ in the suppressed arguments $\mathbf{r}-\mathbf{r}_s$. The curl of (36) equals

$$\nabla\times\langle\Psi\rangle=\nabla\times\phi+\rho\int_{V-\mathfrak{V}}\nabla\times\langle U_s\rangle_s\,d\mathbf{r}_s+\rho\int_{\mathfrak{V}}[\nabla\times\langle\Psi_s\rangle_s-\nabla\times\langle\Phi_s\rangle_s]\,d\mathbf{r}_s, \tag{38}$$

where the surface integrals $\pm\int\hat{\mathbf{n}}\times\langle\Psi_s\rangle_s\,d\mathfrak{S}$ that arose in the interchange of $\nabla\times$ and $\int d\mathbf{r}_s$ canceled by continuity of $\hat{\mathbf{n}}\times\langle\Psi_s\rangle_s$ on $\mathfrak{S}$ (the surface traced by the centers of the scatterers closest to $\mathbf{r}$, with $\mathbf{r}-\mathbf{r}_s=\mathbf{r}'$ as the full set of scatterer's surface points). Similarly, from the curl of (37),

$$k^{-2}\nabla\times(\tilde{\mathfrak{B}}\cdot\nabla\times\langle\Psi\rangle)=\tilde{\mathfrak{C}}\cdot\langle\Psi\rangle=\phi+\rho\int_{V-\mathfrak{V}}\langle U_s\rangle_s\,d\mathbf{r}_s+\rho\int_{\mathfrak{V}}[\mathfrak{C}'\langle\Psi_s\rangle_s-\langle\Phi_s\rangle_s]\,d\mathbf{r}_s, \tag{39}$$

where the surface integrals that arose in the interchange of $\nabla\times$ and $\int d\mathbf{r}_s$ canceled by continuity of $\hat{\mathbf{n}}\times\langle\Psi_s^M\rangle_s$ on $\mathfrak{S}$. If $\langle\Psi\rangle=\mathbf{E}$ in (36), then (37)–(39) with $\tilde{\mathfrak{B}}=\tilde{\mu}^{-1}$, $\tilde{\mathfrak{C}}=\tilde{\epsilon}$ correspond to H, B, D in terms of E; alternatively if $\langle\Psi\rangle=\mathbf{H}$, then (37)–(39) with $\tilde{\mathfrak{B}}=\tilde{\epsilon}^{-1}$ and $\tilde{\mathfrak{C}}=\tilde{\mu}$ correspond to E, D, B in terms of H. All follow directly from the form, say $F[\langle\Psi\rangle]$ in (36), which applies for $\langle\Psi\rangle$ equal to any of the four electromagnetic field vectors, e.g., (36)–(39) correspond to the sequence $F[\mathbf{E}]$, $ikF[\mathbf{H}]=F[\boldsymbol{\mu}^{-1}\cdot\nabla\times\mathbf{E}]$, $ikF[\mathbf{B}]=F[\nabla\times\mathbf{E}]$, $F[\mathbf{D}]=F[\tilde{\epsilon}\cdot\mathbf{E}]$.

From (36) and (39)

$$(\tilde{\mathfrak{C}}-\tilde{\mathfrak{I}})\cdot\langle\Psi\rangle=\rho(\mathfrak{C}'-1)\int_{\mathfrak{V}}\langle\Psi_s\rangle_s\,d\mathbf{r}_s=\mathfrak{L}, \tag{40}$$

and from (37) and (38)

$$(\tilde{\mathfrak{B}}-\tilde{\mathfrak{I}})\cdot\nabla\times\langle\Psi\rangle=\rho(\mathfrak{B}'-1)\int_{\mathfrak{V}}\nabla\times\langle\Psi_s\rangle_s\,d\mathbf{r}_s=-ik\,\mathfrak{L}^M, \tag{41}$$

with $\mathfrak{L}$, $\mathfrak{L}^M$ essentially as for (5).

From the first equality in (39) we observe that $\nabla\cdot(\tilde{\mathfrak{C}}\cdot\langle\Psi\rangle)=0$; the result follows from the divergence of the final form, since the divergence of ϕ and of the integrands vanish and the additional surface contributions that arise in the interchange of $\nabla\cdot$ and $\int d\mathbf{r}_s$ cancel by $\hat{\mathbf{n}}\cdot(\langle U_s\rangle_s+\langle\Phi_s\rangle_s)=\mathfrak{C}'\hat{\mathbf{n}}\cdot\langle\Psi_s\rangle_s$ on $\mathfrak{S}$. Similarly from (38), $\nabla\cdot\nabla\times\langle\Psi\rangle$ is identically zero, as is the divergence of the various curls on the right, and the surface integrals $\pm\int\hat{\mathbf{n}}\cdot\nabla\times\langle\Psi_s\rangle_s\,d\mathfrak{S}$ cancel by continuity. On the other hand, the divergence of $\langle\Psi\rangle$ of (36), or of $\tilde{\mathfrak{B}}\cdot\nabla\times\langle\Psi\rangle$ of (37), equals the nonvanishing difference of the surface integrals that arise in the interchange of $\nabla\cdot$ and $\int d\mathbf{r}_s$; equivalently, from (40) and (41), respectively

$$\nabla\cdot\langle\Psi\rangle=\rho(\mathfrak{C}'-1)\int\hat{\mathbf{n}}\cdot\langle\Psi_s\rangle_s\,d\mathfrak{S}(\mathbf{r}_s)=-\nabla\cdot\mathfrak{L}, \tag{42}$$

$$\nabla\cdot(\tilde{\mathfrak{B}}\cdot\nabla\times\langle\Psi\rangle)=\rho(\mathfrak{B}'-1)\int\hat{\mathbf{n}}\cdot\nabla\times\langle\Psi_s\rangle_s\,d\mathfrak{S}(\mathbf{r}_s)=-ik\nabla\cdot\mathfrak{L}^M. \tag{43}$$

To reduce (40) and (41), we use (35) for $\langle\Psi\rangle$, i.e., $\sum\mathbf{A}_i\exp(i\mathbf{K}_i\cdot\mathbf{r})$, and $\mathfrak{L}=\sum\mathbf{L}_i\exp(i\mathbf{K}_i\cdot\mathbf{r})$; similarly we express $\langle\Psi_s(\mathbf{r}-\mathbf{r}_s)\rangle_s$ as a corresponding set of terms, each with the appropriate translation property $\mathbf{f}^i(\mathbf{r}_s)=\mathbf{f}^i(0)\exp(i\mathbf{K}_i\cdot\mathbf{r}_s)$,

$$\langle\Psi_s(\mathbf{r}-\mathbf{r}_s)\rangle_s=\sum\Psi^i(\mathbf{r}-\mathbf{r}_s)\exp(i\mathbf{K}_i\cdot\mathbf{r}_s)=\sum\Psi^i(\mathbf{r}')\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\exp(i\mathbf{K}_i\cdot\mathbf{r}), \tag{44}$$

where $\mathbf{r}-\mathbf{r}_s=\mathbf{r}'$ is the local vector from the center $\mathbf{r}_s$. Substituting into (40), we obtain

$$(\tilde{\mathfrak{C}}-\tilde{\mathfrak{I}})\cdot\mathbf{A}_i=\rho(\mathfrak{C}'-1)\int_{\mathfrak{V}}\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\,\Psi^i(\mathbf{r}')=\mathbf{L}_i,\qquad\int_{\mathfrak{V}}=\int d\mathfrak{V}(\mathbf{r}'),\quad\hat{\mathbf{K}}_i\cdot\tilde{\mathfrak{C}}\cdot\mathbf{A}_i=0, \tag{45}$$

where the transversality condition $\hat{\mathbf{K}}\cdot\tilde{\mathfrak{C}}\cdot\mathbf{A}=0$ follows from $\nabla\cdot(\tilde{\mathfrak{C}}\cdot\langle\Psi\rangle)=0$. Similarly from (41), and $\nabla_r\times\Psi^i(\mathbf{r}-\mathbf{r}_s)=\nabla_{r'}\times\Psi^i(\mathbf{r}')$, we have

$$i(\tilde{\mathfrak{B}}-\tilde{\mathfrak{I}})\cdot\mathbf{K}_i\times\mathbf{A}_i=[i(\tilde{\mathfrak{B}}-\tilde{\mathfrak{I}})\times\mathbf{K}_i]\cdot\mathbf{A}_i=\rho(\mathfrak{B}'-1)\int_{\mathfrak{V}}\exp(-i\mathbf{K}_i\cdot\mathbf{r}')\nabla\times\Psi^i(\mathbf{r}')=-ik\mathbf{L}_i^M,\quad\nabla=\nabla_{r'}. \tag{46}$$

Operating on (45) with $k^2(I-\hat{\mathbf{K}}_i\hat{\mathbf{K}}_i)\circ$, and on (46) with $-i\mathbf{K}_i\times$ and adding the results, we obtain in terms of $\tilde{\boldsymbol{\phi}}_T(-\mathbf{K})=(\tilde{\mathbf{I}}-\hat{\mathbf{K}}\hat{\mathbf{K}})\exp(-i\mathbf{K}\cdot\mathbf{r})=\tilde{\mathbf{I}}_T(\mathbf{K})\,\phi(-\mathbf{K})$, $\tilde{\mathbf{I}}_T(\mathbf{K}_i)\circ\mathbf{A}_i=\mathbf{A}_{Ti}$,

$$\begin{aligned}(K_i^2-k^2)\mathbf{A}_{Ti}&=\rho\int_{\mathfrak{V}}[(\mathfrak{C}'-1)k^2\tilde{\phi}_T(-\mathbf{K}_i)\cdot\Psi^i+(\mathfrak{B}'-1)(\nabla\times\tilde{\boldsymbol{\phi}}_T)\cdot(\nabla\times\Psi^i)]\\&=k^2(\mathbf{L}_{Ti}-\eta\hat{\mathbf{K}}_i\times\mathbf{L}_i^M)\\&=-(\rho/c_o)[\![\tilde{\boldsymbol{\phi}}_T(-\mathbf{K}_i),\Psi^i]\!]=-(\rho/c_o)\,\mathfrak{G}[\![\mathbf{K}_i]\!],\end{aligned} \tag{47}$$

where the final form is the volume representation of a transverse vector scattering amplitude, $\hat{\mathbf{K}}\cdot\mathfrak{G}[\![\hat{\mathbf{K}}]\!]=0$. From (45) in terms of $\tilde{\boldsymbol{\phi}}_K(-\mathbf{K})=\hat{\mathbf{K}}\hat{\mathbf{K}}\phi(-\mathbf{K})$, $\mathbf{K}_i\mathbf{K}_i\cdot\mathbf{A}_i=\mathbf{A}_{Ki}$, we also have

$$k^2\mathbf{A}_{Ki}=-\rho(\mathfrak{C}'-1)k^2\int_{\mathfrak{V}}\tilde{\boldsymbol{\phi}}_K(-\mathbf{K}_i)\cdot\Psi^i=(\rho/c_o)[\![\tilde{\boldsymbol{\phi}}_K(-\mathbf{K}_i),\Psi^i]\!], \tag{48}$$

where the last form, a longitudinal scattering amplitude, follows from $\nabla\times\tilde{\boldsymbol{\phi}}_K=0$.

If we add $\mathbf{A}_T$ of (47) and $\mathbf{A}_K$ of (48), we obtain

$$\begin{aligned}(K^2-k^2)\mathbf{A}&=-(\rho/c_o)[\![\tilde{\boldsymbol{\phi}}(-\mathbf{K}),\Psi^i]\!],\\\tilde{\boldsymbol{\phi}}(-\mathbf{K})&=(\tilde{\mathbf{I}}+\nabla\nabla/k^2)\,\phi(-\mathbf{K})=(I-\eta^2\hat{\mathbf{K}}\hat{\mathbf{K}})\,\phi\\&=[\tilde{\mathbf{I}}_T-(\eta^2-1)\hat{\mathbf{K}}\hat{\mathbf{K}}]\,\phi,\end{aligned} \tag{49}$$

corresponding to $(\eta^2-1)\mathbf{A}=\mathbf{L}_T-(\eta^2-1)\mathbf{L}_K+\eta\hat{\mathbf{K}}\times\mathbf{L}^M$. Alternatively, if we subtract (48) from (47), we have

$$\begin{aligned}(K^2-k^2)\mathbf{A}_T-k^2\mathbf{A}_K&=-(\rho/c_o)[\![\tilde{\mathbf{I}}\phi(-\mathbf{K}),\Psi^i]\!]\\&=k^2(\mathbf{L}+\eta\hat{\mathbf{K}}\times\mathbf{L}^M).\end{aligned} \tag{50}$$

The $[\![\,,\,]\!]$ forms in (49) and (50) correspond to scattering amplitudes having both longitudinal and transverse components.

We may write $\Psi^i=\tilde{\psi}^i\cdot\mathbf{A}_i$, etc., corresponding to

$$\langle\Psi_s\rangle_s=\sum\Psi^i\exp(i\mathbf{K}\cdot\mathbf{r}_s)=\sum\tilde{\psi}^i\cdot\mathbf{A}_i\exp(i\mathbf{K}_i\cdot\mathbf{r}_s)$$

$$= \sum \tilde{\Psi}^i(\mathbf{r}') \cdot \Psi_i(\mathbf{r}_s),$$

$$\langle \Phi_s \rangle_s = \sum \tilde{\Phi}^i \cdot \Psi_i(\mathbf{r}_s), \quad \langle \mathbf{U}_s \rangle_s = \sum \tilde{\mathfrak{U}}^i \circ \Psi_i(\mathbf{r}_s),$$

such that $\tilde{\Psi}^i, \tilde{\Phi}^i, \tilde{\mathfrak{U}}^i$ satisfy (10)–(15) as for an equivalent single obstacle (but although $\nabla \cdot \tilde{\Psi}^i = 0$, etc., it is not necessary that $\tilde{\Psi}^i \cdot \hat{\mathbf{K}}_i$ vanish). Similarly

$$\mathbf{G}(\mathbf{r}_s; \hat{\mathbf{r}}) = \sum \mathbf{G}^i(\mathbf{r}_s; \hat{\mathbf{r}}) = \sum \tilde{\mathbf{g}}^i \cdot \Psi_i(\mathbf{r}_s),$$

$$\mathbf{G}^i(z_s : \hat{\mathbf{r}}) = \mathfrak{G}^i(\mathbf{k}_r) \exp(i\Gamma_i z_s) = \tilde{\mathbf{g}}^i(\mathbf{k}_r \,|\, \mathbf{K}_i) \cdot \mathbf{A}_i \exp(i\Gamma_i z_s) \tag{51}$$

with $\tilde{\mathbf{g}}^i(\mathbf{k}_r \,|\, \mathbf{K}) = \{\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r), \tilde{\mathfrak{U}}^i\} = [\![\tilde{\boldsymbol{\phi}}(-\mathbf{k}_r), \tilde{\Psi}^i]\!]$, etc. (transverse left, but not necessarily right), as well as $\tilde{\mathbf{g}}(-\mathbf{k}_a \,|\, \mathbf{K}) = \{\tilde{\boldsymbol{\phi}}(\mathbf{k}_a), \tilde{\mathfrak{U}}^i\} = -\{\tilde{\mathbf{u}}_a, \tilde{\boldsymbol{\phi}}^i\}$ with $\tilde{\boldsymbol{\phi}}(\mathbf{k}_a) + \tilde{\mathbf{u}}_a$ as the dyadic solution for a conventional isolated scatterer and arbitrary $\hat{\mathbf{k}}_a$ as in (21)ff.

Thus, substituting $\Psi^i = \tilde{\Psi}^i \cdot \mathbf{A}_i$ in (47)–(50), we obtain

$$[(\eta^2 - 1)\tilde{\mathbf{I}}_T + c\,\tilde{\mathbf{g}}_T] \cdot \mathbf{A} = 0, \quad \tilde{\mathbf{g}}_T = [\![\tilde{\boldsymbol{\phi}}_T(-\mathbf{K}), \tilde{\Psi}^i]\!],$$

$$c = i4\pi\rho/k^3, \quad \tilde{\boldsymbol{\phi}}_T = (\tilde{\mathbf{I}} - \hat{\mathbf{K}}\hat{\mathbf{K}})\,\phi = \tilde{\mathbf{I}}\phi - \tilde{\boldsymbol{\phi}}_K,$$

$$[-\hat{\mathbf{K}}\hat{\mathbf{K}} + c\,\tilde{\mathbf{g}}_K] \cdot \mathbf{A} = 0, \quad \tilde{\mathbf{g}}_K = [\![\tilde{\boldsymbol{\phi}}_K, \tilde{\Psi}^i]\!],$$

$$[(\eta^2 - 1)\tilde{\mathbf{I}} + c\,\tilde{\mathbf{g}}] \cdot \mathbf{A} = 0, \quad \tilde{\mathbf{g}} = [\![\tilde{\boldsymbol{\phi}}_T - (\eta^2 - 1)\tilde{\boldsymbol{\phi}}_K, \tilde{\Psi}^i]\!],$$

$$[\eta^2\,\tilde{\mathbf{I}}_T - \tilde{\mathbf{I}} + c\,\tilde{\mathbf{g}}'] \cdot \mathbf{A} = 0, \quad \tilde{\mathbf{g}}' = [\![\tilde{\mathbf{I}}\phi, \tilde{\Psi}^i]\!]. \tag{52}$$

The forms (47)–(50) and (52) are vector and tensor analogs of the scalar form (1 : 28), and of related representations.[3–11] As for the scalar case, they do not lead directly to determinate dispersion equations but are of interest for interpretation and for constructing approximations. We may also obtain (47)–(50) and the associated extinction and cancelation relations analogous to the scalar forms in (1 : 27) by Gauss–Green procedures applied directly to (36), essentially as discussed in the paragraph after (1 : 55). We decompose the integral over $V - \mathfrak{V}$ into integrals over the boundary layers (say $V_l, V_{l'}$), and over the remainder V_B in which $\langle \mathbf{U}_s \rangle_s = \sum \mathbf{U}^i(\mathbf{r} - \mathbf{r}_s) \exp(i\mathbf{K}_i \cdot \mathbf{r}_s)$. Then we reduce the integrals

$$\mathbf{J} = \int_{V_B} \mathbf{U}^i(\mathbf{r} - \mathbf{r}_s) \exp(i\mathbf{K} \cdot \mathbf{r}_s)\, d\mathbf{r}_s$$

$$= \int (\tilde{\mathbf{I}} \exp(i\mathbf{K} \cdot \mathbf{r}_s)) \cdot \mathbf{U}^i\, d\mathbf{r}_s = \mathbf{J}_T(\tilde{\boldsymbol{\phi}}_T) + \mathbf{J}_K(\tilde{\boldsymbol{\phi}}_K) \tag{53}$$

to surface integrals over the layer surfaces ($z_s = l, d - l'$) and over $\mathfrak{S}$ (centered on $\mathbf{r}$). We have

$$\mathbf{J}_T = \int_{V_B} \tilde{\boldsymbol{\phi}}_T \cdot \mathbf{U}^i\, d\mathbf{r}_s$$

$$= \int_s [(\tilde{\boldsymbol{\phi}}_T \times \hat{\mathbf{n}}) \cdot (\nabla \times \mathbf{U}^i) - (\nabla \times \tilde{\boldsymbol{\phi}}_T) \cdot (\hat{\mathbf{n}} \times \mathbf{U}^i)]\, dS(\mathbf{r}_s)/(K^2 - k^2),$$

$$\mathbf{J}_K = \int_{V_B} \tilde{\boldsymbol{\phi}}_K \cdot \mathbf{U}^i\, d\mathbf{r}_s$$

$$= \hat{\mathbf{K}} \int_{V_B} \phi \hat{\mathbf{K}} \cdot \mathbf{U}^i\, d\mathbf{r}_s = (\hat{\mathbf{K}}/iK) \int (\nabla\phi) \cdot \mathbf{U}^i\, d\mathbf{r}_s$$

$$= (\hat{\mathbf{K}}/iK) \int \hat{\mathbf{n}} \cdot \mathbf{U}^i \phi\, dS, \tag{54}$$

for which we used $(\nabla\phi) \cdot \mathbf{U} = \nabla \cdot (\phi \mathbf{U})$ with $\nabla \cdot \mathbf{U} = 0$. The integrals over the layer surfaces plus $\boldsymbol{\phi}/\rho$ give the extinction ($\phi \mathbf{F} = 0$) and cancellation ($\phi' \mathbf{F}' = 0$) relations

$$\mathbf{F}(l) = 0 = \hat{\mathbf{y}} + c \int_0^l \mathbf{G}(z_s : \hat{\mathbf{k}}) \exp(-i\gamma z_s)\, dz_s$$

$$+ ic \exp(-i\gamma l) \sum \frac{\mathbf{G}^i(l : \hat{\mathbf{k}})}{\Gamma_i - \gamma}$$

$$\mathbf{F}'(l') = 0 = c \int_{d-l'}^{d} \mathbf{G}(z_s : \hat{\mathbf{k}}') \exp(i\gamma z_s)\, dz_s$$

$$- ic \exp[i\gamma(d - l')] \sum \frac{\mathbf{G}^i(d - l' : \hat{\mathbf{k}}')}{\Gamma_i + \gamma} \tag{55}$$

which also follow more directly from a Cartesian decomposition of $\mathbf{U}^i$. The terms Ψ_i/ρ, the integrals over $\mathfrak{S}$ and the corresponding volume integrals in (36)

$$\int_{\mathfrak{V}} [\Psi^i - \Phi^i] \exp(i\mathbf{K} \circ \mathbf{r}_s)\, d\mathbf{r}_s = \mathbf{I}(\Psi^i) - \mathbf{I}(\Phi^i)$$

combine to give the boundary independent coefficients of $\exp(i\mathbf{K} \circ \mathbf{r})$ as in (49). Thus writing $\mathbf{U}^i = \Psi^i - \Phi^i$ in (54) and indicating the corresponding surface integrals by $\mathbf{J}(\Psi^i) - \mathbf{J}(\Phi^i)$, we obtain $\mathbf{J}(\Phi^i) + \mathbf{I}(\Phi^i) = 0$ on converting $\mathbf{J}(\Phi^i)$ to the corresponding volume integral over $\mathfrak{V}$ and using the continuity of $\hat{\mathbf{n}} \times \Phi^i$, $\hat{\mathbf{n}} \times \nabla \times \Phi^i$ and $\hat{\mathbf{n}} \cdot \Phi^i$ on $\mathfrak{S}$ (as appopriate for a free-space source term). Converting $\mathbf{J}(\Psi^i)$ from a exterior integral over $\mathfrak{S}$ to an interior surface integral by using (13) as well as $\hat{\mathbf{n}} \cdot \Psi^i(k) = \mathfrak{C}' \hat{\mathbf{n}} \cdot \Psi^i(K')$, and then to the corresponding internal volume integral, we reduce (36) to the set

$$\mathbf{A}_i/\rho = \mathbf{J}_i(\Psi^i) + \mathbf{I}_i(\Psi^i)$$

$$= \frac{1}{K_i^2 - k^2}$$

$$\times \int_{\mathfrak{V}} [\tilde{\boldsymbol{\phi}}_T^i \circ \Psi^i(\mathfrak{C}' k^2 - K_i^2) + (\mathfrak{B}' - 1)(\nabla \times \tilde{\boldsymbol{\phi}}_T^i) \circ (\nabla \times \Psi^i)]$$

$$- \int_{\mathfrak{V}} \mathfrak{C}'\, \tilde{\boldsymbol{\phi}}^i \circ \Psi^i + \int_{\mathfrak{V}} (\tilde{\boldsymbol{\phi}}_K^i + \tilde{\boldsymbol{\phi}}_T^i) \circ \Psi^i,$$

$$\tilde{\boldsymbol{\phi}}^i = \tilde{\boldsymbol{\phi}}(-\mathbf{K}_i), \tag{56}$$

which is equivalent to (49).

From (33), (34), and (51) we obtain $\mathfrak{T}$ and $\mathfrak{R}$, and after using (55)

$$\mathfrak{T} = c \int_{d-l'}^{d} \mathbf{G}(z_s : \hat{\mathbf{k}}) \exp(-i\gamma z_s)\, dz_s$$

$$- ic \exp[-i\gamma(d - l')] \sum \frac{\mathbf{G}^i(d - l' : \hat{\mathbf{k}})}{(\Gamma_i - \gamma)},$$

$$\mathfrak{R} = c \int_0^l \mathbf{G}(z_s : \hat{\mathbf{k}}') \exp(i\gamma z_s)\, dz_s$$

$$+ ic \exp(i\gamma l) \sum \frac{\mathbf{G}^i(l; \hat{\mathbf{k}}')}{(\Gamma_i + \gamma)}. \tag{57}$$

If the translational property holds for all $0 \leq z_s \leq d$, then we set $l = l' = 0$ in (55) and (57) and obtain a determinate system for $\mathfrak{R}$ and $\mathfrak{T}$ and $\mathbf{A}_i$ once η_i and $\hat{\mathbf{A}}_i$ are known [from a dispersion equation as indicated after (6)]. In particular, for a semi-infinite distribution $z \geq 0$, we obtain the reflection coefficient in the term

$$\mathfrak{R}=R\hat{\mathbf{y}}+R'\hat{\boldsymbol{\alpha}}'=\frac{ic\,\mathfrak{G}_1(\mathbf{k}')}{\Gamma_1+\gamma}+\frac{ic\,\mathfrak{G}_2(\mathbf{k}')}{\Gamma_2+\gamma},$$

$$\mathfrak{G}_i(\mathbf{k}')=\tilde{g}(\mathbf{k}'\,|\,\mathrm{K}_i)\cdot\mathbf{A}_i,\quad c=\frac{2\pi\rho}{\gamma k},\quad \hat{\boldsymbol{\alpha}}'=\hat{\mathbf{y}}\times\hat{\mathbf{k}}', \tag{58}$$

and the extinction relation

$$\hat{\mathbf{y}}+\frac{ic\,\mathfrak{G}_1(\mathbf{k})}{\Gamma_1-\gamma}+\frac{ic\,\mathfrak{G}_2(\mathbf{k})}{\Gamma_2-\gamma}=0,\quad \mathfrak{G}_i(\mathbf{k})=\tilde{\mathbf{g}}(\mathbf{k}\,|\,\mathrm{K}_i)\cdot\mathbf{A}_i, \tag{59}$$

which are the vector analogs of (1 : 40). We could compare these forms with results for reflection from a uniform anisotropic half-space [based on (6)] to specify $\tilde{\mathfrak{C}}$ and $\tilde{\mathfrak{B}}$ in terms of $\tilde{\mathbf{g}}$. We discussed this procedure for the simpler scalar problem,[1] and the existing explicit forms (1 : 37)ff apply not only for scalar bulk parameters, but also for the case $\tilde{\mathfrak{C}}=\mathbb{C}\,\tilde{\mathrm{I}}$ and $\tilde{\mathfrak{B}}=\sum\hat{\mathbf{z}}_n\hat{\mathbf{z}}_m\mathfrak{B}_{mn}$ with $\hat{\mathbf{z}}_1,\hat{\mathbf{z}}_2,\hat{\mathbf{z}}_3=\hat{\mathbf{z}},\hat{\mathbf{x}},\hat{\mathbf{y}}$, provided that $\mathfrak{B}_{n3}=\mathfrak{B}_{3n}=\mathfrak{B}_3\delta_{3n}$. Thus if one bulk parameter is a scalar, and the other is a dyadic with a principal axis perpendicular to the plane of incidence, then we can work with $\langle\boldsymbol{\Psi}\rangle=\langle\Psi\rangle\hat{\mathbf{y}}$; a scalar development suffices for $\langle\Psi\rangle$ and the mate $\langle\boldsymbol{\Psi}^M\rangle$ follows from (2). We discuss this further after we have obtained and applied dispersion equations based on the average of G of (28).

4. DISPERSION RELATION

Essentially as before,[1] we start with the ensemble average of (28) with scatterer t fixed,

$$\langle\mathrm{G}_t(\hat{\mathbf{r}})\rangle_t=\tilde{\mathbf{g}}_t(\hat{\mathbf{r}},\hat{\mathbf{k}})\cdot\hat{\mathbf{y}}\exp(i\mathbf{k}\cdot\mathbf{r}_t)$$

$$+\rho\int d\mathbf{r}_s\,f(\mathbf{R}_{ts})\int_c\tilde{\mathbf{g}}_t(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)\cdot\langle\mathrm{G}_s(\hat{\mathbf{r}}_c)\rangle_{st}\exp(i\mathbf{k}_c\cdot\mathbf{R}_{ts}),$$

$$\mathbf{R}_{ts}=\mathbf{r}_t-\mathbf{r}_s, \tag{60}$$

where $f(\mathbf{R})$ is zero for $R<b(\hat{\mathbf{R}})$, say for $\mathbf{R}$ in $v(\mathbf{b})$ with v the exclusion volume. We replace $\langle\mathbf{G}_s\rangle_{st}$, the average over all variables except $\mathbf{r}_s$ and $\mathbf{r}_t$, by $\langle\mathbf{G}_s\rangle_s$, and for identical scatterers, we use

$$\mathbf{G}(\mathbf{r}_t;\hat{\mathbf{r}})=\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})\cdot\hat{\mathbf{y}}\exp(i\mathbf{k}\cdot\mathbf{r}_t)+\rho$$

$$\int_{V-v}d\mathbf{r}_s\,f(\mathbf{R}_{ts})\int_c\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)\cdot\mathbf{G}(\hat{\mathbf{r}}_s;\hat{\mathbf{r}}_c)\exp(i\mathbf{k}_c\cdot\mathbf{R}_{ts}). \tag{61}$$

As indicated for the scalar case (1 : 58), more complete forms may be obtained by alternative procedures. See particularly Vezzetti and Keller[11] for detailed discussion of electromagnetic dipoles.

Proceeding as for (1 : 59), we write $\int_{V-v}f=\int+\int(f-1)$; for $\int$ we take boundary layers $(V_L,V_{L'})$ into account and use $V-v=V_L+V_{L'}+V'$; for $\int(f-1)$, because $f\sim1$ as $R\sim\infty$, we replace V by V_∞ (all space) and integrate over $V_\infty-v=V''$. In V' and V'' we use (51), i.e., $\mathbf{G}(\mathbf{r}_s;\mathbf{r}_c)=\sum\mathfrak{G}^i(\mathbf{k}_c)\exp(i\mathrm{K}_i\cdot\mathbf{r}_s)$, and apply Green's theorem to the Cartesian components to reduce $\int_{V'}$ to integrals over the layer surfaces $(z_s=L,d-L')$ and exclusion surface $\mathbf{S}(\mathbf{b})$. The coefficients of $\phi(\mathbf{r}_t)$ and $\phi'(\mathbf{r}_t)$, i.e., the resulting terms $\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})\cdot\mathbf{F}\phi(\mathbf{r}_t)=0$ and $\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}}')\cdot\mathbf{F}'\phi'(\mathbf{r}_t)=0$, correspond to extinction and cancellation as in (55). The coefficients of $\exp(i\mathrm{K}_i\cdot\mathbf{r}_t)$ provide dispersion equations that determine K_i; we express these in terms of integrals over the exclusion surface $\mathbf{S}(\mathbf{b})$ and the depleted volume $V_\infty-v(\mathbf{b})$, such that $\mathbf{S}(\mathbf{b})=S\hat{\mathbf{n}}$ with $\hat{\mathbf{n}}$ outward from $v(\mathbf{b})$.

Essentially as before[1] we identify

$$\mathfrak{U}(\mathbf{k}_r,k\mathbf{R};\mathrm{K}_i)=\mathfrak{U}_i=\int_c\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)\cdot\mathfrak{G}(\mathbf{k}_c\,|\,\mathrm{K}_i)\exp(i\mathbf{k}_c\cdot\mathbf{R})$$

$$=\int_c\mathbf{F}(\mathbf{k}_r,\mathbf{k}_c\,|\,\mathrm{K}_i)\exp(i\mathbf{k}_c\cdot\mathbf{R}) \tag{62}$$

as a radiative solution of $(\nabla_R^2+k^2)\,\mathfrak{U}=0$ with $\mathbf{F}$ as the associated radiation amplitude. [In distinction to the vector solutions of (10), in general, $\nabla_R\cdot\mathfrak{U}\neq0$.] In terms of the conventional Green's surface integral form, we have

$$\mathbf{F}(\mathbf{k}_r,\mathbf{k}_a\,|\,\mathrm{K}_i)=\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{r}}_a)\cdot\mathfrak{G}(\mathbf{k}_a)$$

$$=c_o\int_{S(b)}[\exp(-i\mathbf{k}_a\cdot\mathbf{R})\,\partial_n\mathfrak{U}_i-\mathfrak{U}_i\partial_n\exp(-i\mathbf{k}_a\cdot\mathbf{R})]\,dS(\mathbf{b})$$

$$\equiv\{[\exp(-i\mathbf{k}_a\cdot\mathbf{R}),\ \mathfrak{U}_i]\}_b, \tag{63}$$

and we introduce an analog

$$\mathbf{F}\{\mathbf{k}_r,\mathrm{K}\,|\,\mathrm{K}\}\equiv\{[\exp(-i\mathrm{K}\cdot\mathbf{R}),\mathfrak{U}_i]\}_b. \tag{64}$$

For either case, we have $\hat{\mathbf{r}}\cdot\mathbf{F}=0$. In terms of $\mathfrak{U}$ we write the dispersion equation as

$$\mathfrak{G}(\mathbf{k}_r\,|\,\mathrm{K}_i)=-[\rho/(K_i^2-k^2)\,c_o]\{[\exp(-i\mathrm{K}_i\cdot\mathbf{R}),\mathfrak{U}_i]\}$$

$$+\rho\int_{V_\infty-v}[f(\mathbf{R})-1]\exp(-i\mathrm{K}_i\cdot\mathbf{R})\,\mathfrak{U}_i\,d\mathbf{R}. \tag{65}$$

Equivalently, in terms of $\mathbf{F}$, we have

$$\mathfrak{G}(\mathbf{k}_r\,|\,\mathrm{K}_i)=-[\rho\mathbf{F}\{\mathbf{k}_r,\mathrm{K}_i\,|\,\mathrm{K}_i\}/(K_i^2-k^2)\,c_o]$$

$$+\int_c M(\mathbf{k}_c,\mathrm{K}_i)\,\mathbf{F}(\mathbf{k}_r,\mathbf{k}_c,\,|\,\mathrm{K}_i),$$

$$M(\mathbf{k},\mathrm{K})=\rho\int_{V_\infty-v}[f(\mathbf{R})-1]\exp[i(\mathbf{k}-\mathrm{K})\cdot\mathbf{R}]\,d\mathbf{R}. \tag{66}$$

Equations (65) and (66) are vector analogs of (1 : 60) and (1 : 64), and the corresponding versions of the earlier approximations (1 : 65)—(1 : 68) may be developed directly.

Thus the analog of (1 : 65), appropriate for $\rho\approx0$, $M=0$ is

$$[(K^2-k^2)\,\tilde{\mathrm{I}}_T(\hat{\mathbf{k}})+(\rho/c_o)\,\tilde{\mathbf{g}}(\hat{\mathbf{k}},\hat{\mathbf{k}})]\cdot\mathfrak{G}=\tilde{\mathrm{N}}\cdot\mathfrak{G}=0,$$

$$\tilde{\mathrm{I}}_T(\hat{\mathbf{k}})=\tilde{\mathrm{I}}-\hat{\mathbf{k}}\hat{\mathbf{k}}=\hat{\boldsymbol{\alpha}}_1\hat{\boldsymbol{\alpha}}_1+\hat{\boldsymbol{\alpha}}_2\hat{\boldsymbol{\alpha}}_2,\quad\hat{\mathbf{k}}\times\hat{\boldsymbol{\alpha}}_1=\hat{\boldsymbol{\alpha}}_2. \tag{67}$$

Since $\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})$ is transverse left and right, the dispersion equation $|\tilde{\mathrm{N}}(\eta)|=0$, in general, involves a 2×2 determinant. The result obtained originally by Reiche[4] corresponds to $\tilde{\mathbf{g}}=g\,\tilde{\mathrm{I}}_T(\hat{\mathbf{k}})$ with g as the value for a spherical dipole. If the scatterers preserve the incident polarization in the sense $\tilde{\mathbf{g}}(\hat{\mathbf{k}},\hat{\mathbf{k}})=\sum g_i(\hat{\mathbf{k}})\,\hat{\boldsymbol{\alpha}}_i\hat{\boldsymbol{\alpha}}_i=\sum\mathbf{g}(\hat{\mathbf{k}},\hat{\mathbf{k}}:\hat{\boldsymbol{\alpha}}_i)\,\hat{\boldsymbol{\alpha}}_i$, then $|\tilde{\mathrm{N}}|=N_1N_2$ is a product of two factors, each of the form $N_i=K_i^2-k^2+(\rho/c_o)g_i(\hat{\mathbf{k}})=0$ discussed for (1 : 65). To take more implicit account of the anisotropy, we replace (67) by

$$[(K^2-k^2)\,\tilde{\mathrm{I}}_T(\hat{\mathrm{K}})+(\rho/c_o)\,\tilde{\mathbf{g}}(\hat{\mathrm{K}},\hat{\mathrm{K}})]\cdot\mathfrak{G}(\hat{\mathrm{K}}\,|\,\hat{\mathrm{K}})=0,$$

$$\tilde{\mathrm{I}}_T(\hat{\mathrm{K}})=\tilde{\mathrm{I}}-\hat{\mathrm{K}}\hat{\mathrm{K}}=\hat{\boldsymbol{\beta}}_1\hat{\boldsymbol{\beta}}_1+\hat{\boldsymbol{\beta}}_2\hat{\boldsymbol{\beta}}_2,\quad\hat{\mathrm{K}}\times\hat{\boldsymbol{\beta}}_1=\hat{\boldsymbol{\beta}}_2 \tag{68}$$

If the scatterers preserve polarization, $\tilde{\mathbf{g}}(\hat{\mathrm{K}},\hat{\mathrm{K}})=\sum g_i(\hat{\mathrm{K}})\,\hat{\boldsymbol{\beta}}_i\hat{\boldsymbol{\beta}}_i$, then

$$K_i^2-k^2=-(\rho/c_o)\,g_i(\mathbf{K})=-\rho ik\sigma_o g_i,$$

$$2\,\mathrm{Im}K_i\approx-\rho\sigma_o\mathrm{Re}g_i=\rho(\sigma_A^i+\sigma_S^i) \tag{69}$$

and similarly $\mathrm{Re}K_i/k\approx1+(\rho\sigma_o/2k)\,\mathrm{Im}g_i$; these general-

ize the leading term results for spherical dipoles obtained by Rayleigh.[3] To include correlations, for $\mathrm{Im}g \gg \mathrm{Re}g$, essentially as in (1:66), we use

$$2\,\mathrm{Im}K_i \approx \rho\sigma_A^i + \rho\sigma_O\,\mathfrak{M}[\,|\mathbf{g}(\hat{\mathbf{r}},\hat{\mathbf{K}}:\hat{\boldsymbol{\beta}}_i)|^2\,W(\hat{\mathbf{r}},\hat{\mathbf{K}})],$$
$$W(\hat{\mathbf{r}},\hat{\mathbf{K}}) = 1 + \rho\int[f(\mathbf{R})-1]\exp[ik(\hat{\mathbf{r}}-\hat{\mathbf{K}})\cdot\mathbf{R}]\,d\mathbf{R}, \tag{70}$$

which for average spacing between centers small compared to λ reduces to[12]

$$2\,\mathrm{Im}K_i \approx \rho(\sigma_A^i + \sigma_S^i W), \quad W \equiv 1 + \rho\int[f(\mathbf{R})-1]\,d\mathbf{R} \tag{71}$$

to lowest order in k.

Similarly, starting from

$$\mathfrak{G}(\mathbf{k}_K) \approx -(1/\Delta)\,\tilde{\mathbf{g}}(\hat{\mathbf{K}},\hat{\mathbf{K}})\cdot\mathfrak{G}(\mathbf{k}_K) + \int_c M\,\tilde{\mathbf{g}}(\hat{\mathbf{K}},\hat{\mathbf{r}}_c)\cdot\mathfrak{G}(\mathbf{k}_c),$$
$$\mathbf{k}_K = k\hat{\mathbf{K}}, \quad \Delta = (K^2-k^2)\,c_o/\rho,$$

we use $\mathfrak{G}(\mathbf{k}_c) = \tilde{\mathbf{g}}(\mathbf{k}_c|\,\mathrm{K})\cdot\mathbf{A} = \tilde{\mathbf{g}}(\mathbf{k}_c|\,\mathrm{K})\cdot\tilde{\mathbf{g}}^{-1}(\mathbf{k}_K|\,\mathrm{K})\cdot\mathfrak{G}(\mathbf{k}_K)$ to construct the analog of (1:68),

$$\{\Delta\tilde{\mathbf{I}} + [\tilde{\mathbf{I}} - \int_c M\,\tilde{\mathbf{g}}(\hat{\mathbf{K}},\hat{\mathbf{r}}_c)\cdot\tilde{\mathbf{g}}(\mathbf{k}_c|\,\mathrm{K})\cdot\tilde{\mathbf{g}}^{-1}(\mathbf{k}_K|\,\mathrm{K})]^{-1}\cdot$$
$$\cdot\,\tilde{\mathbf{g}}(\hat{\mathbf{K}},\hat{\mathbf{K}})\}\cdot\mathfrak{G}(\mathbf{k}_K) = 0, \tag{72}$$

which we may reduce further if $\tilde{\mathbf{g}}(\mathbf{k}_r|\,\mathrm{K}) \approx \tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{K}})$. The leading terms

$$[\Delta\tilde{\mathbf{I}} + \tilde{\mathbf{g}}(\hat{\mathbf{K}},\hat{\mathbf{K}}) + \int_c M(\mathbf{k}_c,\mathbf{K})\,\tilde{\mathbf{g}}(\hat{\mathbf{K}},\hat{\mathbf{r}}_c)\cdot\tilde{\mathbf{g}}(\hat{\mathbf{r}}_c,\hat{\mathbf{K}})]\cdot\mathfrak{G}(\mathbf{k}_K) \approx 0 \tag{73}$$

give the results in (68)—(71) under the same restrictions.

In the following, we separate variables in (66) to obtain more complete results for η for special cases. To facilitate using existing results, we identify Ψ with $\mathbf{E}$ in general, and as before[2] we refer to functions and coefficients labeled by $\mathbf{B}, b$ as electric (e) and by $\mathbf{C}, c$ as magnetic (m).

5. APPLICATIONS

In terms of vector spherical harmonics

$$\mathbf{B}_n^m(\hat{\mathbf{r}}) = (\hat{\theta}\partial_\theta + \varphi\partial_\varphi/\sin\theta)\,Y_n^m(\hat{\mathbf{r}}) = \hat{\mathbf{r}}\times\mathbf{C}_n^m(\hat{\mathbf{r}}),$$
$$Y_n^m(\hat{\mathbf{r}}) = P_n^m(\cos\theta)\exp(im\varphi), \quad P_n^{-m} = P_n^m D_n^m, \tag{74}$$
$$D_n^m = (-1)^m(n-m)!/(n+m)!$$

we write the dyadic isolated scattering amplitude, essentially in the notation of (2:86)ff as

$$\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}[\mathbf{C}_n^m(\hat{\mathbf{r}})\,\mathbf{c}_{nm}(\hat{\mathbf{k}}) + \mathbf{B}_n^m(\hat{\mathbf{r}})\,\mathbf{b}_{nm}(\hat{\mathbf{k}})],$$
$$\mathbf{c}_{nm}(\hat{\mathbf{k}}) = \sum_{\nu\mu}[\alpha_{n\nu}^{m\mu}\mathbf{C}_\nu^{-\mu}(\hat{\mathbf{k}}) + \beta_{n\nu}^{m\mu}\,\mathbf{B}_\nu^{-\mu}(\hat{\mathbf{k}})] = \mathbf{c}(\alpha,\beta), \tag{75}$$
$$\mathbf{b}_{nm}(\hat{\mathbf{k}}) = \mathbf{c}(\gamma,\delta),$$

where the four sets of scalar coefficients $\alpha,\beta,\gamma,\delta$ are independent of directions. From the reciprocity theorem $\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \tilde{\mathbf{g}}^c(-\hat{\mathbf{k}},-\hat{\mathbf{r}})$, we have $a_{\nu n}^{-\mu-m} = (-1)^{n+\nu}a_{n\nu}^{m\mu}$, $\gamma_{\nu n}^{-\mu-m} = (-1)^{n+\nu+1}\beta_{n\nu}^{m\mu}$ with $a = \alpha$ or $a = \delta$. The vector multiple scattered amplitude has the form

$$\mathfrak{G}(\hat{\mathbf{k}}_r) = \sum_{nm}[\mathbf{C}_n^m(\hat{\mathbf{r}})\,C_{nm} + \mathbf{B}_n^m(\hat{\mathbf{r}})\,B_{nm}] = \tilde{\mathbf{g}}(\mathbf{k}_r|\,\mathrm{K})\cdot\mathbf{A}. \tag{76}$$

We substitute (75) and (76) into (62) and proceed as for (2:136) to parallel the development (1:146)ff.

We rewrite the scalar products of the vector spherical harmonics that arise in $\int_c$ as finite sets of scalar harmonics,

$$\mathbf{C}_p^{-q}\cdot\mathbf{C}_r^t = \mathbf{B}_p^{-q}\cdot\mathbf{B}_r^t = \sum_l d_{l1}\begin{pmatrix}-q & ; & t\\ p & & r\end{pmatrix}Y_l^{t-q},$$
$$\mathbf{B}_p^{-q}\cdot\mathbf{C}_r^t = -\mathbf{C}_p^{-q}\cdot\mathbf{B}_r^t = \sum_l d_{l2}\begin{pmatrix}-q & ; & t\\ p & & r\end{pmatrix}Y_l^{t-q}. \tag{77}$$

The present coefficients d_{l1}, d_{l2} may be constructed from the known coefficients[13] d_l in the expansion

$$Y_p^{-q}Y_s^t = \sum_l d_l\begin{pmatrix}-q & ; & t\\ p & & s\end{pmatrix}Y_l^{t-q},$$

where $p+s+l$ is even and l changes by steps of 2 from $|p-s|$ (or from $|t-q|$ if it is the larger) to $p+s$. Thus, the d_{l1} follow from

$$\mathbf{C}_n^{-m}\cdot\mathbf{C}_r^t = mt\,Y_n^{-m}Y_r^t - \tfrac{1}{2}(r-t+1)(r+t)\,Y_n^{-m+1}Y_r^{t+1}$$
$$-\tfrac{1}{2}(n+m+1)(n-m)\,Y_n^{-m-1}Y_r^{t+1}, \tag{78}$$

and the d_{l2} from

$$\mathbf{B}_n^{-m}\cdot\mathbf{C}_r^t$$
$$= (i/2)(n-m)[(n-m+1)\,Y_{n-1}^{-m-1}Y_r^{t+1} + (r-t)\,Y_{n-1}^{-m}Y_r^t]$$
$$-(i/2)(r+t)[(r+t-1)\,Y_n^{-m+1}Y_{r-1}^{t-1} + (n+m)\,Y_n^{-m}Y_{r-1}^t], \tag{79}$$

such that all terms of (78) and (79) depend on $\exp[i(t-m)\varphi]$. We construct (78) and (79) from

$$\mathbf{C}_n^{-m}\cdot\mathbf{C}_r^t = D_n^m\left[\frac{mt}{\sin^2\theta}\,P_n^mP_r^t + \partial_\theta P_n^m\partial_\theta P_r^t\right]\exp[i(t-m)\varphi],$$
$$\mathbf{B}_n^{-m}\cdot\mathbf{C}_r^t = \frac{iD_n^m}{\sin\theta}\,[mP_n^m\partial_\theta P_r^t + tP_r^t\partial_\theta P_n^m]\exp[i(t-m)\varphi],$$

by successive applications of the recursive relations for P_n^m.

Using (77) within $\mathfrak{U}$ of (65) for radially symmetric statistics, we generate the corresponding pair of lattice-function analogs

$$\begin{pmatrix}-q & | & t\\ p & | & s\end{pmatrix}_i = \mathfrak{E}_i\begin{pmatrix}-q & | & t\\ p & | & s\end{pmatrix} = \sum_l d_{li}\begin{pmatrix}-q & ; & t\\ p & & s\end{pmatrix}Y_l^{t-q}(\hat{\mathbf{K}})\,T_l,$$

$$T_n = \eta^n/\Delta - \mathfrak{H}_n, \quad \Delta = (\eta^2-1)/c, \quad c = i4\pi\rho/k^3,$$
$$\mathfrak{H}_n = 4\pi\rho\int_0^\infty[f(R)-1]\,j_n(KR)\,h_n^{(1)}(kR)\,R^2\,dR, \tag{80}$$

with T_n as discussed for (1:148). As before,[1] we consider situations where the role of η in $\mathfrak{H}_n$ is minor and isolate major contributions in forms $\eta(\mathfrak{H})$ which may then be refined by iteration. We rule out $\eta^2 = 1$ from the start, as well as special eigenfunction solutions of systems which do not yield a leading term of order one (plus corrections) for η. Thus, the analog of the algebraic system (1:150) is the coupled pair

$$C_{nm} = -\sum\left\{\alpha_{n\nu}^{m\mu}\left[\mathfrak{E}_1\begin{pmatrix}-\mu & | & t\\ \nu & | & s\end{pmatrix}C_{st} - \mathfrak{E}_2B_{st}\right.\right.$$
$$\left.\left.+\,\beta[\mathfrak{E}_2C + \mathfrak{E}_1B]\right\} = C(\alpha,\beta), \quad B = C(\gamma,\delta), \tag{81}$$

where the scheme for the indices is shown only once.

Radially symmetric scatterers

For radially symmetric obstacles the β, γ coefficients vanish, and $\alpha_{n\nu}^{m\mu} = c_n(-1)^m \delta_{m\nu}\delta_{m\mu}$; similarly for $\delta_{n\nu}^{m\mu}$ in terms of b_n, with, e.g., c_n and b_n as in (2:98). The isolated scattering amplitude reduces to

$$\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \sum_{nm}[c_n \mathbf{C}_n^m(\hat{\mathbf{r}})\,\mathbf{C}_n^{-m}(\hat{\mathbf{k}}) + b_n \mathbf{B}_n^m(\hat{\mathbf{r}})\,\mathbf{B}_n^{-m}(\hat{\mathbf{k}})](-1)^m$$
$$= \sum_n [c_n \tilde{C}_n(\hat{r},\hat{k}) + b_n \tilde{B}_n(\hat{r},\hat{k})],$$
$$\tilde{\mathbf{B}}_n(\hat{\mathbf{r}},\hat{\mathbf{k}}) = (\tilde{\mathbf{I}} - \hat{\mathbf{r}}\hat{\mathbf{r}})\cdot\tilde{\mathbf{Q}}_n\cdot(\tilde{\mathbf{I}} - \hat{\mathbf{k}}\hat{\mathbf{k}}), \quad \tilde{\mathbf{C}}_n = -\hat{\mathbf{r}}\times\tilde{\mathbf{Q}}_n\times\hat{\mathbf{k}},$$
$$\tilde{\mathbf{Q}}_n = (\tilde{\mathbf{I}}\partial + \hat{\mathbf{k}}\hat{\mathbf{r}}\partial^2)\,P_n(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}), \quad \partial = \partial/\partial\hat{\mathbf{k}}\cdot\hat{\mathbf{r}},$$
$$a_n = \frac{a_n'}{1 - a_n'/d_n}, \quad d_n = \frac{2n+1}{n(n+1)}, \quad a_n = b_n, c_n,$$
$$\tilde{\mathbf{g}}(\hat{\mathbf{k}},\hat{\mathbf{k}}) = (\tilde{\mathbf{I}} - \hat{\mathbf{k}}\hat{\mathbf{k}})\,\tfrac{1}{2}\textstyle\sum n(n+1)(c_n + b_n), \tag{82}$$

such that for lossless scatterers, $-d_n \mathrm{Re} a_n = |a_n|^2$. Then (81) simplifies to

$$C_{nm} = -(-1)^m c_n \sum \left[\mathfrak{E}_1 \begin{pmatrix} -m & t \\ n & s \end{pmatrix} C_{st} - \mathfrak{E}_2 B_{st}\right],$$
$$B_{nm} = -(-1)^m b_n \sum \left[\mathfrak{E}_2 \begin{pmatrix} -m & t \\ n & s \end{pmatrix} C_{st} + \mathfrak{E}_1 B_{st}\right], \tag{83}$$

where (except for special eigenvalues which we discount), $C_{nm} = C_n(-1)^m \mathbf{C}_n^{-m}(\hat{\mathbf{K}})\cdot\mathbf{A}$ and $B_{nm} = B_n(-1)^m \mathbf{B}_n^{-m}(\hat{\mathbf{K}})\cdot\mathbf{A}$, i.e., $\tilde{\mathbf{g}}(\mathbf{k}_r|\mathrm{K})$ of $\mathfrak{G} = \tilde{\mathbf{g}}\cdot\mathbf{A}$ in (76) is the same form as $\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})$ of (82) with c_n, b_n replaced by C_n, B_n and $\hat{\mathbf{k}}$ by $\hat{\mathbf{K}}$. Thus

$$\tilde{\mathbf{g}}(\mathbf{k}_r|\mathrm{K}) = \sum [C_n \tilde{\mathbf{C}}_n(\hat{\mathbf{r}},\hat{\mathbf{k}}) + B_n \tilde{\mathbf{B}}_n(\hat{\mathbf{r}},\hat{\mathbf{K}})],$$
$$\tilde{\mathbf{g}}(\mathbf{k}_K|\mathrm{K}) = (\tilde{\mathbf{I}} - \hat{\mathbf{K}}\hat{\mathbf{K}})\,\tfrac{1}{2}\textstyle\sum n(n+1)(C_n + B_n). \tag{84}$$

This forward scattered amplitude preserves the incident polarization; the internal field is transverse, and the dispersion equation involves a single factor in η^2 with η independent of direction.

If we take $\hat{\mathbf{K}} = \hat{\mathbf{z}}'$ as the axis of the spherical harmonics, then from (74) and

$$2\partial_\theta Y_n^m = [(n-m+1)(n+m)\,P_n^{m-1} - P_n^{m+1}]\exp(im\varphi),$$
$$(2\cot\theta)\,\partial_\varphi Y_n^m = i[(n-m+1)(n+m)\,P_n^{m-1} + P_n^{m+1}]$$
$$\times\exp(im\varphi),$$

with $Y_\nu^\mu(\hat{\mathbf{z}}') = P_\nu^\mu(1) = \delta_{\mu 0}$, we see that the only nonvanishing harmonics are

$$\mathbf{C}_n^{-1}(\hat{\mathbf{z}}') = (i\hat{\mathbf{x}}' + \hat{\mathbf{y}}')/2 = -i\,\mathbf{B}_n^{-1}(\hat{\mathbf{z}}'),$$
$$\mathbf{C}_n^{1}(\hat{\mathbf{z}}') = n(n+1)(i\hat{\mathbf{x}}' - \hat{\mathbf{y}}')/2 = i\,\mathbf{B}_n^{1}(\hat{\mathbf{z}}'). \tag{85}$$

Thus the only coupling symbols of (80) that can arise are $\left(\begin{smallmatrix} \pm 1 & \mp 1 \\ n & s \end{smallmatrix}\right)_i$ and $\left(\begin{smallmatrix} \pm 1 & \pm 1 \\ n & s \end{smallmatrix}\right)_i$, and the second set vanishes [because its terms are proportional to $P_i^2(1) = 0$]. Consequently, (83) simplifies to

$$C_n/c_n = \sum_{s\neq 1}^{\infty}\left[\begin{pmatrix} -1 & 1 \\ n & s \end{pmatrix}_1 C_s - i\begin{pmatrix} -1 & 1 \\ n & s \end{pmatrix}_2 B_s\right],$$
$$B_n/b_n = \sum\left[\begin{pmatrix} -1 & 1 \\ n & s \end{pmatrix}_1 B_s - i\begin{pmatrix} -1 & 1 \\ n & s \end{pmatrix}_2 C_s\right], \tag{86}$$

i.e., to essentially one form. The leading symbols equal

$$\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_1 = -\tfrac{1}{3}(2T_0 + T_2), \quad \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_2 = -iT_1,$$
$$\begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}_1 = -\tfrac{3}{5}(3T_1 + 2T_3) = 3\begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}_1,$$
$$\begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}_2 = -i3T_2 = 3\begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}_2,$$
$$\begin{pmatrix} -1 & 1 \\ 2 & 2 \end{pmatrix}_1 = -3(\tfrac{2}{5}T_0 + \tfrac{1}{7}T_2 + \tfrac{16}{35}T_4),$$
$$\begin{pmatrix} -1 & 1 \\ 2 & 2 \end{pmatrix}_2 = -i\tfrac{3}{5}(T_1 + 4T_3), \tag{87}$$

such that $\left(\begin{smallmatrix} -1 & 1 \\ n & s \end{smallmatrix}\right)_i = \left(\begin{smallmatrix} -1 & 1 \\ n & n \end{smallmatrix}\right)_i s(s+1)/n(n+1)$.

For pure electric or magnetic dipoles, with a_1 equal to either b_1 or c_1, we have

$$\frac{1}{a_1} - \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_1 = \frac{1}{a_1} + \tfrac{1}{3}(2T_0 + T_2)$$
$$= \left[\frac{1}{a_1} - \tfrac{1}{3}(2c + 2\mathfrak{H}_0 + \mathfrak{H}_2)\right] + \frac{c\eta^2}{\eta^2 - 1}$$
$$\equiv \frac{1}{\mathfrak{A}_1} + \frac{c\eta^2}{\eta^2 - 1} = 0.$$

Thus

$$\frac{\eta^2 - 1}{\eta^2 + 2} = \frac{-a_1 c}{3[1 - a_1'\tfrac{1}{3}(2\mathfrak{H}_0 + \mathfrak{H}_2)]},$$
$$a_1 = \frac{a_1'}{1 - a_1'^{\frac{2}{3}}}, \quad a_1' = b_1', c_1', \tag{88}$$

which, with $\eta'^2 = \mathbb{E}'$ equal to ϵ' or μ', and $a_1' \approx i(ka)^3(\mathbb{E}' - 1)/(\mathbb{E}' + 2)$, gives the Maxwell–Clausius result for spheres and $k \to 0$, i.e., $(\mathbb{E} - 1)/(\mathbb{E} + 2) \to \rho\mathfrak{V}(\mathbb{E}' - 1)/(\mathbb{E}' + 2)$, $\mathfrak{V} = \frac{4}{3}\pi a^3$. We rewrite (88) as

$$\eta^{-2} = 1 + c\,\mathfrak{A}_1, \quad \mathfrak{A}_1 = a_1/(1 - a_1\mathfrak{H}_{11}),$$
$$\mathfrak{H}_{11} = \tfrac{1}{3}(2c + 2\mathfrak{H}_0 + \mathfrak{H}_2), \quad \eta^2 - 1 = -c\,\mathfrak{A}_1\eta^2,$$
$$-\Delta = -\frac{(\eta^2 - 1)}{c} = \mathfrak{A}_1\eta^2 = g[\![\mathbf{K}|\mathbf{K}]\!],$$
$$-\frac{\Delta}{\eta^2} = \mathfrak{A}_1 = \frac{a_1'}{\{1 - a_1'\tfrac{1}{3}[2c + 2(\mathfrak{H}_0 + 1) + \mathfrak{H}_2]\}}, \tag{89}$$

which differs from the corresponding result (1:157) for the scalar problem in that $\mathfrak{H}_{11}$ is a different linear combination of c, $\mathfrak{H}_0$, $\mathfrak{H}_2$.

Similarly for pure electric or magnetic quadrupoles, with $a_2 = b_2, c_2$ we have $1 - a_2\left(\begin{smallmatrix} -1 & 1 \\ 2 & 1 \end{smallmatrix}\right)_1 = 0$; from (87), we write $\left(\begin{smallmatrix} -1 & 1 \\ 2 & 1 \end{smallmatrix}\right)_1 = -3(\eta^4\Delta^{-1} - \mathfrak{H}_{22})$, and essentially as for (1:58),

$$\eta^2 - 1 = -c\,\mathfrak{A}_2\eta^4 3, \quad \mathfrak{A}_2 = \frac{a_2}{1 - a_2 3\mathfrak{H}_{22}},$$
$$\mathfrak{H}_{22} = \tfrac{1}{35}[c(14 + 19\eta^2) + 14\mathfrak{H}_0 + 5\mathfrak{H}_2 + 16\mathfrak{H}_4]. \tag{90}$$

More generally for pure 2η poles,

$$\frac{1}{a_n} = \begin{pmatrix} -1 & 1 \\ n & n \end{pmatrix}_1 = -\tfrac{1}{2}n(n+1)\left[\frac{\eta^{2n}}{\Delta} - \mathfrak{H}_{nn}\right],$$
$$\tfrac{1}{2}n(n+1)\mathfrak{H}_{nn} = H_{nn} = -\sum d_{i1}\begin{pmatrix} -1 & 1 \\ n & n \end{pmatrix}\left[\frac{c(\eta^{2n} - \eta^i)}{\eta^2 - 1} + \mathfrak{H}_i\right],$$

where we used $\mathbf{C}_n^{-1}(\hat{\mathbf{z}})\cdot\mathbf{C}_n^1(\hat{\mathbf{z}}) = \sum d_{i1}\left(\begin{smallmatrix} -1 & ; & 1 \\ n & ; & n \end{smallmatrix}\right) P_i(1) = \sum d_{i1}$ $= -\tfrac{1}{2}n(n+1)$. Thus

$$\eta^2 - 1 = -c\,\mathfrak{A}_n \eta^{2n} n(n+1)/2, \quad \mathfrak{A}_n = \frac{a_n}{1 - a_n H_{nn}} \tag{91}$$

If we integrate the equalities in (77) over the unit sphere, and use the orthogonality relations (2:33), we obtain $d_{0i}\binom{-m}{n};\binom{\mu}{\nu} = (-1)^m \delta_{i1}\delta_{n\nu}\delta_{m\mu}/d_n$ with

$$d_{01}\begin{pmatrix} -m & ; & m \\ n & & n \end{pmatrix} = \frac{(-1)^m}{d_n} = (-1)^m \frac{n(n+1)}{2n+1}$$

as the only nonvanishing coefficients of order zero. Thus, as illustrated in the above by $\mathfrak{H}_{11} = H_{11}$ and $3\mathfrak{H}_{22} = H_{22}$, in all cases H_{nn} includes $(c + \mathfrak{H}_0)/d_n$. In terms of the radiationless coefficients a'_n of (82), we write the analog of (1:159) as

$$\eta^2 - 1 = -c\,\mathfrak{A}_n \eta^{2n} n(n+1)/2,$$

$$\mathfrak{A}_n = \frac{a'_n}{1 - a'_n(\mathfrak{H}_0 + 1 + H'_{nn})/d_n}, \quad H_{nn} = (\mathfrak{H}_0 + H'_{nn})/d_n. \tag{92}$$

For lossless scatterers ($\mathrm{Re}\,a'_n = 0$), $-d_n \mathrm{Re}\,\mathfrak{A}_n = |\mathfrak{A}_n|^2 W_{nn}$, $W_{nn} = 1 + \mathrm{Re}(\mathfrak{H}_0 + H'_{nn})$. For dipoles and $W_{11} \approx W$, with W as in (71), we reduce (89) to

$$\eta^2 - 1 = c\,\mathfrak{A}_1 \approx c a''_1/(1 - \tfrac{2}{3} c a''_1),$$
$$a''_1 \approx a'_1/(1 - \tfrac{2}{3} a'_1 W), \quad c = i4\pi\rho/k^3. \tag{89'}$$

If each scatterer is both an electric plus magnetic dipole, then, from (86),

$$\left[\frac{1}{c_1} - \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_1\right] C_1 + i \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_2 B_1 = 0,$$
$$i \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_2 C_1 + \left[\frac{1}{b_1} - \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_1\right] B_1 = 0. \tag{93}$$

Introducing $\mathfrak{C}_1 = \mathfrak{A}_1(c_1)$ and $\mathfrak{B}_1 = \mathfrak{A}_1(b_1)$ in terms of the form $\mathfrak{A}_1(a_1)$ in (89), we have

$$\left[\frac{1}{c_1} + \tfrac{1}{3}(2T_0 + T_2)\right]\left[\frac{1}{b_1} + \tfrac{1}{3}(2T_0 + T_2)\right] - T_1^2$$
$$= \left(\frac{1}{\mathfrak{C}_1} + \frac{\eta^2}{\Delta}\right)\left(\frac{1}{\mathfrak{B}_1} + \frac{\eta^2}{\Delta}\right) - \eta^2\left(\frac{1}{\Delta} - \frac{\mathfrak{H}_1}{\eta}\right)^2 = 0, \tag{94}$$

from which

$$-\frac{\Delta}{\eta^2} = \left(\frac{1}{\eta^2} - 1\right)\frac{1}{c} = \frac{\mathfrak{C}_1 + \mathfrak{B}_1 + \mathfrak{C}_1\mathfrak{B}_1(c + 2\mathfrak{H}_1/\eta)}{1 - \mathrm{C}_1\mathfrak{B}_1\mathfrak{H}_1^2}. \tag{95}$$

In particular, if we neglect $\mathfrak{H}_1$ (or retain only $\mathfrak{H}_0 \approx W - 1$), then

$$\frac{1}{\eta^2} \approx (1 + c\mathfrak{B}_1)(1 + c\mathfrak{C}_1) = \frac{1}{\eta_e^2} \cdot \frac{1}{\eta_m^2} \tag{96}$$

with η_e and η_m as in (89) or (89'). An alternative development for small spheres is given by Mathur and Yeh.[14] More general dipoles are analyzed by Vezzetti and Keller.[11]

For subsequent generalizations, we rewrite the original system (93) as

$$\frac{C_1}{\mathfrak{C}_1} + \frac{\eta^2}{\Delta}(C_1 + B_1) - B_1\eta^2\bar{h}_{11} = 0,$$
$$\frac{B_1}{\mathfrak{B}_1} + \frac{\eta^2}{\Delta}(C_1 + B_1) - C_1\eta^2\bar{h}_{11} = 0, \tag{97}$$
$$\eta^2\bar{h}_{11} = \frac{c\eta}{\eta + 1} + \mathfrak{H}_1.$$

Introducing $m_1 = \eta^2\mathfrak{C}_1$, $p_1 = \eta^2\mathfrak{B}_1$, $M_1/C_1 = P_1/B_1 = -\Delta/(C_1 + B_1)$, (essentially as before for the scalar case) we obtain

$$P_1 = p_1(1 + M_1\bar{h}_{11}), \quad M_1 = m_1(1 + P_1\bar{h}_{11}),$$
$$\bar{h}_{11} = \frac{c}{\eta(\eta+1)} + \frac{\mathfrak{H}_1}{\eta^2}, \quad -\Delta = M_1 + P_1 \tag{98}$$

in the form (1:89). Thus

$$P_1 = p_1(1 + m_1\bar{h}_{11})/D, \quad M_1 = m_1(1 + p_1\bar{h}_{11})/D,$$
$$D = 1 - m_1 p_1 \bar{h}_{11}^2, \tag{99}$$

and

$$-\Delta = \frac{p_1 + m_1 + 2p_1 m_1 \bar{h}_{11}}{1 - m_1 p_1 \bar{h}_{11}^2} = P_1 + M_1 = g[\![\mathbf{K}|\mathbf{K}]\!]. \tag{100}$$

Similarly, if each scatterer is an electric dipole plus electric quadrupole, then from (86) and (87)

$$\left[\frac{1}{b_1} - \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}_1\right] B_1 - \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}_1 B_2 = 0,$$
$$-\begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}_1 B_1 + \left[\frac{1}{b_2} - \begin{pmatrix} -1 & 1 \\ 2 & 2 \end{pmatrix}_1\right] B_2 = 0,$$
$$-\frac{1}{3}\begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}_1 = -\begin{pmatrix} -1 & 1 \\ 2 & 1 \end{pmatrix}_1 = \frac{\eta^3}{\Delta} - \mathfrak{H}_{12}, \tag{101}$$

$$\mathfrak{H}_{12} = \tfrac{1}{5}(3c\eta + 3\mathfrak{H}_1 + 2\mathfrak{H}_3).$$

Thus with $\mathfrak{B}_i$ in the form $\mathfrak{A}_i$ of (89) and (90), we have

$$B_1\left(\frac{1}{\mathfrak{B}_1} + \frac{\eta^2}{\Delta}\right) + B_2 3\left(\frac{\eta^3}{\Delta} - \mathfrak{H}_{12}\right) = 0,$$
$$B_1\left(\frac{\eta^3}{\Delta} - \mathfrak{H}_{12}\right) + B_2\left(\frac{1}{\mathfrak{B}_2} + \frac{3\eta^4}{\Delta}\right) = 0. \tag{102}$$

In terms of $p_1 = \mathfrak{B}_1\eta^2$, $p_2 = 3\mathfrak{B}_2\eta^4$, $\mathbb{B}_1 = B_1\eta$, $\mathbb{B}_2 = 3B_2\eta^2$, $h_{12} = \mathfrak{H}_{12}/\eta^3$, we obtain

$$\frac{\mathbb{B}_1}{p_1} + \frac{\mathbb{B}_1 + \mathbb{B}_2}{\Delta} - \mathbb{B}_2 h_{12} = 0,$$
$$\frac{\mathbb{B}_2}{p_2} + \frac{\mathbb{B}_1 + \mathbb{B}_2}{\Delta} - \mathbb{B}_1 h_{12} = 0. \tag{103}$$

Introducing $P_i = -\Delta\,\mathbb{B}_i/(\mathbb{B}_1 + \mathbb{B}_2)$, we have essentially as for (98)

$$P_1 = p_1(1 + P_2 h_{12}), \quad P_2 = p_2(1 + P_1 h_{12}),$$
$$h_{12} = \frac{1}{5\eta^3}(3c\eta + 3\mathfrak{H}_1 + 2\mathfrak{H}_3), \quad P_1 + P_2 = -\Delta, \tag{104}$$

and consequently

$$P_1 = p_1(1 + p_2 h_{12})/D, \quad P_2 = p_2(1 + p_1 h_{12})/D,$$
$$D = 1 - p_1 p_2 h_{12}^2,$$
$$-\Delta = \frac{p_1 + p_2 + 2p_1 p_2 h_{12}}{1 - p_1 p_2 h_{12}^2} = P_1 + P_2 = g[\![\mathbf{K}|\mathbf{K}]\!]. \tag{105}$$

The dominant term of h_{12} is $3c/5\eta^2$ and for some purposes (e.g., b_2 near resonance) we may use

$$\frac{1}{\eta^2} \approx 1 + \frac{[c(\mathfrak{B}_1 + 3\mathfrak{B}_2\eta^2 + \mathfrak{B}_1\mathfrak{B}_2\eta^2 c 18/5)]}{1 - \mathfrak{B}_1\mathfrak{B}_2\eta^2 c^2 27/25}. \tag{106}$$

The corresponding results for the case of magnetic dipoles plus magnetic quadrupoles are obtained on replacing $b, \mathfrak{B}, p$ by $c, \mathfrak{C}, m$.

If each scatterer is a magnetic dipole (c_1) plus electric quadrupole (b_2), we proceed essentially as for (95) to obtain

$$-\frac{\Delta}{\eta^2}=\frac{\mathbb{C}_1+3\mathfrak{B}_2\eta^2+\mathbb{C}_1 3\mathfrak{B}_2\eta^2(c+2\mathfrak{H}_2/\eta^2)}{1-\mathbb{C}_1\mathfrak{B}_2 3\mathfrak{H}_2^2}\,. \tag{107}$$

The analog of (100) is

$$-\Delta=\frac{m_1+p_2+2m_1p_2\bar{h}_{12}}{1-m_1p_2\bar{h}_{12}^2}\,,\quad \bar{h}_{12}=\frac{c}{\eta(\eta+1)}+\frac{\mathfrak{H}_2}{\eta^3} \tag{108}$$

Similarly for b_1 and c_2 we obtain the same forms in terms of $\mathfrak{B}_1$, $\mathbb{C}_2$ and p_1, m_2.

If each scatterer is an electric plus magnetic quadrupole, we obtain

$$-\Delta=\frac{3\eta^4[\mathfrak{B}_2+\mathbb{C}_2+3\mathfrak{B}_2\mathbb{C}_2\eta^2(c+2\bar{\mathfrak{H}}_{22}/\eta^3)]}{1-9\mathfrak{B}_2\mathbb{C}_2\bar{\mathfrak{H}}_{22}^2}\,,$$
$$\bar{\mathfrak{H}}_{22}=\tfrac{1}{5}(c\eta+\mathfrak{H}_1+4\mathfrak{H}_3) \tag{109}$$

or, equivalently,

$$-\Delta=\frac{m_2+p_2+2m_2p_2\bar{h}_{22}}{1-m_2p_2\bar{h}_{22}^2}\,,$$
$$\bar{h}_{22}=\frac{c}{\eta(\eta+1)}+\frac{\bar{\mathfrak{H}}_{22}}{\eta^4}\,. \tag{110}$$

To generalize the developments that led to (100), (104), (108), and (110), we rewrite (86) in terms of block matrices as

$$\left[\begin{pmatrix}c^{-1}&0\\0&b^{-1}\end{pmatrix}-\begin{pmatrix}\mathfrak{E}_1-i\mathfrak{E}_2\\-i\mathfrak{E}_2\quad\mathfrak{E}_1\end{pmatrix}\right]\begin{pmatrix}C\\B\end{pmatrix}=0,$$
$$c^{-1}=[c_n^{-1}\delta_{n\nu}],\quad b^{-1}=[b_n^{-1}\delta_{n\nu}],\quad \mathfrak{E}_i=\left[\begin{pmatrix}-1&1\\n&\nu\end{pmatrix}_i\right] \tag{111}$$

with $C=[C_\nu]$, $B=[B_\nu]$ as column matrices. In terms of

$$\frac{m_n}{\mathbb{C}_n}=\frac{p_n}{\mathfrak{B}_n}=\eta^{2n}\frac{n(n+1)}{2},\quad \frac{\mathbb{C}_n}{C_n}=\frac{\mathbb{B}_n}{B_n}=\eta^n\frac{n(n+1)}{2}$$

we have

$$\mathbb{C}_n\left(\frac{1}{\Delta}+\frac{1}{m_n}\right)+\sum{}'\,\mathbb{C}_\nu\left(\frac{1}{\Delta}-h_{m\nu}\right)+\sum\mathbb{B}_\nu\left(\frac{1}{\Delta}-\bar{h}_{n\nu}\right)=0,$$
$$\mathbb{B}_n\left(\frac{1}{\Delta}+\frac{1}{p_n}\right)+\sum{}'\,\mathbb{B}_\nu\left(\frac{1}{\Delta}-h_{n\nu}\right)+\sum\mathbb{C}_\nu\left(\frac{1}{\Delta}-\bar{h}_{m\nu}\right)=0. \tag{112}$$

Introducing $M_n/\mathbb{C}_n=P_n/\mathbb{B}_n=-\Delta/\sum(\mathbb{C}_m+\mathbb{B}_m)$, we obtain

$$M_n=m_n[1+\textstyle\sum_\nu' M_\nu h_{m\nu}+\sum_\nu P_\nu\bar{h}_{n\nu}],$$
$$P_n=p_n[1+\textstyle\sum{}' P_\nu h_{m\nu}+\sum M_\nu\bar{h}_{n\nu}], \tag{113}$$
$$\textstyle\sum(P_n+M_n)=-\Delta,$$

which provides the generalization of (1 : 89).

The present h's may be constructed by procedures similar to that in (1 : 160). Thus from

$$\mathbf{C}_n^{-1}(\hat{\mathbf{z}})\cdot\mathbf{C}_\nu^1(\hat{\mathbf{z}})=-\tfrac{1}{2}\nu(\nu+1)=\sum d_{i1}\begin{pmatrix}-1&;&1\\n&;&\nu\end{pmatrix},$$
$$\mathbf{B}_n^{-1}(\hat{\mathbf{z}})\cdot\mathbf{C}_\nu^1(\hat{\mathbf{z}})=-(i/2)\,\nu(\nu+1)=\sum d_{i2}\begin{pmatrix}-1&;&1\\n&;&\nu\end{pmatrix},$$

we write

$$-\frac{2}{\nu(\nu+1)}\begin{pmatrix}-1&1\\n&\nu\end{pmatrix}_1=\frac{\eta^{n+\nu}}{\Delta}-\mathfrak{H}_{n\nu},$$
$$\mathfrak{H}_{n\nu}=-\frac{2}{\nu(\nu+1)}\sum d_{i1}\begin{pmatrix}-1&;&1\\n&;&\nu\end{pmatrix}\left[\frac{c(\eta^{n+\nu}-\eta^t)}{\eta^2-1}+\mathfrak{H}_t\right],$$
$$h_{n\nu}=\frac{\mathfrak{H}_{n\nu}}{\eta^{n+\nu}};\quad\frac{i2}{\nu(\nu+1)}\begin{pmatrix}-1&1\\n&\nu\end{pmatrix}_2=\frac{\eta^{n+\nu-1}}{\Delta}-\bar{\mathfrak{H}}_{n\nu},$$
$$\bar{\mathfrak{H}}_{n\nu}=\frac{2i}{\nu(\nu+1)}\sum d_{i2}\begin{pmatrix}-1&;&1\\n&;&\nu\end{pmatrix}\left[\frac{c(\eta^{n+\nu-1}-\eta^t)}{\eta^2-1}+\mathfrak{H}_t\right],$$
$$\bar{h}_{n\nu}=\frac{c}{\eta(\eta+1)}+\frac{\bar{\mathfrak{H}}_{n\nu}}{\eta^{n+\nu}}\,. \tag{114}$$

Aligned nonspherical scatterers

For aligned triaxial ellipsoids (or for other scatterers having the same reflection and inversion symmetries) with major axes $\mathbf{a}_i$ along $\hat{\boldsymbol{\zeta}}_i$ (in an orthogonal basis $\hat{\boldsymbol{\zeta}}_1, \hat{\boldsymbol{\zeta}}_2, \hat{\boldsymbol{\zeta}}_3=\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\eta}}$) we take $\hat{\boldsymbol{\zeta}}_1$ as the axis of the spherical harmonics Y_n^m. If the scatterers are either purely electric ($\alpha=b$, $\beta=\delta=0$) or purely magnetic ($\delta=c$, $\alpha=\beta=\gamma=0$), we proceed essentially as before for the scalar case. Then, with $A=B$ or C and with $a=b$ or c, (81) reduces to

$$A_n^m=-\sum a_{n\nu}^{m\mu}\begin{pmatrix}-\mu&t\\\nu&s\end{pmatrix}_1 A_s^t,$$
$$\sum=\sum_{\nu\mu}\sum_{st},\quad\sum_{\nu\mu}=\sum_{\nu=0}^{\infty}\sum_{\mu=-\nu}^{\nu}, \tag{115}$$

which is the full analog of (1 : 150). Similarly $\tilde{\mathbf{g}}$ of (75) involves only either CC or BB, and $\mathfrak{G}$ of (76) only C or B. The isolated scattering coefficients $b_{n\nu}^{m\mu}$ or $c_{n\nu}^{m\mu}$ satisfy

$$a_{n\nu}^{m\mu}=a_{\nu n}^{-\mu-m}=a_{n\nu}^{-m-\mu}D_n^m/D_\nu^\mu,\quad a_{n\nu}^{m-\mu}=a_{n\nu}^{-m\mu}D_n^mD_\nu^\mu,$$
$$D_n^m=(-1)^m(n-m)!/(n+m)!, \tag{116}$$

where $n-\nu$ and $m-\mu$ are even. These relations follow from reciprocity and inversion symmetry, i.e., from $\tilde{\mathbf{g}}(-\hat{\mathbf{r}},-\hat{\mathbf{k}})=\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})$, and from reflection symmetry. If we reflect in $z=0$, we replace $\hat{\mathbf{r}}=\hat{\mathbf{r}}(\theta,\varphi)$, $\hat{\mathbf{k}}=\hat{\mathbf{r}}(\beta,\gamma)$ by $\hat{\mathbf{r}}'=\hat{\mathbf{r}}(\pi-\theta,\varphi)$, $\hat{\mathbf{k}}'=\hat{\mathbf{r}}(\pi-\beta,\gamma)$, and require $(\tilde{\mathbf{I}}-2\hat{\mathbf{z}}\hat{\mathbf{z}})\cdot\tilde{\mathbf{g}}(\hat{\mathbf{r}}',\hat{\mathbf{k}}')\cdot(\tilde{\mathbf{I}}-2\hat{\mathbf{z}}\hat{\mathbf{z}})=\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})$; similarly, if we reflect in $y=0$, we replace $\hat{\mathbf{r}},\hat{\mathbf{k}}$ by $\hat{\mathbf{r}}''=\hat{\mathbf{r}}(\theta,-\varphi)$, $\hat{\mathbf{k}}''=\hat{\mathbf{r}}(\beta,-\gamma)$ and require $(\tilde{\mathbf{I}}-2\hat{\mathbf{y}}\hat{\mathbf{y}})\cdot\tilde{\mathbf{g}}(\hat{\mathbf{r}}'',\hat{\mathbf{k}}'')\cdot(\tilde{\mathbf{I}}-2\hat{\mathbf{y}}\hat{\mathbf{y}})=\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})$.

Thus, with $\mathbf{X}_{nm}^e=\mathrm{Re}\mathbf{B}_n^m$, $\mathbf{X}_{nm}^o=\mathrm{Im}\mathbf{B}_n^m$ (or similarly for $\mathbf{C}_n^m$), we proceed as for the scalar case and reduce the scattering dyadic to

$$\tilde{\mathbf{g}}(\hat{\mathbf{r}},\hat{\mathbf{k}})=\sum[a_{n\nu}^{m\mu+}\mathbf{X}_{nm}^e(\hat{\mathbf{r}})\,\mathbf{X}_{nm}^e(\hat{\mathbf{k}})+a_{n\nu}^{m\mu-}\mathbf{X}_{nm}^o\,\mathbf{X}_{\nu\mu}^o],$$
$$\sum=\sum_{nm}\sum_{\nu\mu},\quad\sum_{nm}=\sum_{n=0}^{\infty}\sum_{m=0}^{n},$$
$$a_{n\nu}^{m\mu\pm}=(a_{n\nu}^{m\mu}D_\nu^\mu\pm a_{n\nu}^{m-\mu})\epsilon_m\epsilon_\mu/2,\quad\epsilon_m=2-\delta_{m0}. \tag{117}$$

Similarly, the corresponding multiple scattered amplitude is

$$\mathfrak{G}(\hat{\mathbf{k}}_r)=\sum[A_n^{m+}\mathbf{X}_{nm}^e(\hat{\mathbf{r}})+A_n^{m-}\mathbf{X}_{nm}^o(\hat{\mathbf{r}})],$$
$$A_n^{m\pm}=\begin{Bmatrix}1\\i\end{Bmatrix}(A_n^m\pm A_n^{-m}D_n^m)\epsilon_m/2. \tag{118}$$

Using $X_{nn}^e=P_n^m\cos m\varphi$, $X_{nm}^o=P_n^m\sin m\varphi$ for the corresponding scalar harmonics, we rewrite (115) in terms

of $X^i_{l,\,t\pm\mu}(\hat{\mathbf{K}})$ with $i=e,o$ as

$$-A^{m\mu}_n=\tfrac{1}{2}\sum a^{m\mu\pm}_{n\nu}\left\{A^{t\pm}_s\left[\begin{pmatrix}-\mu\\ \nu\end{pmatrix}\begin{matrix}|\,t\\ |\,s\end{matrix}\Big)^e_1\pm\begin{pmatrix}\mu\\ \nu\end{pmatrix}\begin{matrix}|\,t\\ |\,s\end{matrix}\Big)^e_1\right]+A^{t\mp}_s\left[\pm\begin{pmatrix}-\mu\\ \nu\end{pmatrix}\begin{matrix}|\,t\\ |\,s\end{matrix}\Big)^o_1+\begin{pmatrix}\mu\\ \nu\end{pmatrix}\begin{matrix}|\,t\\ |\,s\end{matrix}\Big)^o_1\right]\right\},$$

$$\begin{pmatrix}\mu\\ \nu\end{pmatrix}\begin{matrix}|\,t\\ |\,s\end{matrix}\Big)^i_1=\sum_l d_{l1}\begin{pmatrix}\mu\\ \nu\end{pmatrix}\begin{matrix};\,t\\ \,\,\,s\end{matrix}\Big)X^i_{l,\,t+\mu}T_l,$$

$$\begin{pmatrix}-\mu\\ \nu\end{pmatrix}\begin{matrix}|\,t\\ |\,s\end{matrix}\Big)^i_1 D^\mu_\nu=\sum d_{l1}\begin{pmatrix}-\mu\\ \nu\end{pmatrix}\begin{matrix};\,t\\ \,\,\,s\end{matrix}\Big)X^i_{l,\,t-\mu}T_l, \tag{119}$$

which differs from (1:164) only in that d_{l1} replaces the earlier d_l.

In particular, if we retain only dipole terms, the initial forms follow by inspection of (1:165), i.e., with $a^{00}_{11}=a_\xi$, $a^{11+}_{11}=a_\xi=a^+_1$, $a^{11-}_{11}=a_\eta=a^-_1$, we have

$$-A^0_1/a_\xi=A^0_1\begin{pmatrix}0|0\\ 1|1\end{pmatrix}_1+A^+_1\begin{pmatrix}0|1\\ 1|1\end{pmatrix}^e_1+A^-_1\begin{pmatrix}0|1\\ 1|1\end{pmatrix}^o_1,$$

$$-A^\pm_1/a^\pm_1=A^0_1\begin{pmatrix}1|0\\ 1|1\end{pmatrix}^e_1+A^\pm_1\tfrac{1}{2}\left[\begin{pmatrix}-1|1\\ 1|1\end{pmatrix}^e_1\pm\begin{pmatrix}1|1\\ 1|1\end{pmatrix}^e_1\right]+A^\mp_1\tfrac{1}{2}\begin{pmatrix}1|1\\ 1|1\end{pmatrix}^o_1, \tag{120}$$

where $A^+_1=A^1_1+A^{-1}_1D^1_1$, $A^-_1=i(A^1_1-A^{-1}_1D^1_1)$ and $D^1_1=-\tfrac{1}{2}$. The present special symbols follow from

$$\begin{pmatrix}0|0\\ 1|1\end{pmatrix}_1=\tfrac{2}{3}(T_0-T_2Y^0_2),\quad\begin{pmatrix}0|1\\ 1|1\end{pmatrix}_1=-\tfrac{1}{3}T_2Y^1_2,$$

$$\begin{pmatrix}-1|1\\ 1|1\end{pmatrix}_1=D^1_1\tfrac{2}{3}(2T_0+T_2Y^0_2),\quad\begin{pmatrix}1|1\\ 1|1\end{pmatrix}_1=-\tfrac{1}{3}Y^2_2T_2. \tag{121}$$

In terms of direction cosines $\alpha_i=\hat{\mathbf{K}}\cdot\hat{\boldsymbol{\zeta}}_i$, we write

$$Y^0_2=\tfrac{1}{2}(3\cos^2\beta-1)=\tfrac{1}{2}(3\alpha_1^2-1),$$

$$Y^1_2=3\sin\beta\cos\beta\exp(i\gamma)=3\alpha_1(\alpha_2+i\alpha_3),$$

$$Y^2_2=3\sin^2\beta\exp(i2\gamma)=3(\alpha_2^2-\alpha_3^2+i2\alpha_2\alpha_3)=3(\alpha_2+i\alpha_3)^2,$$

to obtain

$$-A^0_1/a_\xi=A^0_1\tfrac{2}{3}[T_0-T_2\tfrac{1}{2}(3\alpha_1^2-1)]-A^+_1\alpha_1\alpha_2T_2-A^-_1\alpha_1\alpha_3T_2,$$

$$-A^+_1/a_\xi=-A^0_1\alpha_1\alpha_2T_2+A^+_1\{\tfrac{1}{3}[2T_0+\tfrac{1}{2}T_2(3\alpha_1^2-1)]-\tfrac{1}{2}(\alpha_2^2-\alpha_3^2)\,T_2\}-A^-_1\alpha_2\alpha_3T_2,$$

$$-A^-_1/a_\eta=-A^0_1\alpha_1\alpha_3T_2-A^+_1\alpha_2\alpha_3+A^-_1\{\tfrac{1}{3}[2T_0+\tfrac{1}{2}(3\alpha_1^2-1)\,T_2]+\tfrac{1}{2}(\alpha_2^2-\alpha_3^2)\,T_2\}.$$

Introducing simpler notation $\mathbf{A}_i$ and a_i (with $i=1,2,3$) for the coefficients, we reduce the system to

$$\mathbf{A}_i\left[\frac{1}{a_i}+\tfrac{2}{3}(T_0-T_2)+T_2\right]=\mathbf{A}_i\left[\frac{1}{\mathfrak{A}_i}+T_2\right]=T_2\alpha_i\sum\mathbf{A}_j\alpha_j=T_2\alpha_i\,\mathbf{A}\cdot\hat{\mathbf{K}},$$

$$\mathbf{A}=\sum\mathbf{A}_i\hat{\boldsymbol{\zeta}}_i,\quad\hat{\mathbf{K}}=\sum\alpha_i\hat{\boldsymbol{\zeta}}_i,$$

$$T_2=\frac{\eta^2}{\Delta}-\mathfrak{H}_2,\quad\Delta=\frac{\eta^2-1}{c},\quad c=\frac{i4\pi\rho}{k^3},$$

$$\mathfrak{A}_i=\frac{a_i}{1-\frac{2}{3}a_i(c+\mathfrak{H}_0-\mathfrak{H}_2)}=\frac{a'_i}{1-\frac{2}{3}a'_i(c+\mathfrak{H}_0+1-\mathfrak{H}_2)}, \tag{122}$$

where $a'_i\equiv a_i/(1+\frac{2}{3}a_i)$. If we multiply through by $\hat{\boldsymbol{\zeta}}_i/T_2$ and sum over i, then in terms of $\tilde{\mathfrak{A}}=\sum\mathfrak{A}_i\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$,

$$\sum\left(\frac{1}{\mathfrak{A}_iT_2}+1\right)\mathbf{A}_i\hat{\boldsymbol{\zeta}}_i=[(\tilde{\mathfrak{A}}T_2)^{-1}+\tilde{\mathbf{I}}]\cdot\mathbf{A}=(\tilde{\mathbf{I}}+\tilde{\mathfrak{A}}T_2)\cdot\tilde{\mathfrak{A}}^{-1}\cdot\mathbf{A}/T_2=\hat{\mathbf{K}}\hat{\mathbf{K}}\cdot\mathbf{A}, \tag{123}$$

$$(\tilde{\mathbf{I}}-\hat{\mathbf{K}}\hat{\mathbf{K}})\cdot\mathbf{A}=\tilde{\mathbf{I}}_T\cdot\mathbf{A}=-(\tilde{\mathfrak{A}}T_2)^{-1}\cdot\mathbf{A},\quad\hat{\mathbf{K}}\cdot\tilde{\mathfrak{A}}^{-1}\cdot\mathbf{A}=0. \tag{124}$$

From the last equality in (123) we form $\tilde{\mathfrak{A}}^{-1}\cdot\mathbf{A}/T_2=(\tilde{\mathbf{I}}+\tilde{\mathfrak{A}}T_2)^{-1}\cdot\hat{\mathbf{K}}(\hat{\mathbf{K}}\cdot\mathbf{A})$, and use (124) to obtain

$$\hat{\mathbf{K}}\cdot(\tilde{\mathbf{I}}+\tilde{\mathfrak{A}}T_2)^{-1}\cdot\hat{\mathbf{K}}=0,$$

$$\sum\frac{\alpha_i^2}{1+\mathfrak{A}_iT_2}=0,\quad T_2=\frac{\eta^2c}{\eta^2-1}-\mathfrak{H}_2, \tag{125}$$

where the sum form is essentially a generalization of Fresnel's equation for wave normals in anisotropic media. The form $\hat{\mathbf{K}}\cdot\tilde{\mathbf{N}}^{-1}\cdot\hat{\mathbf{K}}=0$ with $\tilde{\mathbf{N}}=\tilde{\mathbf{I}}+\tilde{\mathfrak{A}}T_2$ is equivalent to $|\tilde{\mathbf{I}}_T\cdot\tilde{\mathbf{N}}\cdot\tilde{\mathbf{I}}_T|=0$ obtained directly from (124).

For small k we use $H_0+1\approx W$ and drop H_2. Then

$$\sum\frac{\alpha_i^2}{\eta^2(1+c\,\mathfrak{A}_i)-1}=0,$$

$$\mathfrak{A}_i\approx\frac{a''_i}{1-\frac{2}{3}a''_ic},\quad a''_i=\frac{a'_i}{1-\frac{2}{3}a'_iW}, \tag{126}$$

where $a''_i\approx a'_i(1+\frac{2}{3}a'_iW)$ includes the reduction in radiation losses compared to the isolated value $a_i=a'_i/(1-\frac{2}{3}a'_i)\approx a'_i(1+\frac{2}{3}a'_i)$; to lowest order in k, $a'_i=O(k^3)$ is imaginary for lossless scatterers. For electric dipoles $\tilde{\mathfrak{A}}=\tilde{\mathfrak{A}}(\tilde{b}'')$, we write

$$\sum\frac{\alpha_i^2}{\eta^{-2}-\epsilon_i^{-1}}=0,\quad\tilde{\boldsymbol{\epsilon}}=\sum\epsilon_i\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i=[\tilde{\mathbf{I}}+c\,\tilde{\mathfrak{A}}(\mathfrak{b}'')]^{-1}. \tag{127}$$

If $k\to0$, or if radiation losses are negligible, then for lossless scatterers (127) reduces to Fresnel's equation, and the standard optical development is fully applicable. We may also obtain $\tilde{\boldsymbol{\epsilon}}$ directly from (124) in the form

$$\tilde{\mathbf{I}}_T\cdot[\tilde{\mathbf{I}}+(c\,\tilde{\mathfrak{A}})^{-1}(\eta^2-1)/\eta^2]\cdot\mathbf{A}=\tilde{\mathbf{I}}_T\cdot\{[\tilde{\mathbf{I}}+(c\,\tilde{\mathfrak{A}})^{-1}](\eta^2-1)+\tilde{\mathbf{I}}\}\cdot\mathbf{A}/\eta^2=0 \tag{128}$$

by comparison with the corresponding form in (7) for $\mathbf{L}=\mathbf{P}$, $\mathbf{L}^M=0$:

$$\tilde{\mathbf{I}}_T\cdot[-(\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}})^{-1}(\eta^2-1)+\tilde{\mathbf{I}}]\cdot\mathbf{P}=0. \tag{129}$$

Thus $\tilde{\mathbf{I}}+(c\,\tilde{\mathfrak{A}})^{-1}=(1-\tilde{\boldsymbol{\epsilon}})^{-1}$, from which $\tilde{\boldsymbol{\epsilon}}=(I+c\,\tilde{\mathfrak{A}})^{-1}$ as in (127). Similarly for magnetic dipoles $\tilde{\mathfrak{A}}(\tilde{\mathfrak{C}}'')$, we use $\mathbf{L}=\mathbf{M}$ and obtain $\tilde{\boldsymbol{\mu}}=[\tilde{\mathbf{I}}+c\,\tilde{\mathfrak{A}}(\tilde{\mathfrak{C}}'')]^{-1}$.

If the scatters are ellipsoids with parameter ϵ', or $\tilde{\boldsymbol{\epsilon}}'=\sum\epsilon'_i\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$ for greater generality, then to order k^3,

$$b'_i=b'_i(\epsilon'_i)=\frac{ik^3\mathfrak{V}}{4\pi}\left[\frac{\epsilon'_i-1}{1+q'_i(\epsilon'_i-1)}\right]=\frac{ik^3\mathfrak{V}}{4\pi}\,\mathfrak{N}_i(\epsilon'_i),\quad\sum q'_i=1, \tag{130}$$

with $\mathfrak{V}$ as the ellipsoid's volume and q'_i as the usual elliptic integral depolarization factor. Thus, in the low-frequency limit,

$$\epsilon_i^{-1} = 1 + c\,\mathfrak{A}_i(b_i') \to 1 - \frac{\rho \mathfrak{B}\mathfrak{K}_i(\epsilon_i')}{1 + \frac{2}{3}\rho \mathfrak{B}\mathfrak{K}_i(\epsilon_i')} = \epsilon_i^{-1}(\epsilon_i'), \tag{131}$$

which reduces to Maxwell's form for spheres if $q_i' = \frac{1}{3}$. Similarly for the magnetic case, $c_i' = b_i'(\mu_i')$ and $\mu_i^{-1} = \epsilon_i^{-1}(\mu_i')$ in terms of the forms in (130), (131). To include radiation losses and the packing effects for spherically symmetric correlations, we use (126) in terms of $b_i''(b_i')$ as in (130), and approximate W by[12]

$$W \approx \frac{(1-w)^4}{(1+2w)^2}, \quad w = \tfrac{4}{3}\pi\left(\frac{\mathfrak{b}}{2}\right)^3 \tag{132}$$

with $\mathfrak{b}$ as the radius of the exclusion sphere (i.e., $\mathfrak{b}$ is the minmum separation of scatterer centers).

To obtain forms for η^2 with $\mathfrak{H}_2$ retained, we rewrite (123) and (124) as

$$[\tilde{\mathbf{I}}_T \cdot \tilde{\mathfrak{A}} T_2 + \tilde{\mathbf{I}}] \cdot (\tilde{\mathfrak{A}} T_2)^{-1} \cdot \mathbf{A} = 0, \quad \tilde{\mathbf{I}}_T \cdot (\tilde{\mathbf{I}} + T_2 \tilde{\mathfrak{A}}) \cdot \tilde{\mathbf{I}}_T \cdot \mathbf{A} = 0,$$
$$I_T = I - \hat{\mathbf{K}}\hat{\mathbf{K}} = \hat{\beta}\hat{\beta} + \hat{\gamma}\hat{\gamma}, \tag{133}$$

where $\mathbf{A}_T \propto \tilde{\mathfrak{A}}^{-1} \cdot \mathbf{A}$ corresponds to A of (6), e.g., for $\mathbf{A} \propto \mathbf{P}$, we have $\mathbf{A} = \mathbf{E}$. With $\tilde{\mathfrak{A}}$ as a dipole dyadic, $\tilde{\mathbf{I}}_T \cdot \tilde{\mathfrak{A}} \cdot \tilde{\mathbf{I}}_T = \tilde{\mathbf{g}}(\tilde{\mathfrak{A}}) = \tilde{\mathbf{g}}_t$ represents a fully transverse forward scattered electromagnetic dyadic amplitude. Thus we may write $[\tilde{\mathbf{I}}_T + (\eta^2\Delta^{-1} - \mathfrak{H}_2)\tilde{\mathbf{g}}_t] \cdot \mathbf{A} = 0$, from which we factor $\Delta^{-1}(\tilde{\mathbf{I}} - \mathfrak{H}_2 \tilde{\mathbf{g}}_t) \cdot$ to obtain

$$[\tilde{\mathbf{I}}_T \Delta + \eta^2 (\tilde{\mathbf{I}} - \mathfrak{H}_2 \tilde{\mathbf{g}})^{-1} \cdot \tilde{\mathbf{g}}_t] \cdot \mathbf{A}$$
$$= [\tilde{\mathbf{I}}_T \Delta + \tilde{\mathbf{g}}[\![\mathbf{K}|\mathbf{K}]\!]] \cdot \mathbf{A} = 0, \quad \Delta = (\eta^2 - 1)/c \tag{134}$$

and thereby isolate $\tilde{\mathbf{g}}[\![\mathbf{K}|\mathbf{K}]\!] = \tilde{\mathbf{g}}_T$ as in (52). For spherical symmetry $(\tilde{\mathbf{I}} - \mathfrak{H}_2 \tilde{\mathfrak{A}})^{-1} \cdot \tilde{\mathfrak{A}}$ with $\tilde{\mathfrak{A}} = \mathfrak{A}\tilde{\mathbf{I}}$ and $\mathfrak{A}$ in the form (122), reduces to $\tilde{\mathbf{I}}\mathfrak{A}/(1 - \mathfrak{A}\mathfrak{H}_2) = \mathfrak{A}_1 \tilde{\mathbf{I}}$ where $\mathfrak{A}_1$ is the form (89) for a spherical dipole.

We reduce the sum form in (125) by a different procedure to facilitate subsequent developments. Clearing the denominator yields

$$1 + T_2 \sum (1 - \alpha_i^2)\mathfrak{A}_i + T_2^2 \sum \alpha_i^2 \mathfrak{A}_{i+1}\mathfrak{A}_{i+2}$$
$$\equiv 1 + T_2 S(\tilde{\mathfrak{A}}) + T_2^2 D(\tilde{\mathfrak{A}}) = 0, \tag{135}$$

where i, $i+1$, $i+2$ follow cyclically. Thus,

$$\Delta^2(1 - \mathfrak{H}_2 S + \mathfrak{H}_2^2 D) + \Delta\eta^2(S - 2\mathfrak{H}_2 D) + \eta^4 D = 0, \tag{136}$$

which, to lowest order in k, is a quadratic for $(1 - \eta^{-2})$.

To display the results (135) and (136) in terms of scattering amplitudes, for an isolated electric dipole $(\tilde{\mathbf{b}})$ we introduce[2]

$$\tilde{\mathbf{g}}^b(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = (\tilde{\mathbf{I}} - \hat{\mathbf{r}}\hat{\mathbf{r}}) \cdot \tilde{\mathbf{b}} \cdot (\tilde{\mathbf{I}} - \hat{\mathbf{K}}\hat{\mathbf{K}}) = (\hat{\theta}\hat{\theta} + \hat{\varphi}\hat{\varphi}) \cdot \tilde{\mathbf{b}} \cdot (\hat{\beta}\hat{\beta} + \hat{\gamma}\hat{\gamma}), \tag{137}$$

and for a magnetic dipole $(\tilde{\mathbf{c}})$,

$$\tilde{\mathbf{g}}^c(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = -(\tilde{\mathbf{I}} \times \hat{\mathbf{r}}) \cdot \tilde{\mathbf{c}} \cdot (\tilde{\mathbf{I}} \times \hat{\mathbf{K}})$$
$$= -(-\hat{\theta}\hat{\varphi} + \hat{\varphi}\hat{\theta}) \cdot \tilde{\mathbf{c}} \cdot (-\hat{\beta}\hat{\gamma} + \hat{\gamma}\hat{\beta}). \tag{138}$$

In the forward direction, $\hat{\mathbf{r}} = \hat{\mathbf{K}}$,

$$\tilde{\mathbf{g}}(\hat{\mathbf{K}}, \hat{\mathbf{K}}) = \hat{\beta}\hat{\beta} g_{\beta\beta} + (\hat{\beta}\hat{\gamma} + \hat{\gamma}\hat{\beta}) g_{\beta\gamma} + \hat{\gamma}\hat{\gamma} g_{\gamma\gamma}$$
$$= I_T(\hat{\mathbf{K}}) \cdot \tilde{\mathbf{g}} \cdot I_T(\hat{\mathbf{K}}), \tag{139}$$

where, e.g., $g_{\beta\beta}^b = \hat{\beta} \cdot \tilde{\mathbf{g}}^b \cdot \hat{\beta} = \hat{\beta} \cdot \tilde{\mathbf{b}} \cdot \hat{\beta}$ but $g_{\beta\beta}^c = \hat{\beta} \cdot \tilde{\mathbf{g}}^c \cdot \hat{\beta}$ $= \hat{\gamma} \cdot \tilde{\mathbf{c}} \cdot \hat{\gamma}$, etc. With $\hat{\mathbf{K}} = \sum \alpha_i \hat{\boldsymbol{\xi}}_i$, we have

$$\hat{\beta} = [-(1 - \alpha_1^2)\hat{\boldsymbol{\xi}}_1 + \alpha_1\alpha_2 \hat{\boldsymbol{\xi}}_2 + \alpha_1\alpha_3 \hat{\boldsymbol{\xi}}_3]/(1 - \alpha_1^2)^{1/2}$$
$$\hat{\gamma} = (-\alpha_3 \hat{\boldsymbol{\xi}}_2 + \alpha_2 \hat{\boldsymbol{\xi}}_3)/(1 - \alpha_1^2)^{1/2},$$

and consequently

$$\hat{\beta} \cdot \tilde{\mathbf{a}} \cdot \hat{\beta} + \hat{\gamma} \cdot \tilde{\mathbf{a}} \cdot \hat{\gamma} = \sum (1 - \alpha_i^2) a_i = g_{\beta\beta} + g_{\gamma\gamma} = S(\tilde{\mathbf{a}}),$$
$$(\hat{\beta} \cdot \tilde{\mathbf{a}} \cdot \hat{\beta})(\hat{\gamma} \cdot \tilde{\mathbf{a}} \cdot \hat{\gamma}) - (\hat{\beta} \cdot \tilde{\mathbf{a}} \cdot \hat{\gamma})^2 = \sum \alpha_i^2 a_{i+1} a_{i+2} = g_{\beta\beta} g_{\gamma\gamma} - g_{\beta\gamma}^2$$
$$= |\tilde{\mathbf{g}}| = D(\tilde{\mathbf{a}}) \tag{140}$$

with $\tilde{\mathbf{a}} = \tilde{\mathbf{b}}$ or $\tilde{\mathbf{c}}$, and $\tilde{\mathbf{g}} = \tilde{\mathbf{g}}^b$ or $\tilde{\mathbf{g}}^c$.

Thus (135) may be rewritten in terms of $\tilde{\mathbf{g}}(\tilde{\mathfrak{A}}) = \tilde{\mathbf{g}}_t = \tilde{\mathbf{g}}$ as

$$\Delta^2 |\tilde{\mathbf{I}}_T - \mathfrak{H}_2 \tilde{\mathbf{g}}| + \Delta\eta^2 [g_{\beta\beta}(1 - \mathfrak{H}_2 g_{\gamma\gamma}) + g_{\gamma\gamma}(1 - \mathfrak{H}_2 g_{\beta\beta})$$
$$+ 2\mathfrak{H}_2 g_{\beta\gamma}^2] + \eta^4 |\tilde{\mathbf{g}}| = 0,$$
$$|\tilde{\mathbf{I}}_T \Delta + (\eta^2 - \Delta\mathfrak{H}_2)\tilde{\mathbf{g}}_t| = 0, \tag{141}$$

or equivalently as

$$|\tilde{\mathbf{I}}_T \Delta + \eta^2 (1 - \mathfrak{H}_2 \tilde{\mathbf{g}}_t)^{-1} \cdot \tilde{\mathbf{g}}_t| = 0 \tag{142}$$

as obtained directly from (134).

In general, the low frequency result is a quadratic for η^{-2}. If $\hat{\eta} = \hat{\mathbf{y}}$ and one of the axes of the ellipsoid is perpendicular to the plane of incidence (z, x or ζ, ξ) for the slab distribution discussed for (30)ff, then with $\hat{\mathbf{k}} = \hat{\mathbf{z}}\cos\alpha_0 + \hat{\mathbf{x}}\sin\alpha_0$, we have $\hat{\mathbf{K}} = \hat{\mathbf{z}}\cos\beta_0 + \hat{\mathbf{x}}\sin\beta_0$ $= \alpha_1 \hat{\boldsymbol{\xi}}_1 + \alpha_2 \hat{\boldsymbol{\xi}}_2$, i.e., i.e., $\alpha_3 = 0$ and the corresponding polarization vectors are $\hat{\beta} = -\alpha_2 \hat{\boldsymbol{\xi}} + \alpha_1 \hat{\boldsymbol{\xi}}$, $\hat{\gamma} = \hat{\mathbf{y}}$. Thus $g_{\beta\gamma} = 0$ and (135) and (141) reduce to the uncoupled form

$$[1 + T_2(\alpha_2^2 \mathfrak{A}_1 + \alpha_1^2 \mathfrak{A}_2)](1 + T_2 \mathfrak{A}_3)$$
$$= (1 + T_2 g_{\beta\beta})(1 + T_2 g_{yy}) = 0. \tag{143}$$

For the wave polarized along $\hat{\mathbf{y}}$, i.e., for $\mathbf{E} = E\hat{\mathbf{y}}$, $\tilde{\mathfrak{A}} = \tilde{\mathfrak{A}}(\tilde{\mathbf{b}})$,

$$\frac{1}{\eta^2} = 1 + \frac{c\mathfrak{A}_3}{1 - \mathfrak{H}_2 \mathfrak{A}_3} = \frac{1}{\epsilon_3}. \tag{144}$$

For the other case, we have essentially the same structure as in (1:120)ff with $\hat{\mathbf{K}} \cdot \tilde{\mathbf{B}} \cdot \hat{\mathbf{K}}$ replaced by $\hat{\beta} \cdot \tilde{\epsilon}^{-1} \cdot \hat{\beta}$, i.e., with $\alpha_1^2 + \alpha_2^2 = 1$, $\mathbf{E} = \mathbf{E}(\hat{\mathbf{z}}, \hat{\mathbf{x}})$,

$$\frac{1}{\eta^2} = 1 + \frac{c(\alpha_2^2 \mathfrak{A}_1 + \alpha_1^2 \mathfrak{A}_2)}{D}$$
$$= \alpha_2^2 \left(1 + \frac{c\mathfrak{A}_1}{D}\right) + \alpha_1^2 \left(1 + \frac{c\mathfrak{A}_2}{D}\right) = \hat{\beta} \cdot \tilde{\epsilon}^{-1} \cdot \hat{\beta},$$
$$D = 1 - \mathfrak{H}_2(\alpha_2^2 \mathfrak{A}_1 + \alpha_1^2 \mathfrak{A}_2). \tag{145}$$

Proceeding as before, we obtain the same forms for η^2 and $K\cos\beta_0$ as in (1:123)ff with $B_{ij} = \mathbf{z}_i \cdot \tilde{\mathbf{B}} \cdot \hat{\mathbf{z}}_0$ replaced by $\epsilon_{ij}/|\tilde{\epsilon}_\rho| = \epsilon_{ij}/(\epsilon_{11}\epsilon_{22} - \epsilon_{12}^2)$, $\hat{\mathbf{z}}_1 = \hat{\mathbf{z}}$, $\hat{\mathbf{z}}_2 = \hat{\mathbf{x}}$.

If each scatterer has both electric and magnetic properties, we work with the coupled equations for the C's and B's as in (81). When symmetrized for ellipsoids each set contains term analogous to (119) as well as cross terms. We write the corresponding isolated scattering amplitude as

$$\tilde{\mathbf{g}}=\tilde{\mathbf{g}}_b+\tilde{\mathbf{g}}_c+\tilde{\mathbf{g}}_x,$$

$$\tilde{\mathbf{g}}_x=\sum\sum(\beta_{n\nu}^{m\mu}\mathbf{C}_n^m\mathbf{B}_\nu^{-\mu}+\gamma_{n\nu}^{m\mu}\mathbf{B}_n^m\mathbf{C}_\nu^{-\mu}), \tag{146}$$

where $\tilde{\mathbf{g}}_b$ is the form (117) in terms of b and $\mathbf{B}$, and $\tilde{\mathbf{g}}_c$ involves $c,\mathbf{C}$. Reciprocity and inversion symmetry require $\beta_{n\nu}^{m\mu}(-1)^{n+\nu+1}=\beta_{n\nu}^{m\mu}$ and similarly for $\gamma_{n\nu}^{m\mu}$; thus $n-\nu$ is odd. From reflection symmetry in $z=0$, it follows that $m+\mu$ is even, and from reflection symmetry in $y=0$, we have $\beta_{n\nu}^{m\mu}=-\beta_{n\nu}^{-m-\mu}D_n^m/D_\nu^\mu$, $\beta_{n\nu}^{m-\mu}=-\beta_{n\nu}^{-m\mu}D_n^mD_\nu^\mu$. Using $\beta^\pm$ and $\gamma^\pm$ for the same form as $a^\pm$ of (117), we obtain

$$\begin{aligned}\tilde{\mathbf{g}}_x=i\sum(&\beta_{n\nu}^{m\mu+}\,\mathrm{Im}\mathbf{C}_n^m\,\mathrm{Re}\mathbf{B}_\nu^\mu-\beta^-\,\mathrm{Re}\mathbf{C}\,\mathrm{Im}\mathbf{B})\\&+i\sum(\gamma^+\,\mathrm{Im}\mathbf{B}\,\mathrm{Re}\mathbf{C}-\gamma^-\,\mathrm{Re}\mathbf{B}\,\mathrm{Im}\mathbf{C}),\end{aligned} \tag{147}$$

which excludes $\beta_{n\nu}^{00}$ and $\gamma_{n\nu}^{00}$ (because $\mathbf{C}_n^0$ and $\mathbf{B}_n^0$ are real). Thus the leading terms involve β_{12}^{02}, β_{12}^{11}, etc., i.e., there are no pure dipole effects.

If the individual scatterers are characterized solely by $\alpha=c$ and $\delta=b$ coefficients, the appropriate form of (81) is

$$\begin{aligned}-C_n^m&=\sum c_{n\nu}^{m\mu}\left[\mathfrak{E}_1\begin{pmatrix}-\mu & t\\ \nu & s\end{pmatrix}C_s^t-\mathfrak{E}_2B_s^t\right],\\ -B_n^m&=\sum b_{n\nu}^{m\mu}\left[\mathfrak{E}_1B_s^t+\mathfrak{E}_2C_s^t\right],\quad \mathfrak{E}_t=\begin{pmatrix}-\mu & t\\ \nu & s\end{pmatrix}_t.\end{aligned} \tag{148}$$

Retaining only dipole terms $n=\nu=1$, we generalize (120) by using the values of the symbols in (121) and the additional values

$$\begin{aligned}&\begin{pmatrix}0&1\\1&1\end{pmatrix}_2=T_1(-iY_1^1),\\ &\begin{pmatrix}0&-1\\1&1\end{pmatrix}_2=D_1^1T_1(-iY_1^1)^*,\quad \begin{pmatrix}1&-1\\1&1\end{pmatrix}_2=T_1iP_1^0\end{aligned} \tag{149}$$

with $Y_1^1=P_1^1\exp(i\gamma)=\sin\beta\exp(i\gamma)=\alpha_2+i\alpha_3$, and $P_1^0=\cos\beta=\alpha_1$. Thus with (120) representing C_1^m, the equation for $-C_1^0/c_\epsilon$ has the additional terms

$$\begin{aligned}-B_1^1\begin{pmatrix}0&1\\1&1\end{pmatrix}_2-B_1^{-1}\begin{pmatrix}0&-1\\1&1\end{pmatrix}_2&=-\left[B_1^+\begin{pmatrix}0&1\\1&1\end{pmatrix}_2^0-B_1^-\begin{pmatrix}0&1\\1&1\end{pmatrix}_2^e\right]\\&=-(B_1^+\alpha_3-B_1^-\alpha_2)T_2\end{aligned} \tag{150}$$

and that for $-C_1^\pm/c_1^\pm$ now includes

$$\begin{aligned}&-\frac{1}{2}\begin{Bmatrix}1\\i\end{Bmatrix}\left\{\frac{1}{D_1^1}\left[B_1^0\begin{pmatrix}-1&0\\1&1\end{pmatrix}_2+B_1^1\begin{pmatrix}-1&1\\1&1\end{pmatrix}_2\right]\right.\\&\qquad\left.\pm\left[B_1^0\begin{pmatrix}1&0\\1&1\end{pmatrix}_2+B_1^{-1}\begin{pmatrix}1&-1\\1&1\end{pmatrix}_2\right]\right\}\\&=-\frac{1}{2}\begin{Bmatrix}1\\i\end{Bmatrix}T_1\{B_1^0[(iY_1^1)^*\pm iY_1^1]+2\alpha_1T_1i(B_1^1\mp B_1^{-1}D_1^1)\}\\&=-(\mp B_1^0\alpha_{\frac{3}{2}}\pm B_1^\mp\alpha_1)T_1.\end{aligned} \tag{151}$$

The analogs for B/b contain the corresponding terms of (150) and (151) with B replaced by $-C$.

Thus, in terms of $\mathbb{C}_t=C_1^0,C_1^+,C_1^-$ and $\mathbb{B}_t=B_1^0,B_1^+,B_1^-$, etc., as before, we obtain the generalization of (122)

$$\begin{aligned}&\mathbb{C}_t(\mathbb{C}_t^{-1}+T_2)-T_2\alpha_t\sum\mathbb{C}_j\alpha_j-(\mathbb{B}_{t+1}\alpha_{t+2}-\mathbb{B}_{t+2}\alpha_{t+1})T_1=0,\\&\mathbb{B}_t(\mathfrak{B}_t^{-1}+T_2)-T_2\alpha_t\sum\mathbb{B}_j\alpha_j+(\mathbb{C}_{t+1}\alpha_{t+2}-\mathbb{C}_{t+2}\alpha_{t+1})T_1=0,\end{aligned} \tag{152}$$

where $\mathbb{C}_t(c_t)$ and $\mathfrak{B}_t(b_t)$ correspond to the form $\mathfrak{A}_t(a_t)$ of (122). Multiplying through by $\hat{\boldsymbol{\xi}}_t$ and introducing $\tilde{\mathbb{C}}=\sum\mathbb{C}_t\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t$, $\tilde{\mathfrak{B}}=\sum\mathfrak{B}_t\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t$, and $\mathbb{C}=\sum\mathbb{C}_t\hat{\boldsymbol{\xi}}_t$, $\mathbb{B}=\sum\mathbb{B}_t\hat{\boldsymbol{\xi}}_t$, we obtain the generalization of (124),

$$\begin{aligned}&\tilde{\mathbb{C}}^{-1}\cdot\mathbb{C}+T_2(\tilde{\mathbf{I}}-\hat{\mathbf{K}}\hat{\mathbf{K}})\cdot\mathbb{C}+T_1(\tilde{\mathbf{I}}\times\hat{\mathbf{K}})\cdot\mathbb{B}=0,\quad \hat{\mathbf{K}}\cdot\tilde{\mathbb{C}}^{-1}\cdot\mathbb{C}=0,\\&\tilde{\mathfrak{B}}^{-1}\cdot\mathbb{B}+T_2(\tilde{\mathbf{I}}-\hat{\mathbf{K}}\hat{\mathbf{K}})\cdot\mathbb{B}-T_1(\tilde{\mathbf{I}}\times\hat{\mathbf{K}})\cdot\mathbb{C}=0,\quad \hat{\mathbf{K}}\cdot\tilde{\mathbb{B}}^{-1}\cdot\mathbb{B}=0,\end{aligned} \tag{153}$$

where the T_1 terms couple the electric and magnetic effects.

At low frequencies with $\tilde{\mathbb{C}}(c'')$ and $\tilde{\mathfrak{B}}(b'')$ in the form $\tilde{\mathfrak{A}}(a'')$ as in (126), we drop $\mathfrak{G}_2$ and $\mathfrak{G}_1$ to obtain

$$\begin{aligned}&(\eta^2-1)(c\tilde{\mathbb{C}})^{-1}\cdot\mathbb{C}+\eta^2\mathbb{C}_T+\eta\hat{\mathbf{K}}\times\mathbb{B}=0,\\&(\eta^2-1)(c\tilde{\mathfrak{B}})^{-1}\cdot\mathbb{B}+\eta^2\mathbb{B}_T-\eta\hat{\mathbf{K}}\times\mathbb{C}=0,\\&\mathbb{C}_T=(\tilde{\mathbf{I}}-\hat{\mathbf{K}}\hat{\mathbf{K}})\cdot\mathbb{C}=\tilde{\mathbf{I}}_T\cdot\mathbb{C},\quad \mathbb{B}_T=\tilde{\mathbf{I}}_T\cdot\mathbb{B}.\end{aligned} \tag{154}$$

Equivalently,

$$\begin{aligned}&(\eta^2-1)\tilde{\mathbf{I}}_T\cdot[(c\tilde{\mathbb{C}})^{-1}+\tilde{\mathbf{I}}]\cdot\mathbb{C}+\mathbb{C}_T+\eta\hat{\mathbf{K}}\times\mathbb{B}=0,\\&(\eta^2-1)\tilde{\mathbf{I}}_T\cdot[(c\tilde{\mathfrak{B}})^{-1}+\tilde{\mathbf{I}}]\cdot\mathbb{B}+\mathbb{B}_T-\eta\hat{\mathbf{K}}\times\mathbb{C}=0,\end{aligned} \tag{155}$$

which corresponds to the $\mathbf{L},\mathbf{L}^M$ form in (7) with $\mathbb{B}=\mathbf{P}=(\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}})\cdot\mathbf{E}$, $\mathbb{C}=\mathbf{M}=(\tilde{\boldsymbol{\mu}}-\tilde{\mathbf{I}})\cdot\mathbf{H}$ as the polarizabilities, and the parameters specified by

$$\tilde{\boldsymbol{\epsilon}}=(\tilde{\mathbf{I}}+c\tilde{\mathfrak{B}})^{-1},\quad \tilde{\boldsymbol{\mu}}=(\tilde{\mathbf{I}}+c\tilde{\mathbb{C}})^{-1}. \tag{156}$$

Thus corresponding to (6) in the form $(\eta^2\hat{\mathbf{K}}\times\tilde{\boldsymbol{\mu}}^{-1}\times\hat{\mathbf{K}}+\tilde{\mathbf{I}}_T\cdot\tilde{\boldsymbol{\epsilon}})\cdot\mathbf{E}=0$, we have

$$|\eta^2[\hat{\mathbf{K}}\times(\tilde{\mathbf{I}}+c\tilde{\mathbb{C}})\times\hat{\mathbf{K}}]\cdot(\tilde{\mathbf{I}}+c\tilde{\mathfrak{B}})+\tilde{\mathbf{I}}_T|=0 \tag{157}$$

in terms of a 2×2 determinant in the plane transverse to $\hat{\mathbf{K}}$.

Reverting to (153), we clear the inverses to obtain the form

$$\begin{aligned}&(\tilde{\mathbf{I}}+T_2\tilde{\mathfrak{A}}\cdot\tilde{\mathbf{I}}_T)\cdot\mathbf{L}-T_1\tilde{\mathfrak{A}}\cdot(\hat{\mathbf{K}}\times\tilde{\mathbf{L}}^M)=0,\\&\tilde{\mathfrak{A}}=\begin{Bmatrix}\tilde{\mathfrak{B}}\\\tilde{\mathbb{C}}\end{Bmatrix},\quad \mathbf{L}=\begin{Bmatrix}\mathbb{B}\\\mathbb{C}\end{Bmatrix}=\begin{Bmatrix}\mathbf{P}\\\mathbf{M}\end{Bmatrix},\quad \mathbf{L}^M=\begin{Bmatrix}\mathbf{M}\\-\mathbf{P}\end{Bmatrix},\end{aligned} \tag{158}$$

where if we regard the braces after $\mathbf{L}$ as enclosing a 1×2 column matrix, then $\mathbf{L}^M=S\mathbf{L}$ with $S_{11}=S_{22}=0$, $S_{12}=-S_{21}=1$. Explicitly, we write

$$\begin{aligned}&\mathbb{C}+T_2\tilde{\mathbb{C}}\cdot\tilde{\mathbf{I}}_T\cdot\mathbb{C}+T_1(\tilde{\mathbb{C}}\times\hat{\mathbf{K}})\cdot\mathbb{B}=0,\\&\mathbb{B}+T_2\tilde{\mathfrak{B}}\cdot\tilde{\mathbf{I}}_T\cdot\mathbb{B}-T_1\tilde{\mathfrak{B}}\cdot(\hat{\mathbf{K}}\times\mathbb{C})=0.\end{aligned} \tag{159}$$

Operating on the first equation with $\hat{\mathbf{K}}\times=(\tilde{\mathbf{I}}\times\hat{\mathbf{K}})\cdot$, and on the second with $\tilde{\mathbf{I}}_T=(\tilde{\mathbf{I}}-\hat{\mathbf{K}}\hat{\mathbf{K}})=-(\tilde{\mathbf{I}}\times\hat{\mathbf{K}})\cdot(\tilde{\mathbf{I}}\times\hat{\mathbf{K}})$, we rewrite the result in terms of scattering amplitudes, $\tilde{\mathbf{g}}_m=-\hat{\mathbf{K}}\times\tilde{\mathbb{C}}\times\hat{\mathbf{K}}$, $\tilde{\mathbf{g}}_e=\tilde{\mathbf{I}}_T\cdot\tilde{\mathfrak{B}}\cdot\tilde{\mathbf{I}}_T$ as

$$\begin{aligned}&(\tilde{\mathbf{I}}+T_2\tilde{\mathbf{g}}_m)\cdot(\hat{\mathbf{K}}\times\mathbb{C})-T_1\tilde{\mathbf{g}}_m\cdot\mathbb{B}_T=0,\\&-T_1\tilde{\mathbf{g}}_e\cdot(\hat{\mathbf{K}}\times\mathbb{C})+(\tilde{\mathbf{I}}+T_2\tilde{\mathbf{g}}_e)\cdot\mathbb{B}_T=0.\end{aligned} \tag{160}$$

The low-frequency form of (160) is

$$\begin{aligned}&(\Delta\tilde{\mathbf{I}}_T+\eta^2\tilde{\mathbf{g}}_m)\cdot(\hat{\mathbf{K}}\times\mathbb{C})-\eta\tilde{\mathbf{g}}_m\cdot\mathbb{B}_T=0,\\&-\eta\tilde{\mathbf{g}}_e\cdot(\hat{\mathbf{K}}\times\mathbb{C})+(\Delta\tilde{\mathbf{I}}_T+\eta^2\tilde{\mathbf{g}}_e)\cdot\mathbb{B}_T=0,\end{aligned} \tag{161}$$

and we may eliminate $\mathbf{K}\times\mathbb{C}$,

$$\eta^{-1}\Delta[\Delta\tilde{\mathbf{I}}_T+\eta^2(\tilde{\mathbf{g}}_e+\tilde{\mathbf{g}}_m+c\tilde{\mathbf{g}}_m\cdot\tilde{\mathbf{g}}_e)]\cdot(\tilde{\mathbf{g}}_e^{-1}\cdot\mathbb{B}_T)=0, \tag{162}$$

and identify $\tilde{\mathbf{g}}_T[\![\mathbf{K}|\mathbf{K}]\!]$ as $\eta^2(\tilde{\mathbf{g}}_e+\tilde{\mathbf{g}}_m+c\tilde{\mathbf{g}}_m\cdot\tilde{\mathbf{g}}_e)$. Rewriting (162) as

$$-(\eta\Delta/c)[\tilde{\mathbf{I}}_T\eta^{-2}-(\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_m)\cdot(\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_e)]\cdot(\tilde{\mathbf{g}}_e^{-1}\cdot\mathbf{B}_T)=0, \tag{163}$$

we obtain the dispersion equation in the form

$$|\tilde{\mathbf{I}}_T\eta^{-2}-(\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_m)\cdot(\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_e)|=|\tilde{\mathbf{I}}_T\eta^{-2}+(\hat{\mathbf{K}}\times\mu^{-1}\times\hat{\mathbf{K}})\cdot\tilde{\epsilon}^{-1}|=0. \tag{164}$$

To include the $\mathfrak{H}_i$ terms of T_i, we may recast (160) in tensor versions of the forms that led to (95) and (100). Thus with $\tilde{\mathfrak{G}}_\tau=(\tilde{\mathbf{I}}-\mathfrak{H}_2\tilde{\mathbf{g}}_\tau)^{-1}\cdot\tilde{\mathbf{g}}_\tau$ for $\tau=m,e$, we eliminate $\hat{\mathbf{K}}\times\mathbb{C}$ to obtain the analog of (95),

$$|\Delta\tilde{\mathbf{I}}_T+\eta^2(1-\mathfrak{H}_1^2\tilde{\mathfrak{G}}_m\cdot\tilde{\mathfrak{G}}_e)^{-1}\cdot[\tilde{\mathfrak{G}}_e+\tilde{\mathfrak{G}}_m+\tilde{\mathfrak{G}}_m\cdot\tilde{\mathfrak{G}}_e(c+2\mathfrak{H}_1/\eta)]|=0. \tag{165}$$

Similarly the analog of (100) in terms of $\tilde{\Gamma}_\tau=\eta^2\tilde{\mathfrak{G}}_\tau$ is

$$|\Delta\tilde{\mathbf{I}}_T+(\tilde{\mathbf{I}}-\bar{h}_{11}^2\tilde{\Gamma}_m\cdot\tilde{\Gamma}_e)^{-1}\cdot(\tilde{\Gamma}_m+\tilde{\Gamma}_e+2\bar{h}_{11}\tilde{\Gamma}_m\cdot\tilde{\Gamma}_e)|=0,$$
$$\bar{h}_{11}=\frac{c}{\eta(\eta+1)}+\frac{\mathfrak{H}_1}{\eta^2}. \tag{166}$$

The determinantal equation corresponding to (160), i.e., $|(\tilde{\mathbf{g}}_e^{-1}+T_2\tilde{\mathbf{I}})\cdot(\tilde{\mathbf{g}}_m^{-1}+T_2\tilde{\mathbf{I}})-T_1^2\tilde{\mathbf{I}}|=0$, may be rewritten as

$$(|\tilde{\mathbf{I}}+T_2\tilde{\mathbf{g}}_m|-|T_1\tilde{\mathbf{g}}_m|)(|\tilde{\mathbf{I}}+T_2\tilde{\mathbf{g}}_e|-|T_1\tilde{\mathbf{g}}_e|)+|T_1(\tilde{\mathbf{g}}_e-\tilde{\mathbf{g}}_m)|=0, \tag{167}$$

where, e.g., the function within the initial parenthesis equals $1+T_2(g^m_{\beta\beta}+g^m_{\gamma\gamma})+(T_2^2-T_1^2)|\tilde{\mathbf{g}}_m|$, $|\tilde{\mathbf{g}}_m|=g^m_{\beta\beta}g^m_{\gamma\gamma}-(g^m_{\beta\gamma})^2$. For low frequencies, we replace T_2 and T_1 by $\eta^2/\Delta=c/(1-\eta^{-2})$ and $\eta/\Delta=c/\eta(1-\eta^{-2})$ to obtain

$$[\eta^{-2}-|\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_m|][\eta^{-2}-|\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_e|]+\eta^{-2}c^2|\tilde{\mathbf{g}}_e-\tilde{\mathbf{g}}_m|=0. \tag{168}$$

Thus, (168) or (164) corresponds to a quadratic for η^{-2} with

$$\eta_1^{-2}\eta_2^{-2}=|\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_m||\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_e|,$$
$$\eta_1^{-2}+\eta_2^{-2}=|\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_m|+|\tilde{\mathbf{I}}+c\,\tilde{\mathbf{g}}_e|-c^2|\tilde{\mathbf{g}}_e-\tilde{\mathbf{g}}_m|$$
$$=2+c(g^e_{\beta\beta}+g^e_{\gamma\gamma}+g^m_{\beta\beta}+g^m_{\gamma\gamma})+c^2(g^e_{\beta\beta}g^m_{\beta\beta}+g^e_{\gamma\gamma}g^m_{\gamma\gamma}-2g^e_{\beta\gamma}g^m_{\beta\gamma}). \tag{169}$$

From (167) for $\alpha_3=0$, $\hat{\eta}=\hat{\mathbf{y}}$, for which case $\tilde{\mathbf{g}}^e=\hat{\beta}\hat{\beta}(\mathfrak{B}_1\alpha_2^2+\mathfrak{B}_2\alpha_1^2)+\hat{\mathbf{y}}\hat{\mathbf{y}}\mathfrak{B}_3$ and $\tilde{\mathbf{g}}_m=\hat{\beta}\hat{\beta}\,\mathbb{C}_3+\hat{\mathbf{y}}\hat{\mathbf{y}}(\mathbb{C}_1\alpha_2^2+\mathbb{C}_2\alpha_1^2)$ are diagonal, we obtain two equations of the form

$$(1/g_i^m+T_2)(1/g_i^e+T_2)-T_1^2=0,\quad i=\beta,y,\quad g_\tau=\hat{\mathbf{i}}\cdot\tilde{\mathbf{g}}_\tau\cdot\hat{\mathbf{i}}.$$

Thus,

$$-\frac{(\eta^2-1)}{c}=-\Delta=\eta^2\frac{g_i^e+g_i^m+(c-2\mathfrak{H}_2+2\mathfrak{H}_1/\eta)g_i^eg_i^m}{1-\mathfrak{H}_2(g_i^e+g_i^m)+(\mathfrak{H}_2^2-\mathfrak{H}_1^2)g_i^eg_i^m}$$
$$=\eta^2\frac{\bar{g}_i^e+\bar{g}_i^m+(c+2\mathfrak{H}_1/\eta)\bar{g}_i^e\bar{g}_i^m}{1-\mathfrak{H}_1^2\bar{g}_i^e\bar{g}_i^m}, \tag{170}$$

where the second form with $\bar{g}_\tau=g_\tau/(1-\mathfrak{H}_2g_\tau)$ facilitates comparison with (95), the analog for spherical symmetry. For low frequencies, $\eta_i^{-2}=(1+cg_i^e)(1+cg_i^m)$,

$$\eta_\beta^{-2}=[1+c(\mathfrak{B}_1\alpha_2^2+\mathfrak{B}_2\alpha_1^2)][1+c\mathbb{C}_3]=\hat{\beta}\cdot\tilde{\epsilon}^{-1}\cdot\hat{\beta}/\mu_3$$
$$=\hat{\mathbf{K}}\cdot\tilde{\epsilon}\cdot\hat{\mathbf{K}}/|\tilde{\epsilon}_p|\mu_3,\quad \mathbf{H}=H\hat{\mathbf{y}},$$
$$\eta_y^{-2}=[1+c\mathfrak{B}_3][1+c(\mathbb{C}_1\alpha_2^2+\mathbb{C}_2\alpha_1^2)]=\hat{\beta}\cdot\tilde{\mu}^{-1}\cdot\hat{\beta}/\epsilon_3$$
$$=\hat{\mathbf{K}}\cdot\tilde{\mu}\cdot\hat{\mathbf{K}}/|\tilde{\mu}_p|\epsilon_3,\quad \mathbf{E}=E\hat{\mathbf{y}} \tag{171}$$

where $\tilde{\epsilon}_p=I_T(\hat{\mathbf{y}})\cdot\tilde{\epsilon}\cdot I_T(\hat{\mathbf{y}})$ as indicated for (145). Thus the earlier scalar development (1:128)ff applies for $H\hat{\mathbf{y}}$ with $C=\mu_3$, $\tilde{\mathbf{B}}=\tilde{\epsilon}_p/|\tilde{\epsilon}_p|$ and for $E\hat{\mathbf{y}}$ with $C=\epsilon_3$, $\tilde{\mathbf{B}}=\tilde{\mu}_p/|\tilde{\mu}_p|$.

Nonspherical correlations

For aligned elliptic cylinders and elliptic correlations (with nonconfocal, nonsimilar, and nonparallel scatterer and exclusion surfaces) the earlier[1] scalar development in terms of Mathieu functions also covers the electromagnetic problems for incident H or E parallel to the cylinders axes ($\hat{\mathbf{y}}=\hat{\eta}$) and $\hat{\mathbf{K}}=\alpha_1\hat{\boldsymbol{\zeta}}+\alpha_2\hat{\boldsymbol{\xi}}$. We indicate the essentials for the low frequency case, and then generalize the results to ellipsoids by inspection.

For the two-dimensional scalar problem of the dipole $\tilde{\mathbf{a}}=a_1\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}+a_2\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}$ with semiaxes $\mathfrak{a}_i$ along $\hat{\boldsymbol{\zeta}}_i$ and physical parameter $\tilde{\mathbf{B}}'=\sum B_i'\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$, the isolated forward scattering amplitude to lowest order in k is

$$g(\hat{\mathbf{K}},\hat{\mathbf{K}})=\hat{\mathbf{K}}\cdot\tilde{\mathbf{a}}\cdot\hat{\mathbf{K}}=a_1\alpha_1^2+a_2\alpha_2^2,\quad \hat{\mathbf{K}}=\alpha_1\hat{\boldsymbol{\zeta}}+\alpha_2\hat{\boldsymbol{\xi}},$$
$$a_i\approx a_i'=-\frac{i\pi k^2\mathfrak{a}_1\mathfrak{a}_2(B_i'-1)}{4[1+q_i'(B_i'-1)]}, \tag{172}$$

where the depolarization factors equal $q_1'=\mathfrak{a}_2/(\mathfrak{a}_1+\mathfrak{a}_2)$, $q_2'=\mathfrak{a}_1/(\mathfrak{a}_1+\mathfrak{a}_2)=1-q_1'$. For the analogous electromagnetic problem specified by $\tilde{\epsilon}'=\sum\epsilon_i'\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$ with H parallel to $\hat{\mathbf{y}}$, in terms of $\tilde{b}=b_1\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}+b_2\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}$, we have

$$g^b=\hat{\boldsymbol{\beta}}\cdot\tilde{\mathbf{b}}\cdot\hat{\boldsymbol{\beta}}=b_2\alpha_1^2+b_1\alpha_2^2,\quad \hat{\boldsymbol{\beta}}=-\alpha_2\hat{\boldsymbol{\zeta}}+\alpha_1\hat{\boldsymbol{\xi}},$$
$$b_i\approx b_i'=\frac{i\pi k^2\mathfrak{a}_1\mathfrak{a}_2(\epsilon_i'-1)}{4[1+q_i'(\epsilon_i'-1)]} \tag{173}$$

such that g^b follows from g by using $B_i'=1/\epsilon_i'$, $B_2'=1/\epsilon_1'$ (i.e., $\tilde{B}'=\tilde{\epsilon}_p'/|\tilde{\epsilon}_p'|$ with $\tilde{\epsilon}_p'=\epsilon_1'\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}+\epsilon_2'\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}$) and $q_1'=1-q_2'$, $q_2'=1-q_1'$. Thus, $\alpha_1^2a_1+\alpha_2^2a_2=\alpha_1^2b_2+\alpha_2^2b_1$, and g^b follows from g by replacing a_1 by b_2 and a_2 by b_1. Similarly, the earlier multiple scattered form (1:140), say $\hat{\mathbf{K}}\cdot\tilde{\mathfrak{A}}\cdot\hat{\mathbf{K}}$, corresponds to an electromagnetic form $\hat{\boldsymbol{\beta}}\cdot\tilde{\mathfrak{B}}\cdot\hat{\boldsymbol{\beta}}$ with $\mathfrak{A}_{ii}=\mathfrak{B}_{jj}$, $\mathfrak{A}_{ij}=-\mathfrak{B}_{ji}$.

From (1:136)ff, for an exclusion ellipse characterized by q_i with semiaxes along $\hat{\boldsymbol{\zeta}}_i^b$ (such that $\hat{\boldsymbol{\zeta}}_i\cdot\hat{\boldsymbol{\zeta}}_i^b=\cos\omega$), we have, for example,

$$\mathfrak{A}_{11}=\frac{a_1(1-ca_2q_{22})}{|1-c\tilde{\mathbf{a}}\cdot\tilde{\mathbf{q}}|},$$
$$\mathfrak{A}_{12}=\frac{a_1a_2cq_{12}}{|1-c\tilde{\mathbf{a}}\cdot\tilde{\mathbf{q}}|},\qquad c=c_2=\frac{i4\rho}{k^2}, \tag{174}$$

where $\tilde{\mathbf{q}}=\sum q_i\hat{\boldsymbol{\zeta}}_i^b\hat{\boldsymbol{\zeta}}_i^b=\sum q_{ij}\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_j$. The elements $q_{22}=q_1\sin^2\omega+q_2\cos^2\omega=(1-q_2)\sin^2\omega+(1-q_1)\cos^2\omega=1-q_{11}$ and $q_{21}=q_{12}=(q_2-q_1)\cos\omega\sin\omega$ are unaltered if we replace q_2 by $1-q_1$ and q_1 by $1-q_2$. Using $a_1a_2=b_1b_2$ and $q_1q_2=(1-q_2)(1-q_1)$ we rewrite the denominator of (174), $1-ca_1q_{11}-ca_2q_{22}+c^2|\tilde{\mathbf{a}}||\tilde{\mathbf{q}}|=1-cb_2(1-q_{22})-cb_1(1-q_{11})+c^2|\tilde{\mathbf{b}}||I-\tilde{\mathbf{q}}|$. Consequently, for the electromagnetic case, we may write $\mathfrak{A}_{11}=\mathfrak{B}_{22}$, $-\mathfrak{A}_{12}=\mathfrak{B}_{21}$ such that

$$\mathfrak{B}_{22}=\frac{b_2[1-cb_1(1-q_{11})]}{|\tilde{\mathbf{I}}-c\tilde{\mathbf{b}}\cdot(\tilde{\mathbf{I}}-\tilde{\mathbf{q}})|},\quad \mathfrak{B}_{21}=\frac{-b_2b_1cq_{12}}{|\tilde{\mathbf{I}}-c\tilde{\mathbf{b}}\cdot(\tilde{\mathbf{I}}-\tilde{\mathbf{q}})|},\tag{175}$$

and the earlier form (1:140) for $\hat{\mathbf{K}}\cdot\tilde{\mathfrak{A}}\cdot\hat{\mathbf{K}}$ equals $\hat{\boldsymbol{\beta}}\cdot\tilde{\mathfrak{B}}\cdot\hat{\boldsymbol{\beta}}$ with

$$\tilde{\mathfrak{B}}=\tilde{\mathbf{b}}\cdot[\tilde{\mathbf{I}}+c(\tilde{\mathbf{I}}-\tilde{\mathbf{q}})\cdot\tilde{\mathfrak{B}}].\tag{176}$$

Thus (176) emerges from the Mathieu function development for the corresponding electromagnetic problem for H along $\hat{\mathbf{y}}$, i.e., for $\mathbf{E}=\mathbf{E}(\hat{\mathbf{z}},\hat{\mathbf{x}})$. If we incorporate the result for $\mathbf{E}=E\hat{\mathbf{y}}$, then

$$\begin{aligned}&\tilde{\mathbf{b}}=b_1\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}+b_2\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}+b_3\hat{\mathbf{y}}\hat{\mathbf{y}},\quad b_3=\tfrac{1}{4}i\pi k^2a_1a_2(\epsilon_3'-1),\\&\tilde{\mathbf{g}}^e=\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}[\hat{\boldsymbol{\beta}}\cdot(b_1\hat{\boldsymbol{\zeta}}\hat{\boldsymbol{\zeta}}+b_2\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}})\cdot\hat{\boldsymbol{\beta}}]+b_3\hat{\mathbf{y}}\hat{\mathbf{y}},\end{aligned}\tag{177}$$

and (176) applies with $q_3=0$ (as appropriate for the infinite elliptic cylinder).

Essentially as before, we use the form (176) for the corresponding three-dimensional electromagnetic problem of a triaxial ellipsoid. We incorporate the leading terms for radiation losses within the distribution by replacing $\tilde{\mathbf{b}}\approx\tilde{\mathbf{b}}'$ by

$$\tilde{\mathbf{b}}''=\sum b_i''\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i,\quad b_i''\approx b_i'/(1-\tfrac{2}{3}b_i'W),\quad \tilde{\mathbf{b}}''=\tilde{\mathbf{b}}''(\tilde{\boldsymbol{\epsilon}}')\tag{178}$$

with b_i' as in (130) and W corresponding to ellipsoidally symmetric correlations. Thus with $c=c_3=i4\pi\rho/k^3$,

$$\tilde{\mathfrak{B}}=[\tilde{\mathbf{I}}-c\tilde{\mathbf{b}}''\cdot(I-\tilde{\mathbf{q}})]^{-1}\cdot\tilde{\mathbf{b}}'',\quad \tilde{\boldsymbol{\epsilon}}^{-1}=\tilde{\mathbf{I}}+c\tilde{\mathfrak{B}}.\tag{179}$$

We rewrite (176) as $\tilde{\boldsymbol{\epsilon}}^{-1}-\tilde{\mathbf{I}}=c\tilde{\mathbf{b}}''\cdot[\tilde{\mathbf{I}}+(\tilde{\mathbf{I}}-\tilde{\mathbf{q}})\cdot(\tilde{\boldsymbol{\epsilon}}^{-1}-\tilde{\mathbf{I}})]$, from which

$$\begin{aligned}&\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}}=-c\tilde{\mathbf{b}}''\cdot[\tilde{\mathbf{I}}+\tilde{\mathbf{q}}\cdot(\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}})],\\&\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}}=-(\tilde{\mathbf{I}}+c\tilde{\mathbf{b}}''\cdot\tilde{\mathbf{q}})^{-1}\cdot\tilde{\mathbf{b}}''c,\end{aligned}\tag{180}$$

i.e., the same form as in (1:143) for the scalar problem. The left side of the relation

$$(\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\tilde{\mathbf{q}}\cdot(\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}})]^{-1}=-c\tilde{\mathbf{b}}''\tag{181}$$

determines an ellipsoid with shape $\tilde{\mathbf{q}}$ (corresponding to the exclusion ellipsoid) and parameter $\tilde{\boldsymbol{\epsilon}}$ specified by symmetric tensors with noncoincident principal axes. If the scattering losses are negligible, then the form

$$\begin{aligned}&(\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\tilde{\mathbf{q}}\cdot(\tilde{\boldsymbol{\epsilon}}-\tilde{\mathbf{I}})]^{-1}=w_0(\tilde{\boldsymbol{\epsilon}}'-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\tilde{\mathbf{q}}'\cdot(\tilde{\boldsymbol{\epsilon}}'-\tilde{\mathbf{I}})]^{-1},\\&w_0=\rho\mathfrak{B}\end{aligned}\tag{182}$$

generalizes the scalar relation[15] applied earlier for birefringence considerations.

In particular if the axes of $\hat{\mathbf{q}}$, $\hat{\mathbf{q}}'$ and $\tilde{\boldsymbol{\epsilon}}'$ coincide, we obtain

$$\epsilon_i-1=\frac{w_0(\epsilon_i'-1)}{[1+(\epsilon_i'-1)(q_i'-w_0q_i)]},\tag{183}$$

which differs from the earlier result[15] in that q_i is not necessarily equal to q_i'. Whereas $(1-w_0)q_i$ is positive, the present $q_i'-w_0q_i$ may be positive or negative. Now there are three distinct bases of anisotropy for coherent propagation: the obstacle's intrinsic anisotropy represented by its material parameter $\tilde{\boldsymbol{\epsilon}}'$, the effects of its form represented by the shape parameter $\tilde{\mathbf{q}}'$, and the effects of the distribution represented by the shape $\tilde{\mathbf{q}}$ of the exclusion region. The corresponding η^2 is determined by $\hat{\mathbf{K}}\cdot(\eta^{-2}\tilde{\mathbf{I}}-\tilde{\boldsymbol{\epsilon}}^{-1})\cdot\hat{\mathbf{K}}=0$. If a principal axis of $\tilde{\boldsymbol{\epsilon}}$ is along $\hat{\mathbf{y}}$, then for incidence perpendicular to $\hat{\mathbf{y}}$, we have $\eta_y^2=\epsilon_3$ for $\mathbf{E}=E\hat{\mathbf{y}}$, and $\eta_\beta^{-2}=\hat{\boldsymbol{\beta}}\cdot\tilde{\boldsymbol{\epsilon}}^{-1}\cdot\hat{\boldsymbol{\beta}}$ for $\mathbf{H}=H\hat{\mathbf{y}}$, essentially as discussed for (144) and (145); for this case (1:123) applies.

Similarly, for obstacles specified by $\tilde{\boldsymbol{\mu}}'$ and $\tilde{\mathbf{c}}''=\tilde{\mathbf{b}}''(\tilde{\boldsymbol{\mu}}')$ in the form (179), we have

$$\begin{aligned}&\tilde{\mathfrak{C}}=\tilde{\mathbf{c}}''\cdot[\tilde{\mathbf{I}}+c(\tilde{\mathbf{I}}-\tilde{\mathbf{q}})\cdot\tilde{\mathfrak{C}}],\\&\tilde{\mathfrak{C}}=[\tilde{\mathbf{I}}-c\tilde{\mathbf{c}}''\cdot(\tilde{\mathbf{I}}-\tilde{\mathbf{q}})]^{-1}\cdot\tilde{\mathbf{c}}'',\quad \tilde{\boldsymbol{\mu}}^{-1}=\tilde{\mathbf{I}}+c\tilde{\mathfrak{C}},\end{aligned}\tag{184}$$

or, equivalently,

$$(\tilde{\boldsymbol{\mu}}-\tilde{\mathbf{I}})\cdot[\tilde{\mathbf{I}}+\tilde{\mathbf{q}}\cdot(\tilde{\boldsymbol{\mu}}-\tilde{\mathbf{I}})]^{-1}=-c\tilde{\mathbf{c}}''.\tag{185}$$

The corresponding η^2 for obstacles specified by both $\tilde{\boldsymbol{\epsilon}}'$ and $\tilde{\boldsymbol{\mu}}'$ is determined by $|\tilde{\mathbf{I}}_T\eta^{-2}+\hat{\mathbf{K}}\times\tilde{\boldsymbol{\mu}}^{-1}\times\hat{\mathbf{K}}\cdot\tilde{\boldsymbol{\epsilon}}^{-1}|=0$. If $\tilde{\boldsymbol{\epsilon}}$ and $\tilde{\boldsymbol{\mu}}$ each have an axis along $\hat{\mathbf{y}}$, then for incidence perpendicular to $\hat{\mathbf{y}}$ we obtain the forms in (171) but the principal axes of $\tilde{\boldsymbol{\epsilon}}_\rho$ and $\tilde{\boldsymbol{\mu}}_\rho$ are not necessarily parallel. For this case, the scalar development[1] for the slab distribution suffices for the transmitted and reflected fields, etc., in terms of the form (1:128), etc.

We considered ellipsoids for illustrative purposes, but the initial forms (65) and (81), as well as special cases and results may be used for other scatterers. Thus, if each ellipsoid is replaced by an identical pair of similarly aligned ellipsoids with fixed separation, then the pair scatterer has the same inversion and reflection symmetries as the single ellipsoid, and the simplified systems as in (117) and (147) apply. Results for the single pair were given earlier,[2] with detailed considerations for pairs of dipoles. Thus (2:141), in terms of, e.g., (2:198) or (2:210) for appropriate identical pairs of electric or magnetic dipoles, may be used in (147). However, if the elements of a pair scatterer differ in physical properties or shape or alignment, then we require (81). In particular if one element of the pair scatterer is an electric dipole ($\tilde{\mathbf{b}}=\sum b_i\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$) and the other a magnetic dipole ($\tilde{\mathbf{c}}=\sum c_i\hat{\boldsymbol{\zeta}}_i\hat{\boldsymbol{\zeta}}_i$) with $\hat{\boldsymbol{\zeta}}_1=\hat{\boldsymbol{\zeta}}$ as the axis of the pair, then the solution for the composite obstacle is given by (2:212)–(2:226). The scattering amplitude for the composite does not have inversion symmetry in the origin (the midpoint of the line joining the dipoles centers) or reflection symmetry in $\zeta=0$, but does have reflection symmetry in $\zeta_2=0$ and $\zeta_3=0$, e.g., as shown by the special case of spherical dipoles (2:175)–(2:188). Results based on (81), (2:141), and (2:179) provide a relatively simple case of polarizabilities $\mathbf{P},\mathbf{M}$ such that each depends on both $\mathbf{E}$ and $\mathbf{H}$, i.e., $\mathbf{P}=\tilde{\mathbf{P}}_e\cdot\mathbf{E}+\tilde{\mathbf{P}}_m\cdot\mathbf{H}$, $\mathbf{M}=\tilde{\mathbf{M}}_m\cdot\mathbf{H}+\tilde{\mathbf{M}}_e\cdot\mathbf{E}$. Similar interrelations for the general ellipsoid, and for a symmetrical pair such that each element consists of both an electric plus a magnetic dipole as in (2:227)ff, follow from (147).

[1] V. Twersky, "Coherent Scalar Field in Pair-Correlated Random Distributions of Aligned Scatterers," J. Math. Phys. 18, 2468 (1977).

[2] V. Twersky, "Multiple Scattering of Electromagnetic Waves by Arbitrary Configurations," J. Math. Phys. 8, 589–610 (1967).

[3]Lord Rayleigh, "On The Transmission of Light Through an Atmosphere Containing Small Particles in Suspension, and on the origin of the Blue of the Sky," Phil. Mag. **47**, 375–83 (1899).
[4]F. Reiche, "Zur Theorie der Dispersion in Gasen und Dampfen," Ann. Phys. 50 (1), 121 (1916).
[5]L. L. Foldy, "The Multiple Scattering of Waves," Phys. Rev. 67 (2), 107–19 (1945).
[6]M. Lax, "Multiple Scattering of Waves," Rev. Mod. Phys. **23**, 287–310 (1951); "The Effective Field in Dense Systems," Phys. Rev. (2) **88**, 621–629 (1952).
[7]V. Twersky, "On Scattering of Waves by Random Distributions," J. Math. Phys. **3**, 700–15, 724–34 (1962).
[8]J. B. Keller, "Wave Propagation in Random Media," in *Proc. Symp. Appl. Math., Vol.* 13, (Amer. Math. Soc., Providence, R.I., 1962), pp. 227–46; "Stochastic Equations and Wave Propagation in Random Media," *ibid.*, Vol. 16 (1964), pp. 145–70.
[9]V. Twersky, "On Propagation in Random Media of Discrete Scatterers," in Ref. 8, Vol. 16 (1964), pp. 84–116.
[10]V. Twersky, "Propagation Parameters in Random Distribution of Scatterers", J. de Anal. Math. **30**, 498–511 (1976).
[11]D. J. Vezzetti and J. B. Keller, "Refractive Index, Dieletic Constant, and Permeability for Waves in a Polarizable Medium," J. Math. Phys. **8**, 1861–70 (1967).
[12]V. Twersky, "Transparency of pair-correlated, random distributions of small scatterers with applications to the cornea," J. Opt. Soc. Am. **65**, 524–30 (1975).
[13]B. Friedman and J. Russek, "Addition Theorems for Spherical Waves," Quart. Appl. Math. **12**, 13–23 (1954); S. Stein, Quart. Appl. Math. **19**, 15 (1961).
[14]N. C. Mathur and K. C. Yeh, "Multiple Scattering of Electromagnetic Waves by Random Scatterers of Finite Size," J. Math. Phys. **5**, 1619–28 (1964).
[15]V. Twersky, "Form and Intrinsic Birefringence," J. Opt. Soc. Am. **65**, 239–45 (1975).

A no-go theorem concerning the cluster decomposition property of direct-interaction scattering theories*

U. Mutze

Sektion Physik der Universität München, Germany
(Received 14 March 1977)

An axiomatic framework for relativistic direct-interaction single channel scattering theories is formulated in terms of two representations U' and U of the Poincaré group. The infinitesimal generators $\mathbf{P}$ (momentum), $\mathbf{J}$ (angular momentum), E (energy), $\mathbf{N}$ ("boost") of U, and $\mathbf{P}'$, $\mathbf{J}'$, E', $\mathbf{N}'$ of U' are assumed to be related by the formulas of Bakamjian, Thomas, and Foldy: $\mathbf{P}'=\mathbf{P}$, $\mathbf{J}'=\mathbf{J}$, $E'=(M'^2+\mathbf{P}^2)^{1/2}$, and $\mathbf{N}'=(E'\mathbf{X}+\mathbf{X}E')/2+\mathbf{P}\times(\mathbf{J}-\mathbf{X}\times\mathbf{P})(M'+E')^{-1}$, where $\mathbf{X}$ is the center-of-mass position operator of U given by $\mathbf{X}=\mathbf{T}-\mathbf{P}\times(\mathbf{J}-\mathbf{T}\times\mathbf{P})M^{-1}(M+E)^{-1}$ with $\mathbf{T}=(E^{-1}\mathbf{N}+\mathbf{N}E^{-1})/2$ and M' is a positive operator that commutes with $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{X}$. Then, it is proved that, within the above-mentioned framework, the Møller operators $W_\pm=\lim \exp(-itE')\exp(itE)$ for $t\to\pm\infty$ cannot satisfy the cluster decomposition property (also known as separability) except for interactions that vanish if any one of the particles is removed to infinity.

I. INTRODUCTION AND SUMMARY OF THE RESULTS

This paper[1] contributes to a topic that is usually referred to as the direct-interaction approach to relativistic particle dynamics[2–19] or as the theory of action at a distance. Its scope is limited to such direct-interaction theories that describe only elastic scattering, irrespectively of how large the center of mass energy may be. Though departing obviously from reality, such theories may be useful for phenomenological approximations and, further, for studying the relations between the principles of quantum physics and the principles of relativity. An axiomatic characterization of the theories under consideration will be given in items (c) and (e). The matter to be discussed particularly concerns two further axioms: one [see (f)(2)] expressing a rather evident physical property, the other [see (d)(2)] giving the theory an especially simple structure that is suggested by direct analogy with nonrelativistic theory. On an informal level these axioms may be expressed as follows.

(a) The motion of the particles is separable in the sense that two subsystems are dynamically independent if they are separated by a sufficiently large distance.

(b) The motion of the particles is decomposable into a center of mass motion and an internal motion such that the interaction affects the internal motion only.

The result of this paper is formulated in (g); it states that these two axioms contradict each other[20] except for interactions that shrink away whenever one of the particles is removed to infinity; for a system of three or more particles such interactions are clearly unphysical. This result may be considered as sharpening a rather cryptic argument of Foldy.[21] The fact that recent work on direct-interaction theory[12,15,18] does not adopt the critical axiom (d)(2) indicates that experts know about difficulties associated with it. To reveal the mathematical reasons for such difficulties is the purpose of this paper.

a)This work was supported in part by the Deutsche Forschungsgemeinschaft.

We turn now to exposing the announced axiomatic framework; the ideas involved therein are those of general scattering theory as described in Refs. 22 and 23 and those of direct-interaction theory as described in Ref. 8. Some of these ideas will be indicated in later comments. Throughout the remainder of this section, Definition 1 (see Sec. II) and the notational convention following it will be adopted. Sections II and III are formally independent of the rest of this section.

(c) Let U be a positive rep. of $\mathcal{P}^1$ and let S be a bounded operator on H. Assume that there exists a positive rep. U' on H such that

(1) $U'(-t)U(t)\to W_\epsilon(U',U)=:W_\epsilon$ as $t\to\epsilon\infty$, for $\epsilon=+$ and $\epsilon=-$, where W_ϵ is a bounded operator on H;

(2) $U'(g)W_\epsilon=W_\epsilon U(g)$ for all $g\in\mathcal{P}$, $\epsilon\in\{+,-\}$;

(3) $S=W_-^*W_+$.

Then we call the pair (U,S) a *scattering system* and S is called the *scattering operator* of the system. U' is said to be a *dynamical description* of the system and W_+ and W_- are said to be *wave operators* of the system.

The reps. U' and U should be interpreted as describing respectively the relativistic transformation properties of the states and of the asymptotic configurations[23] of the scattering states.

(d) Consider the following conditions applying to positive reps. U' and U on H:

(1) $U'(g)=U(g)$ for all $g\in\mathcal{E}:=\{(a,A)\in\mathcal{P}:a^0=0,\ A\in\mathrm{SU}(2)\}$.

(2) In addition to (1), we have $\mathbf{X}=\mathbf{X}'$, where $\mathbf{X}$ and $\mathbf{X}'$ are the Newton–Wigner position operators of U and U' resp. (see Theorem 1).

Then, a dynamical description U' of a scattering system (U,S) is said to be *barycentric* if (2) is satisfied.

To describe scattering of *particles* some more structure has to be added:

(e) Let $(U_i)_{i\in I}$ be a finite family of irreducible positive reps.; for any subset $c\subset I$ (the case $c=I$ being included) choose a tensor product $H_c=\otimes_{i\in c}H_i$, define $U_c:=\otimes_{i\in c}U_i$, let (U_c,S_c) be a scattering system such that

 0022-2488/78/1901-0231$1.00

S_c is the identity operator if c is a one-element set, and let U'_c be a dynamical description of (U_c, S_c). Then, the family $(U_c, S_c)_{c\subset I}$ is called a *particle scattering system* and the family $(U'_c)_{c\subset I}$ is called a *dynamical description* of this system.

Separability, as characterized in (a), demands that the scattering operators S_c of subsystems result from the scattering operator S_I of the whole system by a spatial separation limit:

(f) In addition to the assumptions in (e), let π: $=(c_1, \ldots, c_k)$ be any collection of pairwise disjoint subsets of I, and put $c := c_1 \cup \cdots \cup c_k$. Let $H_c = H_{c_1} \otimes \cdots \otimes H_{c_k}$ be the tensor product that is uniquely determined by requiring $\otimes_{i\in c}\psi_i = (\otimes_{i\in c_1}\psi_i)\otimes\cdots\otimes(\otimes_{i\in c_k}\psi_i)$ for all families $(\psi_i)_{i\in c}$ with $\psi_i \in H_i$ for all $i \in c$. Then the particle system (e) is said to be π-separable if

(1) $U_{c_1}(d\mathbf{a}_1)\otimes\cdots\otimes U_{c_k}(d\mathbf{a}_k) S_c U_{c_1}(-d\mathbf{a}_1)\otimes\cdots\otimes U_{c_k}(-d\mathbf{a}_k) \to S_{c_1}\otimes\cdots\otimes S_{c_k}$ weakly as $|d|\to\infty$, for all $(\mathbf{a}_1, \ldots, \mathbf{a}_k) \in R^{3k}$ with $\mathbf{a}_i \neq \mathbf{a}_j$ for all $i \neq j$.

Further, the dynamical description $(U'_c)_{c\subset I}$ is said to be *π-separable* if

(2) $(W_\epsilon(U'_c, U_c) - W_\epsilon(U'_{c_1}, U_{c_1})\otimes\cdots\otimes W_\epsilon(U'_{c_k}, U_{c_k}))$ $\times U_{c_1}(d\mathbf{a}_1)\otimes\cdots\otimes U_{c_k}(d\mathbf{a}_k) \to 0$ as $|d|\to\infty$, for all $\epsilon \in \{+,-\}$ and all $\mathbf{a}_1, \ldots, \mathbf{a}_k$ as in (1).

By definition, "*separable*" means "π-separable for all collections π as considered above."

Now the precise result of the present paper can be stated.

(g) Adopt the assumption in (e) and (f), choose an element $i \in I$, denote by $\{i\}'$ the set $\{j \in I : j \neq i\}$, and assume the following conditions to be satisfied:

(1) The dynamical description $(U'_c)_{c\subset I}$ is $(\{i\}, \{i\}')$-separable.

(2) The pairs $(U'_{\{i\}}, U_{\{i\}})$ and $(U'_{\{i\}'}, U_{\{i\}'})$ of reps. satisfy (d)(1).

(3) The pair (U'_I, U_I) of reps. satisfies (d)(2).

Then the scattering operators $S_{\{i\}'}$ and $S_{\{i\}}$ are the identity operators.

If the particle scattering system is moreover separable we conclude at once from (f)(1) that S_c is the identity operator for all $c \subset \{i\}'$. Similarly, we find $S_c = 1$ for all $c \subset I$ with $c \neq I$ if the particle scattering system (e) is assumed to possess a separable dynamical description $(U'_c)_{c\subset I}$ such that the pair (U'_I, U_I) satisfies (d)(2) and the pairs (U'_c, U_c) with $c \neq I$ satisfy (d)(1). To summarize this in a few words: An n-particle scattering system that allows a barycentric and separable dynamical description shows at most a pure n-particle interaction.

Several comments on the items (c)—(g) should be made.

Comments on (c): Obviously, the W_ϵ are isometric operators; the ranges of W_+ and W_- coincide iff S is unitary. The wave operators are said to be asymptotically complete if each of them maps H onto the orthocomplement of the eigenvectors of the mass operator M' of U'. The present notion of a scattering system is similar to that of a simple scattering system introduced in Ref. 22. There are two distinctions: In (c) asymptotic completeness is not required while, on the contrary, the requirements of relativity are taken into account by introducing two reps. of $\mathcal{P}$ instead of two reps. of the time translation group. Condition (2)[24] assures covariance: If the state ψ has the outgoing asymptotic configuration ψ_{out} (i.e., $\psi = W_-\psi_{\text{out}}$), then $U'(g)\psi$ has the outgoing configuration $U(g)\psi_{\text{out}}$. The Poincaré invariance of the scattering operator (i.e., $\mathrm{SU}(g) = U(g)S$ for all $g \in \mathcal{P}$) follows at once from (2). A restricted form of (2) is satisfied as a consequence of (1): $U'(t)W_\epsilon = W_\epsilon U(t)$ for all time translations t. The objects U and S being given, neither U' nor the W_ϵ are determined uniquely: As pointed out by Coester,[25] the definition $\tilde{U}'(g) := V^*U'(g)V$, where V is a unitrary operator on H such that $U(t)VU(-t) \to 1$ as $|t| \to \infty$ (a large class of such operators can easily be constructed), yields the wave operators $\tilde{W}_\epsilon = V^*W_\epsilon$ and, hence, $S = W_-^*W_+ = \tilde{W}_-^*\tilde{W}_+$ and $\tilde{U}'(g)\tilde{W}_\epsilon = \tilde{W}_\epsilon U(g)$ for all $g \in \mathcal{P}$. The term "dynamical description" is chosen for U' since U' allows understanding the mapping $\psi_{\text{in}} \mapsto \psi_{\text{out}} := S\psi_{\text{in}}$ as the result of a dynamical evolution: $U(t)U'(-2t)U(t)\psi_{\text{in}} \to \psi_{\text{out}}$ weakly as $t \to \infty$.

Comments on (d): The following points will be discussed in the sequel: (i) How a positive rep. can be characterized in terms of more elementary mathematical objects. (ii) Characterization of the sets $S_i(U)$: $=\{U' : U'$ and U satisfy condition (i)$\}$ for $i = 1, 2$. (iii) Characterization of $S_2(U)$ in the case that U is the rep. that is associated with a free n-particle system. (iv) Conditions (1) and (2) and the axioms of scattering systems. (v) Relations to the literature.

(i) Let U be a positive rep. of $\mathcal{P}$ on a Hilbert space H. Then we are able to write H as a tensor product $H_{\text{in}}\otimes L^2(R^3)$ of an internal space H_{in} and a one-particle momentum space $L^2(R^3)$ such that

(α) for all $g = (0, \mathbf{a}, A) \in \mathcal{E}$ we have $U(g) = U_{\text{in}}(A) \otimes U_{cm}(\mathbf{a}, A)$, where U_{in} is a rep. of SU(2) and U_{cm} is the rep. of $\mathcal{E}$ that is given by $(U_{cm}(\mathbf{a}, A)\psi)(\mathbf{p}) = \exp(-i\mathbf{a}\cdot\mathbf{p}) \times \psi(A^{-1}\mathbf{p})$ and

(β) the Newton—Wigner position operator $\mathbf{X}$ of U is of the form $1\otimes i\nabla_{\mathbf{p}}$. Let us call such a tensor product a barycentric factoring of H (with respect to U). Now the following proposition can be proved.

(γ) A bounded or a self-adjoint operator commutes with both $\mathbf{P}$ and $\mathbf{X}$ iff it is of the form $A_{\text{in}}\otimes 1$. Such an operator will be called an *internal operator*; examples are the mass operator $M = M_{\text{in}}\otimes 1$ and the spin operator $\mathbf{S} = \mathbf{J} - \mathbf{X}\times\mathbf{P}$, a convenient definition of which is $\exp(2i\mathbf{u}\cdot\mathbf{S}) = U_{\text{in}}(\exp(i\mathbf{u}\cdot\boldsymbol{\sigma}))\otimes 1$. The rep. U can be reconstructed from U_{in} and M_{in} in the following sense.

(δ) Let U_{in} be a rep. of SU(2) on a Hilbert space H_{in}, let M_{in} be a strictly positive operator that commutes with U_{in}, and let $H = H_{\text{in}}\otimes L^2(R^3)$ be a tensor product. Then, the formulas in (α) define a rep of $\mathcal{E}$, which can be uniquely extended to a positive rep. U such that (β) holds and $M_{\text{in}}\otimes 1$ is the mass operator of U. Formally, this extension is effected by the *Bakamjian—Thomas—Foldy formulas*[26] $E = (M^2 + \mathbf{P}^2)^{1/2}$, $\mathbf{N} = \frac{1}{2}(E\mathbf{X} + \mathbf{X}E) + \mathbf{P} \times (\mathbf{J} - \mathbf{X}\times\mathbf{P})(M + E)^{-1}$ for the generators of U, where $\mathbf{P}$, $\mathbf{J}$ (the generators belonging to $\mathcal{E}$), and $\mathbf{X} = 1\otimes i\nabla_{\mathbf{p}}$ are known. A rigorous and direct construction of U can be

obtained by a spectral representation with respect to $M_{\rm in}$ and $U_{\rm in}$.

(ii) Let U be a positive rep. and let $H = H_{\rm in} \otimes L^2(R^3)$ be a barycentric factoring of its Hilbert space. Obviously, the latter is also a barycentric factoring with respect to U' provided U' belongs to $S_2(U)$. Hence the mass operator M' of U' is of the form $M'_{\rm in} \otimes 1$ and, by (δ), the prescription $U' \mapsto M'_{\rm in}$ defines a bijective mapping from $S_2(U)$ onto the set of all strictly positive operators that commute with $U_{\rm in}$. This provides some control over $S_2(U)$. The set $S_1(U)$ is closely related to $S_2(U)$: Any $U' \in S_1(U)$ can be obtained from a member of $S_2(U)$ by a unitary transformation that commutes with $U(g)$ for all $g \in \mathcal{E}^{27}$ and is arbitrary otherwise. Thus $S_2(U)$ is a proper subset of $S_1(U)$.

(iii) Let U be the rep. that is associated with a free n-particle system. For the sake of simplicity let us assume the particles to be spinless and pairwise distinguishable. Then, a complete set of observables is given by the total momentum $\mathbf{P} = \mathbf{P}_1 + \cdots + \mathbf{P}_n$ and the rest frame momenta $\mathbf{K}_1, \ldots, \mathbf{K}_{n-1}$ that are given by $\mathbf{K}_j = (\mathbf{P}_j - (E_j - (M+E)^{-1}\mathbf{P}\cdot\mathbf{P}_j)M^{-1}\mathbf{P}$ (note that $\mathbf{K}_1 + \cdots + \mathbf{K}_n = 0$). Let us choose a spectral representation such that these operators become multiplication operators with respect to the arguments $\mathbf{p}, \mathbf{k}_1, \ldots, \mathbf{k}_{n-1}$ of the wavefunctions. Then it is easily shown that $\mathbf{X}$ becomes $i\nabla_{\mathbf{p}}$ and that $(\epsilon)L^2(R^{3n}) = L^2(R^{3(n-1)}) \otimes L^2(R^3)$, $(\psi \otimes \varphi)(\mathbf{p}, \mathbf{k}_1, \ldots, \mathbf{k}_{n-1}) = \varphi(\mathbf{p})\psi(\mathbf{k}_1, \ldots, \mathbf{k}_{n-1})$ defines a barycentric factoring. Hence, the internal operators are the operators that "do not act on the variable p." As shown above, the mass operator M' of $U' \in S_2(U)$ is an internal operator. Therefore, the "interaction term" $V = M' - M$ acts only on $\mathbf{k}_1, \ldots, \mathbf{k}_{n-1}$. The operator M' being given, the whole rep. U' can be constructed as shown in (i). The analogy between V and a nonrelativistic potentiallike interaction term is obvious: Replacing the rest frame momenta $\mathbf{K}_j$ by their nonrelativistic counterparts $\mathbf{K}_j = \mathbf{P}_j - m_j \times (m_1 + \cdots + m_n)^{-1}\mathbf{P}$ (m_j is the mass of the jth particle) and the mass operator $(m_1^2 + \mathbf{K}_1^2)^{1/2} + \cdots + (m_n^2 + \mathbf{K}_n^2)^{1/2}$ by the operator of nonrelativistic internal energy $(2m_1)^{-1}\mathbf{K}_1^2 + \cdots + (2m_n)^{-1}\mathbf{K}_n^2$, we see that V corresponds to the nonrelativistic interaction energy, which acts on $\mathbf{k}_1, \ldots, \mathbf{k}_{n-1}$ and not on $\mathbf{p}$. With the nonrelativistic meaning of $\mathbf{p}, \mathbf{k}_1, \ldots, \mathbf{k}_{n-1}$ the operator $i\nabla_{\mathbf{p}}$ is the center of mass position $(m_1\mathbf{X}_1 + \cdots + m_n\mathbf{X}_n)(m_1 + \cdots + m_n)^{-1}$ and the tensor product (ϵ) describes the classical procedure of separating the center of mass motion. An analogous interpretation should be adopted for a barycentric factoring in the relativistic case.

(iv) First we assume that the reps. U' and U in (c) satisfy (1). Then (c)(1) is equivalent to a less special relation: $U'(-da)U(da) \to W_\epsilon$ as $d \to \epsilon\infty$, for all translations $a \in R^4$ with $a^0 > 0$. If (1) holds for the reps. U'_{c_l}, U_{c_l}, $l \in \{1, \ldots, k\}$ in (f), the limit relation (f)(2) is equivalent to the more tractable one (see Theorem 5)

$$U_{c_1}(d\mathbf{a}_1) \otimes \cdots \otimes U_{c_k}(d\mathbf{a}_k)\, W_\epsilon U_{c_1}(-d\mathbf{a}_1) \otimes \cdots \otimes U_{c_k}(-d\mathbf{a}_k) \to W_\epsilon(U'_{c_1}, U_{c_1}) \otimes \cdots \otimes W_\epsilon(U'_{c_k}, U_{c_k}) \quad \text{as} \quad |d| \to \infty.$$

Now let us assume that U' and U satisfy (c)(1), (c)(2), and (2) and let us choose a barycentric factoring (with respect to both U' and U). Then the operators M', M, and W_ϵ are internal operators [the latter one by (c)(2) and Theorem 3] and hence we get operators $M'_{\rm in}$, $M_{\rm in}$, and $W_{\epsilon,{\rm in}}$. The property (c)(1) suggests the relation $\exp(-itM'_{\rm in}) \times \exp(itM_{\rm in}) \to W_{\epsilon,{\rm in}}$ as $t \to \epsilon\infty$, a rigorous derivation of which, however, is not known to the author. Conversely, let us assume the following condition (ξ) instead of (c)(1) and (c)(2).

(ξ) $\exp(-itM'_{\rm in})\exp(itM_{\rm in}) \to A_\epsilon$ as $t \to \epsilon\infty$ with a bounded operator A_ϵ. Then, we can prove by Theorem 3 that the operator $B_\epsilon := A_\epsilon \otimes 1$ satisfies $B_\epsilon U'(g) = U(g)B_\epsilon$ for all $g \in \mathcal{P}$. According to the invariance principle for wave operators[28] we expect $\exp[-itf(M'_{\rm in})] \times \exp[itf(M_{\rm in})] \to A_\epsilon$ as $t \to \epsilon\infty$ for all strictly increasing differentiable functions f from $[0, \infty)$ to R under rather weak assumptions on $M'_{\rm in}$ and $M_{\rm in}$. Therefore, let us assume that the preceding relation holds for any of the functions $f_c : x \mapsto (x^2 + c^2)^{1/2}$, $c \geq 0$. Then, we can prove[29] $U'(-t)U(t) \to B_\epsilon = W_\epsilon(U', U)$ as $t \to \epsilon\infty$. Thus the conditions (c)(1) and (c)(2) are shown to be satisfied. In conclusion we may roughly say that (d)(2) simplifies the structure of a scattering system by reducing it to the internal space of a barycentric factoring.

(v) Besides the scattering theoretical interpreation of U' and U adopted here, a different interpretation is conceivable: U is associated with a noninteracting system, the only meaning of which is to serve for the starting point of a mental (or mathematical) construction that is to make the system an interacting one by adding "interaction terms" to the infinitesimal generators of U, thus giving rise to a new rep. U' of $\mathcal{P}$.[30] This point of view is similar to that of Refs. 2–4, and 12. In Ref. 2, Dirac discussed three possibilities of choosing the stable subgroup $\{g \in \mathcal{P} : U'(g) = U(g)\}$; choosing the Euclidean group [as in (1) and (2)] characterizes the *instant form* of dynamics in Dirac's terminology. This instant form of dynamics is well known from the Lagrangian field theory where, in the case of nonderivative coupling, the modifications of the energy–momentum tensor due to interaction change neither the momentum nor the angular momentum. Moreover, nonrelativistic potential interaction always fits the instant form. Therefore, it is quite natural that the predominant part of the literature on direct interaction is based on the instant form; Refs. 3–8 and 19 may serve as examples. More recently, however, direct interaction theories based on a special version of Dirac's *point form* were investigated by Sokolov[16–18] (see also Ref. 15, Sec. II). In these theories the following condition (1)′ is satisfied instead of (1).

(1)′ $U'(0, A) = U(0, A)$ for all $A \in \mathrm{SL}(2, C)$ and $\mathbf{P}'M'^{-1} = \mathbf{P}M^{-1}$.

The discussion in (i) and (ii) shows that condition (2) gives an abstract characterization of those reps. U' that can be obtained from U by the classical Bakamjian–Thomas–Foldy formulas.[26] In Ref. 4 these formulas are used in the first of two steps: Let U be the rep. of a free n-particle system. Then, a rep. belonging to $S_2(U)$, called the "reduced representation," is constructed and finally the "physical representation" is obtained by a unitary transformation. This physical rep. belongs to $S_1(U)$ but generally not to $S_2(U)$.[31] The scattering theory associated with reps. U' and U satisfying (2) is investigated in Ref. 8. The analog to condi-

tion (2) in a theory satisfying (1)′ instead of (1) is obtained by adding to (1)′ the condition $M\mathbf{X}=M'\mathbf{X}'$.[32]

Comments on (e): The tensor product structure of U allows constructing the observables of a free n-particle system (where n is the cardinality of c) as an irreducible set of operators on the Hilbert space H_c.[33] Therefore, the generalized eigenvectors that are needed for defining the S matrix and for calculating the cross sections are given canonically. The question of how to introduce observables of interacting particles such that the above-mentioned channel observables result by some kind of asymptotic limit[34] will not be discussed in this paper. The framework given by (c)—(e) may easily be extended by introducing a countable number of scattering channels; it is not evident, however, how to carry over the result of this paper to such an extended framework. Finally, it should be noted that the tensor products introduced in (e) are given canonically if the H_t are spaces of square integrable functions; this remark applies also to (f).

Comments on (f): If the $S_{c_1},\ldots,S_{c_k}$ are unitrary, the weak limit condition (1) is equivalent to a strong one. Hence (1) expresses just the *cluster property* formulated in Ref. 8. Condition (2) is easily shown to imply (1)[35] but the converse cannot be true generally. The most natural way of translating the concept of separability from the S-operator level to the level of dynamics seems to be provided by the condition (2)′ that is obtained from (2) by replacing the wave operators $W_\epsilon(U'_c,U_c)$, $W_\epsilon(U'_{c_1},U_{c_1}),\ldots,W_\epsilon(U'_{c_k},U_{c_k})$ by the operators $U'_c(g)$, $U'_{c_1}(g),\ldots,U'_{c_k}(g)$, $g\in P$. If in (2)′ the limit condition referring to $U'(t)$ is assumed to be satisfied uniformly with respect to the time translation t,[36] we can easily show that (2) and (1) are satisfied. Condition (2)′ comes very close to the *separability condition* of Sokolov[37] and is just a formalization of that of Foldy and Krajcik.[12]

Comments on (g): The condition that plays the most striking role in (g) is (3). In fact, it can be shown, by a simple model describing two interacting particles together with a free one, that the conclusion of (g) becomes wrong if in (g)(3), requiring (d)(2) is replaced by requiring only (d)(1). The only consequence of (g) for direct-interaction theories discussed in the literature seems to be the conclusion that the three-particle model of Ref. 8 cannot satisfy the separation property (f)(2) of the wave operators if the momenta of the particles are defined as those of the free particles the construction has started with.

The following two sections are devoted to the mathematical arguments leading to the result stated above. For the sake of transparency these are presented in the definition-theorem-proof style. The steps to be noticed particularly are given by Theorem 2, Theorem 4, and Lemma 1.

II. AUXILIARY INFORMATION ABOUT THE NEWTON-WIGNER POSITION OPERATOR

Let us confine our interest to particles with nonzero mass; then only reps. of the following kind occur:

Definition 1. A rep. of P is said to be positive if its energy operator E and its mass square operator M^2 are both positive and the spectrum of M^2 has a strictly positive lower bound.

Note that any tensor product of positive reps. is again positive. In all that follows, U (with or without an index, etc.) will denote a positive rep. of P and with U we associate some objects which will always be denoted by the symbols (marked by the same index as U) used in the following list:

1. H, the Hilbert space on which the $U(g)$, $g\in P$, are operators;

2. E, P^j, J^j, N^j, $j\in\{1,2,3\}$, the self-adjoint generators of U defined by $U(a,1)=\exp(ia^0E-i\mathbf{a}\cdot\mathbf{P})$,[38] $U(0,\exp(i\mathbf{u},\sigma))=\exp(2i\mathbf{u}\cdot\mathbf{J})$, $U(0,\exp(\mathbf{u}\cdot\sigma))=\exp(2i\mathbf{u}\cdot\mathbf{N})$ for all $a\in R^4$, $\mathbf{u}\in R^3$;

3. $M=(E^2-\mathbf{P}^2)^{1/2}$, the mass operator;

4. $\mathbf{X}$, the (Newton—Wigner) position operator defined in Theorem 1.

The following three theorems are given without proof because their formal content can easily be verified by formal calculation using the Lie algebra of P and the technicalities of an exact proof would take too much space.

Theorem 1: Let U be a positive rep. of P. Then there is a dense linear subset $D\subset H$ with the following properties:

(i) D is invariant under the operators E, P^j, J^j, N^j, E^{-1}, M^{-1}, and $(M+E)^{-1}$. Any real linear combination of the generators, considered as an operator on D, is essentially self-adjoint.

(ii) Any real linear combination of the three operators defined on D by $\mathbf{T}-\mathbf{P}\times(\mathbf{J}-\mathbf{T}\times\mathbf{P})M^{-1}(M+E)^{-1}$ with $\mathbf{T}=\frac{1}{2}(E^{-1}\mathbf{N}+\mathbf{N}E^{-1})$ is essentially self-adjoint.

Further, the self-adjoint extensions of the three operators defined in (ii) form a triple X of self-adjoint operators which are the same for any dense linear subset D with the properties (i) and (ii); on any such subspace we have the *Bakamjian—Thomas—Foldy formula*:

$$\mathbf{N}=\tfrac{1}{2}(E\mathbf{X}+\mathbf{X}E)+\mathbf{P}\times(\mathbf{J}-\mathbf{X}\times\mathbf{P})(M+E)^{-1}.$$

For any $\mathbf{a}$, $\mathbf{b}\in R^3$ the self-adjoint operators $\mathbf{a}\cdot\mathbf{X}$ and $\mathbf{b}\cdot\mathbf{X}$ commute with each other and with M. Finally we have

$$U(a)\mathbf{b}\cdot\mathbf{X}U(-a)=\mathbf{b}\cdot\mathbf{X}+a^0\mathbf{b}\cdot\mathbf{P}E^{-1}-\mathbf{a}\cdot\mathbf{b},$$

for all $a\in R^4$, $\mathbf{b}\in R^3$. $\mathbf{X}$ is the *Newton—Wigner position operator* of U.

Theorem 2: Let $U=U_1\otimes U_2$ be a tensor product of positive representations. Then for all $\mathbf{a},\mathbf{b}\in R^3$ we have

$$(1\otimes U_2(\mathbf{a}))\mathbf{b}\cdot\mathbf{X}(1\otimes U_2(\mathbf{a}))^{-1}=\mathbf{b}\cdot\mathbf{X}-\mathbf{b}\cdot\mathbf{Z}(\mathbf{a})$$

with

$$\mathbf{Z}(\mathbf{a}):=\mathbf{a}(E_2E^{-1}+\mathbf{V}\cdot\mathbf{v})-(\mathbf{a}\cdot\mathbf{V})\mathbf{v},$$

$$\mathbf{V}:=\mathbf{P}E^{-1},\quad \mathbf{v}:=(E_2\mathbf{P}_1-E_1\mathbf{P}_2)M^{-1}(M+E)^{-1},$$

where the generators with indices 1 and 2 are those of $U_1\otimes 1$ and $1\otimes U_2$ respectively. The components of $\mathbf{V}$ and $\mathbf{v}$ are bounded operators.

Theorem 3: Let U and U' be positive reps. of $\mathcal{P}$. A bounded operator $B: H \to H'$ satisfies $U'(g)B = BU(g)$ for all $g \in \mathcal{P}$ iff $U'(a,A)B = BU(a,A)$ for all $a \in R^4$, $A \in \mathrm{SU}(2)$, and $\mathbf{b}\cdot\mathbf{X}'B \supset B\mathbf{b}\cdot\mathbf{X}$ for all $\mathbf{b} \in R^3$.

III. CONSEQUENCES OF SEPARABILITY

The essential argument in this paper is:

Theorem 4: Let $U = U_1 \otimes U_2$ be a tensor product of positive reps., where the mass operator of U_2 is a multiple of the identity operator. Let W be a bounded operator that commutes with $U(\mathbf{a})$ and with $\mathbf{b}\cdot\mathbf{X}$ for all $\mathbf{a}, \mathbf{b} \in R^3$. Suppose that there is a bounded operator W_∞ such that

$$(1 \otimes U_2(d\mathbf{a}))\, W (1 \otimes U_2(d\mathbf{a}))^{-1} \psi \to W_\infty \psi \quad \text{as} \quad d \to \infty,$$

for all $\mathbf{a} \in R^3 \setminus \{0\}$ and all $\psi \in H$. Then W_∞ commutes with $U(a)$ for all $a \in R^4$.

Proof: Choose $\mathbf{a}, \mathbf{b} \in R^3 \setminus \{0\}$ and for all $d \in R$ put $T(d) := 1 \otimes U_2(d\mathbf{a})$ and $W_d := T(d) W T(-d)$. Since W commutes with the self-adjoint operator $\mathbf{b}\cdot\mathbf{X}$, it also commutes with the bounded operator $(d^{-1}\mathbf{b}\cdot\mathbf{X} + i)^{-1}$. Consequently, W_d commutes with $R_d := T(d)(d^{-1}\mathbf{b}\cdot\mathbf{X} + i)^{-1} T(-d) = (d^{-1} T(d)\mathbf{b}\cdot\mathbf{X} T(-d) + i)^{-1} = (d^{-1}\mathbf{b}\cdot\mathbf{X} - \mathbf{b}\cdot\mathbf{Z}(\mathbf{a}) + i)^{-1}$, where $\mathbf{b}\cdot\mathbf{Z}(\mathbf{a})$ is the bounded operator introduced in Theorem 2. As is easily shown, we have $R_d \to (-\mathbf{b}\cdot\mathbf{Z}(\mathbf{a}) + i)^{-1}$ as $d \to \infty$. Since $\|W_d\| \leq \|W\|$ and $\|R_d\| \leq 1$ for all $d \in R$, we obtain $0 = \lim[W_d, R_d] = [W_\infty, \lim R_d]$ for $d \to \infty$. Therefore, and since $\mathbf{b}$ was arbitrary, W_∞ commutes with the three components of $\mathbf{Z}(\mathbf{a})$. Further, the bounded operators $P^j(\mathbf{P}\cdot\mathbf{P})^{-1/2} = P^j |\mathbf{P}|^{-1}$, $j \in \{1,2,3\}$, commute with W and with $T(d)$ and hence with W_∞. Taken together, these statements yield that W_∞ commutes with the bounded operator $\mathbf{P}\cdot\mathbf{Z}(\mathbf{a})\,|\mathbf{P}|^{-1} = (\mathbf{a}\cdot\mathbf{P})\,|\mathbf{P}|^{-1} E_2 E^{-1}$. Since the operator $(\mathbf{a}\cdot\mathbf{P})\,|\mathbf{P}|^{-1}$ is injective, we easily conclude that W_∞ commutes with $E_2 E^{-1}$. To proceed further, we observe that W_∞ commutes with $1 \otimes U_2(\mathbf{c})$ for all $\mathbf{c} \in R^3$ and hence, since the mass operator M_2 of $1 \otimes U_2$ is a multiple of the identity, with the bounded operator $E_2^{-1} = (M_2^2 + \mathbf{P}_2^2)^{-1/2}$. Consequently W_∞ commutes with $E_2^{-1} E_2 E^{-1}$ and hence with E and with $U(t)$ for all $t \in R$. Thus the theorem is proved.

For applying Theorem 4 to a scattering theory as in Sec. I it is convenient to keep the rep. U fixed and to consider the set of all the wave operators that are obtained if U' varies over all reps. which are allowed by the conditions (c)(1), (c)(2), (d)(2) in Sec. I. These operators are characterized anew in point (iv) of the following definition. To be able to state some facts in due generality, in points (i), (ii), and (iii) three more general kinds of asymptotic operators are introduced.

Definition 2: Let U be a positive rep. and W a bounded operator on H. Then W is said to be

(i) a generalized wave operator for U, if there is a self-adjoint operator E' on H and a Borel function $f: R^4 \to R$ such that $\exp(-itE') \exp[itf(E, P^1, P^2, P^3)] \to W$ as $t \to \infty$;

(ii) a wave operator for U if there is a positive rep. U' and $\epsilon \in \{+, -\}$ such that

$$U'(-t)U(t) \to W \quad \text{as} \quad t \to \epsilon\infty \quad \text{and} \quad U'(g)W = WU(g)$$

for all $g \in \mathcal{P}$;

(iii) an instant form wave operator for U if (ii) is satisfied with U' being such that U' and U coincide as reps. of the Euclidean group $\mathcal{E} := \{(a,A) \in \mathcal{P} : a^0 = 0, A \in \mathrm{SU}(2)\}$;

(iv) a barycentric wave operator for U if (ii) is satisfied with U being such that U' and U coincide as reps. of $\mathcal{E}$ and $\mathbf{X}' = \mathbf{X}$.

Obviously, any wave operator is a generalized wave operator (for fixed U). Two lemmas will give the facts that will be needed about these operators.

Lemma 1: Let W be a generalized wave operator for U. Suppose that W commutes with $U(a)$ for all $a \in R^4$. Then W is the identity operator.

Proof: Put $f(E, \mathbf{P}) =: F$ and $W(t) := \exp(-itE')\exp(itF)$. For all $\tau \in R$ we have $\exp(itE')W(\tau) = \exp[-i(\tau - t)E'] \times \exp(itF) = W(\tau - t)\exp(itF)$ and, by $\tau \to \infty$, $\exp(itE')W = W\exp(itF)$. Since W commutes with $U(a)$, it commutes with $\exp(itF)$. Thus $W^2 = \lim \exp(-itE')\exp(itF)W = \lim \exp(-itE')W\exp(itF) = \lim \exp(-itE')\exp(itE')W = W$ for $t \to \infty$. Finally, since W as a limit of unitary operators is isometric, we have $1 = W^*W = W^*W^2 = (W^*W)W = W$.

Lemma 2: Let W_1 and W_2 be wave operators for U_1 and U_2 resp. and let the time direction ϵ in condition (ii) of Definition 2 be the same for W_1 and W_2. Then $W_1 \otimes W_2$ is a wave operator for $U_1 \otimes U_2$.

Proof: Let U_j' be such that condition (ii) of Definition 2 is satisfied for W_j, U_j', U_j, $j \in \{1,2\}$. Put $U := U_1 \otimes U_2$, $U' := U_1' \otimes U_2'$, $W := W_1 \otimes W_2$. Obviously $U'(g)W = WU(g)$ for all $g \in \mathcal{P}$. For all $\varphi_1 \in H_1$ and $\varphi_2 \in H_2$ we easily show $U'(-t)U(t)\varphi_1 \otimes \varphi_2 \to W\varphi_1 \otimes \varphi_2$ as $t \to \epsilon\infty$. Since $U'(-t)U(t)$ is unitary and the $\varphi_1 \otimes \varphi_2$'s form a total set in $H_1 \otimes H_2$, we find $U'(-t)U(t) \to W$ as $t \to \epsilon\infty$. Thus the lemma is proved.

Now we are able to prove the main result.

Theorem 5: Let $U = U_1 \otimes U_2$ be a tensor product of positive reps., where the mass operator of U_2 is a multiple of the identity operator. Let W be a barycentric wave operator for U and let W_∞ be a generalized wave operator for U. Finally, suppose

$$(1 \otimes U_2(d\mathbf{a}))W(1 \otimes U_2(d\mathbf{a}))^{-1}\psi \to W_\infty \psi \quad \text{as} \quad d \to \infty,$$

for all $\mathbf{a} \in R^3 \setminus \{0\}$ and all $\psi \in H$.

Then W_∞ is the identity operator.

Proof: Theorem 3 shows that W commutes with $U(\mathbf{a})$ and $\mathbf{b}\cdot\mathbf{X}$ for all $\mathbf{a}, \mathbf{b} \in R^3$. By Theorem 4, W_∞ commutes with $U(a)$ for all $a \in R^4$ and Lemma 1 shows $W_\infty = 1$.

Theorem 6: The assertion of Theorem 5 remains valid if all assumptions concerning W_∞ are replaced by:

(i) W_∞ is of the form $W_{\infty,1} \otimes W_{\infty,2}$, where $W_{\infty,i}$ are instant form wave operators for U_i with the same time direction (see Lemma 2);

(ii) $(W - W_\infty)(1 \otimes U_2(d\mathbf{a}))\psi \to 0$ as $d \to \infty$, for all $\mathbf{a} \in R^3 \setminus \{0\}$ and all $\psi \in H$.

Proof: By Lemma 2 we infer from (i) that W_∞ is a

wave operator for U. Further, by (i), W_∞ commutes with $1\otimes U_2(\mathbf{a})$ for all $\mathbf{a}\in R^3$. Thus (ii) implies that the limit condition in Theorem 5 is satisfied. Hence the assumptions of Theorem 5 are satisfied.

Obviously, Theorem 6 implies the statement (g) in Sec. I.

ACKNOWLEDGMENTS

I am grateful to Professor H. J. Meister for his constant interest and for valuable suggestions, to Professor R. Haag for having drawn my attention to the cluster problem, to Professor F. Coester for many valuable comments on previous versions of this paper, and to Dr. R. H. Brennich and Dr. D. Castrigiano for many clarifying discussions. The financial support of the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

[1]The following conventions are adopted throughout this paper: The terms Hilbert space, operator on H, bounded operator on H, and representation (or rep.) are used respectively instead of complex separable Hilbert space, H-valued linear mapping defined on a linear subset of H, continuous linear mapping $H\to H$, and continuous unitary representation. Operator limits, if not specified otherwise are strong limits. The symbols 0 and 1 besides denoting numbers, may denote zero vectors and identity operators respectively. $\mathcal{P}$ denotes the quantum mechanical Poincaré group without inversions, i.e., the topological space $R^4\times \mathrm{SL}(2,C)$ together with the law of multiplication $(a,A)(a',A')=(a+Aa',AA')$, where the action of $\mathrm{SL}(2,C)$ on R^4 is given by $(Ax)^\mu=\frac{1}{2}\mathrm{Tr}(A(x^0+\mathbf{x}\cdot\sigma)A^*\sigma^\mu)$ for all $x=(x^0,x^1,x^2,x^3)=(x^0,\mathbf{x})\in R^4$, $\mu\in\{0,1,2,3\}$, where $\sigma^0=1$ and σ is the triple of Pauli matrices. The elements $(t,\mathbf{0},1)$, $(a^0,\mathbf{a},1)$, and $(0,\mathbf{a},1)$ of $\mathcal{P}$ are abbreviated as t, a, and $\mathbf{a}$ resp., if they appear as arguments of a representation.

[2]P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[3]B. Bakamjian and L. H. Thomas, Phys. Rev. **92**, 1300 (1953).
[4]L. L. Foldy, Phys. Rev. **122**, 275 (1961).
[5]R. Fong and J. Sucher, J. Math. Phys. **5**, 456 (1964).
[6]T. F. Jordan, A. J. Macfarlane, and E. C. G. Sudarshan, Phys. Rev. **133**, B487 (1964).
[7]L. B. Rédei, J. Math. Phys. **6**, 487 (1965).
[8]F. Coester, Helv. Phys. Acta **38**, 7 (1965).
[9]H. Osborn, Phys. Rev. **176**, 1514 (1968).
[10]M. K. Liou, Phys. Rev. D **9**, 1091 (1974).
[11]R. A. Krajcik and L. L. Foldy, Phys. Rev. D **10**, 1777 (1974).
[12]L. L. Foldy and R. A. Krajcik, Phys. Rev. D **12**, 1700 (1975).
[13]F. Coester, S. C. Pieper, and F. J. D. Serduke, Phys. Rev. C **11**, 1 (1975).
[14]F. Coester and A. Ostebee, Phys. Rev. C **11**, 1836 (1975).
[15]F. Coester and P. Havas, Phys. Rev. D **14**, 2556 (1976).
[16]S. N. Sokolov, Theor. Math. Phys. **24**, 799 (1976).
[17]S. N. Sokolov, Serpukov preprint IHEP 74-133.
[18]S. N. Sokolov, Serpukov preprints IHEP 74-134, 135 and further papers of Sokolov cited there.
[19]M. Pauri and G. M. Prosperi, J. Math. Phys. **17**, 1468 (1976).
[20]This was recognized by the author only after many unsuccessfull attempts to construct n-particle models with two-body forces satisfying these axioms.
[21]See the comments on (d), especially Ref. 31.
[22]J. M. Jauch, Helv. Phys. Acta **31**, 661 (1958).
[23]W. Brenig and R. Haag, Fortschr. Phys. **7**, 183 (1959).
[24]Cf. Ref. 5, Sec. 5, Ref. 6, Eqs. (16) and (17).
[25]See Ref. 8, Eqs. (13)–(15). See also Ref. 16, Eqs. (5)–(7).
[26]Cf. Ref. 3, Eq. (3.12), Ref. 4, Eqs. (74)–(75), and Ref. 8, Sec. II.
[27]The most general form of such transformations is determined in A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962), below Theorem 3.
[28]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed., pp. 545 and 579.
[29]By Proposition 4 (ii) on p. 160 of J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1969).
[30]It seems worthwhile to note that a rep. of $\mathcal{P}$, without being supplemented by an enumeration of observables or something like this, is too poor a structure for describing interaction or dynamics.
[31]In Ref. 4 a conjecture is formulated concerning the generality of the solution discussed in Sec. VII of this reference. If this conjecture were true one would easily derive the incorrect equality $S_1(U)=S_2(U)$. Therefore, the very fact that separability excludes U' from belonging to $S_2(U)$ is only indirectly inferred from this reference.
[32]The relation between U' and U given by this condition arises if U' and U are obtained by inducing (in the sense of Mackey) from such reps. V' and V of the subgroup $\{(a,A):\mathbf{a}=0, A\in \mathrm{SU}(2)\}$ that satisfy $V'(0,A)=V(0,A)$ for all $A\in\mathrm{SU}(2)$. This relation was proposed and studied in 1974 by R. H. Brennich (unpublished).
[33]If the particles are not pairwise distinguishable, only the eigenvectors of certain symmetrization operators represent physical states.
[34]See Sec. 7 of Ref. 22.
[35]See. Eq. (19) of Ref. 8.
[36]For a proof of this property in the framework of nonrelativistic n-particle scattering see W. Hunziker, J. Math. Phys. **6**, 6 (1965).
[37]Sokolov's form in Ref. 17 is not confined to the single channel case and also takes into account separation in a timelike direction.
[38]Let $\mathbf{A}=(A^1,A^2,A^3)$ be a triple of operators on H and $\mathbf{a}\in R^3$. Then $\mathbf{a}\cdot\mathbf{A}$ denotes the operator $a^1A^1+a^2A^2+a^3A^3$ if this operator is not closable, and the closure of this operator otherwise.

Integral equations for even parity perturbations of hot perfect fluid nonrotating neutron stars[a)]

P. Cazzola

Istituto di Fisica dell'Università, 35100 Padova, Italy

L. Lucaroni

Istituto di Fisica dell'Università, 44100 Ferrara, Italy

R. Semenzato[b)]

Laboratorio di Radioastronomia del C.N.R., Istituto di Fisica dell'Università, 40100 Bologna, Italy and Unità di Ricerca del C.N.R. di Padova-Asiago, Istituto di Fisica and Istituto di Astronomia, Università di Padova, 35100 Padova, Italy
(Received 12 April 1977)

In the present paper, in the framework of the linearized Einstein equations, systems of two integral equations are set up, which describe coupled sound and gravitational waves associated with small perturbations of hot nonrotating neutron stars made of a perfect fluid. Suitable majorizations are given for the successive iterations in view of obtaining the asymptotic distribution of large eigenfrequencies.

1. INTRODUCTION

In recent years a great deal of attention was devoted to the problem of small perturbations of highly collapsed stars in the framework of general relativity, and also in connection with convective and vibrational stability.[1-12]

The perturbations are usually decomposed into multipoles and are treated in the framework of linearized Einstein equations, even if the question of the legitimacy of such linearization is still open.[13-17] In the present context we will consider nonrotating neutron stars, consisting of a perfect fluid, with a polytropic behavior of the equation of state near the surface.

In 1973, Detweiler and Ipser,[18] gave sufficient conditions for the stability of single eigenmodes for such configurations. However, it is not sufficient to make sure that the eigenfrequencies ω_n do not lie in the lower complex ω plane, since stability with respect to an arbitrary perturbation also requires completeness of eigenmodes, which is still an open question; as far as we know, this is true not only for neutron stars, but also for rotating black holes.[19,20]

It seems plausible that the solution of the problem of completeness requires some information about the distribution of the eigenfrequencies ω_n in the complex ω plane, in particular for $\omega \to \infty$, and about possible accumulation points. Such an analysis may also be of interest in connection with other problems, i.e., emission of gravitational waves. The purpose of this paper, together with successive ones, is to determine the asymptotic distribution, for $\omega \to \infty$, of the eigenfrequencies in the ω plane. To this aim, by means of a considerable amount of work, we write a system of coupled differential equations for the Lagrangian variation of the pressure and for a suitable gravitational amplitude. This system is of the second order in each variable and describes coupled sound and gravitational waves. The equations were obtained in the framework of a gauge first introduced by two of us[21]; this result cannot be achieved in the Regge–Wheeler gauge, since, as shown by Thorne and Campolattaro,[22] in that gauge one obtains a fifth order system. Successively, the differential equations, together with the boundary conditions at $r=0$, are transformed into two systems of coupled integral equations. In the star's interior the first system, whose amplitudes are labeled with the letter "g," is obtained from a zero order iteration which essentially describes a purely gravitational wave not coupled to matter. On the other hand, the second system, whose amplitudes are labeled with the letter "s," is obtained from a zero order iteration which describes a pure sound wave not coupled to the gravitational field. The physical solution (which is determined up to an arbitrary multiplicative constant) is given by a suitable linear combination of the above solutions, satisfying the condition that the Lagrangian variation of the pressure vanishes on the surface. The iterative series for solutions of type "s" or of type "g" are here suitably majorized in view of obtaining the distribution of the eigenfrequencies for $\omega \to \infty$. These indeed are related to the zeroes of the Wronskian constructed with the internal and external gravitational amplitudes.[23,24] Our treatment is in some respects similar to the one described by Newton[25] regarding the analytical properties of eigenfunctions and eigenfrequencies of the Schrödinger equation in potential theory. However, it should be noted that particular difficulties arise in the present problem from the fact that the sound velocity, the energy density and the pressure, expressed as functions of the distance r from the star center, in general vanish as a power of $r_0 - r$, where r_0 is the star radius. These difficulties lead us to dividing the interval $(0, r_0)$ into two parts: $(0, \bar{r})$ and $(\bar{r}, r_0)$, where different analytical expressions hold for the zero order iteration of type "s" and for a kernel appearing in the integral equations. These expressions, however, tend to each other in the

[a)]Work supported in part by the Consiglio Nazionale delle Ricerche.
[b)]Permanent address: Istituto di Fisica "G. Galilei" Via Marzolo, 8 35100 Padova, Italy.

 0022-2488/78/1901-0237$1.00

limit $\omega \to \infty$, whatever may be the value of $\bar{r}$, provided that $\bar{r} \neq 0, \neq r_0$, and is independent of ω.

The plan of this paper is the following: In Sec. 2 we sketch the procedure followed by us in order to obtain the coupled differential equations. These are transformed into integral equations of type "g" and "s" in Sec. 3. In Sec. 4 suitable majorizing quantities for the iterations and series of iterations are obtained, in which the ω dependence is explicitly exhibited. In Sec. 5 a similar procedure is carried out for the external solution, in the upper ω plane.

2. THE COUPLED EQUATIONS

In the framework of a given frequency ω, for a given multipole l, let us introduce the amplitude η defined by

$$\Delta p = \eta(r)\, P_l(\cos\theta) \exp(i\omega t), \tag{2.1}$$

where Δp is the Lagrangian variation of the pressure; η is connected to the amplitude α for the Eulerian variation of the pressure by

$$\eta = \alpha + p'\xi, \tag{2.2}$$

where ξ is the amplitude for the radial component of the fluid displacement defined by

$$\xi^r = \xi(r)\, P_l(\cos\theta) \exp(i\omega t). \tag{2.3}$$

$\eta(r)$ must satisfy the condition[22]:

$$\eta(r_0) = 0, \tag{2.4}$$

where r_0 is the star radius.

Let us further introduce the gravitational amplitude

$$u = z r^{-2} = r^{-2}[\exp(-\nu)\, y - \exp(-\lambda)\,((\nu'/2) + r^{-1}) h_1], \tag{2.5}$$

where y and h_1 have been defined in Ref. 21.

Here we want to obtain two coupled equations for the amplitudes u and η. In order to do this, first consider Eqs. (1.9), (1.10), and (B4) of Ref. 21. By algebraic elimination of the amplitudes H, H_2 one obtains

$$\left(\frac{r^{-2}}{2}\, l(l+1)[\exp(\lambda)\sigma\varphi - 2r^{-2}] - \varphi \exp(-\nu)\,\omega^2 r^{-1}\right) h_1$$
$$+ [r^{-2}\sigma l(l+1) - 2r^{-1}\exp(-(\lambda+\nu))\,\omega^2]\, h_1'$$
$$+ 2r^{-1}\exp(-2\nu)\,\omega^2 y - l(l+1)\, r^{-2}\exp(-\nu)\, y'$$
$$+ 8\pi\omega^2(\rho+p)\exp(\lambda-\nu)\,\xi = 0. \tag{2.6}$$

Moreover, consider the following combination of the above equation with Eq. (B2) of Ref. 21:

$$\frac{\nu'}{2}\exp(-\lambda)(2.6) - r^{-2}\omega^2\exp(-\nu)(\mathrm{B2}) = 0, \tag{2.7}$$

which reads explicitly:

$$\exp(-\lambda)[\nu' r^{-1} l(l+1) - 4\omega^2\exp(-\nu)]\, u'$$
$$- [2\omega^2\exp(-(\lambda+\nu))(3\nu' + 2r^{-1}) + r^{-1} l(l+1)$$
$$\times (2\omega^2\exp(-\nu) - (\nu')^2\exp(-\lambda) - 2r^{-1}\nu'\exp(-\lambda))]\, u$$
$$+ 16\pi r^{-1}\exp(-\nu)\,\omega^2\eta - \omega^4 r^{-2} F h_1 = 0, \tag{2.8}$$

where

$$F = 4\exp(-\lambda + 2\nu) + \omega^{-2}[(\nu')^2\exp(-\lambda) + 32\pi p$$
$$+ 8r^{-2}(1 - \exp(-\lambda))]\exp(-(\lambda+\nu))$$
$$- l(l+1)\,\omega^{-4}(\nu')^2 r^{-2}\exp(-2\lambda). \tag{2.9}$$

Now, from (1.17) of Ref. 21 and (2.8) one obtains

$$u'' + \left(\frac{5}{2}\nu' - \frac{\lambda'}{2} + 2r^{-1}\frac{\Phi}{\omega^4 F}\left[\nu' l(l+1) r^{-1} - 4\omega^2\exp(-\nu)\right]\right) u'$$
$$+ \Big(r^{-1}(\lambda' + 5\nu' + 2r^{-1}) + (\nu')^2 - 2r^{-2}\exp(\lambda) + \omega^2\exp(\lambda-\nu)$$
$$- r^{-2}\exp(\lambda)\, l(l+1) + \frac{\Phi}{\omega^4 F}\,[r^{-1} l(l+1)(2r^{-1}\nu' - (\nu')^2$$
$$+ 2\omega^2\exp(\lambda-\nu)) - 2\omega^2\exp(-\nu)(3\nu' + 2r^{-1})]\Big)\, u$$
$$+ \frac{\Phi r^{-1}}{F}\, 16\pi\omega^{-2}\exp(\lambda-\nu)\,\eta = 0. \tag{2.10}$$

The above equation is one of the two coupled equations in the variables u and η.

In order to get the other it is necessary to write down the equations of motion $\delta(T_{ik}{}^{;k}) = 0$ which for $i = 1, 2, 4$, after elimination of angular and time dependent factors, explicitly reads

$$(\rho+p)\exp(-\nu)\, H + (\rho+p)\,\omega^2\exp(\lambda-\nu)\,\xi$$
$$- (\nu'/2)(\alpha+\beta) - \alpha' = 0, \tag{2.11}$$

$$(\rho+p)\exp(-\nu)\, y + (\rho+p)\,\omega^2 r^2\exp(-\nu)\,\zeta - \alpha = 0, \tag{2.12}$$

$$\beta + [\rho' + (\rho+p)(2r^{-1} + \lambda'/2)]\,\xi + (\rho+p)\,\xi'$$
$$- l(l+1)(\rho+p)\,\zeta - (\rho+p)\, H_2/2 = 0, \tag{2.13}$$

where H_2 and H have been defined in Ref. 21, ζ, the azimuthal amplitude of the fluid displacement, is defined by

$$\xi^\theta = \zeta(r)\,\partial_\theta(P_l(\cos\theta))\exp(i\omega t), \tag{2.14}$$

and β is the amplitude for the Eulerian variation of the energy density.[21]

By algebraic elimination of ζ, α, y from (2.2), (2.5), (2.12), and (2.13), we obtain

$$\frac{\nu'}{2}(\rho+p)\left[\frac{H_2}{2} - (\rho+p)\,\xi' - \left(2r^{-1} + \frac{\lambda'}{2}\right)\xi\right.$$
$$\left. - l(l+1)\,\omega^{-2}\exp(\nu)(u + r^{-2}\sigma h_1)\right] + \frac{\nu'}{2}\, l(l+1)\, r^{-2}$$
$$\times\exp(\nu)\,\omega^{-2}(\eta - p'\xi) - \beta\frac{\nu'}{2}\,\rho'\xi = 0 \tag{2.15}$$

and by algebraic elimination of α, ξ' from (2.11), (2.15), and the derivative of (2.2) with respect to r, we obtain

$$\eta' + \left[\frac{\nu''}{2}(\rho+p) + p'\left(2r^{-1} + \frac{\lambda'}{2}\right) - \omega^2\exp(\lambda-\nu)(\rho+p)\right.$$
$$\left. - \frac{\nu'}{2}\, l(l+1)\, r^{-2}\omega^{-2}\exp(\nu)\, p'\right]\,\xi$$
$$+ \frac{\nu'}{2}\,(1 + l(l+1)\,\omega^{-2} r^{-2}\exp(\nu))\,\eta$$

$$-l(l+1)r^{-2}\omega^{-2}\frac{\nu'}{2}(\rho+p)y$$

$$-(\rho+p)\exp(-\nu)H+\frac{\nu'}{4}(\rho+p)H_2=0. \tag{2.16}$$

Moreover, the following combination of Eqs. (1.10)–(1.13) of Ref. 21,

$$\exp(\nu)\left(\frac{r}{4}[(1.11)+\exp(\lambda)(1.12)]-\frac{r^{-1}}{2}(1.13)\right.$$

$$\left.-\frac{l(l+1)}{4}r^{-1}\exp(\lambda)(1.10)\right)=0, \tag{2.17}$$

together with (2.5), gives

$$H=\frac{\exp(\nu)}{2}r^{-1}(1+\nu' r)H_2+\frac{l(l+1)}{2}r\exp(\lambda+\nu)u$$

$$-4\pi r\exp(\lambda+\nu)\alpha. \tag{2.18}$$

By algebraic elimination of H, y, and α from (2.2), (2.5), (2.16), and (2.18) we then get

$$\eta'+\left(\nu'+\frac{\lambda'}{2}+\frac{\nu'}{2}l(l+1)\omega^{-2}r^{-2}\exp(\nu)\right)\eta$$

$$+\left(\frac{\nu''}{2}-\omega^2\exp(\lambda-\nu)+\frac{(\nu')^2}{4}-\nu' r^{-1}+\frac{(\nu')^2}{4}l(l+1)r^{-2}\omega^{-2}\right.$$

$$\left.\times\exp(\nu)\right)(\rho+p)\xi-(\rho+p)\frac{r^{-1}}{2}\left(1+\frac{\nu' r}{2}\right)H_2-\frac{l(l+1)}{2}$$

$$\times(\rho+p)(\nu'\omega^{-2}\exp(\nu)+r\exp(\lambda))u$$

$$-\frac{\nu'}{2}l(l+1)\exp(\nu-\lambda)\left(\frac{\nu'}{2}+r^{-1}\right)(\rho+p)\omega^{-2}h_1=0. \tag{2.19}$$

Now the Eulerian amplitude β can be expressed in terms of η and ξ as follows,[22]

$$\beta=\frac{\eta}{v_s^2}-\rho'\xi, \tag{2.20}$$

where v_s is the velocity of sound.

By algebraic elimination of β and $H+4\pi r[\exp(\lambda+\nu)]\alpha$ from (1.13) of Ref. 21 and from (2.18) and (2.20), one gets

$$H_2'=2r^{-2}l(l+1)h_1+r\exp(\lambda)l(l+1)u+(\lambda'-2r^{-1})H_2$$

$$-8\pi r\exp(\lambda)(v_s^{-2}\eta-\rho'\xi). \tag{2.21}$$

In addition, by algebraic elimination of α, y, y', and h_1' from (B2) and (1.10) of Ref. 21, (2.2), (2.5), and its derivative with respect to r, one gets

$$\exp(-\lambda)[16\pi p-2\omega^2\exp(-\nu)-\nu'(\sigma+3r^{-1}\exp(-\lambda))]h_1$$

$$+8\pi r\eta+4\pi\nu' r(\rho+p)\xi-2r^2\exp(-\lambda)u'$$

$$-[2\exp(-\lambda)+3\nu' r\exp(-\lambda)-l(l+1)]ru$$

$$-\frac{\nu'}{2}\exp(-\lambda)H_2=0. \tag{2.22}$$

Finally by algebraic elimination of h_1, h_1', H_2, H_2', ξ, and ξ' from (1.10) of Ref. 21, and from (2.8), (2.15), (2.21), (2.22), (2.19), and its derivative with respect to r, one gets

$$\eta''+\left\{\nu'-\frac{\lambda'}{2}+2r^{-1}-\frac{\rho'+p'}{\rho+p}+4\frac{\exp[-(2\lambda+\nu)]}{F}\omega^{-2}\right.$$

$$\times\left[\frac{\nu'}{2}\omega^2\exp(\lambda-\nu)-\nu''\left(2r^{-1}+\frac{\nu'}{4}\right)\right.$$

$$-(\nu')^2\left(\frac{5}{4}r^{-1}-\frac{\lambda'}{8}\right)+\nu'\left(\tfrac{5}{4}\lambda' r^{-1}+2r^{-2}\right)$$

$$\left.-l(l+1)\exp(\nu)\omega^{-2}r^{-2}\nu'\left(\frac{\nu'\lambda'}{4}-\frac{3}{8}(\nu')^2-\frac{\nu''}{2}+\frac{\nu' r^{-1}}{2}\right)\right]\left.\right\}\eta'$$

$$+\left\{\frac{\nu''}{2}+\frac{3}{2}r^{-1}(2\nu'+\lambda')-\frac{\nu'}{2}\left(r^{-1}+\lambda'+\frac{\rho'+p'}{\rho+p}\right)\right.$$

$$-\omega^{-2}r^{-2}l(l+1)\exp(\nu)\left(\frac{\rho'+p'}{\rho+p}\frac{\nu'}{2}+\frac{\nu'}{2}(\nu'+\lambda')\right.$$

$$\left.\left.-\lambda' r^{-1}+2r^{-2}(\exp(\lambda)-1)\right)-\exp(\lambda)l(l+1)r^{-2}\right.$$

$$+v_s^{-2}\left[\omega^2\exp(\lambda)G+\frac{\nu'+\lambda'}{2}\left(r^{-1}+\frac{\nu'}{2}\right)\right]$$

$$+\frac{1}{F}\left(\frac{4\exp(-\lambda)}{\omega^2\exp(\nu+\lambda)}\left(\nu'+\frac{\lambda'}{2}+\frac{\nu'}{2}l(l+1)\exp(\nu)\omega^{-2}r^{-2}\right)\right.$$

$$\times\left[\nu'\left(\tfrac{5}{4}\lambda' r^{-1}+2r^{-2}\right)-\nu''\left(2r^{-1}+\frac{\nu'}{4}\right)-(\nu')^2\left(\frac{5}{4}r^{-1}-\frac{\lambda'}{8}\right)\right.$$

$$-l(l+1)\exp(\nu)\omega^{-2}r^{-2}\nu'\left(\frac{\nu'\lambda'}{4}-\frac{3}{8}(\nu')^2-\frac{\nu''}{2}+\frac{\nu' r^{-1}}{2}\right)$$

$$\left.+\frac{\nu'}{2}\omega^2\exp(\lambda-\nu)\right]+16\pi(\rho+p)r\exp(-\nu)\omega^{-2}$$

$$\times\left[r^{-2}\nu' l(l+1)+\omega^{-2}r^{-2}\exp(\nu)l(l+1)\right.$$

$$\times\left(\nu' r^{-2}+(\nu')^2 r^{-1}\exp(-\lambda)-\sigma\left(\nu''+\frac{(\nu')^2}{2}-\frac{\nu'\lambda'}{2}\right)\right)$$

$$\left.\left.-\frac{(\nu')^3}{4}l^2(l+1)^2\omega^{-4}r^{-4}\exp(2\nu-\lambda)\right]\right)\left.\right\}\eta$$

$$+l(l+1)\omega^{-4}(\rho+p)F^{-1}\left[\exp(-2\lambda)\left\{\nu''\left[3(\nu')^2\right.\right.\right.$$

$$-4r^{-1}\left(\nu'+\frac{\lambda'}{2}+r^{-1}\right)-2\nu'\lambda'\left.\right]+\nu' r^{-1}(\lambda'+2r^{-1})^2$$

$$\left.+(\nu')^2\left[(\nu')^2+\frac{(\lambda')^2}{2}-3\nu' r^{-1}-\tfrac{3}{2}\nu'\lambda'-2r^{-2}\right]\right\}$$

$$\left.+\exp(-\lambda)r^{-2}l(l+1)(2\nu''-2\nu' r^{-1}-\lambda'\nu'+2(\nu')^2)\right]u=0, \tag{2.23}$$

where

$$G=\exp(-\nu)-\omega^{-2}\exp(-\lambda)\left[\frac{\nu''}{2}-\nu' r^{-1}\right.$$

$$\left.+\frac{(\nu')^2}{4}\omega^{-2}r^{-2}\exp(\nu)l(l+1)\right]. \tag{2.24}$$

The above equation and (2.10) form a system of coupled equations in the variables u, η.

A remarkable feature of the system is that there are no coupling terms in u', η' since they cancel out.

This fact strongly simplifies the whole treatment.

3. INTEGRAL EQUATIONS FOR THE INTERNAL SOLUTION

In the following we will consider hot neutron stars with a polytropic behavior near the surface; in this case one has for $r \to r_0$,[22]

$$\rho \sim (r_0 - r)^N, \quad p \sim (r_0 - r)^{N+1} \quad (N > 0), \tag{3.1}$$

$$v_s = \left(\frac{\gamma p}{\rho + p}\right)^{1/2} \sim (r_0 - r)^{1/2}. \tag{3.2}$$

Equations (2.10) and (2.23) can be respectively written in the form

$$O_1 u = V_{11} u + V_{12} \eta, \tag{3.3}$$

$$O_2 \eta = V_{21} u + V_{22} \eta, \tag{3.4}$$

where O_1, O_2 are the following differential operators:

$$O_1 = \frac{d^2}{dr^2} + B_1 \frac{d}{dr} + \frac{B_1^2}{4} + \frac{B_1'}{2} - \left(\frac{\nu' - \lambda'}{4}\right)^2 - \frac{\nu'' - \lambda''}{4} + \omega^2 \exp(\lambda - \nu) - \frac{l(l+1)}{W_g^2} \exp(\lambda - \nu), \tag{3.5}$$

B_1 being the coefficient of u' in (2.10),

$$W_g = \int_0^r \exp\left(\frac{\lambda - \nu}{2}\right) dr', \tag{3.6}$$

$$O_2 = \frac{d^2}{dr^2} + B_2 \frac{d}{dr} + \frac{B_2^2}{4} + \frac{B_2'}{2} - \left(\frac{\nu' - \lambda'}{4} + \frac{v_s'}{2v_s}\right)^2 - \frac{\nu'' - \lambda''}{4} - \frac{v_s''}{2v_s} + \frac{1}{2}\left(\frac{v_s'}{v_s}\right)^2 + \omega^2 \frac{\exp(\lambda - \nu)}{v_s^2} - \left(\theta(\bar{r} - r) \frac{l(l+1)}{W_s^2} + \theta(r - \bar{r}) \frac{\bar{l}(\bar{l}+1)}{(W_s - W_0)^2}\right) \frac{\exp(\lambda - \nu)}{v_s^2}, \tag{3.7}$$

$$W_s = \int_0^r \exp\left(\frac{\lambda - \nu}{2}\right) \frac{dr'}{v_s}, \quad W_0 = W_s(r_0), \tag{3.8}$$

B_2 being the coefficient of η' in (2.23), $\theta(x)$ the step function, $\bar{r}$ an arbitrary fixed ω-independent value of r, such that $0 < \bar{r} < r_0$,

$$\bar{l} = N + \tfrac{1}{2}, \tag{3.9}$$

$$V_{11} = l(l+1)\exp(\lambda)[r^{-2} - \exp(-\nu) W_g^{-2}] - \frac{r^{-1}}{2}(\lambda' + 3\nu') + \left(\frac{\Phi}{2\omega^4 F}[l(l+1)\nu' r^{-1} - 4\omega^2 \exp(-\nu)]\right)^2 + \left(\frac{\Phi}{2\omega^4 F}[l(l+1)\nu' r^{-1} - 4\omega^2 \exp(-\nu)]\right)' + \frac{\Phi}{2\omega^4 F}\left[l(l+1) r^{-1}\left(\frac{9}{2}(\nu')^2 - \frac{\lambda'\nu'}{2} - 2\nu' r^{-1} - 4\omega^2 \exp(\lambda - \nu)\right) + 2\omega^2 \exp(-\nu)(\nu' + \lambda')\right], \tag{3.10}$$

$$V_{12} = -D_1, \quad V_{21} = -D_2, \tag{3.11}$$

D_1, D_2 being respectively the coefficients of η in (2.10) and of u in (2.23);

$$V_{22} = \left[\frac{3}{4}\left(\frac{p' + \rho'}{p + \rho}\right)^2 - \frac{1}{2}\frac{p'' + \rho''}{p + \rho} + \frac{1}{4}\left(\frac{v_s'}{v_s}\right)^2 - \frac{v_s''}{2v_s} - \theta(r - \bar{r}) \frac{\bar{l}(\bar{l}+1)\exp(\lambda - \nu)}{v_s^2 (W_s - W_0)^2}\right] - \left[\frac{\rho' + p'}{\rho + p} r^{-1} + \frac{\tilde{B}_2}{2}\frac{\rho' + p'}{\rho + p} + \frac{(\nu' - \lambda')v_s'}{4v_s}\right] + \left[\tilde{B}_2 r^{-1} + \frac{\tilde{B}_2^2}{4} + \frac{\tilde{B}_2'}{2} - \left(\frac{\nu' - \lambda'}{4}\right)^2 - \frac{\nu'' - \lambda''}{4}\right] + \frac{\omega^2 \exp(\lambda - \nu)}{v_s^2} - \theta(\bar{r} - r) \frac{l(l+1)\exp(\lambda - \nu)}{v_s^2 W_s^2} - C_2, \tag{3.12}$$

where

$$\tilde{B}_2 = B_2 - 2r^{-1} + \frac{\rho' + p'}{\rho + p} \tag{3.13}$$

and C_2 is the coefficient of η in (2.23); we observe that B_2 is finite for $0 \le r \le r_0$ and vanishes as r for $r \to 0$. Near the surface the first four terms in the first square bracket of (3.12) diverge as $(r_0 - r)^{-2}$ owing to (3.1) and (3.2); this is the case also for the last term in the same bracket since, owing to (3.2) and (3.8),

$$(W_s - W_0)^2 \sim r_0 - r \quad (r \to r_0). \tag{3.14}$$

In addition, taking into account (3.9), it can be easily realized that these terms cancel out one another so that their sum diverges only as $(r_0 - r)^{-1}$.

The terms in the second bracket diverge also as $(r_0 - r)^{-1}$, whereas those in the third bracket are finite for $0 \le r \le r_0$; $\omega^2 \exp(\lambda - \nu)/v_s^2$ cancels exactly with a term appearing in C_2, as it can be seen from (2.23) and (2.24). Finally, taking into account the behavior of $\exp(-\nu)/W_s^2$ for $r \to 0$, the sum of $-\theta(\bar{r} - r) \times [\exp(\lambda - \nu)] l(l+1)/W_s^2 v_s^2$ with the term $r^{-2} l(l+1)\exp(\lambda)$ appearing in $-C_2$, is easily found to be finite for $0 \le r \le r_0$.

The other terms of C_2, as it can be seen by inspection, are finite for $r \ne r_0$, and diverge at most as $(r_0 - r)^{-1}$ for $r \to r_0$; it follows that the same properties hold also for V_{22}.

Observing the dependence on ω^2 of the different terms of V_{22}, one can state in addition that

$$|V_{22}| < C(r_0 - r)^{-1}, \tag{3.15}$$

where C is a positive constant, independent of r and ω.

The system of equations

$$O_1 u = 0, \quad O_2 \eta = 0 \tag{3.16}$$

allows the two solutions

$$u = u_g^{(0)} \equiv \phi_g \bar{j}_l(x_g), \quad \eta = \eta_g^{(0)} \equiv 0, \tag{3.17}$$

$$\begin{cases} \eta = \eta_s^{(0)} \equiv \phi_s[\theta(\bar{r} - r)\bar{j}_l(x_s) \\ \qquad + \theta(r - \bar{r})(\alpha_+ \bar{h}_l^{(+)}(X) + \alpha_- \bar{h}_l^{(-)}(X))], \\ u = u_s^{(0)} \equiv 0, \end{cases} \tag{3.18}$$

$$x_g = \omega W_g(r), \quad x_s = \omega W_s(r), \quad X = \omega(W_s(r) - W_0), \tag{3.19}$$

$$\phi_g = r^{-1}\exp(-\nu)\exp\left(-\frac{1}{2}\int_0^r \frac{\Phi}{\omega^4 F}[l(l+1)\nu' r^{-1} - 4\omega^2 \exp(-\nu)]\, dr'\right), \tag{3.20}$$

$$\phi_s = r^{-1}\exp[(\nu-\lambda)/4](\rho+p)^{1/2}v_s^{1/2}\exp\left(-\tfrac{1}{2}\int_0^r \tilde{B}_2\,dr'\right), \tag{3.21}$$

$$\bar{j}_\nu(z) = \left(\frac{\pi z}{2}\right)^{1/2} J_{\nu+1/2}(z), \quad \bar{h}_\nu^{(\pm)}(z) = \left(\frac{\pi z}{2}\right)^{1/2} H_{\nu+1/2}^{(\pm)}, \tag{3.22}$$

where $J_{\nu+1/2}(z)$ and $H^{(\pm)}_{\nu+1/2}(z)$ are Bessel and Hankel functions; α_+ and α_- are constants determined in such a way that $\eta_s^{(0)}$ and $(d/dr)\,\eta_s^{(0)}$ be continuous at the "junction" point $\bar{r}$. In the following (3.17), and (3.18) will be denoted as "zero order gravitational and matter field solutions," respectively; these solutions are both regular for $r=0$ but (3.18) does not satisfy in general the condition (2.4).

In addition Eqs. (3.16) also allow the irregular solutions

$$u = u_{\rm irr}^{(0)} \equiv \phi_g \bar{n}_l(x_g), \tag{3.23}$$

$$\eta = \eta_{\rm irr}^{(0)} \equiv \phi_s[\theta(\bar{r}-r)\bar{n}_l(x_s) + \theta(r-\bar{r})(\beta_+\bar{h}_l^{(+)}(X) + \beta_-\bar{h}_l^{(-)}(X))], \tag{3.24}$$

$$\bar{n}_\nu(z) = \left(\frac{\pi z}{2}\right)^{1/2} N_{\nu+1/2}(z), \tag{3.25}$$

where $N_{\nu+1/2}$ are Neumann functions and β_+ and β_- are constants determined in such a way that $\eta_{\rm irr}^{(0)}$ and $(d/dr)\,\eta_{\rm irr}^{(0)}$ be continuous at the junction point $\bar{r}$.

Note that both in (3.18) and (3.24) the divergences arising from $\bar{h}_l^{(\pm)}(X)$ (of the type $\sim X^{-\bar{l}}$) are exactly compensated by the factor $[(\rho+p)v_s]^{1/2}$, whose behavior for $r\to r_0$ is $\sim(W_s-W_0)^{\bar{l}}$, owing to (3.1), (3.2), and (3.9).

Let us consider now the following two systems of integral equations:

$$\begin{aligned} u_g &= u_g^{(0)} + \int_0^r g_1(r,r')[V_{11}u_g + V_{12}\eta_g]_{r'}\,dr', \\ \eta_g &= \int_0^r g_2(r,r')[V_{21}u_g + V_{22}\eta_g]_{r'}\,dr', \end{aligned} \tag{3.26}$$

$$\begin{aligned} u_s &= \int_0^r g_1(r,r')[V_{11}u_s + V_{12}\eta_s]_{r'}\,dr', \\ \eta_s &= \eta_s^{(0)} + \int_0^r g_2(r,r')[V_{21}u_s + V_{22}\eta_s]_{r'}\,dr', \end{aligned} \tag{3.27}$$

$$g_1(r,r') = -\frac{u_g^{(0)}(r)\,u_{\rm irr}^{(0)}(r') - u_{\rm irr}^{(0)}(r)\,u_g^{(0)}(r')}{[W(u_g^{(0)}, u_{\rm irr}^{(0)})]_{r'}}, \tag{3.28}$$

$$g_2(r,r') = -\frac{\eta_s^{(0)}(r)\,\eta_{\rm irr}^{(0)}(r') - \eta_{\rm irr}^{(0)}(r)\,\eta_s^{(0)}(r')}{[W(\eta_s^{(0)}, \eta_{\rm irr}^{(0)})]_{r'}}, \tag{3.29}$$

where $W(\psi_1,\psi_2)$ denotes the Wronskian with respect to r constructed with the functions ψ_1, ψ_2.

The solutions of (3.26) or (3.27) also satisfy the system (3.3) and (3.4). In the next section it will be shown that they may be obtained by successive iterations and that they are regular at the origin.

The solutions of (3.26) and (3.27) will be denoted repsectively as the "gravitational solution" and the "matter field solution."

The physical solutions satisfying (2.4) is obtained by the combination

$$u = u_g + \mu u_s, \tag{3.30}$$

$$\eta = \eta_g + \mu\eta_s, \tag{3.31}$$

where

$$\mu = -\frac{\eta_g(r_0)}{\eta_s(r_0)}. \tag{3.32}$$

4. MAJORIZATION OF THE ITERATIVE INTERNAL SOLUTIONS

Successive iterations can be given for the systems (3.26) and (3.27),

$$u^{(n+1)} = \int_0^r g_1(r,r')(V_{11}u^{(n)} + V_{12}\eta^{(n)})_{r'}\,dr', \tag{4.1}$$

$$\eta^{(n+1)} = \int_0^r g_2(r,r')(V_{21}u^{(n)} + V_{22}\eta^{(n)})_{r'}\,dr', \tag{4.2}$$

$$\begin{cases} u^{(0)} \equiv u_g^{(0)} \\ \eta^{(0)} \equiv 0 \end{cases} \quad \text{or} \quad \begin{cases} u^{(0)} \equiv 0 \\ \eta^{(0)} \equiv \eta_s^{(0)}. \end{cases} \tag{4.3}$$

We want to obtain suitable majorizations of the single iterations and of their sums. To this aim, in both cases, it is convenient to introduce $\tilde{u}$, $\tilde{\eta}$ defined by

$$u = \tilde{u}r^{-1}[L(|x_s|)]^{l+1}\exp(|\mathrm{Im}\omega|\,W_s), \tag{4.4}$$

$$\eta = \tilde{\eta}r^{-1}(\rho+p)^{1/2}v_s^{1/2}\{\theta(\bar{r}-r)[L(|x_s|)]^{l+1} + \theta(r-\bar{r})[L(|X|)]^{-\bar{l}}\}\exp(|\mathrm{Im}\omega|\,W_s), \tag{4.5}$$

where

$$L(x) = \frac{x}{1+x}. \tag{4.6}$$

Let us introduce $\tilde{u}_g^{(n)}$, $\tilde{\eta}_g^{(n)}$, $\tilde{u}_s^{(n)}$, and $\tilde{\eta}_s^{(n)}$ in analogy with (4.4) and (4.5).

Moreover, let us define ($j=1,2$):

$$\tilde{V}_{j1} = V_{j1}\exp[(\nu-\lambda)/2]\,v_s[L(|x_s|)]^{l+1} \times[\delta_{j1} + \delta_{j2}(\rho+p)^{-1/2}v_s^{-1/2}], \tag{4.7}$$

$$\tilde{V}_{j2} = V_{j2}v_s\exp[(\nu-\lambda)/2]\{[L(|x_s|)]^{l+1}\,\theta(\bar{r}-r) + \theta(r-\bar{r})[L(|X|)]^{-\bar{l}}\}[\delta_{j2} + \delta_{j1}(\rho+p)^{1/2}v_s^{1/2}], \tag{4.8}$$

$$\tilde{g}_1(r,r') = g_1(r,r')\frac{r}{r'}\,[L(|x_s|)]^{-l-1} \times\exp[|\mathrm{Im}\omega|\,(W_s(r') - W_s(r))], \tag{4.9}$$

$$\tilde{g}_2(r,r') = g_2(r,r')\frac{r}{r'}\exp[|\mathrm{Im}\omega|\,(W_s(r') - W_s(r))] \times\frac{[(\rho+p)^{1/2}v_s^{1/2}]_{r'}}{(\rho+p)^{1/2}v^{1/2}}\quad\{[L(|x_s|)]^{l+1}\,\theta(\bar{r}-r) + [L(|X|)]^{-\bar{l}}\,\theta(r-\bar{r})\}^{-1}. \tag{4.10}$$

Then we have

$$\begin{bmatrix} \tilde{u}_g^{(n)} \\ \tilde{\eta}_g^{(n)} \end{bmatrix} = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_{n-1}=1}^{2}\int_0^{W_s}\begin{bmatrix} \tilde{g}_1(r,r_1)\,\tilde{V}_{1i_1}(r_1) \\ \tilde{g}_2(r,r_1)\,\tilde{V}_{2i_1}(r_2) \end{bmatrix} dW_{s1} \times\int_0^{W_{s1}}\tilde{g}_{i_1}(r_1,r_2)\,\tilde{V}_{i_1i_2}(r_2)\,dW_{s2}\cdots \times\int_0^{W_{sn-1}}\tilde{g}_{i_{n-1}}(r_{n-1},r_n)\,\tilde{V}_{i_{n-1}1}(r_n)\,\tilde{u}_g^{(0)}(r_n)\,dW_{sn}, \tag{4.11}$$

and

$$\begin{bmatrix}\tilde{u}_s^{(n)}\\ \tilde{\eta}_s^{(n)}\end{bmatrix} = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_{n-1}=1}^{2}\int_0^{W_s}\begin{bmatrix}\tilde{g}_1(r,r_1)\,\tilde{V}_{1i_1}(r_1)\\ \tilde{g}_2(r,r_1)\,\tilde{V}_{2i_1}(r_1)\end{bmatrix} dW_{s1}$$

$$\times\int_0^{W_{s1}}\tilde{g}_{i_1}(r_1,r_2)\,\tilde{V}_{i_1i_2}(r_2)\,dW_{s2}\times\cdots$$

$$\times\int_0^{W_{sn-1}}\tilde{g}_{i_{n-1}}(r_{n-1},r_n)\,\tilde{V}_{i_{n-1}2}\,\tilde{\eta}_s^{(0)}(r_n)\,dW_{sn}. \qquad (4.12)$$

Hereafter we will denote with $C, C_1, C_2, \cdots$ positive constants which are independent of r and ω.

As shown in the Appendix, we have then for $|\omega|$ large enough

$$|\tilde{g}_i(r,r')\,\tilde{V}_{ik}(r')| \le \frac{C}{|\omega|} \quad (i,k=1,2), \qquad (4.13)$$

$$|\tilde{u}_g^{(0)}| < C, \quad |\tilde{\eta}_s^{(0)}| < C. \qquad (4.14)$$

From (4.11) it follows

$$\begin{bmatrix}|u_g^{(n)}|\\ |\eta_g^{(n)}|\end{bmatrix} < \frac{C^{n+1}}{|\omega|^n}\int_0^{W_s} dW_{s1}\int_0^{W_{s1}} dW_{s2}\cdots\int_0^{W_{sn-1}} dW_{sn}$$

$$= \frac{C^{n+1}W_s^n}{|\omega|^n n!}. \qquad (4.15)$$

It follows that

$$\left|\sum_{n=m}^{\infty}\begin{bmatrix}\tilde{u}_g^{(n)}\\ \tilde{\eta}_g^{(n)}\end{bmatrix}\right| \le \sum_{n=m}^{\infty}\begin{bmatrix}|\tilde{u}_g^{(n)}|\\ |\tilde{\eta}_g^{(n)}|\end{bmatrix} \le C\sum_{n=m}^{\infty}\frac{C^nW_s^n}{n!\,|\omega|^n}$$

$$< \frac{C^{m+1}W_s^m}{m!\,|\omega|^m}\left(1+\frac{W_sC}{(m+1)|\omega|}+\frac{W_s^2C^2}{(m+1)(m+2)|\omega|^2}+\cdots\right)$$

$$< \frac{C^{m+1}W_0^m}{m!\,|\omega|^m}\left(1+\frac{W_0C}{|\omega|}+\frac{W_0^2C^2}{2!\,|\omega|^2}+\cdots\right) \qquad (4.16)$$

$$= \frac{C^{m+1}W_0^m}{m!\,|\omega|^m}\exp\left(\frac{CW_0}{|\omega|}\right).$$

The same majorizations (4.15), (4.16) hold altogether for $|\tilde{u}_s^{(n)}|$, $|\tilde{\eta}_s^{(n)}|$, as seen from (4.12), (4.13), and (4.14).

The series of the iterations converge uniformly in r for $|\omega|$ large enough and give the solutions of the respective systems of integral equations (4.1), (4.2). From (4.5), (3.18), and (4.16) written down for $\tilde{\eta}_s$, recalling the majorization (A20), we obtain in the limit $|\omega|\to\infty$, and $|\mathrm{Im}\omega|$ large enough

$$\left|\sum_{m=1}^{\infty}\eta_s^{(m)}(r)\right| < \frac{C}{|\omega|}\,|\eta_s^{(0)}(r)|\,; \qquad (4.17)$$

it follows that $\eta_s^{(0)}$ gives the dominant contribution to η_s in the above limit.

The above relation suggests that matter field waves are decoupled from gravitational waves, in the limit $\omega\to\infty$, up to the surface. However, a similar relation cannot be obtained for u_g, as seen from (3.17), (4.16), (4.4), and (A24).

In order to construct the Wronskian one also needs the derivative of the gravitational amplitudes u_s and u_g.

To this aim, from (4.1), (3.28), (3.23), and (3.17) one gets for $u_g^{(n)}$ or $u_s^{(n)}$, by derivation,

$$\frac{du^{(n)}}{dW_g} = \omega\left[\left(\phi_g^{-1}\frac{d\phi_g}{dW_g}+\frac{l+1}{W_g}\right)\frac{u^{(n)}}{\omega} - \int_0^r g_1^{l+1,l}(r,r')(V_{11}u^{(n-1)}+V_{12}\eta^{(n-1)})_{r'}\,dr'\right], \qquad (4.18)$$

where

$$g_1^{l+1,l}(r,r') = -\frac{u_{g,l+1}^{(0)}(r)\,u_{\mathrm{irr}}^{(0)}(r') - u_{\mathrm{irr},l+1}^{(0)}(r)\,u_g^{(0)}(r')}{[W(u_g^{(0)},u_{\mathrm{irr}}^{(0)})]_{r'}}, \qquad (4.19)$$

$$u_{g,l+1}^{(0)} = \phi_g\bar{j}_{l+1}(x_g), \quad u_{\mathrm{irr},l+1}^{(0)} = \phi_g\bar{n}_{l+1}(x_g).$$

A majorization similar to (4.15), (4.16) can now be obtained for $du^{(n)}/dW_g$. In fact, $g_1^{l+1,l}$ is majorized as in (4.13) together with (4.9), where g_1 is replaced by $g_1^{l+1,l}$ (see the Appendix). Then the second term in (4.17) is majorized according to (4.16), with a procedure which is completely similar to (4.11), (4.12). Since the quantity which multiplies $u^{(n)}$ in (4.18) is finite for $|x_g| > C$, with the help of (A5) it follows that

$$\left|\frac{du^{(n)}}{dW_g}\right| < \frac{C^{n+2}W_s^n}{|\omega|^{n-1}n!},$$

$$\left|\sum_{n=m}^{\infty}\frac{du^{(n)}}{dW_g}\right| < \frac{C^{m+1}W_0^m}{m!\,|\omega|^{m-1}}\times\exp\left(\frac{CW_0}{|\omega|}\right). \qquad (4.20)$$

5. INTEGRAL EQUATION FOR THE EXTERNAL SOLUTION

In this paper we shall confine ourselves to considering the region of the complex ω plane defined by $\mathrm{Im}\omega \ge 0$; furthermore, since the eigenfrequencies are symmetrically distributed with respect to the imaginary axis[23] it is sufficient to consider the region $\mathrm{Re}\omega \ge 0$.

For $r > r_0$ the system (3.3), (3.4) reduces to a single second order equation

$$O_1u = V_{11}u. \qquad (5.1)$$

The "ingoing" solution of the above equation satisfies the boundary condition[23]

$$\lim_{r\to\infty}u_-\exp(-i\omega W_g) = 1. \qquad (5.2)$$

The Eq. (5.1), together with this condition, is equivalent to the single integral equation

$$u_- = u_-^{(0)} + \int_\infty^r g_1(r,r')\,V_{11}(r')\,u_-(r')\,dr',$$

$$u_-^{(0)} = \phi_g\bar{h}_l^{(+)}(x_g), \qquad (5.3)$$

$$g_1(r,r') = \frac{1}{2i\omega}\left[\exp\left(\frac{\nu-\lambda}{2}\right)\right]_{r=r'}\frac{\phi_g(r)}{\phi_g(r')}$$

$$\times[\bar{h}_l^{(+)}(x_g)\,\bar{h}_l^{(-)}(x_g') - \bar{h}_l^{(-)}(x_g)\,\bar{h}_l^{(+)}(x_g')], \qquad (5.4)$$

where $x_g' = x_g(r')$ [the above expression is an alternative form of $g_1(r,r')$ given by (3.28), suitable for external solutions].

The successive iterations of this equation are

$$u_-^{(n)} = \int_\infty^r g_1(r, r_1)\, V_{11}(r_1)\, dr_1 \int_\infty^{r_1} g_1(r_1, r_2)\, V_{11}(r_2)\, dr_2 \times \cdots \times \int_\infty^{r_{n-1}} g_1(r_{n-1}, r_n)\, V_{11}(r_n)\, \phi_\varepsilon(r_n)\, \bar{h}_l^{(+)}(x_\varepsilon(r_n))\, dr_n . \tag{5.5}$$

For $r > r_0$ the functions ν', λ', and ϕ appearing in the definition (3.10) of V_{11} are given by

$$\nu' = -\lambda' = \frac{2M}{r^2} \exp(\lambda), \tag{5.6}$$

$$\Phi = 4\frac{M}{r^4}\left(5 - 9\frac{M}{r}\right), \tag{5.7}$$

$$\exp(-\lambda) = 1 - \frac{2M}{r} > 0, \tag{5.8}$$

where M is the mass of the star. Then, as seen from (3.10), V_{11} is of the form

$$V_{11} = \frac{1}{r^2} \sum_{p,q=0}^{2} c_{pq}\left(\frac{1}{\omega^2}, M\right) \frac{\exp(\lambda)}{F^q} P_{pq}\left(\frac{1}{r}, M\right), \tag{5.9}$$

where c_{pq}, P_{pq} are polynomials respectively in $1/\omega^2$ and $1/r$ whose coefficients depend on M,

$$F = \exp(\lambda)\left[4 + \omega^{-2}\left(\frac{16M}{r^3} - \frac{28M^2}{r^4}\right) - \omega^{-4} l(l+1)\left(\frac{4M^2}{r^6} - \frac{8M^3}{r^7}\right)\right]. \tag{5.10}$$

The successive iterations $u_-^{(n)}$ for u_- as obtained from (5.3), (5.4), and (5.5) can be written in the form

$$u_-^{(n)} = \frac{1}{(2i\omega)^n} \sum_{\sigma_1} \sum_{\sigma_2} \cdots \sum_{\sigma_n} \int_\infty^r \sigma_1 (\phi_\varepsilon(r)/\phi_\varepsilon(r_1)) \times \bar{h}_l^{(\sigma_1)}(\omega W_\varepsilon)\, \bar{h}_l^{(-\sigma_1)}(\omega W_1)\, \bar{V}_{11}(r_1)\, dr_1 \times \int_\infty^{r_1} \sigma_2 (\phi_\varepsilon(r_1)/\phi_\varepsilon(r_2))\, \bar{h}_l^{(\sigma_1)}(\omega W_1)\, \bar{h}_l^{(-\sigma_2)}(\omega W_2)\, \bar{V}_{11}(r_2)\, dr_2 \times \cdots \int_\infty^{r_{n-1}} \sigma_n \phi_\varepsilon(r_{n-1})\, \bar{h}_l^{(\sigma_n)}(\omega W_{n-1})\, \bar{h}_l^{(-\sigma_n)}(\omega W_n) \times \bar{V}_{11}(r_n)\, \bar{h}_l^{(+)}(\omega W_n)\, dr_n,$$

$$W_m = W_\varepsilon(r_m), \quad \bar{V}_{11} = \exp(-\lambda)\, V_{11}, \tag{5.11}$$

where $\sigma_1, \sigma_2, \sigma_3, \cdots$ assume the determinations plus and minus. Let us consider the product of exponentials arising from the functions $\bar{h}_l^{(\sigma_i)}$ which appear in the integrand relative to a single term of the summation over $\sigma_1, \sigma_2, \ldots, \sigma_n$; this is of the form

$$\exp\{i\omega[W_n + \sigma_n(W_{n-1} - W_n) + \sigma_{n-1}(W_{n-2} - W_{n-1}) + \cdots + \sigma_1(W_\varepsilon - W_1)]\} = \exp\{i\omega[2W_{m_\nu} - 2W_{m_{\nu-1}} + 2W_{m_{\nu-2}} - \cdots + (-1)^{\nu-1}(2W_{m_1} - W_\varepsilon)]\}, \tag{5.12}$$

where m_i are positive integers depending on the choice of $\sigma_1, \sigma_2, \ldots, \sigma_n$, such that $1 \le m_i \le n$, $m_{i+1} \ge m_i$.

By introducing further

$$y_m = W_m - W_{m-1}, \quad y_1 = W_1 - W_\varepsilon \tag{5.13}$$

we can set the argument of the above exponential, respectively for even and odd ν, in the form

$$2(y_{m_\nu} + y_{m_\nu - 1} + \cdots + y_{m_{\nu-1}+1}) + 2(y_{m_{\nu-1}} + y_{m_{\nu-1}-1} + \cdots + y_{m_{\nu-2}+1}) + \cdots + 2(y_{m_2} + y_{m_2 - 1} + \cdots + y_{m_1+1}) + W_\varepsilon(r) \tag{5.14}$$

or

$$2(y_{m_\nu} + y_{m_\nu - 1} + \cdots + y_{m_{\nu-1}+1}) + 2(y_{m_{\nu-1}} + y_{m_{\nu-1}-1} + \cdots + y_{m_{\nu-2}+1}) + 2(y_{m_1} + y_{m_1 - 1} + \cdots + y_1) + W_\varepsilon(r), \tag{5.15}$$

where the indices of the y's inside a single parenthesis are consecutive.

With reference to (5.11) let us introduce

$$K_l^{(\pm)}(z) = \exp(\mp iz)\, \bar{h}_l^{(\pm)}(z). \tag{5.16}$$

Then we can write $u_-^{(n)}$ in the form

$$u_-^{(n)} = \frac{\exp(i\omega W_\varepsilon)}{(2i\omega)^n} \sum_{\sigma_1} \sum_{\sigma_2} \cdots \sum_{\sigma_n} \int_\infty^r \sigma_1 K_l^{(\sigma_1)} K_l^{(-\sigma_1)} \times \exp(2i\omega\theta_1 y_1)\, \bar{V}_{11}\, dr_1 \int_\infty^{r_1} \sigma_2 K_l^{(\sigma_2)} K_l^{(-\sigma_2)} \times \exp(2i\omega\theta_2 y_2)\, \bar{V}_{11}\, dr_2 \cdots \int_\infty^{r_{n-1}} \sigma_n K_l^{(\sigma_n)} K_l^{(-\sigma_n)} \times \exp(2i\omega\theta_n y_n)\, \bar{V}_{11} K_l^{(+)}\, dr_n, \tag{5.17}$$

where the arguments of the various functions are omitted for sake of shortness and $\theta_i = \theta_i(\sigma_1, \sigma_2, \ldots, \sigma_n)$ are zero or one according to the determination of $\sigma_1, \sigma_2, \ldots, \sigma_n$.

From the definition of W_ε, for $r > r_0$, one gets

$$W_m = W_\varepsilon(r_m) = W_\varepsilon(r) + \eta_m + 2M \ln\left(1 + \frac{\eta_m}{r - 2M}\right), \tag{5.18}$$

$$\eta_m = r_m - r, \quad r - 2M > r_0 - 2M > 0. \tag{5.19}$$

From the above equalities, it is easily shown that

$$|W_m| \ge |W_\varepsilon(r)| > r_0. \tag{5.20}$$

Since $K_l^{(\pm)}(\omega W_m)$ given by (5.16) are polynomials in $1/\omega W_m$, whose coefficients depend on l only, from the above inequality, we get for $|\omega|$ large enough

$$|K_l^{(\pm)}(W_m)| < C. \tag{5.21}$$

Furthermore, recalling (5.8), (5.9), and (5.10) it is elementary to show that

$$|\exp[\pm\lambda(r_m)]| < C, \quad |F(r_m)|^{\pm 1} < C, \quad |V_{11}(r_m)| < C. \tag{5.22}$$

From the above inequalities and from (5.9) it follows also that

$$|V_{11}(r_m)| < C/|r_m|^2. \tag{5.23}$$

It must be remarked that the constants C appearing in the above inequalities are independent not only of ω but also of $r_1, r_2, \ldots, r_{m-1}$.

In addition, we have from (5.13), (5.16), (5.19), and (5.20)

$$y_m = r_m - r_{m-1} + 2M \ln\left(\frac{r_m - 2M}{r_{m-1} - 2M}\right) . \tag{5.24}$$

Thus the exponential factors appearing in the integrand of (5.18), for $\mathrm{Im}\omega \geq 0$, are such that

$$\left|\exp(2i\omega\theta_m y_m)\right| = \exp(-2\,\mathrm{Im}\omega\,\theta_m y_m) \leq 1. \tag{5.25}$$

Finally, we want to obtain a majorization for $u_-^{(n)}$ analogous to (4.16). Let us introduce the function

$$F^{(\mu,\nu)}(r,\omega) = \left|K_l^{(\mu)}(\omega W_\varepsilon) K_l^{(\nu)}(\omega W_\varepsilon)\, \bar{V}_{11}(r,\omega)\right|. \tag{5.26}$$

Then, from (5.17) we obtain

$$\begin{aligned}\left|u_-^{(n)}\right| \leq{}& \frac{\exp(-\mathrm{Im}\omega W_\varepsilon)}{|2\omega|^n} \sum_{\sigma_1,\sigma_2,\cdots,\sigma_n} \left|K_l^{(\sigma_1)}(\omega W_\varepsilon)\right| \\ &\times \int_\infty^r F^{(-\sigma_1,\sigma_2)}(r_1,\omega)\, dr_1 \int_\infty^{r_1} F^{(-\sigma_2,\sigma_3)}(r_2,\omega)\, dr_2 \\ &\times \cdots \int_\infty^{r_{n-1}} F^{(-\sigma_n,+)}(r_n,\omega)\, dr_n . \end{aligned} \tag{5.27}$$

From (5.21), (5.22), and (5.23) it follows that the last integral in the above inequality is majorized by a constant C independent of r_{n-1}.

Then from the identity

$$\begin{aligned}&\int_\infty^r f(r_1)\, dr_1 \int_\infty^{r_1} f(r_2)\, dr_2 \cdots \int_\infty^{r_{n-1}} f(r_n)\, dr_n \\ &\quad = \frac{\left(\int_\infty^r f(r')\, dr'\right)^n}{n!} \end{aligned} \tag{5.28}$$

and from (5.21) and (5.27) we get

$$\left|u_-^{(n)}\right| \leq \frac{\exp(-\mathrm{Im}\omega W_\varepsilon)\, C^{n+1}}{|2\omega|^n n!} , \tag{5.29}$$

$$\begin{aligned}\left|R_m\right| = \left|\sum_{n=m}^{\infty} u_-^{(n)}\right| &\leq \frac{\exp(-\mathrm{Im}\omega W_\varepsilon)}{|\omega|^m m!}\, C^{m+1} \\ &\times \left[1 + \frac{C}{|\omega|(m+1)} + \frac{C^2}{|\omega|^2(m+1)(m+2)} + \cdots\right] \\ &< \frac{\exp(-\mathrm{Im}\omega W_\varepsilon)}{|\omega|^m m!}\, C^{m+1} \exp(C/|\omega|) \\ &< C\, \frac{\exp(-\mathrm{Im}\omega W_\varepsilon)}{|\omega|^m} . \end{aligned} \tag{5.30}$$

Straightforwardly, a similar majorization can be performed for

$$\begin{aligned}\frac{du_-^{(n)}}{dW_\varepsilon} = \omega\Bigg\{&\left(\frac{l+1}{\omega W_\varepsilon} + \frac{1}{\phi_\varepsilon}\frac{d\phi_\varepsilon}{dW_\varepsilon}\right) u_-^{(n)} - \frac{1}{2i\omega} \\ &\times \int_\infty^r \left[\bar{h}_{l+1}^{(+)}(\omega W_\varepsilon)\, \bar{h}_l^{(-)}(\omega W_\varepsilon(r')) - \bar{h}_{l+1}^{(-)}(\omega W_\varepsilon)\right. \\ &\left.\times \bar{h}_l^{(+)}(\omega W_\varepsilon(r'))\right] (\bar{V}_{11} u_-^{(n-1)})_{r'}\, dr'\Bigg\} . \end{aligned} \tag{5.31}$$

We obtain

$$\left|\sum_{n=m}^{\infty} \frac{du_-^{(n)}}{dW_\varepsilon}\right| \leq \frac{C \exp(-\mathrm{Im}\omega W_\varepsilon)}{|\omega|^{m-1}} . \tag{5.32}$$

APPENDIX

Throughout this Appendix it is understood that some of the majorizations which follow (e.g., $|V_{11}| < C$) hold "for $|\omega|$ large enough," whereas $g_1(r,r')$, $g_2(r,r')$ are considered in the domain $0 \leq r' \leq r \leq r_0$. In addition, primed and unprimed quantities denote functions of r', r, respectively.

In order to obtain the inequalities (4.15), (4.16), (4.17), (4.20), (5.29), and (5.32), use is made of (4.13) which must be justified. In potential theory, majorizations for the Green functions have been given.[26,27] We note, however, that in the present problem things are somewhat more complicated. For example, $g_2(r,r')$, given by (3.29) together with (3.18) and (3.24), is constructed with the Hankel functions of argument $\omega(W_s - W_0)$, which diverge for $r \to r_0$. In order to get such majorizations, let us consider the following relations implying spherical Bessel functions:

$$j_\nu(z) = R_\nu(z)\sin z + S_\nu(z)\cos z \tag{A1}$$

$$n_\nu(z) = R_\nu(z)\cos z - S_\nu(z)\sin z \tag{A2}$$

where $R_\nu(z)$, $S_\nu(z)$ are polynomials in $1/z$ of degree $l+1$, whose coefficients depend on l, if ν is a nonnegative integer.[28] More generally, j_ν, n_ν, R_ν, and S_ν are assigned in terms of power series expansions in z,[29] whereas, for $z \to \infty$ asymptotic expansions in $1/z$ hold for R_ν, S_ν. In any case R_ν, S_ν behave like $z^{\nu-1}$, $\sim z^{-1}$, respectively, for $z \to 0$, $z \to \infty$, and are majorized by a constant C_2 for $|z| > C_1$, as seen from their expansions in powers of z. Equations (A1) and (A2) imply

$$\begin{aligned}-W_1(r')\, g_1^{l+p,\,l}(r,r') &= \phi_\varepsilon \phi'_\varepsilon [\bar{j}_{l+p}(x_\varepsilon)\, \bar{n}_l(x'_\varepsilon) - \bar{n}_{l+p}(x_\varepsilon)\, \bar{j}_l(x'_\varepsilon)] \\ &= x_\varepsilon x'_\varepsilon \phi_\varepsilon \phi'_\varepsilon [(S_{l+p} R'_l - R_{l+p} S'_l)\cos(x_\varepsilon - x'_\varepsilon) \\ &\quad + (R_{l+p} R'_l + S_{l+p} S'_l)\sin(x_\varepsilon - x'_\varepsilon)], \end{aligned} \tag{A3}$$

where $p = 0, 1$.

Now the Wronskian W_1 is explicitly given by $-\omega\phi_\varepsilon^2 \times \exp[(\lambda - \nu)/2]$; moreover ϕ_ε, given by (3.20), satisfies (as seen by inspection) the inequalities $C_1 \leq r|\phi_\varepsilon| \leq C_2$. It follows that $|\phi_\varepsilon/\phi'_\varepsilon| < C$; then (A3) leads to the majorization

$$\left|g_1^{l+p,\,l}(r,r')\right| < C\,|\omega|^{-1} \exp[\,|\mathrm{Im}\omega|\,(W_\varepsilon - W'_\varepsilon)] \tag{A4}$$

for $|x_\varepsilon| \geq |x'_\varepsilon| > C$.

In order to also include the case $|x'_\varepsilon| \ll 1$, $|x_\varepsilon| \geq C$, it is sufficient to supply a factor $[L(|x'_\varepsilon|)]^{-l}$ [see (4.6)] in the above equation; indeed, in this way, the correct divergence x_ε^{-l} appearing in the second member of (A3) is accounted for. So we have

$$\begin{aligned}\left|g_1^{l+p,\,l}(r,r')\right| &< C\,|\omega|^{-1} \exp[\,|\mathrm{Im}\omega|\,(W_\varepsilon - W'_\varepsilon)] \\ &\quad \times [L(|x'_\varepsilon|)]^{-l}\, r'/r, \quad |x_\varepsilon| > C. \end{aligned} \tag{A5}$$

When $p = 0$, a similar argument leads to[26,27]

$$\begin{aligned}\left|g_1(r,r')\right| &< C\,|\omega|^{-1} \exp[\,|\mathrm{Im}\omega|\,(W_\varepsilon - W'_\varepsilon)][L(|x'_\varepsilon|)]^{l+1} \\ &\quad \times [L(|x'_\varepsilon|)]^{-l}\, r'/r . \end{aligned} \tag{A6}$$

As regards g_2, let us generalize its definition (3.29) by introducing

$$g_2^{l+p,\,l+q}(r,r') = -\frac{\eta_{s\,l+p}^{(0)}(r)\, \eta_{\mathrm{irr}\,l+q}^{(0)}(r') - \eta_{\mathrm{irr}\,l+p}^{(0)}(r)\, \eta_{s\,l+q}^{0}(r')}{[W(\eta_s^{(0)}, \eta_{\mathrm{irr}}^{(0)})]_{r'}} , \tag{A7}$$

where, in analogy with (3.18), (3.24),

$$\eta^{(0)}_{s,\bar{l}+p} = \phi_s[\theta(\bar{r}-r)\bar{j}_{\bar{l}+p}(x_s) + \theta(r-\bar{r})(\alpha_+ \bar{h}^{(+)}_{\bar{l}+p}(X) + \alpha_- \bar{h}^{(-)}_{\bar{l}+p}(X))], \tag{A8}$$

$$\eta^{(0)}_{\mathrm{irr},\,l+p} = \phi_s[\theta(\bar{r}-r)\,\bar{n}_{l+p}(x_s) + \theta(r-\bar{r})(\beta_+ \bar{h}^{(+)}_{l+p}(X) + \beta_- \bar{h}^{(-)}_{l+p}(X))]. \tag{A9}$$

Note that ϕ_s, given by (3.21), is such that $|\phi_s/\phi_s'| < C$. Then the following inequalities similar to (A5) and (A6) are deduced for $r < \bar{r}$:

$$|g_2^{l+p,\,l}(r,r')| < C|\omega|^{-1}\exp[|\mathrm{Im}\omega|(W_s - W_s')] \times [L(|x_s'|)]^{-l}\, r'/r, \quad |x_s| > C, \tag{A10}$$

$$|g_2(r,r')| < C|\omega|^{-1}\exp[|\mathrm{Im}\omega|(W_s - W_s')] \times [L(|x_s|)]^{l+1}[L(|x_s'|)]^{-l}\, r'/r. \tag{A11}$$

Let us further analyze the case $\bar{r} < r' \le r \le r_0$. Recalling the relationship between the Hankel, Bessel, and Neumann functions we get from (A7)–(A9) after a short calculation

$$g_2^{l+p,\,l+q}(r,r') = \frac{1}{\omega}\frac{\phi_s}{\phi_s'}[\bar{j}_{\bar{l}+p}(X)\bar{n}_{\bar{l}+q}(X') - \bar{n}_{\bar{l}+p}(X)\bar{j}_{\bar{l}+q}(X')](\exp(\nu-\lambda)/2)v_s)_{r=r'}. \tag{A12}$$

Note that

$$|X'| = |x_0 - x_s'| \ge |x_0 - x_s| = |X|. \tag{A13}$$

So in (A12) the roles of the primed and unprimed variables are interchanged with respect to those appearing in (A10) and (A11) and l is replaced by $\bar{l}$. Further, we observe that $X' - X = \omega(W_s - W_s')$ and that near the star surface $v_s \sim (W_s - W_0)$, $\phi_s \sim (W_s - W_0)^l$ owing to (3.21), (3.1), (3.2), (3.14), and (3.9). Then the following inequalities similar to (A10) and (A11) are deduced:

$$|g_2^{\bar{l},\,l+q}(r,r')| < C|\omega|^{-1}(W_0 - W_s)^{\bar{l}} \times \exp[|\mathrm{Im}\omega|(W_s - W_s')][L(|X|)]^{-\bar{l}}, \quad |X'| > C, \tag{A14}$$

$$|g_2(r,r')| < C|\omega|^{-1}[L(|X'|)]^{\bar{l}+1}[L(|X|)]^{-\bar{l}} \times \exp[|\mathrm{Im}\omega|(W_s - W_s')](W_0 - W_s)^{\bar{l}}(W_0 - W_s')^{-(\bar{l}-1)}. \tag{A15}$$

Let us consider further the case $r' \le \bar{r} \le r$. Introducing in (3.29) the expressions for α_+, α_-, β_+, and β_- [obtained by imposing that $\eta_s^{(0)}$ and $\eta_{\mathrm{irr}}^{(0)}$, given by (3.18), (3.24), be continuous at the junction point $\bar{r}$ together with their derivatives] we get with a short calculation

$$g_2(r,r') = \omega\left(\frac{\exp[(\lambda-\nu)/2]}{v_s}\right)_{\bar{r}} \times \{g_2(\bar{r},r')[g_{2+}^{l,\,l+1}(r,\bar{r}) - Ag_2(r,\bar{r})] - g_{2-}^{l+1,\,l}(\bar{r},r')g_2(r,\bar{r})\}, \tag{A16}$$

$$A = \left(\frac{\bar{l}+1}{X(\bar{r})} - \frac{l+1}{x_s(\bar{r})}\right), \tag{A17}$$

$$g_{2-}^{l+1,\,l}(\bar{r},r') = \lim_{r\to\bar{r}_-} g_2^{l+1,\,l}(r,r'), \qquad g_{2+}^{l,\,l+1}(r,\bar{r}) = \lim_{r'\to\bar{r}_+} g_2^{l,\,l+1}(r,r'). \tag{A18}$$

Clearly $g_{2-}^{l+1,\,l}$ allows a majorization similar to (A10) since $|x_s(\bar{r})| > C$; on the other hand $g_{2+}^{l,\,l+1}$ allows a majorization similar to (A14), since $|X'(\bar{r})| > C$. From (A16) it follows that

$$|g_2(r,r')| < C|\omega|^{-1}\exp[|\mathrm{Im}\omega|(W_s - W_s')] \times (W_0 - W_s)^{\bar{l}}[L(|x_s'|)]^{-l}[L(|X|)]^{-\bar{l}}\, r'. \tag{A19}$$

Recalling the properties of ν, λ for $r \to 0$ one can see by inspection that the expression between square brackets appearing in V_{21} [see (2.23)] is majorized by a constant C for $0 \le r \le r_0$, together with $r^{-2} - \exp(-\nu)W_g^{-2}$ and Φr^{-1}, which appear respectively in (3.10) and in Ref. 21. Then let us consider the majorizations of V_{ik} given by (3.10)–(3.12). In addition, F, given by (2.9), satisfies $C_1 < |F| < C_2$ for $|\omega|$ large enough. It follows $|V_{11}| < C$, $|V_{12}| < C|\omega|^{-2}$.

Recalling further the properties (3.1), from (3.11) we get $|V_{21}| < C(W_0 - W_s)^{2\bar{l}-1}|\omega|^{-4}$, whereas we have $V_{22} < C(W_0 - W_s)^{-2}$ owing to (3.15), (3.14). Now from the majorizations already given for g_1, g_2, and V_{ik}, together with (4.7)–(4.10), majorizations (4.13) easily follow for $|\omega|$ large enough.

Finally, we want to prove that [for $\omega \to \infty$ and such that W_s is sufficiently far from zeroes of $j_l(z)$] $\tilde{\eta}_s^{(0)}$, given by (4.5), satisfies

$$C_1 < |\tilde{\eta}_s^{(0)}| < C_2. \tag{A20}$$

Indeed the Bessel functions satisfy[25]

$$|\bar{j}_l(\omega W_s)| < C_1[L(|\omega W_s|)]^{l+1}\exp(|\mathrm{Im}\omega|W_s); \tag{A21}$$

on the other hand, recalling (A1), with the above specifications for ω, one obtains

$$|\bar{j}_l(\omega W_s)| > C_2[L(|\omega W_s|)]^{l+1}\exp(|\mathrm{Im}\omega|W_s). \tag{A22}$$

With similar arguments, for $r > \bar{r}$, recalling the definition of α_+, α_-, one gets

$$C_1 < |(\alpha_+ \bar{h}_l^{(+)}(X) + \alpha_- \bar{h}_l^{(-)}(X))| \times \exp[-|\mathrm{Im}\omega|W_s][L(|X|)]^{\bar{l}} < C_2. \tag{A23}$$

From (A21)–(A23), together with (4.5), one obtains (A20). Furthermore, a relation similar to (A21) holds for $\bar{j}_l(x_g)$

$$|\bar{j}_l(\omega W_g)| < C[L(|\omega W_g|)]^{l+1}\exp(|\mathrm{Im}\omega|W_g). \tag{A24}$$

[1] K.S. Thorne and A. Campolattaro, Astrophys. J. **152**, 673 (1967).
[2] K.S. Thorne, Phys. Rev. Lett. **21**, 320 (1968).
[3] K.S. Thorne, Astrophys. J. **158**, 1 (1969).
[4] K.S. Thorne, Astrophys. J. **158**, 997 (1969).
[5] R. Price and K.S. Thorne, Astrophys. J. **155**, 163 (1969).
[6] A. Campolattaro and K.S. Thorne, Astrophys. J. **159**, 847 (1970).
[7] J.R. Ipser, K.S. Thorne, Astrophys. J. **181**, 181 (1973).
[8] P. Cazzola and L. Lucaroni, *Proc. Symp. E.A.U. Athens*, 1972 (Springer, New York, 1974), Vol. 3.
[9] S.L. Detweiler, Astrophys. J. **197**, 203 (1975).
[10] S.L. Detweiler, Astrophys. J. **201**, 440 (1975).
[11] J.R. Ipser, Astrophys. J. **199**, 220 (1975).
[12] R. Semenzato, thesis, Padova (1975).
[13] A. Fisher and J. Marsden, Bull. Ann. Math. Soc. **79**, 997 (1973).

[14]A. Fisher and J. Marsden, Proc. Symp. Pure Math., Vol. XXVII.
[15]V. Moncrief, J. Math. Phys. **16**, 493 (1975).
[16]V. Moncrief, J. Math. Phys. **16**, 1556 (1975).
[17]V. Moncrief, J. Math. Phys. **17**, 1893 (1976).
[18]S. L. Detweiler and J. R. Ipser, Astrophys. J. **185**, 685 (1973).
[19]J. B. Hartle and D. C. Wilkins, Commun. Math. Phys. **38**, 47 (1974).
[20]J. M. Stewart, Proc. R. Soc. **344**, 81 (1975).
[21]L. Battiston, P. Cazzola, and L. Lucaroni, Nuovo Cimento B **3**, 295 (1971).
[22]K. S. Thorne and A. Campolattaro, Astrophys. J. **149**, 591 (1967).
[23]P. Cazzola and L. Lucaroni, Phys. Rev. D **6**, 950 (1972).
[24]P. Cazzola and L. Lucaroni, Phys. Rev. D **10**, 2038 (1974).
[25]R. G. Newton, J. Math. Phys. **1**, 319 (1960).
[26]N. Levinson, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **25**, No. 9 (1949).
[27]R. G. Newton, Phys. Rev. **100**, 412 (1955).
[28]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1964), Vol. 1.
[29]G. N. Watson, *Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1952).

Identification of the velocity operator in an irreducible unitary representation of the Poincaré group for imaginary mass or zero mass and variable helicity

Thomas F. Jordan

Department of Physics, University of Minnesota–Duluth, Duluth, Minnesota 55812[a]
and Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627[b]
(Received 18 May 1977)

For an irreducible unitary representation of the Poincaré group with imaginary mass or zero mass and variable helicity it is shown, as it was previously for positive mass or zero mass and fixed helicity, that the velocity operator can be identified by its transformations under the Poincaré group together with the assumption that its different components commute with each other. In particular, there is no way to use an unconventional velocity operator to avoid the tachyon interpretation of imaginary-mass representations.

In a previous paper,[1] which provides an introduction to this one, as well as in an earlier paper by Ahmad and Wigner[2] which provided the idea, it was shown that in an irreducible unitary representation of the Poincaré group for positive mass or zero mass and fixed helicity ("discrete spin"), the velocity operator can be identified by its transformations under the Poincaré group together with the assumption that its components commute with each other. Here the same is done for imaginary mass or zero mass and variable helicity ("continuous spin"). Again we find the conventional velocity operator is the only possibility. In particular there is no way to use an unconventional velocity operator to avoid the tachyon interpretation of imaginary-mass representations.

Let P denote the generator for space translations. The generator for time translation is

$$H=\epsilon(\mathbf{P}^2+m^2)^{1/2},$$

where $\epsilon=\pm 1$ is fixed if m is zero and variable if m is imaginary. For the rotation and Lorentz transformation generators J and K we use the "standard helicity" form written first for zero mass by Lomont and Moses[3] and then for positive and imaginary mass by Moses.[4] We use the conventions and notations of Lomont and Moses.[3] The representation of the little group, the two-dimensional Lorentz group for imaginary mass or the two-dimensional Euclidean group for zero mass, has generators S, T_2, T_3 where S is the helicity and the commutation relations of T_2 and T_3 with S are those of the y and z components of a vector with a generator for rotations around the x axis. The irreducible unitary representation of the Poincaré group is spanned by eigenkets $|\mathbf{p},\epsilon,s\rangle$ of $\mathbf{P}, H$, and S. The eigenvalues of S may be the integers $s=0,\pm1,\pm2,\cdots$ or the half-integers $s=\pm\frac{1}{2},\pm\frac{3}{2},\cdots$. In addition, for each $k=0,\frac{1}{2},1,\frac{3}{2},\cdots$ there is an imaginary-mass representation where $s=k,k+1,k+2,\cdots$ and one where $s=-k,-k-1,-k-2,\cdots$. The operators T_2, T_3 are given by

$$(T_2\pm iT_3)|\mathbf{p},\epsilon,s\rangle=t_\pm(s)|\mathbf{p},\epsilon,s\pm1\rangle, \tag{1}$$

where $t_\pm$ are functions of s that are characteristic of the representation and are zero only when s is the maximum or minimum eigenvalue of S so there is no $|\mathbf{p},\epsilon,s\pm1\rangle$ in the representation. Let κ be 1 for zero-mass representations. For imaginary-mass representations let κ be the positive number such that κ^2 is $-m^2$. From the formulas Lomont and Moses[3] and Moses[4] write for J and K we can write the Pauli–Lubanski 4-vector in terms of $\mathbf{P}, \epsilon, S, T_2, T_3$,

$$\mathbf{P}\cdot\mathbf{J}=PS, \tag{2}$$

$$H\mathbf{J}+\mathbf{P}\times\mathbf{K}=H\hat{P}S+\kappa\epsilon(\mathbf{E}_2T_2+\mathbf{E}_3T_3), \tag{3}$$

where

$$P=(\mathbf{P}^2)^{1/2},$$

$$\hat{P}=\mathbf{P}/P,$$

$$\mathbf{E}_2=(P_3/P,\ P_2P_3/P(P+P_1),\ -P_2^2/P(P+P_1)-P_1/P),$$

$$\mathbf{E}_3=(-P_2/P,\ P_3^2/P(P+P_1)+P_1/P,\ -P_2P_3/P(P+P_1)).$$

The vectors $\hat{P}, \mathbf{E}_2, \mathbf{E}_3$ are orthonormal and $\hat{P}\times\mathbf{E}_2$ is $\mathbf{E}_3$, etc.

Here we will show that the only possible velocity operator V in the irreducible unitary representation of the Poincaré group is $\mathbf{V}=\mathbf{P}/H$ because there is no other operator that transforms correctly under the Poincaré group and has components that commute with each other. Let us be more specific about the transformations. We assume V is translation-invariant and time-independent, rotates as a vector, and Lorentz transforms as a time-independent velocity should, as characterized by the commutation relations[1]

$$(1/i)[V_j,K_k]=V_jV_k-\delta_{jk} \tag{4}$$

for $j,k=1,2,3$.

Let

$$U_0=(1-\mathbf{V}^2)^{-1/2},\quad \mathbf{U}=\mathbf{V}(1-\mathbf{V}^2)^{-1/2}.$$

From the Lorentz transformations of V, described by the commutation relations (4), it follows that U_0, U transform as a 4-vector, in particular that

$$(1/i)[U_0,\mathbf{K}]=-\mathbf{U}. \tag{5}$$

Also, from our assumptions about the transformations of V, it follows that U_0 and U are translation invariant

[a] Permanent.
[b] 1976–77.

 0022-2488/78/1901-0247$1.00

and time-independent, U_0 is rotation-invariant, and $\mathbf{U}$ rotates as a vector.

From Eqs. (3) and (5) we have

$$\kappa\epsilon P\mathbf{E}_3(1/i)[U_0, T_2] - \kappa\epsilon P\mathbf{E}_2(1/i)[U_0, T_3]$$
$$= (1/i)[U_0, \mathbf{P}\times(H\mathbf{J} + \mathbf{P}\times\mathbf{K})] = \mathbf{P}\times(\mathbf{P}\times(1/i)[U_0, \mathbf{K}])$$
$$= -\mathbf{P}\times(\mathbf{P}\times\mathbf{U}) = P^2\mathbf{U} - \mathbf{P}(\mathbf{P}\cdot\mathbf{U}). \tag{6}$$

From Eq. (2) we see that U_0 commutes with S. Now $\mathbf{P}$, ϵ, S are a complete set of commuting operators so, since U_0 commutes with $\mathbf{P}$, ϵ, S, it must be a function of them. Also $\mathbf{P}\cdot\mathbf{U}$ must be a function of $\mathbf{P}$, ϵ, S; it commutes with $\mathbf{P}$ and $\mathbf{H}$ and is rotation-invariant so from Eq. (2) it follows that it commutes with S. From Eqs. (1) we find that for a function f of $\mathbf{P}$, ϵ, S

$$[f(S), T_2 \pm iT_3]|s\rangle = [f(s\pm1) - f(s)]t_\pm(s)|s\pm1\rangle$$
$$= [f(S) - f(S\mp1)](T_2 \pm iT_3)|s\rangle, \tag{7}$$
$$[f(S), T_2] = \tfrac{1}{2}[2f(S) - f(S+1) - f(S-1)]T_2$$
$$+ (i/2)[f(S+1) - f(S-1)]T_3,$$
$$[f(S), T_3] = \tfrac{1}{2}[2f(S) - f(S+1) - f(S-1)]T_3$$
$$- (i/2)[f(S+1) - f(S-1)]T_2.$$

The dependence on $\mathbf{p}$, ϵ is suppressed here because it plays no role. Then from Eq. (6) we see $\mathbf{U}$ is of the form

$$\mathbf{U} = \mathbf{A} + \mathbf{B}_2T_2 + \mathbf{B}_3T_3, \tag{8}$$

with $\mathbf{A}$, $\mathbf{B}_2$, $\mathbf{B}_3$ functions of $\mathbf{P}$, ϵ, S. Since U_0 also is a function of $\mathbf{P}$, ϵ, S, we have isolated the dependence on T_2 and T_3.

From Eq. (2) we find

$$[U_j, S] = P^{-1}\sum_{k=1}^{3} P_k[U_j, J_k] = P^{-1}\sum_{k=1}^{3} P_k i\epsilon_{jkl}U_l$$

for $j = 1, 2, 3$ or

$$(1/i)[\mathbf{U}, S] = \hat{P}\times\mathbf{U}. \tag{9}$$

Substituting the form (8) for U, we get

$$-\mathbf{B}_2T_3 + \mathbf{B}_3T_2 = \hat{P}\times\mathbf{A} + \hat{P}\times\mathbf{B}_2T_2 + \hat{P}\times\mathbf{B}_3T_3.$$

With this and Eqs. (1) we can set

$$\hat{P}\times\mathbf{A} = 0, \quad \hat{P}\times\mathbf{B}_2 = \mathbf{B}_3, \quad \hat{P}\times\mathbf{B}_3 = -\mathbf{B}_2 \tag{10}$$

because if there is a maximum or minimum value of s, the value of $\mathbf{B}_2 - i\mathbf{B}_3$ at the minimum value of s, or $\mathbf{B}_2 + i\mathbf{B}_3$ at the maximum value of s, does not occur in

$$\mathbf{U} = \mathbf{A} + \tfrac{1}{2}(\mathbf{B}_2 - i\mathbf{B}_3)(T_2 + iT_3)$$
$$+ \tfrac{1}{2}(\mathbf{B}_2 + i\mathbf{B}_3)(T_2 - iT_3)$$

and, since

$$(1/i)[T_2 \pm iT_3, S] = \pm i(T_2 \pm iT_3),$$

it is just these superfluous parts of $\mathbf{B}_2$, $\mathbf{B}_3$ that are not made to satisfy Eqs. (10) as a consequence of Eq. (9). Since $\mathbf{B}_2$ and $\mathbf{B}_3$ are perpendicular to $\mathbf{P}$, we can let

$$\mathbf{B}_2 = B_{22}\mathbf{E}_2 + B_{23}\mathbf{E}_3, \quad \mathbf{B}_3 = B_{32}\mathbf{E}_2 + B_{33}\mathbf{E}_3.$$

Then from Eqs. (10) we get

$$B_{22} = B_{33}, \quad B_{23} = -B_{32}$$

so we have

$$\mathbf{U} = \mathbf{A} + C\mathbf{E}_2T_2 + D\mathbf{E}_3T_2 - D\mathbf{E}_2T_3 + C\mathbf{E}_3T_3 \tag{11}$$

with $\mathbf{A}$, C, D functions of $\mathbf{P}$, ϵ, S. Now we see that Eq. (6) breaks down into

$$\kappa\epsilon(1/i)[U_0, T_2] = PDT_2 + PCT_3,$$
$$\kappa\epsilon(1/i)[U_0, T_3] = -PCT_2 + PDT_3. \tag{12}$$

Both $(H, \mathbf{P})$ and the Pauli–Lubanski 4-vector of Eqs. (2) and (3) have the transformation properties we have required for $(U_0, \mathbf{U})$, but it remains to impose the requirement that the different components commute with each other. From our assumption that the different components of $\mathbf{V}$ commute with each other it follows that U_0 commutes with U. From Eqs. (11) and (12) we see this implies

$$2CDT_2 + (C^2 - D^2)T_3 = 0$$

or

$$(C + iD)^2(T_2 + iT_3) - (C - iD)^2(T_2 - iT_3) = 0.$$

With this and Eqs. (1) we can set C and D equal to zero because if there is a minimum or maximum value of s, it is just the value of $C + iD$ at the minimum value of s, or $C - iD$ at the maximum value of s, that need not be zero as a result of these equations, and it does not occur in

$$\mathbf{U} = \mathbf{A} + \tfrac{1}{2}(\mathbf{E}_2 - i\mathbf{E}_3)(C + iD)(T_2 + iT_3)$$
$$+ \tfrac{1}{2}(\mathbf{E}_2 + i\mathbf{E}_3)(C - iD)(T_2 - iT_3).$$

Then from Eqs. (12) we see U_0 commutes with T_2 and T_3 and thus with all the generators of the irreducible representation of the little group. This implies U_0 is independent of S. So U_0 is a function of $\mathbf{P}$ and ϵ.

The Lorentz transformation generator is of the form[3,4]

$$\mathbf{K} = H\mathbf{Q} + \mathbf{N},$$

where $\mathbf{N}$ is an operator that commutes with $\mathbf{P}$ and ϵ. Here $\mathbf{Q}$ is $i\nabla$ on momentum-space wavefunctions with the invariant inner product. From Eq. (5) we see $\mathbf{U}$ is a function of $\mathbf{P}$ and ϵ. Then $\mathbf{V} = \mathbf{U}/U_0$ is a function of $\mathbf{P}$ and ϵ. From our assumption that $\mathbf{V}$ rotates as a vector it follows that

$$\mathbf{V} = F(P^2, \epsilon)\mathbf{P}$$

with F a function of P^2 and ϵ. Substituting this into the commutation relations (4) we find that the only possibility is

$$F = 1/H.$$

Therefore,

$$\mathbf{V} = \mathbf{P}/H.$$

[1] T. F. Jordan, J. Math. Phys. **18**, 608 (1977).
[2] S. M. W. Ahmad and E. P. Wigner, Nuovo Cimento A **28**, 1 (1975).
[3] J. S. Lomont and H. E. Moses, J. Math. Phys. **3**, 405 (1962).
[4] H. E. Moses, J. Math. Phys. **9**, 2039 (1968).

Uniqueness of solutions to the linearized Boltzmann equation

L. Freede Garbanati and William Greenberg

Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

P. F. Zweifel[a)]

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061
(Received 11 April 1977)

Uniqueness theorems are proved for the linearized Boltzmann equation for both the "exterior" and "interior" problems under generalized Maxwell boundary conditions. The solution space is a weighted L_p space, and agrees with the space in which solutions have previously been constructed.

I. INTRODUCTION

Although the Boltzmann equation is more than 100 years old, only recently have rigorous mathematical treatments of the equation and other types of irreversible statistical mechanics been developed which would parallel corresponding rigorous treatments of equilibrium statistical mechanics as summarized for example in Ruelle's book.[1] Along such lines are investigations of uniqueness and existence of solutions to various forms of the linearized Boltzmann equation, which are still active areas in the mathematical physics literature. A quasirigorous approach to the neutron transport equation, including constructive methods of existence proofs, was developed in 1960 by Case,[2] and is reviewed extensively in a later book.[3] These same techniques were even earlier applied to the linearized Vlasov equation describing plasma oscillations[4,5] and later to the kinetic theory of gases[6–8] and radiative transport in stellar atmospheres.[9]

In the early 1970's strictly rigorous methods for solving these equations were introduced independently by Hangelbroek[10] and Larsen and Habetler.[11] These approaches have been described and compared with one another and the Case method in a review paper.[12]

The purpose of the present study is to prove uniqueness to supplement the rigorous constructive existence proofs mentioned above. Except for a brief remark near the end of the paper, we restrict our attention to the linearized Boltzmann equation describing gas kinetics. Our technique is based on Case's treatment[13] with two major differences. First, we consider the "exterior problem" instead of the "interior problem" studied by Case, although our results can easily be extended to the interior problem as well. More importantly, we believe that the (Hilbert) space used in Case's work is not the appropriate solution space. In particular, rigorous constructive solutions have been obtained[14,15] in a different space [the space $X_p^m(\mathbb{R}^n)$ defined below], and so we will prove uniqueness in that space. Further, the existence of a certain integral, which is crucial to our proof, can be inferred in $X_p^m(\mathbb{R}^n)$, and not in Case's Hilbert space.

For these reasons, let us define the space $X_p^m(\mathbb{R}^n)$, $p>1$, by

$$X_p^m(\mathbb{R}^n)=\bigoplus_{i=1}^{m} X_p(\mathbb{R}^n), \tag{1a}$$

$$X_p(\mathbb{R}^n)=\{f: c_i f\in L_p(\mathbb{R}^n,\mu),\ 1\leq i\leq n\}, \tag{1b}$$

where $L_p(\mathbb{R}^n,\mu)$ is the weighted Banach space with norm

$$||f||=\int_{\mathbb{R}^n}|f|^p\,d\mu(\mathbf{c})=\int_{\mathbb{R}^n}|f|^p\exp(-\mathbf{c}\cdot\mathbf{c})\,d^nc, \tag{1c}$$

[$L_p^m(\mathbb{R}^n,\mu)$ is related to $L_p(\mathbb{R}^n,\mu)$ in analogy with Eq. (1a)].

We call attention to two other attempts to develop uniqueness in a rigorous context. The first is presented in a series of papers by Cercignani and Pao (a bibliography appears on p. 154 of Ref. 7; cf. pp. 140ff of the same reference for a description). Unfortunately, the existence proofs are not constructive. Furthermore, the weighted Banach space necessary for the demonstration of existence encountered in Refs. 14 and 15 does not appear to be a convenient space in which to work (leaving aside questions of physical relevance). The second attempt is due to Giraud,[16] but his techniques are considerably more cumbersome than our simple methods based on Case's work.

II. THE TIME-INDEPENDENT, EXTERIOR PROBLEM

The time-dependent equation is considerably easier to treat than the time-independent, as a quick reading of Ref. 13 indicates, and so we will deal only with the latter. The interior problem, discussed in Ref. 13, is a straightforward modification of the exterior problem, and the appropriate uniqueness theorem for that case will be stated without proof.

The linearized Boltzmann equation can be written in the form

$$\mathbf{c}\cdot\nabla h(\mathbf{c},\mathbf{r})=J(h), \tag{2}$$

where $\mathbf{c}\in\mathbb{R}^n$ is the (dimensionless) gas velocity, the gradient operator is with respect to the position variable $\mathbf{r}$, and the collision integral $J(h)$ is dissipative. This means that

$$\int_{\mathbb{R}^n} h(\mathbf{c},\mathbf{r})J(h(\mathbf{c},\mathbf{r}))\,d\mu(\mathbf{c})\leq 0, \tag{3}$$

[a)]Supported by the National Science Foundation Grant ENG. 75-15882.

 0022-2488/78/1901-0249$1.00

with equality iff $J(h)=0$. (This condition is necessary for the existence of an H-theorem for the Boltzmann equation). Specifically, the unbounded linear transformation $J: X_p^m(\mathbb{R}^n) \to L_1^m(\mathbb{R}^n, \mu)$ is given by

$$J(h) = \int_{\mathbb{R}^n} K(\mathbf{c}, \mathbf{c}')\, h(\mathbf{c}', \mathbf{r})\, d\mu(\mathbf{c}') - \nu(\mathbf{c})h, \tag{4}$$

where $K(\mathbf{c}, \mathbf{c}')$ is the collision kernel and $\nu(\mathbf{c})$ is known as the scattering rate

$$\nu(\mathbf{c}) = \int_{\mathbb{R}^n} K(\mathbf{c}', \mathbf{c})\, d\mu(\mathbf{c}'). \tag{5}$$

The domain of J is the dense subset of $X_p^m(\mathbb{R}^n)$ for which $J(h_\mathbf{r}) \in L_1^m(\mathbb{R}^n, d\mu)$. [Where the spatial variable is held fixed, we shall write $h(\mathbf{c}, \mathbf{r}) = h_\mathbf{r}(\mathbf{c})$; e.g., $h_\mathbf{r} \in X_p^m(\mathbb{R}^n)$.]

In the model considered in Ref. 14, $m=1$, $n=1$, and $K(\mathbf{c}, \mathbf{c}') = 1/\sqrt{\pi}$. The constructive solution of Eq. (2) obtained in that reference requires the space $X_p(\mathbb{R})$. This work has suggested our choice for the domain and range spaces of J in the more general case treated here. As a result of these choices the integral in Eq. (3) may exist as an extended real number.

Time reversal invariance actually requires that $K(\mathbf{c}, \mathbf{c}')$ be a real symmetric function of $\mathbf{c}$ and $\mathbf{c}'$; rotational invariance requires further that it depends only on $\mathbf{c} \cdot \mathbf{c}'$. These facts are well known[7,17]; however, we shall not make use of these conditions in our uniqueness theorems. The following useful result is readily obtained from Hölder's inequality.

Proposition 1: If $K(\mathbf{c}, \mathbf{c}')$ is a polynomial in $\mathbf{c}' \cdot \mathbf{c}$ with no constant term, then $J(h)$ is continuous.

Note also that the left-hand side of Eq. (2) defines a function in $L_1^m(\mathbb{R}^n, \mu)$ if the components of $\nabla h_\mathbf{r}$ are contained in $X_p^m(\mathbb{R}^n)$.

We now consider solutions of Eq. (2) in the exterior of a bounded set $V \subseteq \mathbb{R}^n$ with connected complement and piecewise smooth, orientable boundary. Appropriate boundary conditions will be imposed on ∂V and at the point ∞. We define a solution of Eq. (2) to be a map $h: \mathbb{R}^n \to X_p^m(\mathbb{R}^n)$ with continuous spatial first partial derivatives such that the components of $\nabla h_\mathbf{r} \in X_p^m(\mathbb{R}^n)$. (The continuity of the spatial partial derivatives is used only for the application of Gauss's theorem and so can be weakened.) The boundary conditions which are generally adopted on ∂V are the so-called linearized Maxwell conditions, namely

$$h(\mathbf{c}, \mathbf{r}_s) = (1-\alpha)h(c_\parallel, -\mathbf{c}_\perp, \mathbf{r}_s) + (2\alpha/\pi) \int_{\mathbf{n}_s \cdot \mathbf{c}' > 0} \mathbf{n}_s \cdot \mathbf{c}' h(\mathbf{c}', \mathbf{r}_s)\, d\mu(\mathbf{c}') + h_0(\mathbf{c}, \mathbf{r}_s), \tag{6}$$

when $\mathbf{n}_s \cdot \mathbf{c} < 0$. Here $\mathbf{r}_s \in \partial V$, $0 \le \alpha \le 1$, and $\mathbf{n}_s$ is the outward normal to ∂V at $\mathbf{r}_s$; $\mathbf{c}_\perp$ is the component of $\mathbf{c}$ perpendicular to ∂V at $\mathbf{r}_s$ and $c_\parallel$ is the parallel component. At infinity we require

$$\lim_{|\mathbf{r}| \to \infty} h(\mathbf{c}, \mathbf{r}) = h_\infty(\mathbf{c}), \tag{7a}$$

in the sense

$$\lim_{|\mathbf{r}| \to \infty} \int (\mathbf{n}_s \cdot \mathbf{c})[h(\mathbf{c}, \mathbf{r}) - h_\infty(\mathbf{c})]^2\, dS = 0, \tag{7b}$$

where the integration is carried out over a sphere of fixed radius $|\mathbf{r}|$. We now state

Theorem 1: Subject to conditions (6), (7a), and (7b), Eq. (2) has at most one solution for $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{r} \in \mathbb{R}^n \setminus V$.

Proof: Assume that two solutions h_1 and h_2 exist. Then $h = h_1 - h_2$ obeys Eq. (2) subject to

$$h(\mathbf{c}, \mathbf{r}) = (1-\alpha)h(c_\parallel, -\mathbf{c}_\perp, \mathbf{r}_s) + (2\alpha/\pi) \int_{\mathbf{n}_s \cdot \mathbf{c}' > 0} \mathbf{n}_s \cdot \mathbf{c}' h(\mathbf{c}', \mathbf{r}_s)\, d\mu(\mathbf{c}') \tag{8a}$$

and

$$\lim_{|\mathbf{r}| \to \infty} h(\mathbf{c}, \mathbf{r}) = 0 \tag{8b}$$

[the limit being defined by (7b)].

We now proceed as in Ref. 13, i.e., multiply Eq. (2) by $\exp(-c^2)h(\mathbf{c}, \mathbf{r})$ and integrate over $d^n r$ and $d^n c$. The integral on the left-hand side can be converted into a surface integral over ∂V plus a large sphere of radius $|\mathbf{r}| \to \infty$, by application of Gauss's theorem after the identity $h\nabla h = \frac{1}{2}\nabla h^2$ is employed. By virtue of (8b), the integral over the surface of radius $|\mathbf{r}|$ vanishes as $|\mathbf{r}| \to \infty$.

We thus arrive at Eq. (13) of Ref. 13, except that the order of integration is reversed:

$$\tfrac{1}{2} \int_{\mathbb{R}^n} \int_{\partial V} \mathbf{n}_s \cdot \mathbf{c}\, h^2(\mathbf{c}, \mathbf{r}_s)\, dS\, d\mu(\mathbf{c}) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n \setminus V} h(\mathbf{c}, \mathbf{r}) J(h)\, d^n r\, d\mu(\mathbf{c}). \tag{9}$$

Because of the continuity of $h(\mathbf{c}, \mathbf{r})$ in $\mathbf{r}$, Fubini's theorem applies,[18] and we can carry out the integration over $\mathbf{c}$ first. Dissipativity shows that the right-hand side of Eq. (9) is nonpositive (it may equal $-\infty$). The left-hand side can be simplified, as in Ref. 13, by applying the boundary condition (8a), and we conclude [Eq. (8) of Ref. 13] that

$$\int_{\mathbb{R}^n} \mathbf{n}_s \cdot \mathbf{c} h^2(\mathbf{c}, \mathbf{r}_s)\, d\mu(\mathbf{c}) = [2\alpha(1-\alpha) + \alpha^2]\Big[\int_{\mathbf{n}_s \cdot \mathbf{c} > 0} \mathbf{n}_s \cdot \mathbf{c} h^2(\mathbf{c}, \mathbf{r}_s)\, d\mu(\mathbf{c}) - (2/\pi)\Big|\int_{\mathbf{n}_s \cdot \mathbf{c} > 0} \mathbf{n}_s \cdot \mathbf{c} h(\mathbf{c}, \mathbf{r}_s)\, d\mu(\mathbf{c})\Big|^2\Big]. \tag{10}$$

The second integral always exists via Hölder's inequality.

The first integral either exists, or is $+\infty$, since h is measurable. If the integral exists, we follow Case's reasoning[13] (using the Schwartz inequality) to conclude that Eq. (10) is nonnegative. If it is infinite, the same conclusion is immediate. In either case it follows that both sides of Eq. (9) must vanish. Thus [Eq. (3)]

$$J(h) = 0, \tag{11a}$$

and

$$\int_{\mathbf{n}_s \cdot \mathbf{c} > 0} \mathbf{n}_s \cdot \mathbf{c} h^2(\mathbf{c}, \mathbf{r}_s)\, d\mu(\mathbf{c}) = (2/\pi)\Big|\int_{\mathbf{n}_s \cdot \mathbf{c} > 0} \mathbf{n}_s \cdot \mathbf{c} h(\mathbf{c}, \mathbf{r}_s)\, d\mu(\mathbf{c})\Big|^2. \tag{11b}$$

[These are Eqs. (14) and (15) of Ref. 13.] From the

Schwartz inequality and Eq. (8a) it follows that

$$h(\mathbf{c},\mathbf{r}_s)=h(\mathbf{r}_s), \quad \text{everywhere for almost all } \mathbf{c}, \tag{12}$$

i.e., a function of $\mathbf{r}_s$ alone.

From Eqs. (11a) and (2) we conclude further that $h(\mathbf{c},\mathbf{r})$ is independent of $\mathbf{r}$ along any ray originating at point $\mathbf{r}_s$ in the direction $\mathbf{c}$, for $\mathbf{n}_s\cdot\mathbf{c}>0$. Consider two rays originating at points $\mathbf{r}_s$ and $\mathbf{r}_s'$ and intersecting at $\mathbf{r}$. It follows that $h(\mathbf{r}_s)=h(\mathbf{r}_s')$ and, in fact, that $h(\mathbf{c},\mathbf{r}_s)$ is a constant also independent of $\mathbf{r}_s$. Thus

$$h(\mathbf{c},\mathbf{r})=\text{const}, \quad \text{everywhere for almost all } \mathbf{c}\in\mathbb{R}^n. \tag{13}$$

Finally, from (8b) we conclude

$$h(\mathbf{c},\mathbf{r})=0, \quad \mathbf{r}\in\mathbb{R}^n\backslash V,$$

$$\text{everywhere for almost all } \mathbf{c}\in\mathbb{R}^n. \tag{14}$$

This completes the proof of Theorem 1.

Uniqueness for the initial-boundary value problem is slightly simpler to prove since it is possible to conclude, as in Ref. 13, that

$$\int_{\mathbb{R}^n\backslash V} d^n r \int d\mu(\mathbf{c})\, h^2(\mathbf{c},\mathbf{r},t)=0, \tag{15}$$

which implies

$$h(\mathbf{r},\mathbf{c},t)=0, \quad \mathbf{r}\in\mathbb{R}^n\backslash V,\ 0<t<\infty,$$

$$\text{everywhere for almost all } \mathbf{c}\in\mathbb{R}^n. \tag{16}$$

Here we assume again that h is continuously differentiable in t, and one extra change in order of integration is required.

We may remark that the major portion of the above proof already appears in Ref. 13. However, it was felt necessary to justify certain of the mathematical manipulations in order to make Case's treatment "rigorous." Note in particular that Case chose to work in the space $L_2(\mathbb{R}^3,\mu)$; in fact, Kuščer[17] states specifically that this is the appropriate solution space. However, in this space there is no guarantee that

$$\int_{\mathbf{n}_s\cdot\mathbf{c}>0} \mathbf{n}_s\cdot\mathbf{c}\, h(\mathbf{c},\mathbf{r}_s)\, d\mu(\mathbf{c})$$

exists [see Eq. (10)]. The existence of this integral is crucial to the proof of Theorem 1. Furthermore, the work on the BKG model referred to earlier indicates that our choice of the X_p spaces is appropriate. It is interesting that, in studies of the neutron transport equation, the X_p spaces also entered in a natural way.[19]

We now state without proof

Theorem 2: Eq. (2), subject to condition (6), has, up to an additive constant, at most one solution for $\mathbf{c}\in\mathbb{R}^n$, $\mathbf{r}\in V$. Here $\mathbf{n}_s$ must be interpreted as the *inward* normal at $\mathbf{r}_s$.

The proof proceeds in direct analogy with that of Theorem 1. Extension to the initial boundary value problem is also immediate. The additive constant ambiguity in the interior solution[13] does not exist for the exterior because in the latter the behavior at infinity is specified.

Siewert[20] has raised the question as to the uniqueness of the solution to the equation of radiative transport in a half-space, subject to reflecting boundary conditions at $x=0$ as given by Eq. (6). The relevant equation is

$$\mu\frac{\partial\psi}{\partial x}(x,\mu)=J(\psi), \quad \mu\in[-1,1],\ x\in\mathbb{R}^+, \tag{17a}$$

where

$$J(\psi)=-\psi+\int_{-1}^{1}\psi(x,\mu')\,d\mu', \tag{17b}$$

and solutions ψ are to be sought in $\tilde{X}_p$ with norm

$$\|f\|_p=\{\int_{-1}^{1}|\mu f(\mu)|^p\,d\mu\}^{1/p}.$$

It is trivial to show that $J(\psi)$ is dissipative, and the left-hand side may be treated analogously to the gas case. One concludes that the solution to Eq. (17) is unique. A similar result can also be shown to hold for the full three-dimensional equation

$$\Omega\cdot\nabla\psi=J(\psi),$$

where

$$J(\psi)=-\psi+\int d\Omega' f(\Omega\cdot\Omega')\,\psi(\mathbf{r},\Omega')$$

and

$$\int f(\Omega\cdot\Omega')\,d\Omega'=1.$$

The equation and the notation are identical to those of one-speed neutron transport with $c=1$, as discussed in Ref. 3. The uniqueness proofs for the neutron transport equation, as described there, were primarily the inspiration for Ref. 13, and they may be made rigorous along the same lines as discussed in this paper for the Boltzmann equation.

III. THE ONE-DIMENSIONAL BGK MODEL

The present work was actually motivated by the constructive solutions obtained in Refs. 14 and 15. The BGK model equations considered there are not really linearized versions of the Boltzmann equation since the dependent variable h does not represent the deviation of the gas distribution function from equilibrium, but rather certain moments thereof. For this reason, it is probably necessary to show (although it is stated without proof in Ref. 7 and is presumably well known) that the collision operator is dissipative. The relevant equations can be written

$$c\frac{dh}{dx}=J(h), \quad x,c\in\mathbb{R}, \tag{18}$$

where for the scalar case

$$J_s(h)=\int_{-\infty}^{\infty}h(c',x)\,d\mu(c')-h(c,x), \tag{19}$$

and for the vector case

$$J_v(h)=Q(c)\int_{-\infty}^{\infty}Q^T(c')\,h(c',x)\,d\mu(c')-h(c',x), \tag{20a}$$

where Q is a 2×2 matrix

$$Q(c)=\begin{bmatrix}\sqrt{\frac{2}{3}}(c^2-\frac{1}{2}) & 1\\ \sqrt{\frac{2}{3}}c & 0\end{bmatrix}. \tag{20b}$$

Furthermore, the boundary condition (6) is replaced, it turns out, by the simpler condition

$$h(c,0)=h_0(c)+\alpha h(-c,0),\quad c>0, \tag{21}$$

where $\alpha=0$ corresponds to diffuse reflection and $\alpha=1$ to specular reflection. (We are considering the half-space problem, $x\in\mathbb{R}^+$, $c\in\mathbb{R}$.)

The proof for J_s follows from the Schwartz inequality. For J_v, one proves that

$$Q(c)\int_{-\infty}^{\infty} Q^T(c')h(c',x)d\mu(c'),$$

is a projection, from which the result follows fairly easily. The solutions as constructed in Ref. 14 can be verified to be continuously differentiable (in fact C_∞) in x by application of the Lebesque monotone convergence theorem. The proof of Theorem 1 (and Theorem 2) is readily adapted to the semi-infinite case; in fact, the argument is somewhat simpler, since the gradient found in Eq. (2) is replaced by $\partial h/\partial x$, and one can simply integrate from zero to infinity. However, the continuity properties of these solutions suggest that the more general case treated here should satisfy the continuity conditions imposed on $h(\mathbf{c},\mathbf{r})$ which allow application of Gauss's theorem.

ACKNOWLEDGMENTS

The authors are indebted to Professors Carlo Cercignani, Robert L. Bowden and James Thomas and Mr. William Cameron for several illuminating discussions.

[1]D. Ruelle, *Statistical Mechanics: Rigorous Results*, Mathematical Physics, Monographs, No. 12 (Benjamin, New York, 1974).
[2]K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[3]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[4]N. G. Van Kampen, Physica **21**, 949 (1955).
[5]K. M. Case, Ann. Phys. (N.Y.) **7**, 349 (1959).
[6]C. Cercignani, Ann. Phys. (N.Y.) **20**, 219 (1962).
[7]C. Cercignani, *Mathematical Methods in Kinetic Theory* (Plenum, New York, 1969).
[8]C. Cercignani, *Theory and Application of the Boltzmann Equation* (Elsevier, New York, 1976).
[9]C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966).
[10]R. J. Hangelbroek, Trans. Theory Stat. Phys. **5**, 1 (1976).
[11]E. Larsen and G. J. Habetler, Comm. Pure Appl. Math. **26**, 525 (1973).
[12]W. Greenberg and P. F. Zweifel, Trans. Theory Stat. Phys. **5**, 219 (1976).
[13]K. M. Case, Phys. Fluids **15**, 377 (1972).
[14]R. L. Bowden and L. F. Garbanati, "Constructive Solution of the Linearized Boltzmann Equation, Scalar BGK Model Equation," preprint.
[15]R. L. Bowden and W. C. Cameron, "Functional Analytic Treatment of Coupled Model Equations in Kinetic Theory," preprint.
[16]J. P. Guiraud, "Dissipation and Uniqueness for Classical Solutions of the Linearized Steady Exterior Boltzmann Boundary Value Problem," oral presentation, Conference on Mathemetical Methods in Transport Theory, Oberwolfach, Germany, Oct. 31—Nov. 5, 1976.
[17]I. Kuščer, "Mathematical Properties of Collision Operators," Conference on Mathematical Methods in Transport Theory, Oberwolfach, Germany, Oct. 31—Nov. 5, 1976.
[18]W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966), p. 150.
[19]E. Larsen, S. Sancaktar, and P. F. Zweifel, J. Math. Phys. **16**, 1117 (1975).
[20]C. E. Siewert, private communication.

On the inverse scattering problem for a class of spin–orbit and central potentials[a)]

M. A. Hooshyar

Physics Department, Pahlavi University, Shiraz, Iran
(Received 20 July 1977)

The nonrelativistic scattering of spin-1/2 particles by a certain class of central and spin–orbit potentials is considered. The potential coefficients associated with the phase shifts of all physical angular momenta at a fixed energy are deduced. It is shown that the construction does not lead to a unique solution and the found set of potential coefficients depend on two arbitrary constants. Finally the potential coefficients associated with transparent spin–orbit and central potentials are explicitly constructed.

1. INTRODUCTION

In this work we are interested in the construction of a certain class of spin–orbit and central potentials from the information on phase shifts for all angular momenta at a fixed energy. The class of spin–orbit and central potentials which we will be concerned with is the analog of the class of central potentials which was first studied by Newton.[1] By Newton's class of central potentials we mean those potentials whose associated potential coefficients, defined in terms of a zero-base potential, are zero for nonphysical values of the angular momentum.

An important tool in construction of central potentials from phase shifts at a fixed energy is the so-called Regge[2]–Newton[1] equation. The analog of this equation for spin–orbit and central potentials was first found by Sabatier.[3] The problem of constructing the potential coefficients from the phase shifts for spin–orbit and central potentials at a fixed energy was considered by the author in a previous work.[4] In that work we showed that, for a large class of spin–orbit and central potentials, the information on phase shifts can give us the potential coefficients only up to an unknown multiplicative constant. We were also able to rearrange Sabatier's equations[3] in such a way that the information which we can get for potential coefficients[5] from phase shifts was enough to find the spin–orbit and central potentials associated with the given set of phase shifts at a fixed energy.

The procedure will be as follows. Since we need to depend heavily on the work of Sabatier[3] and on our previous results,[4] in Sec. 2 we review the relevant parts of those works related to the construction of spin–orbit potentials from information on potential coefficients. In Sec. 3 we construct the potential coefficients from the phase shifts at a fixed energy. For the sake of completeness, in that section we also have reviewed the relevant parts of our previous work.[4] The main difference is that in the present work all the fundamental operators and their inverses are explicitly calculable. Section 4 is devoted to the problem of finding potential coefficients for a set of physical phase shifts all being zero. This section helps us to see how our method can be used and it also gives the potential coefficients associated with transparent potentials explicitly. In Sec. 5 we give the summary of the construction procedure.

Restricting ourselves to the class of spin–orbit and central potentials which are the analog of Newton's class of central potentials, just like the central case, enables us to calculate the fundamental operators and their inverses explicitly. Having done so, we are able to show that if[6] $\delta_\lambda^\pm = o(\lambda^{-3})$ for large values of λ, then we can construct the potential coefficients from the phase shifts. It also allows us to see the exact form of the nonuniqueness which exists in this construction. Furthermore, we can explicitly construct a set of potential coefficients which are associated with transparent spin–orbit and central potentials at a fixed energy.

In this work we have limited ourselves to the problem of constructing potential coefficients from phase shift information as a function of the angular momentum at a fixed energy for a class of spin–orbit and central potentials which are the analog of Newton's class of central potentials. The details of construction and the asymptotic properties of the potentials considered in this work will be the subject of a forthcoming communication.

2. A SURVEY OF PREVIOUS RESULTS

The Schrödinger equation and the differential cross section for scattering of spin-$\frac{1}{2}$ particles by central and spin–orbit potentials have the following forms[7]:

$$\Delta\Psi(\mathbf{r}) + [1 - U_c(r) - 2\mathbf{L}\cdot\mathbf{S}U_s(r)]\Psi(\mathbf{r}) = 0,$$
$$I(\theta) = |f(\theta)|^2 + |g(\theta)|^2, \tag{2.1}$$

where

$$f(\theta) = (1/2i)\sum\{(l+1)[\exp(2i\delta_l^+) - 1] + l[\exp(2i\delta_l^-) - 1]\}P_l(\cos\theta)$$

and

$$g(\theta) = (1/2i)\sum [\exp(2i\delta_l^+) - \exp(2i\delta_l^-)]P_l^1(\cos\theta).$$

We are measuring the radial distance r in units of $\lambda\!\!\!^{-}$, the reduced wavelength of the relative motion, which is fixed throughout the following. Writing above equa-

[a)]Supported in part by the Pahlavi University Research Council, Shiraz, Iran.

 0022-2488/78/1901-0253$1.00

tions in terms of the two possible angular momenta $j=l\pm\frac{1}{2}$, we have[3]

$$r^2\frac{d^2}{dr^2}\Psi_\lambda^\pm(r)+r^2[1-V(r)\mp 2\lambda Q(r)]\Psi_\lambda^\pm(r)$$
$$=(\lambda^2-\tfrac{1}{4})\Psi_\lambda^\pm(r), \tag{2.2}$$

where[8]

$$Q(r)=\tfrac{1}{2}U_s(r),\quad V(r)=U_c(r)-Q(r),\quad \lambda=l+\tfrac{1}{2},$$

and $\Psi_\lambda^\pm$ are the regular solutions associated with $j=l\pm\frac{1}{2}$ of the above equations.

The inverse scattering problem is that of finding the potentials $V(r)$ and $Q(r)$ from the asymptotic behavior of $\Psi_\lambda^\pm(r)$. To do so, we first define the following input matrix[4]:

$$G(r,r')=\begin{bmatrix}-e(r,r') & f^-(r,r')\\ f^+(r,r') & -e(r,r')\end{bmatrix}, \tag{2.3}$$

where

$$e(r,r')=\sum_{\lambda\in\Omega}a_\lambda u_\lambda(r)u_\lambda(r'),$$

$$f^\pm(r,r')=\sum_{\lambda\in\Omega}d_\lambda^\pm u_\lambda(r)u_\lambda(r'),$$

$$u_\lambda(r)=(\tfrac{1}{2}\pi r)^{1/2}J_\lambda(r),$$

$$a_\lambda=\begin{cases}2\lambda/\pi & \text{for positive integer } \lambda,\\ 0 & \text{otherwise,}\end{cases}$$

with

$$\Omega=\{\tfrac{1}{2},1,\tfrac{3}{2},2,\tfrac{5}{2},\cdots\}$$

and the potential coefficients $d_\lambda^\pm$ are to be defined later in terms of the phase shifts.[5]

Next we define an auxiliary matrix $H(r,r')$ through the following integral equation:

$$H(r,r')=G(r,r')-\int_0^r H(r,s)G(s,r')s^{-2}\,ds. \tag{2.4}$$

Having defined $H(r,r')$ in terms of $d_\lambda^\pm$ we define the function $t(r)$ to be the solution of the following equation:

$$\frac{d}{dr}t(r)+H_1(r)t(r)+H_2(r)t^2(r)=H_3(r) \tag{2.5}$$

with $\lim_{r\to\infty}t(r)=1$ where

$$H_1(r)=2r^{-2}[H_{22}(r,r)-H_{11}(r,r)],$$

$$H_2(r)=2r^{-2}H_{12}(r,r),\quad\text{and}\quad H_3(r)=2r^{-2}H_{21}(r,r).$$

Having found $t(r)$, we define $Q(r)$, $F^\pm(r)$, and $h_\pm$ by the following:

$$Q(r)=[2rt(r)]^{-1}\frac{d}{dr}t(r) \tag{2.6}$$

$$F^\pm(r)=\exp[\pm\int_0^r sQ(s)\,ds] \tag{2.7}$$

and

$$h_\pm=\exp[\pm\int_0^\infty sQ(s)\,ds]. \tag{2.8}$$

Next define the vector $\mathbf{K}(r,r')=[K^+(r,r'),K^-(r,r')]$ to be

$$\mathbf{K}(r,r')=\mathbf{F}(r)h^{-1}H(r,r')h, \tag{2.9}$$

where

$$\mathbf{F}(r)=[F^+(r),F^-(r)]\quad\text{and}\quad h=\begin{bmatrix}h_+ & 0\\ 0 & h_-\end{bmatrix}.$$

Using the above definitions in (2.4), we get the Sabatier equations[3]

$$\mathbf{K}(r,r')=\mathbf{F}(r)E(r,r')-\int_0^r\mathbf{K}(r,s)E(s,r')s^{-2}\,ds, \tag{2.10}$$

where

$$E(r,r')=h^{-1}G(r,r')h=\begin{bmatrix}-e(r,r') & g^-(r,r')\\ g^+(r,r') & -e(r,r')\end{bmatrix}$$

$$g^\pm(r,r')=\sum_{\lambda\in\Omega}b_\lambda^\pm u_\lambda(r)u_\lambda(r')$$

and[5,8]

$$b_\lambda^\pm=h_\pm^2 d_\lambda^\pm.$$

It has been shown by Sabatier[3] that $\mathbf{K}(r,r')$ defined by Eq. (2.10) is such that the following relation is satisfied:

$$\Psi_\lambda^\pm=F^\pm(r)u_\lambda(r)-\int_0^r ds\,s^{-2}K^\pm(r,s)u_\lambda(s) \tag{2.11}$$

if the potential coefficients are chosen in such a way that the following is valid:

$$K^\pm(r,r)=\tfrac{1}{2}rF^\pm(r)\{\pm r^2Q(r)+\int_0^r ds[s^3Q^2(s)-sV(s)]\}. \tag{2.12}$$

A look at the definition of $Q(r)$ enables us to realize that both of the relations in Eq. (2.12) can be satisfied if we have the following relation:

$$V(r)=r^2Q^2(r)-r^{-1}\frac{d}{dr}[r^{-1}F^-(r)K^+(r,r)$$
$$+r^{-1}F^+(r)K^-(r,r)]. \tag{2.13}$$

Since the central potential $V(r)$ is not known, therefore, Eq. (2.13) is not a condition to be satisfied by the functions involved, but it is the definition of the central potential in terms of the input functions. So by using the above procedure it follows that we are to choose $d_\lambda^\pm$ in such a way that the regular solutions $\Psi_\lambda^\pm$ defined by Eq. (2.11) have the desired asymptotic forms. If that is possible then the potentials $Q(r)$ and $V(r)$ defined by the above procedure are associated with the desired phase shifts. In other words, the inverse scattering problem for spin–orbit and centeral potentials has been solved.

In order to find $d_\lambda^\pm$ from the information on phase shifts at a fixed energy, we use Eq. (2.11) in Eq. (2.10) and find a representation of $\mathbf{K}(r,r')$ in terms of $\Psi_\lambda^\pm(r)$. Using this representation in Eq. (2.11), we find the following relations:

$$\Psi_\lambda^\pm(r)=F^\pm(r)u_\lambda(r)-\sum_{\mu\in\Omega}[\Psi_\mu^\mp(r)b_\mu^\pm-\Psi_\mu^\pm(r)a_\mu]L_\mu^\lambda(r), \tag{2.14}$$

where

$$L_\mu^\lambda(r)=\int_0^r u_\lambda(s)u_\mu(s)s^{-2}\,ds.$$

We have been able to show[4] that $d_\lambda^\pm$ can be found from Eq. (2.14), if the phase shifts are such that the following conditions are satisfied[6]:

$$\int_0^\infty r|Q(r)|dr<\infty,$$

$$Q(r)=o(r^{-3}) \quad \text{for large } r, \tag{2.15}$$

$$c_\lambda^\pm=o(\lambda^{1/3}) \quad \text{for large } \lambda,$$

where

$$c_\lambda^\pm=d_\lambda^\pm-a_\lambda.$$

As was shown before,[4] the $c_\lambda^\pm$ for nonphysical values of λ are arbitrary. A different set of values for them corresponds to different classes of spin–orbit and central potentials which are all associated with the same set of phase shifts for physical values of angular momenta. Therefore, if we are interested only in finding a set of spin–orbit and central potentials which corresponds to a given set of phase shifts, then we are perfectly justified in making a specific choice for $c_\lambda^\pm$, for nonphysical values of angular momenta. In this work we will choose them all to be zero, that is,

$$c_\lambda^\pm=0 \quad \text{for integer values of } \lambda. \tag{2.16}$$

It should be noted that $c_\lambda^\pm$ are the exact analog of the potential coefficients used by Newton[1] in his study of the inverse scattering problem for only central potentials. The choice he made in his work is identical with the one we have made through Eq. (2.16). Therfore, we expect the class of spin–orbit and central potentials to which we are restricting ourselves in this work to be the analog of the class of central potentials studied first by Newton.[1]

3. DETERMINATION OF POTENTIAL COEFFICIENTS

In this section we would like to find the potential coefficients $d_\lambda^\pm$ from the information on the asymptotic behavior of $\Psi_\lambda^\pm(r)$, that is, from the phase shifts for physical values of λ. As was stated before, we are assuming that Eq. (2.16) is satisfied. The procedure to be used is identical with that used in our previous work,[4] but due to the need of defining the relevant operators for our special case, we find it necessary to give a brief review of how we are going to construct $d_\lambda^\pm$ from the phase shifts at a fixed energy.

The starting point for finding $d_\lambda^\pm$ is, of course, Eq. (2.14). We have shown[4] that if Eqs. (2.15) are satisfied then we can use the asymptotic form of Eq. (2.14) when r tends to infinity. Doing so, and equating coefficients of e^{ir} and e^{-ir} separately, we find that

$$B_\lambda^\pm \exp[i(\delta_\lambda^\pm-\lambda\pi/2)]$$

$$=\exp[-i\lambda\pi/2]-\sum_{\mu\in\Omega}\{B_\mu^\mp d_\mu^\pm \exp[i\delta_\mu^\mp]$$

$$-B_\mu^\pm a_\mu \exp[i\delta_\mu^\pm]\}\exp[-i\mu\pi/2]L_\mu^\lambda. \tag{3.1}$$

In deriving Eq. (3.1) we used the following asymptotic forms for large r

$$\Psi_\lambda^\pm \sim h_\pm B_\lambda^\pm \sin[r+\delta_\lambda^\pm-\pi(\lambda-\tfrac{1}{2})/2],$$

$$u_\lambda \sim \sin[r-\pi(\lambda-\tfrac{1}{2})/2], \tag{3.2}$$

$$L_\mu^\lambda(\infty)=L_\mu^\lambda=[\lambda^2-\mu^2]^{-1}\sin\tfrac{1}{2}\pi(\lambda-\mu).$$

Next, in Eq. (3.1), we separate the terms associated with physical values of λ, from those associated with nonphysical values of λ. Writing the result in a matrix notation, we have

$$\varphi_0^\pm=\xi_0-L_0^0[a_0\varphi_0^\mp-a_0\varphi_0^\pm]-L_p^0 d_p^\pm\varphi_p^\mp, \tag{3.3}$$

$$\varphi_p^\pm=\xi_p-L_0^p[a_0\varphi_0^\mp-a_0\varphi_0^\pm]-L_p^p d_p^\pm\varphi_p^\mp, \tag{3.4}$$

where the column vectors $\varphi_0^\pm$ and $\varphi_p^\pm$ are defined in such a way that the elements of $\varphi_0^\pm$ are $B_\lambda^\pm\exp[i(\delta_\lambda^\pm-\lambda\pi/2)]$ with $\lambda\in\Omega_0=\{1,2,3,\cdots\}$ and the elements of $\varphi_p^\pm$ are similar except that $\lambda\in\Omega_p=\{\frac{1}{2},\frac{3}{2},\frac{5}{2},\cdots\}$. ξ_0 is a column vector with elements of $\exp(-i\lambda\pi/2)$ with $\lambda\in\Omega_0$. $d_p^\pm$ and a_0 are diagonal matrices with diagonal elements $d_\lambda^\pm$ and a_μ, where $\lambda\in\Omega_p$ and $\mu\in\Omega_0$. Note that in writing Eq. (3.3) and Eq. (3.4), we have made use of the fact that Eq. (2.16) is satisfied.

Using Eq. (3.3), we find that

$$[1-2a_0L_0^0]a_0(\varphi_0^--\varphi_0^+)=-a_0L_p^0[d_p^-\varphi_p^+-d_p^+\varphi_p^-]. \tag{3.5}$$

It is easy to show that the matrix $(1-2a_0L_0^0)=-2a_0\bar{L}_0^0$ is such that $\bar{L}_0^0$, except for some sign factors, is nothing but the matrix $M^{(1)}$. $M^{(\alpha)}$ is the matrix which was first defined and studied in detail by Sabatier.[10] We find it very remarkable that the matrix M, first defined by Newton,[1] and its generalization $M^{(\alpha)}$ play such an important role in the inverse scattering problem at a fixed energy. In this work we shall be dealing with $M^{(\alpha)}$ for values of α being equal to $\frac{1}{2}$, 1 and $\frac{3}{2}$.

Following Sabatier's results,[10] it is now easy to see that the matrix $\bar{L}_0^0$ has an inverse $\bar{\Gamma}$ and also there is a vector $\bar{v}$ which is annihilated by $\bar{L}_0^0$. The matrix $\bar{\Gamma}$ and the vector $\bar{v}$ have the following forms:

$$\bar{\Gamma}_{2l+\beta}^{2m+1+\beta}=\bar{\Gamma}_{2m+1+\beta}^{2l+\beta}$$

$$=-\frac{4[(2m+1+\beta)^2-\beta^2]\alpha_l^{0(\beta)}\alpha_m^{1(\beta)}}{\pi[(2l+\beta)^2-(2m+1+\beta)^2]}$$

and (3.6)

$$\bar{\Gamma}_{2m+\beta}^{2l+\beta}=\bar{\Gamma}_{2m+1+\beta}^{2l+1+\beta}=0,$$

$$\bar{v}_{2m+\beta}=\alpha_m^{0(\beta)} \quad \text{and} \quad \bar{v}_{2m+1+\beta}=0, \tag{3.7}$$

where

$$\alpha_n^{1(\beta)}=\Gamma(n+1+\beta)\Gamma(\tfrac{1}{2})(2n+1+\beta)/n!\,\Gamma(\tfrac{1}{2}-n)\Gamma(n+\beta+\tfrac{3}{2}),$$

$$\alpha_n^{0(\beta)}=\Gamma(n+\beta)\Gamma(\tfrac{1}{2})(2n+\beta)/n!\,\Gamma(\tfrac{1}{2}-n)\Gamma(n+\beta+\tfrac{1}{2}) \tag{3.8}$$

with $\beta=1$. We should note that $\alpha_n^{0(\beta)}$ and $\alpha_n^{1(\beta)}$ are nothing but the functions α_n^0 and α_n^1 first defined by Sabatier.[10] In this work since we need to work with these functions for different values of β, therefore we need to elevate their importance by adding another index β in their definition. In spite of the simpler appearance the $\bar{\Gamma}$ given above is equal to that of Sabatier.[10] The realization that $\bar{\Gamma}$ can be written in the above form was the crucial point in being able to evaluate the matrices, which are to follow in a closed form.

Having the inverse of $\bar{\Gamma}$ and knowing that a_0^{-1} exists, we find that

$$a_0(\varphi_0^- - \varphi_0^+) = -\bar{\gamma}\bar{v} + \tfrac{1}{2}\bar{\Gamma}L_p^0[d_p^-\varphi_p^+ - d_p^+\varphi_p^-], \tag{3.9}$$

where

$$\bar{\gamma} = \gamma e^{i\theta}$$

with γ and θ being arbitrary finite real constants. Having found $a_0(\varphi_0^- - \varphi_0^+)$, we substitute them in Eq. (3.4) to get an equation containing only terms related to physical values of λ:

$$\varphi_p^\pm = \bar{\eta}_p^\pm - (L_p^p + \bar{S}_p^p)d_p^\pm\varphi_p^\mp + \bar{S}_p^p\, d_p^\mp\varphi_p^\pm \tag{3.10}$$

where

$$\bar{\eta}_p^\pm = \xi_p \pm \bar{\gamma}L_0^p\bar{v}, \quad \bar{S}_p^p = -\tfrac{1}{2}L_0^p\bar{\Gamma}L_p^0 .$$

Making a long but essentially similar calculations as done by Sabatier[10] for finding the inverse of $M^{(\alpha)}$, we find that

$$\bar{S}_\mu^\lambda = -\tfrac{1}{2}[1 + \lambda/2\mu + \mu/2\lambda]L_\mu^\lambda + [\sin\tfrac{1}{2}\pi(\lambda+\mu)]/4\lambda\mu, \tag{3.11}$$

$$\bar{\eta}_\lambda^\pm = \eta_\lambda^\pm \exp[i(\Delta_\lambda^\pm - \pi\lambda/2)], \tag{3.12}$$

where

$$\eta_\lambda^\pm \cos\Delta_\lambda^\pm = 1 \pm \gamma(\sqrt{\pi}/2\lambda)\sin(\pi\lambda/2)\cos(\theta + \lambda\pi/2)$$

and

$$\eta_\lambda^\pm \sin\Delta_\lambda^\pm = \pm\gamma(\sqrt{\pi}/2\lambda)\sin(\pi\lambda/2)\sin(\theta + \lambda\pi/2).$$

Some of the important relations needed to find $\bar{S}_\mu^\lambda$ in a closed form are proven in the Appendix.

Next we write Eq. (3.10) in terms of its elements:

$$B_\lambda^\pm \exp[i(\delta_\lambda^\pm - \lambda\pi/2)]$$

$$= \bar{\eta}_\lambda^\pm - \sum_{\mu\in\Omega_p}[L_\mu^\lambda + \bar{S}_\mu^\lambda]d_\mu^\pm B_\mu^\mp \exp[i(\delta_\mu^\mp - \mu\pi/2)]$$

$$+ \sum_{\mu\in\Omega_p}\bar{S}_\mu^\lambda d_\mu^\mp B_\mu^\pm \exp[i(\delta_\mu^\pm - \mu\pi/2)]. \tag{3.13}$$

If we equate the real and imaginary parts separately, after we have multiplied Eq. (3.13) by $\exp[-i(\delta_\lambda^\pm - \lambda\pi/2)]$, we find that

$$\tilde{\gamma}_\lambda^\pm = \sum_{\mu\in\Omega_p} D_\mu^\mp[(1 + \tan\delta_\mu^\mp\tan\delta_\lambda^\pm)(M_\mu^\lambda + S_\mu^\lambda)$$

$$- (\tan\delta_\mu^\mp - \tan\delta_\lambda^\pm)N_\mu^\lambda] \tag{3.14}$$

$$- \sum_{\mu\in\Omega_p} D_\mu^\pm[(1 + \tan\delta_\mu^\pm\tan\delta_\lambda^\pm)S_\mu^\lambda - (\tan\delta_\mu^\pm - \tan\delta_\lambda^\pm)N_\mu^\lambda]$$

and

$$B_\lambda^\pm = \eta_\lambda^\pm\cos(\Delta_\lambda^\pm - \delta_\lambda^\pm) + \sum_{\mu\in\Omega_p}\{D_\mu^\pm[(\cos\delta_\lambda^\pm + \tan\delta_\mu^\pm\sin\delta_\lambda^\pm)\bar{N}_\mu^\lambda$$

$$+ (\tan\delta_\mu^\pm\cos\delta_\lambda^\pm - \sin\delta_\lambda^\pm)S_\mu^\lambda]$$

$$- D_\mu^\mp[(\cos\delta_\lambda^\pm + \tan\delta_\mu^\mp\sin\delta_\lambda^\pm)N_\mu^\lambda$$

$$+ (\tan\delta_\mu^\mp\cos\delta_\lambda^\pm - \sin\delta_\lambda^\pm)(M_\mu^\lambda + S_\mu^\lambda)]\}, \tag{3.15}$$

where[8]

$$M_\mu^\lambda = i^{\lambda-\mu+1}L_\mu^\lambda = 1/[\mu^2 - \lambda^2] \quad \text{if } \lambda - \mu \text{ odd},$$

$$= 0 \quad \text{otherwise},$$

$$S_\mu^\lambda = i^{\lambda-\mu+1}\bar{S}_\mu^\lambda = -\tfrac{1}{2}M_\mu^\lambda(1 + \lambda/2\mu + \mu/2\lambda),$$

$$N_\mu^\lambda = i^{\lambda-\mu}\bar{S}_\mu^\lambda + \left(\frac{\pi}{4\lambda}\right)\delta_\mu^\lambda = \{\cos[\pi(\lambda-\mu)/2]$$

$$\times \sin[\pi(\lambda+\mu)/2]\}/4\lambda\mu, \tag{3.16}$$

$$\bar{N}_\mu^\lambda = N_\mu^\lambda - (\pi/4\lambda)\delta_\mu^\lambda,$$

$$D_\lambda^\pm = B_\lambda^\pm d_\lambda^\mp\cos\delta_\lambda^\pm,$$

$$\tilde{\gamma}_\lambda^\pm = \eta_\lambda^\pm(\cos\Delta_\lambda^\pm\tan\delta_\lambda^\pm - \sin\Delta_\lambda^\pm).$$

So the problem of finding $d_\lambda^\pm$ is reduced to the problem of finding $D_\lambda^\pm$ from Eq. (3.14). Using the found $D_\lambda^\pm$, we can then find $B_\lambda^\pm$ from Eq. (3.15). Having found $D_\lambda^\pm$ and $B_\lambda^\pm$, using Eq. (3.16), we are then able to find $d_\lambda^\pm$.

Except for giving the explicit form of all the operators involved in our method, what has been presented up to now is the same as in our previous work.[4] At this point we have to depart from the previous method for finding $D_\lambda^\pm$, because the specific choice of $c_\lambda^\pm$ (for nonphysical values of the angular momentum) made here would force severe conditions on $\delta_\lambda^\pm$ in order for the matrices involved to be invertible. The reason is that in Eq. (3.14) the value of $\delta_{1/2}^-$, which corresponds to $j = -\frac{1}{2}$, is not physical and is not given to us by experimentation. Therefore, the value of $\delta_{1/2}^-$ is a free parameter in this analysis. Unfortunately, there seems to be no simple way to use this freedom on $\delta_{1/2}^-$ in Eq. (3.14), in order to make the inversion possible in general. To make use of this freedom, we will follow a method which is essentially the same as that used for eliminating the coefficients associated with other nonphysical values of λ. That is, instead of choosing $\delta_{1/2}^-$ arbitrary, we choose the associated potential coefficient $d_{1/2}^-$ to be arbitrary. Of course, once $d_{1/2}^-$ is fixed, $\delta_{1/2}^-$ is no longer arbitrary; it must be defined in a way that is consistent with the rest of our analysis.

Choosing $d_{1/2}^-$ arbitrarily and doing the analysis which is to follow, one finds that the problem of inversion becomes possible if we choose $d_{1/2}^-$ to have the following value:

$$d_{1/2}^- = -2/\pi . \tag{3.17}$$

With this choice of $d_{1/2}^-$, we note from Eq. (3.13) that the following is satisfied:

$$d_{1/2}^- B_{1/2}^+ \exp[i(\delta_{1/2}^+ - \pi/4)]$$

$$= -\bar{\eta}_{1/2}^+ + \sum_{\mu\in\Omega_p^+}(L_\mu^{1/2} + \bar{S}_\mu^{1/2})d_\mu^+ B_\mu^- \exp[i(\delta_\mu^- - \mu\pi/2)]$$

$$- \sum_{\mu\in\Omega_p^-}\bar{S}_\mu^{1/2}d_\mu^- B_\mu^+ \exp[i(\delta_\mu^+ - \mu\pi/2)], \tag{3.18}$$

where

$$\Omega_p^- = \{\tfrac{3}{2}, \tfrac{5}{2}, \tfrac{7}{2}, \cdots\} \quad \text{and} \quad \Omega_p^+ = \Omega_p.$$

Next we substitute Eq. (3.18) in Eq. (3.13) and get a new form for $B_\lambda^\pm$ in which the factor on the left side of the Eq. (3.18) does not appear:

$$B_\lambda^\pm \exp[i(\delta_\lambda^\pm - \lambda\pi/2)]$$

$$= \bar{\eta}_\lambda^\pm + \Lambda_{1/2}^{\lambda\pm}\bar{\eta}_{1/2}^+ - \sum_{\mu\in\Omega_p^\mp}[\Lambda_\mu^{\lambda+} + \Lambda_{1/2}^{\lambda\pm}\Lambda_\mu^{1/2\pm}]d_\mu^\mp B_\mu^\pm$$

$$\times \exp[i(\delta_\mu^\pm - \mu\pi/2)] - \sum_{\mu\in\Omega_\beta^\pm} [\Lambda_\mu^{\lambda-} + \Lambda_{\frac{1}{2}}^{\lambda\pm}\Lambda_\mu^{1/2\mp}] d_\mu^\pm B_\mu^\mp$$

$$\times \exp[i(\delta_\mu^\mp - \mu\pi/2)], \tag{3.19}$$

where

$$\Lambda_\mu^{\lambda\pm} = \tfrac{1}{2}[L_\mu^\lambda \mp (L_\mu^\lambda + 2\bar{S}_\mu^\lambda)].$$

Next, multiplying Eq. (3.19) by $\exp[-i(\delta_\lambda^\pm - \lambda\pi/2)]$, we find that the imaginary part of the resulting equation for even nonzero values of the angular momentum satisfy the following equations:

$$\sigma_n^\pm = \sum_{m=0}^{\infty} \bar{P}_m^+ \{X_m^n \mp Z_m^n + (T_n^+ \pm T_n^-)[\pm Y_m^n \bar{T}_m^- + (X_m^n \mp Z_m^n)\bar{T}_m^+]\}$$

$$+ \sum_{m=0}^{\infty} \bar{P}_m^- \{\pm Y_m^n + (T_n^+ \pm T_n^-)[\pm Y_m^n \bar{T}_m^+ + (X_m^n \mp Z_m^n)\bar{T}_m^-]\}, \tag{3.20}$$

where

$$T_m^\pm = \tfrac{1}{2}[\tan\delta_{2m+2+1/2}^+ \pm \tan\delta_{2m+2+1/2}^-],$$

$$\bar{T}_m^\pm = \tfrac{1}{2}[\tan\delta_{2m+1+1/2}^+ \pm \tan\delta_{2m+1+1/2}^-],$$

$$\bar{P}_m^\pm = \tfrac{1}{2}[D_{2m+1+1/2}^+ \pm D_{2m+1+1/2}^-],$$

$$\sigma_n^\pm = [1 \mp 1/2(2n+5/2)](T_n^+ \pm T_n^-),$$

$$X_m^n = M_{2m+1+1/2}^{2n+2+1/2}, \quad Z_m^n = M_{2m+3/2}^{1/2}[2(2n+\tfrac{5}{2})]^{-1},$$

and

$$Y_m^n = \{(2m+\tfrac{3}{2})[(2n+\tfrac{5}{2})^2 - \tfrac{1}{4}]/(2n+\tfrac{5}{2})[(2m+\tfrac{3}{2})^2 - \tfrac{1}{4}]\} X_m^n .$$

Doing the same thing for the case when angular momentum is equal to zero, we get

$$\tilde{\Omega}^- \tan\delta_{1/2}^- = \sum_{m=0}^{\infty} \bar{P}_m^+ M_{2m+3/2}^{1/2}, \tag{3.21}$$

$$\tilde{\Omega}^+ = \sum_{m=0}^{\infty} \left[\frac{T_m^- P_m^+ + (T_m^+ - \tan\delta_{1/2}^+)P_m^-}{2m+\frac{5}{2}}\right] - D_{1/2}^-[\tan\delta_{1/2}^- - \tan\delta_{1/2}^+], \tag{3.22}$$

where

$$\tilde{\Omega}^+ = [1 + (\pi/2)^{1/2}\gamma\cos(\theta+\pi/4)]\tan\delta_{1/2}^+ - (\pi/2)^{1/2}\gamma\sin(\theta+\pi/4)$$

$$- \sum_{m=0}^{\infty} \{[M_{2m+3/2}^{1/2}(1 + \bar{T}_m^+ \tan\delta_{1/2}^+) + R_{2m+3/2}^{1/2}\bar{T}_m^- \tan\delta_{1/2}^+]\bar{P}_m^+ + [R_{2m+3/2}^{1/2}(1 + \bar{T}_m^+ \tan\delta_{1/2}^+) + M_{2m+3/2}^{1/2}\bar{T}_m^- \tan\delta_{1/2}^+]\bar{P}_m^-\},$$

$$\tilde{\Omega}^- = 1 - \sum_{m=0}^{\infty} [\bar{P}_m^+ \bar{T}_m^+ + \bar{P}_m^- \bar{T}_m^-] M_{2m+3/2}^{1/2},$$

$$P_n^\pm = \tfrac{1}{2}[D_{2n+2+1/2}^+ \pm D_{2n+2+1/2}^-],$$

and

$$R_\mu^\lambda = [(\lambda^2 + \mu^2)/2\lambda\mu]M_\mu^\lambda .$$

For odd values of angular momenta, from the imaginary part of Eq. (3.19) after it has been multiplied by $\exp[-i(\delta_\lambda^\pm - \lambda\pi/2)]$, we find that the following is to be satisfied:

$$\bar{\sigma}_l^\pm + D_{\frac{1}{2}}^- \tilde{\sigma}_l^\pm = \sum_{m=0}^{\infty} \{\bar{X}_m^l + (\bar{T}_l^+ \pm \bar{T}_l^-)[\bar{X}_m^l T_m^+ - (\bar{Z}_m^l \mp \bar{Y}_m^l)T_m^-]\}P_m^+$$

$$+ \sum_{m=0}^{\infty} \{(\bar{T}_l^+ \pm \bar{T}_l^-)[\bar{X}_m^l T_m^- - (\bar{Z}_m^l \mp \bar{Y}_m^l)T_m^+] - \bar{Z}_m^l \pm \bar{Y}_m^l\}P_m^- , \tag{3.23}$$

where

$$\bar{X}_m^l = M_{2m+1+3/2}^{2l+3/2}, \quad \bar{Z}_m^l = M_{\frac{1}{2}}^{2l+3/2}[2(2m+\tfrac{5}{2})]^{-1},$$

$$\bar{Y}_m^l = \frac{(2l+\frac{3}{2})[(2m+\frac{5}{2})^2 - \frac{1}{4}]}{(2m+\frac{5}{2})[(2l+\frac{3}{2})^2 - \frac{1}{4}]}\bar{X}_m^l ,$$

$$\tilde{\sigma}_l^\pm = [1 + \tan\delta_{\frac{1}{2}}^-(\bar{T}_l^+ \pm \bar{T}_l^-)]/[(2l+\tfrac{3}{2})^2 - \tfrac{1}{4}],$$

$$\bar{\sigma}_l^\pm = [\bar{T}_l^+ \pm \bar{T}_l^-]\left[1 \mp \frac{(\frac{1}{2}\pi)^{1/2}\gamma\sin(\theta+\pi/4)}{2l+\frac{3}{2}\mp\frac{1}{2}}\right]$$

$$+ \sum_{m=0}^{\infty} [(\tilde{\Lambda}_{m+}^{l\pm} - \tilde{\Lambda}_{m-}^{l\pm})\bar{P}_m^+ - (\tilde{\Lambda}_{m+}^{l\pm} + \tilde{\Lambda}_{m-}^{l\pm})\bar{P}_m^-]$$

$$- \frac{1}{2(2l+\frac{3}{2})}\left(\frac{1 - \tilde{\alpha} + (\frac{1}{2}\pi)^{1/2}\gamma\cos(\theta+\pi/4)}{2l+\frac{3}{2}\mp\frac{1}{2}} \pm [\sqrt{2}\pi^{1/2}\gamma\cos(\theta+\pi/4) + 1 + \tilde{\alpha}]\right),$$

$$\tilde{\alpha} = \sum_{m=0}^{\infty} \frac{(2m+\frac{3}{2})[\bar{T}_m^+\bar{P}_m^+ + \bar{T}_m^-\bar{P}_m^-] + [(2m+\frac{3}{2})^2 + \frac{1}{4}][\bar{T}_m^+\bar{P}_m^- + \bar{T}_m^-\bar{P}_m^+]}{(2m+\frac{3}{2})[(2m+\frac{3}{2})^2 - \frac{1}{4}]}$$

$$\tilde{\alpha} = \sum_{m=0}^{\infty} \left(\frac{(\bar{T}_m^+ - \bar{T}_m^-)(\bar{P}_m^+ - \bar{P}_m^-)}{2m+2} - \frac{(\bar{T}_m^+ + \bar{T}_m^-)(\bar{P}_m^+ + \bar{P}_m^-)}{2m+1}\right)$$

and[11]

$$\tilde{\Lambda}_{m\pm}^{l\pm} = \left([\bar{T}_l^- \pm \bar{T}_l^+] + \frac{[2m+\frac{3}{2}\mp\frac{1}{2}]}{2(2m+\frac{3}{2})[2l+\frac{3}{2}\mp\frac{1}{2}]}[\bar{T}_l^+ \pm \bar{T}_l^-]\right) \Big/ 2(2l+\tfrac{3}{2})[2m+\tfrac{3}{2}\pm\tfrac{1}{2}].$$

At this point, it is illuminating to note that Eqs. (3.20), (3.22), and (3.23) are equivalent to Eq. (3.14) for our special choice of $\delta_{1/2}^-$ given by Eq. (3.21). The reason for choosing to work with the former equations is that by making a specific choice for $d_{1/2}^-$ we have been able to eliminate the troublesome coupling between $D_\lambda^\pm$ for even and odd values of the angular momentum. Having done so, we will be able to find the inverses of the matrices involved and to justify the necessary interchange of summations, without having to consider the detailed form of $\delta_\lambda^\pm$ appearing in the above matrices. So the method to be followed is first to find $D_\lambda^\pm$ for odd values of the angular momenta. Having found $D_\lambda^\pm$ for odd values of the angular momentum, we are then in a position to find $\tilde{\Omega}^\pm$, $\delta_{\frac{1}{2}}^-$, $\tilde{\alpha}$, $\tilde{\sigma}^\pm$, $\tilde{\alpha}$, and $\bar{\sigma}_l^\pm$ from Eqs. (3.21), (3.22), and (3.23). Next we can solve for $D_\lambda^\pm$ for nonzero even angular momenta from Eq. (3.23) in

terms of $D^-_{1/2}$. Finally this information is put into Eq. (3.22) from which $D^-_{1/2}$ can be found. Thus we have been able to find $D^-_{1/2}$ and $D^\pm_\lambda$ for all values of $\lambda \neq \frac{1}{2}$. Since we already know the form of $d^-_{1/2}$, we can use all the above information in Eq. (3.15). Doing so and using Eq. (3.16), we then can find the potential coefficients $d^\pm_\lambda$ for all possible values of the angular momentum.

At this point we may also note that our method treats the information on $\delta^+_{1/2}$ in a very different way from the information on $\delta^\pm_\lambda$ for $\lambda \neq \frac{1}{2}$. This we should have expected, because, looking at Eq. (2.2), we note that the $\delta^\pm_\lambda$ for $\lambda \neq \frac{1}{2}$ contain information about the central and spin orbit potentials, while $\delta^+_{1/2}$ contains information only about the central potential. That is the reason for treating the information on $\delta^+_{1/2}$, that is, Eq. (3.22), separately from the information given by the other phase shifts.

We should also note that, for very special values of $\delta^\pm_\lambda$, it may happen that the unknown constants $\delta^-_{1/2}$ and/or $D^-_{1/2}$ drop out of Eq. (3.21) and/or Eq. (3.22). In that case Eq. (3.21) and/or Eq. (3.22) become conditions to be satisfied by the factors involved. Should we not be able to satisfy those conditions, then our method does not work for those special sets of phase shifts. Of course, this is nothing new. We have already seen examples of phase shifts for which the associated central potentials cannot be constructed using the Newton[1] method.

The next step in our method is to evaluate $\bar{P}^\pm_m$ from Eq. (3.20). To do so, we shall keep one of the equations in Eq. (3.20) and instead of the other equality, we use the sum of the two equations in Eq. (3.20). We then find an equivalent relation to Eq. (3.20), which in a matrix notation, has the following form:

$$\sigma^+ = [X - Z + W^+_+ + W^+_-]\bar{P}^+ + [Y + W^-_+ + W^-_-]\bar{P}^-, \tag{3.24}$$

$$\tfrac{1}{2}\sigma^+ + \tfrac{1}{2}\sigma^- = [X + W^+_+]\bar{P}^+ + W^-_-\bar{P}^-, \tag{3.25}$$

where[11]

$$W^\pm_\pm = [T^\mp]X[\bar{T}^\pm] + [T^\pm]\{Y[\bar{T}^\mp] - Z[\bar{T}^\pm]\}.$$

The next step is to find a right inverse for X. We note that we can think of elements of X either as some of the elements of M or $M^{(3/2)}$. But a look at Eq. (3.24) informs us that we can write the following in a simpler form if we think of X to be associated with matrix M. In that case let us define $\tilde{X}$ in the following form:

$$\tilde{X}^m_l = \frac{4}{\pi}(-1)^{m+l+1}\frac{[(2m+1+\frac{1}{2})^2 - \frac{1}{4}]\alpha^{1(1/2)}_m \alpha^{0(1/2)}_{l+1}}{(2m+1+\frac{1}{2})^2 - (2l+2+\frac{1}{2})^2}. \tag{3.26}$$

One can show that[10]

$$X\tilde{X} = (X - Z)\tilde{X} = 1. \tag{3.27}$$

We should also note that there exists a vector v^0 which is annihilated by X. That is

$$Xv^0 = 0, \tag{3.28}$$

where

$$v^0 = (-1)^n \alpha^{1(1/2)}_n.$$

Using the method given first by Newton[1] for the case of a central potential, we find that the matrix $[1 + \tilde{X}W^+_+]$ is invertible if the phase shifts tend to zero fast enough for large values of the angular momentum. It then follows that

$$\bar{P}^+ = \bar{R}\{\tilde{X}[\tfrac{1}{2}\sigma^+ + \tfrac{1}{2}\sigma^- - W^-_-\bar{P}^-] + \gamma^0 v^0\}, \tag{3.29}$$

where

$$\bar{R} = [1 + \tilde{X}W^+_+]^{-1},$$

where γ^0 is some arbitrary constant.

Substituting Eq. (3.28) in Eq. (3.24), we find that $\bar{P}^-$ satisfies the following equation:

$$\tilde{V}^- = (Y + \tilde{W}^-)\bar{P}^-, \tag{3.30}$$

where

$$\tilde{W}^- = W^-_- - W^+_+\bar{R}\tilde{X}W^-_-,$$

$$\tilde{V}^- = \tfrac{1}{2}\sigma^+ - \tfrac{1}{2}\sigma^- - \tfrac{1}{2}W^+_+\bar{R}[\tilde{X}(\sigma^+ + \sigma^-) + 2\gamma^0 v^0] + \sqrt{\pi}\gamma^0/4(2n + \tfrac{5}{2}).$$

Again because of the way $\tilde{W}^-$ depends on $\tan\delta^\pm_\lambda$ we note $Y + W$ is invertible, if Y has an inverse. Keeping in mind that for later calculations $\bar{P}^-_n$ will be required to tend to zero for large values of n, it is possible to show that $\tilde{Y}$ defined below is the desired right inverse of Y:

$$\tilde{Y}^m_l = \frac{4(2l+1+\frac{3}{2})[(2m+\frac{3}{2})^2 - \frac{1}{4}](-1)^{m+l}[(2l+1+\frac{3}{2})^2 - \frac{9}{4}]}{\pi(2m+\frac{3}{2})[(2l+1+\frac{3}{2})^2 - \frac{1}{4}][(2m+\frac{3}{2})^2 - (2l+1+\frac{3}{2})^2]}\alpha^{0(3/2)}_m \alpha^{1(3/2)}_l. \tag{3.31}$$

It also follows that the vector v^1, defined below, is annihilated by Y:

$$v^1_n = [(2n+\tfrac{3}{2})^2 - \tfrac{1}{4}](-1)^n \alpha^{0(3/2)}_n/(2n+\tfrac{3}{2}). \tag{3.32}$$

But since $\bar{P}^-$ has to tend to zero for large values of n, the above nonuniqueness does not lead to nonuniqueness in the definition of $\bar{P}^-$. It follows that

$$\bar{P}^- = [1 + \tilde{Y}\tilde{W}^-]^{-1}\tilde{Y}\tilde{V}^-. \tag{3.33}$$

Clearly Eq. (3.33) is the solution to Eq. (3.30) if necessary interchange of summations is justified. For all the terms containing $\delta^\pm_\lambda$, making suitable conditions on the asymptotic behavior of $\delta^\pm_\lambda$, the interchange of summations can be justified. The only term that may cause a problem is the last one appearing in the definition of $\tilde{V}^-$. Using the relations given in the Appendix, we can explicitly find the value of the term appearing in the left-hand side of the following equation. Having done so, we have

$$\sum_{m=0}^{\infty} Y^n_m \left\{\sum_{l=0}^{\infty} \tilde{Y}^m_l [2l + \tfrac{5}{2}]^{-1}\right\} = [2n + \tfrac{5}{2}]^{-1}. \tag{3.34}$$

Therefore, the interchange of summations is justified, and $\bar{P}^-$ defined by Eq. (3.33) satisfies Eq. (3.30).

Having found $\bar{P}_n^{\pm}$, given by Eqs. (3.33) and (3.29), we are then in a position to find $\delta_{1/2}^-$, $\bar{\alpha}$, $\tilde{\Omega}^{\pm}$, $\bar{\sigma}_n^{\pm}$, and $\tilde{\sigma}_n^{\pm}$ from Eqs. (3.21), (3.22), and (3.23). Noting that $\bar{P}_n^-$ tends to zero and $\bar{P}_n^+$ tends to a constant for large n, we see that the above quantities are all well defined if $\delta_\lambda^{\pm}$ tends to zero rapidly enough for large values of the angular momentum. Knowing the above quantities, we are then in a position to find $P_n^{\pm}$ from Eq. (3.23). Again, keeping one of the equations in Eq. (3.23) as it is and, instead of the other equality, using the sum of the two equations in Eq. (3.23), we find an equivalent relation to Eq. (3.23), which in a matrix notation, has the following form:

$$\bar{\sigma}^+ + D_{1/2}^- \tilde{\sigma}^+ = [\bar{X} + U_+^- + U_+^+]P^+ + [\bar{Y} - \bar{Z} + U_-^- + U_-^+]P^-, \tag{3.35}$$

$$\tfrac{1}{2}(\bar{\sigma}^+ + \bar{\sigma}^-) + \tfrac{1}{2}D_{1/2}^-(\tilde{\sigma}^+ + \tilde{\sigma}^-) = [\bar{X} + U_+^+]P^+ + [-\bar{Z} + U_-^+]P^-, \tag{3.36}$$

where[11]

$$U_{\pm}^{\pm} = [\bar{T}^{\pm}]\{\bar{X}[T^{\pm}] - \bar{Z}[T^{\mp}]\} + [\bar{T}^{\mp}]\bar{Y}[T^{\mp}].$$

In this case, it is easy to see that $\bar{X}$ is related to $M^{(3/2)}$ so the right inverse, $\tilde{Z}$, of $\bar{X}$ exists. Furthermore, we do not have a vector[12] which is annihilated by $\bar{X}$. The right inverse of $\bar{X}$ is given by the following relation:

$$\tilde{Z}_l^m = \frac{4}{\pi}(-1)^{l+m}\frac{[(2m+1+\frac{3}{2})^2 - \frac{9}{4}]\alpha_m^{1(3/2)}\alpha_l^{0(3/2)}}{(2m+1+\frac{3}{2})^2 - (2l+\frac{3}{2})^2}. \tag{3.37}$$

We can then claim that P^+, defined by the following, is the solution to Eq. (3.36):

$$P^+ = \tilde{R}\{\tfrac{1}{2}(\bar{\sigma}^+ + \bar{\sigma}^-) + \tfrac{1}{2}D_{1/2}^-(\tilde{\sigma}^+ + \tilde{\sigma}^-) + [\bar{Z} - U_-^+]P^-\}, \tag{3.38}$$

where

$$\tilde{R} = [1 + \tilde{Z}U_+^+]^{-1}\tilde{Z}.$$

To prove the above statement, we need to justify the necessary interchange of summations. If $\delta_\lambda^{\pm} = o(\lambda^{-3})$ for large values of λ, then, using an argument similar to the one given for Eq. (3.33), we see that the interchange of summation is justified. Therefore, P^+ defined by Eq. (3.38) satisfies Eq. (3.36). Substituting Eq. (3.38) in Eq. (3.35), we find an equation for P^-,

$$\bar{V}^+ + D_{1/2}^-\bar{V}^- = (Y + \tilde{W}^+)P^- \tag{3.39}$$

where

$$\bar{V}^+ = \tfrac{1}{2}[\tilde{\sigma}^+ - \bar{\sigma}^- - \tilde{A}(\tilde{\sigma}^+ + \bar{\sigma}^-)], \quad \bar{V}^- = \tfrac{1}{2}[\tilde{\sigma}^+ - \tilde{\sigma}^- - \tilde{A}(\tilde{\sigma}^+ + \tilde{\sigma}^-)], \quad \tilde{W}^+ = U_-^- + \tilde{A}(\bar{Z} - U_-^+), \quad \text{and} \quad \tilde{A} = U_+^-\tilde{R}.$$

It is easy to write down the form of the right inverse, A, of the matrix Y. We can also see[12] that there is no vector which is annihilated by Y:

$$A_l^m = \frac{4(2m+1+\frac{3}{2})[(2l+\frac{3}{2})^2 - \frac{1}{4}][(2m+1+\frac{3}{2})^2 - \frac{9}{4}](-1)^{l+m}\alpha_m^{1(3/2)}\alpha_l^{0(3/2)}}{\pi[(2m+1+\frac{3}{2})^2 - \frac{1}{4}](2l+\frac{3}{2})[(2m+1+\frac{3}{2})^2 - (2l+\frac{3}{2})^2]}. \tag{3.40}$$

At this point, we should note that we could have written Eq. (3.31) and Eq. (3.40) in simpler forms, had we used the existing relations between $\alpha_l^{0(3/2)}$, $\alpha_l^{1(3/2)}$, $\alpha_l^{0(1/2)}$, and $\alpha_l^{1(1/2)}$. But we have not done so, because the forms given indicate their origin in a more illuminating fashion. We also should note that although the matrix $[1 + A\tilde{W}^+]$ looks similar to other matrices in this analysis, finding its inverse may not be as simple as the inverses of the other similar matrices, even if we assume that all the phase shifts vanish for values of the angular momentum greater than some number. The reason is that in the definition of $\tilde{W}^+$ there appears a term associated with the matrix $\bar{Z}$. Noting that $\bar{Z}$ is a degenerate matrix,[13] we see that with a little extra work $[1 + A\tilde{W}^+]$ can still be inverted. Having done so, we have

$$P^- = \tilde{V}^+ + D_{1/2}^-\tilde{V}^0, \tag{3.41}$$

where

$$\tilde{V}^+ = [1 + A\tilde{W}^+]^{-1}A\bar{V}^+, \quad \tilde{V}^0 = [1 + A\tilde{W}^+]^{-1}A\bar{V}^-.$$

So again P^- defined by Eq. (3.41) is a solution of Eq. (3.39) if the necessary interchange of summations is justified. Using an argument similar to the one given for Eq. (3.33), we can justify the interchange of summations if the arbitrary constants γ and θ satisfy the following equation:

$$\gamma\cos(\theta + \pi/4) = -(1 + \bar{\alpha})(2\pi)^{-1/2}. \tag{3.42}$$

With the above choice of γ and θ the interchange of summations is justified and P^- defined by Eq. (3.41) is the solution to Eq. (3.39).

The last step in our method is to find $D_{1/2}^-$ from Eq. (3.22). To do so, we substitute the values of $P^{\pm}$ given by Eq. (3.41) and Eq. (3.38) into Eq. (3.22), then we are left with the following equation:

$$\bar{\Omega}D_{1/2}^- = \tilde{\Omega}^+ - \sum_{m=0}^{\infty}\frac{T_m^-\tilde{U}_m^+ + (T_m^+ - \tan\delta_{1/2}^+)\bar{V}_m^+}{2(m+\frac{5}{2})} \tag{3.43}$$

where

$$\bar{\Omega}=\tan\delta^{+}_{1/2}-\tan\delta^{-}_{1/2}+\sum_{m=0}^{\infty}\frac{T^{-}_{m}\tilde{U}^{0}_{m}+(T^{+}_{m}-\tan\delta^{+}_{1/2})\tilde{V}^{0}_{m}}{(2m+\frac{5}{2})},\quad \tilde{U}^{+}=\tfrac{1}{2}\tilde{R}[\tilde{\sigma}^{+}+\tilde{\sigma}^{-}+2(\bar{Z}-U^{+}_{-})\tilde{V}^{+}],\quad \tilde{U}^{0}=\tfrac{1}{2}\tilde{R}[\tilde{\sigma}^{+}+\tilde{\sigma}^{-}+2(\bar{Z}-U^{+}_{-})\tilde{V}^{0}].$$

In general $\bar{\Omega}$ is different from zero. Therefore, Eq. (3.43) gives us $D^{-}_{1/2}$. Of course, it may happen that for some special set of phase shifts $\bar{\Omega}$ becomes equal to zero, in that case $D^{-}_{1/2}$ is not given by Eq. (3.43) and (3.43) becomes a condition on the potential coefficients. In that case if Eq. (3.43) is satisfied, then $D^{-}_{1/2}$ can be chosen aribtrarily, and if Eq. (3.43) is not satisfied, then we are dealing with an exceptional set of phase shifts for which our method cannot give any solution. As is customary, in this work we will not be interested in solving the inverse scattering problem for such sets of exceptional phase shifts.

At this point we have solved the problem of finding $D^{\pm}_{\lambda}$ from the information on phase shifts at a fixed energy. Using Eq. (3.15), we find the potential coefficients $d^{\pm}_{\lambda}$. As expected, the information on phase shifts does not give us $d^{\pm}_{\lambda}$ uniquely. The solution depends on two arbitrary constants γ^{0} and θ.

4. TRANSPARENT SPIN-ORBIT POTENTIALS

In order to clarify the method and also because of its own right, we would like to apply the method given in Sec. 3 to the most trivial, but interesting inverse scattering problem for spin–orbit potentials. That is, we would like to find the potential coefficients associated with the set of physical phase shifts which are zero for all the physical values of the angular momentum at a fixed energy.

$$\delta^{\pm}_{2l+2+1/2}=\delta^{+}_{1/2}=\delta^{\pm}_{2l+1+1/2}=0. \tag{4.1}$$

Starting with Eqs. (3.33) and (3.29) and noting that $W^{\pm}$, $U^{\pm}$, and $\sigma^{\pm}$ are zero, we get

$$\bar{P}^{+}_{m}=\gamma^{0}(-1)^{m}\alpha^{1(1/2)}_{m},\quad \bar{P}^{-}_{m}=[2(2m+\tfrac{3}{2})]^{-1}\bar{P}^{+}_{m}. \tag{4.2}$$

In deriving Eq. (4.2) from Eqs. (3.29) and (3.33), use was made of the relations given in the Appendix. In what will follow, we shall make extensive use of the identities given in the Appendix, in order to find the values of all the relevant sums which are involved. Having found $\bar{P}^{\pm}_{n}$, we are then in a position to find $\bar{\alpha}$, $\tilde{\Omega}^{\pm}$, $\delta^{-}_{1/2}$, $\tilde{\sigma}^{\pm}_{l}$, and $\sigma^{\pm}_{l}$ from Eqs. (3.21), (3.22), and (3.23). Their values are

$$\begin{aligned}&\tan\delta^{-}_{1/2}=\gamma^{0}\sqrt{\pi}/2,\quad \tilde{\Omega}^{-}=1,\\ &\tilde{\Omega}^{+}=\tfrac{1}{2}\tan(\theta+\pi/4)+\gamma^{0}(\tilde{\beta}\pi/4-\sqrt{\pi}),\\ &\bar{\sigma}^{\pm}_{l}=-1/4(2l+3/2)(2l+3/2\mp\tfrac{1}{2}),\\ &\bar{\alpha}=0,\quad \tilde{\sigma}^{\pm}_{l}=[(2l+\tfrac{3}{2})^{2}-\tfrac{1}{4}]^{-1},\end{aligned} \tag{4.3}$$

where

$$\tilde{\beta}=16\pi\Gamma(\tfrac{1}{2})[\Gamma(\tfrac{1}{4})]^{-4}.$$

In deriving these relations we have also used the fact that $\gamma\cos(\theta+\pi/4)$ should satisfy Eq. (3.42), that is

$$\gamma\cos(\theta+\pi/4)=-(2\pi)^{-1/2}. \tag{4.4}$$

Using the above results in Eq. (3.38) and Eq. (3.41) we find that:

$$\begin{aligned}&P^{-}_{m}=-\tilde{\beta}(-1)^{m+1}\alpha^{0(1/2)}_{m+1}/4(2m+\tfrac{5}{2}),\\ &P^{+}_{m}=(2/\sqrt{\pi})[D^{-}_{1/2}-\sqrt{\pi}\tilde{\beta}/4](-1)^{m+1}\alpha^{0(1/2)}_{m+1}.\end{aligned} \tag{4.5}$$

The last step in deriving $P^{\pm}_{m}$ is to find the value of $D^{-}_{1/2}$ which is found from Eq. (3.43). Using the fact that the physical phase shifts are all zero, we find that Eq. (3.43) reduces to the following simple form:

$$\gamma^{0}D^{-}_{1/2}=2\gamma^{0}[1-\sqrt{\pi}\tilde{\beta}/4]-(1/\sqrt{\pi})\tan(\theta+\pi/4). \tag{4.6}$$

Having found $\bar{P}^{\pm}_{m}$ and $P^{\pm}_{m}$ from Eqs. (4.2), (4.5), and (4.6), we can then substitute them in Eq. (3.15) to find the associated potential coefficients. The interesting point to be noted is that although the arbitrary constant θ came into consideration only as a phase ambiguity, in the determination of the potential coefficients the associated nonuniqueness is through the factor $\tan(\theta+\pi/4)$, which can influence the potential coefficients just like the other arbitrary constant γ^{0}.

Since in this section we are interested in finding transparent spin–orbit potential, we note that the example we have given is associated with all phase shifts being zero except $\delta^{-}_{1/2}$ which may not be zero. To get a truly transparent spin–orbit potential, we should require that

$$\delta^{\pm}_{\lambda}=0,\quad \lambda\in\Omega_{p}. \tag{4.7}$$

We note that the above condition is satisfied if we choose γ^{0} to be zero. In that case we note that Eq. (4.6) cannot be satisfied in general, except if we require a certain value for θ, that is,

$$\gamma^{0}=0,\quad \theta=-\pi/4. \tag{4.8}$$

If we choose our arbitrary constants with the above values, we have

$$\begin{aligned}&\delta^{\pm}_{\lambda}=0,\quad \bar{P}^{\pm}_{m}=0,\\ &P^{-}_{m}=-[\tilde{\beta}(-1)^{m+1}/4(2m+\tfrac{5}{2})]\alpha^{0(1/2)}_{m+1},\\ &P^{+}_{m}=(2/\sqrt{\pi})[D^{-}_{1/2}-\sqrt{\pi}\tilde{\beta}/4](-1)^{m+1}\alpha^{0(1/2)}_{m+1},\quad d^{-}_{1/2}=-2/\pi,\end{aligned} \tag{4.9}$$

where $D^{-}_{1/2}$ is an arbitrary constant. At this point we substitute Eq. (4.9) in Eq. (3.15) for the value of the angular momentum being zero:

$$B^{+}_{1/2}=(1-\pi/2)D^{+}_{1/2}+\sqrt{\pi}\tilde{\beta}/2-D^{-}_{1/2}. \tag{4.10}$$

Noting that $D^{+}_{1/2}$ is nothing but $d^{-}_{1/2}B^{+}_{1/2}$, we find from Eq. (4.10) the following relation:

$$D^{+}_{1/2}=D^{-}_{1/2}-\sqrt{\pi}\tilde{\beta}/2. \tag{4.11}$$

Let us next write Eq. (4.11) and Eq. (4.9) in a more familiar form:

$$\begin{aligned}&D^{\pm}_{2l+1+1/2}=0,\\ &D^{+}_{2l+1/2}+D^{-}_{2l+1/2}=\gamma^{1}(-1)^{l}\alpha^{0(1/2)}_{l},\\ &D^{+}_{2l+1/2}-D^{-}_{2l+1/2}=-[\tilde{\beta}(-1)^{l}/2(2l+\tfrac{1}{2})]\alpha^{0(1/2)}_{l},\end{aligned} \tag{4.12}$$

where γ^1 is an arbitrary constant and $D^-_{1/2}$ is now defined through γ^1 and is no longer arbitrary by itself. The reason for writing Eqs. (4.9) and (4.11) in the above form is clear. Equation (4.12) is the analog of the familiar potential coefficients[10] associated with only central transparent potentials. From Eq. (4.12) we note that $(D^+_\lambda + D^-_\lambda)$ is the analog of the potential coefficients associated with transparent central potential.

Having found $D^\pm_\lambda$ from Eq. (3.15), we can find the potential coefficients associated with transparent spin–orbit and central potentials:

$$d^\pm_{2l+1+1/2} = 0,$$
$$d^\pm_{2l+1/2} = D^\mp_{2l+1/2}[B^\mp_{2l+1/2}]^{-1}, \tag{4.13}$$

where

$$B^\pm_{2l+1/2} = 1 \mp \frac{1}{2(2l+\frac{1}{2})} - \frac{\pi}{4(2l+\frac{1}{2})} D^\pm_{2l+1/2}.$$

At this point we have found the potential coefficients associated with transparent spin–orbit and central potentials. The found set of coefficients depend on one arbitrary constant γ^0. Since we found the set of potential coefficients in such a roundabout way, it is instructive to check that this set of coefficients indeed satisfies the equation we started with, that is, Eq. (3.14). Writing Eq. (3.14) for the case when all the phase shifts are zero and θ is equal to $-\pi/4$, we get

$$\begin{aligned} 0 &= \sum_m M^{2l+1/2}_{2m+1+1/2}(D^+_{2m+1+1/2} + D^-_{2m+1+1/2}) \\ &\quad \pm \sum_m R^{2l+1/2}_{2m+1+1/2}(D^+_{2m+1+1/2} - D^-_{2m+1+1/2}), \\ &\quad \mp \frac{\gamma\sqrt{2}\pi^{1/2}}{2(2l+1+\frac{1}{2})} \\ &= \sum_m M^{2l+1+1/2}_{2m+1/2}(D^+_{2m+1/2} + D^-_{2m+1/2}) \\ &\quad \pm \sum_m R^{2l+1+1/2}_{2m+1/2}(D^+_{2m+1/2} - D^-_{2m+1/2}). \end{aligned} \tag{4.14}$$

For reasons which will become clear later, we have not used the fact that $\gamma\sqrt{2}\pi^{1/2}$ is equal to -1 in Eq. (4.14). A look at Eq. (4.12) informs us that the first equation in Eq. (4.14) is satisfied and the second equation is reduced to the following form:

$$\frac{\gamma\sqrt{2}\pi^{1/2}}{2(2l+1+\frac{1}{2})} = -\sum_m R^{2l+1+1/2}_{2m+1/2}(D^+_{2m+1/2} - D^-_{2m+1/2}). \tag{4.15}$$

Using Eq. (4.12) and the identities given in the Appendix, it is easy to show that indeed Eq. (4.15) is satisfied if γ is given by Eq. (4.4). In other words our solution is indeed the solution to our original equation.

As the reader may have suspected, we did not give this verification in such detail, for its own sake, but we did so in order to point out that the set of potential coefficients that our method gives for transparent spin–orbit and central potential is not the most general set that one can have. Now that we know of a solution to (4.15), it is very easy to note that we can satisfy Eq. (4.15) for any value of γ if we change the definition of $(D^+_\lambda - D^-_\lambda)$ given by Eq. (4.12) by an appropriate factor of γ. So the most general solution to Eq. (3.14) when θ is equal to $-\pi/4$ is given by the following relations:

$$\begin{aligned} &D^\pm_{2l+1+1/2} = 0, \\ &D^+_{2l+1/2} + D^-_{2l+1/2} = \gamma^1(-1)^l\alpha^{0(1/2)}_l, \\ &D^+_{2l+1/2} - D^-_{2l+1/2} = \gamma\tilde{\beta}(2\pi)^{1/2}(-1)^l\alpha^{0(1/2)}_l/2(2l+\tfrac{1}{2}). \end{aligned} \tag{4.16}$$

One should note that our solution, Eq. (4.12), is a special case of Eq. (4.16), where the arbitrary constant γ in Eq. (4.16) is chosen to be equal to $-(2\pi)^{-1/2}$. The potential coefficients $d^\pm_\lambda$ associated with the most general set of transparent spin–orbit and central potential in the class that we are considering is then given by substituting Eq. (4.16) in Eq. (3.15):

$$d^\pm_{2l+1+1/2} = 0,$$
$$d^\pm_{2l+1/2} = D^\mp_{2l+1/2}[B^\mp_{2l+1/2}]^{-1}, \tag{4.17}$$

where

$$B^\pm_{2l+1/2} = 1 \pm \frac{\gamma(2\pi)^{1/2} - 1}{4(2l+\frac{1}{2})} - \frac{\pi}{4(2l+\frac{1}{2})} D^\pm_{2l+1/2}.$$

In evaluating $D^\pm_\lambda$ from Eq. (4.17) we should use $D^\pm_\lambda$ given by Eq. (4.16). We should also note that in Eq. (4.16) γ^0 and γ are arbitrary constants and in analog to the case of a central potential, we expect γ^0 to be associated with the strength of the transparent central potential and γ to be related to the strength of the transparent spin–orbit potential.

The reason why our method is not able to give the most general solution to Eq. (3.14) for the special case when all the phase shifts are zero is, of course, the fact that in that case the troublesome coupling between odd and even angular momenta does not appear in Eq. (3.14). Therefore, one does not need to make any specific choice of $d^-_{1/2}$ in order that the coupling not to appear in our analysis. An interesting point emerges if we try to find the solution to Eq. (4.15) without making use of our knowledge of a special solution to it. We see that the inhomogeneous term in Eq. (4.15) is not in the domain of the inverse of R, which we can construct. In other words, the solution to Eq. (4.15) cannot be obtained by a simple inversion. We find it remarkable that our method was able to find a special solution to Eq. (4.15) and also the vectors annihilated by M in a systematic fashion.

5. SUMMARY

In this work we were able to construct explicitly the potential coefficients from the phase shifts at a fixed energy, for the case when the potential coefficients for nonphysical values of the angular momentum had a definite form given by Eqs. (2.16) and (3.17). In order to clarify the method of construction of the potential coefficients from the phase shifts at a fixed energy, let us note that the first step in the construction is to find $\bar{P}^\pm$ from Eqs. (3.29) and (3.33). Having found $\bar{P}^\pm_n$, we are in a position using Eq. (3.21), to find the phase shift $\delta^-_{1/2}$, which is associated with a nonphysical value of angular momentum and is not given by experiment. Then we can find $\bar{\alpha}$, $\tilde{\Omega}^\pm$, $\bar{\sigma}^\pm_n$, $\tilde{\alpha}$, and $\tilde{\sigma}^\pm_l$ from Eqs. (3.22)

and (3.23). Next we choose the arbitrary constant γ in such a way that Eq. (3.42) is satisfied. Having done so, we are in a position to find $D^-_{1/2}$ from Eq. (3.43). Finally, substituting $D^-_{1/2}$ into Eqs. (3.38) and (3.41), we can find $P^\pm_n$. $\bar{P}^\pm_n$ and $P^\pm_n$ give us the constants $D^\pm_{2l+2+1/2}$ and $D^\pm_{2l+1+1/2}$. Knowing $D^-_{1/2}$ and $d^-_{1/2}$ enables us to find $D^+_{1/2}$ from Eq. (3.15). Now that we have found $D^\pm_\lambda$ for all values of the angular momentum, substituting them in Eq. (3.15) and using Eq. (3.16) enables us to find the potential coefficients.

As is shown in Sec. 3, the set of potential coefficients found from the phase shifts at a fixed energy is not unique. They depend on two arbitrary constants γ^0 and $\tan(\theta + \pi/4)$.

Having found the potential coefficients, we are then in a position to use the method of Sec. 2 to find the associated spin–orbit and central potentials. The details of constructing spin–orbit and central potentials from potential coefficients, which are associated with the desired phase shifts at a fixed energy, will be the subject of a forthcoming communication.

ACKNOWLEDGMENTS

It is a great pleasure for the author to thank Professor R.G. Newton for many helpful discussions, fruitful suggestions, and stylistic improvements in the manuscript. The hospitability of the Physics Department of Indiana University, where part of this work was done, is also gratefully acknowledged.

APPENDIX

In this appendix we want to give the types of relations and identities that we need most frequently in our evaluations of operators and sums. The method for evaluation of the series involved is exactly the same as that given by Sabatier[10] in his work dealing with the inversion of matrix $M^{(\alpha)}$. The only difference is that in this work we have to perform the operations in a more general form and to apply the method to some new sums. Space limitation does not allow us to show all the series we had to deal with, but the method of evaluation is essentially the same in all cases.

The following functions were defined first by Sabatier[10]:

$$\tilde{S}^{(\beta)}_1(r) = (\tfrac{1}{2}r)^{1/2} J_{1/2+\beta}(r) = \sum_{n=0}^{\infty} \alpha^{1(\beta)}_n J_{2n+1+\beta}(r), \tag{A1}$$

$$\tilde{S}^{(\beta)}_0(r) = (\tfrac{1}{2}r)^{1/2} J_{-1/2+\beta}(r) = \sum_{n=0}^{\infty} \alpha^{0(\beta)}_n J_{2n+\beta}(r). \tag{A2}$$

With the above definitions it follows that[14]

$$\int_0^\infty dr\, r^{-1} J_{\lambda+\beta}(r) \tilde{S}^{(\beta)}_1(r) = \frac{\Gamma(\frac{1}{2})\Gamma(\beta + \frac{1}{2} + \frac{1}{2}\lambda)}{2\Gamma(1 - \frac{1}{2}\lambda)\Gamma(1 + \beta + \frac{1}{2}\lambda)\Gamma(\frac{1}{2} + \frac{1}{2}\lambda)} \quad \text{for } \lambda + 2\beta + 1 > 0, \tag{A3}$$

$$\int_0^\infty dr\, r^{-1} J_{\lambda+\beta}(r) \tilde{S}^{(\beta)}_0(r) = \frac{\Gamma(\frac{1}{2})\Gamma(\beta + \frac{1}{2}\lambda)}{2\Gamma(\frac{1}{2} - \frac{1}{2}\lambda)\Gamma(\beta + \frac{1}{2}\lambda + \frac{1}{2})\Gamma(1 + \frac{1}{2}\lambda)} \quad \text{for } \lambda + 2\beta > 0. \tag{A4}$$

In this work β takes on the values of $\frac{1}{2}$, 1, $\frac{3}{2}$.

Next we define the following functions:

$$\tilde{S}^{\mu(\beta)}_1 = \sum_{n=0}^{\infty} \frac{\alpha^{1(\beta)}_n J_{2n+1+\beta}}{(2n+1+\beta)^2 - (\mu+\beta)^2}, \quad \mu \neq 2n+1, \tag{A5}$$

$$S^{\mu(\beta)}_0 = \sum_{n=0}^{\infty} \frac{\alpha^{0(\beta)}_n J_{2n+\beta}}{(2n+\beta)^2 - (\mu+\beta)^2}, \quad \mu \neq 2n, \tag{A6}$$

which are generalizations of the functions $\bar{S}^{(p)}_0$ and $\bar{S}^{(p)}_1$ of Sabatier.[10] Using the same arguments as given by Sabatier,[10] we get

$$[(\lambda+\beta)^2 - (\mu+\beta)^2] \int_0^\infty dr\, r^{-1} J_{\lambda+\beta} \tilde{S}^{\mu(\beta)}_1 = \tilde{B}^{\lambda(\beta)}_{1\mu} + \int_0^\infty dr\, r^{-1} J_{\lambda+\beta} \tilde{S}^{(\beta)}_1, \tag{A7}$$

$$[(\lambda+\beta)^2 - (\mu+\beta)^2] \int_0^\infty dr\, r^{-1} J_{\lambda+\beta} \tilde{S}^{\mu(\beta)}_0 = \tilde{B}^{\lambda(\beta)}_{0\mu} + \int_0^\infty dr\, r^{-1} J_{\lambda+\beta} \tilde{S}^{(\beta)}_0, \tag{A8}$$

where

$$\tilde{B}^{\lambda(\beta)}_{1\mu} = \left[\tilde{S}^{\mu(\beta)}_1 r \frac{d}{dr} J_{\lambda+\beta} - J_{\lambda+\beta} r \frac{d}{dr} \tilde{S}^{\mu(\beta)}_1 \right]_0^\infty,$$

$$\tilde{B}^{\lambda(\beta)}_{0\mu} = \left[\tilde{S}^{\mu(\beta)}_0 r \frac{d}{dr} J_{\lambda+\beta} - J_{\lambda+\beta} r \frac{d}{dr} \tilde{S}^{\mu(\beta)}_0 \right]_0^\infty. \tag{A9}$$

Using the asymptotic form of the Bessel functions[14] and the fact that we are only interested in those values of λ, μ, and β for which the brackets in Eq. (A9) are zero, for r equal to zero, we get

$$\tilde{B}^{\lambda(\beta)}_{1\mu} = -\frac{2}{\pi} \sum_{n=0}^{\infty} \frac{\alpha^{1(\beta)}_n \sin(\frac{1}{2}\pi)\,[(2n+1+\beta) - (\lambda+\beta)]}{(2n+1+\beta)^2 - (\mu+\beta)^2}, \tag{A10}$$

$$\tilde{B}^{\lambda(\beta)}_{0\mu} = -\frac{2}{\pi} \sum_{n=0}^{\infty} \frac{\alpha^{0(\beta)}_n \sin(\frac{1}{2}\pi)\,[(2n+\beta) - (\lambda+\beta)]}{(2n+\beta)^2 - (\mu+\beta)^2}. \tag{A11}$$

Since series of the above types appear in many parts of this work, we give the method of their evaluation in detail. Using identities for sine function, we note that the above can be written in the following form:

$$\tilde{B}^{\lambda(\beta)}_{1\mu} = -\frac{\sin\frac{1}{2}\pi(1-\lambda)}{\sin\frac{1}{2}\pi(1-\mu)} \times \sum_{n=0}^{\infty} \frac{\alpha^{1(\beta)}_n \sin\frac{1}{2}\pi[(2n+1+\beta) - (\mu+\beta)]}{\frac{1}{2}\pi[(2n+1+\beta)^2 - (\mu+\beta)^2]}, \tag{A12}$$

$$\tilde{B}^{\lambda(\beta)}_{0\mu} = -\frac{\sin\frac{1}{2}\pi\lambda}{\sin\frac{1}{2}\pi\mu} \times \sum_{n=0}^{\infty} \frac{\alpha^{0(\beta)}_n \sin\frac{1}{2}\pi[(2n+\beta) - (\mu+\beta)]}{\frac{1}{2}\pi[(2n+\beta)^2 - (\mu+\beta)^2]}. \tag{A13}$$

Of course, this is valid only if $\sin\frac{1}{2}\pi(1-\mu)$ or $\sin\frac{1}{2}\pi\mu$ are not zero. For cases of interest this condition is satisfied. Using the well-known relation[14]

$$\tfrac{1}{2}\pi(\lambda^2-\mu^2)\int_0^\infty dr\, r^{-1}J_\lambda(r)J_\mu(r)$$

$$=\sin\tfrac{1}{2}\pi(\lambda-\mu)\quad \text{for } \lambda+\mu>0. \tag{A14}$$

We have the following results:

$$\tilde{B}_{1\mu}^{\lambda(\beta)}=-\frac{\sin\frac{1}{2}\pi(\lambda-1)}{\sin\frac{1}{2}\pi(\mu-1)}\int_0^\infty dr\, r^{-1}J_{\mu+\beta}\tilde{S}_1^{(\beta)}, \tag{A15}$$

$$\tilde{B}_{0\mu}^{\lambda(\beta)}=-\frac{\sin\frac{1}{2}\pi\lambda}{\sin\frac{1}{2}\pi\mu}\int_0^\infty dr\, r^{-1}J_{\mu+\beta}\tilde{S}_0^{(\beta)}. \tag{A16}$$

Except for our generalization to noninteger values and the constants which Sabatier did not write down explicitly (because in his case the above became zero) the values we have found are identical with Eqs. (2.20) and (2.21) of Sabatier.[10]

As a last example let us consider one of the sums we had to evaluate in Sec. 4, that is

$$\tilde{\beta}=\sum_{n=0}^{\infty}\frac{2\alpha_\pi^{1(1/2)}(-1)^n}{\pi(2n+\frac{3}{2})^2}\,. \tag{A17}$$

It is easy to write the above in standard form, and note that $\tilde{\beta}$ is nothing but $-\tilde{B}_{1-1/2}^{0(1/2)}$. Using Eqs. (A15) and (A3) for β equal to $\frac{1}{2}$, we get

$$\tilde{\beta}=\sqrt{2}\int_0^\infty dr\, r^{-1}J_{-1/2+1/2}(r)\tilde{S}_1^{(1/2)}(r)$$

$$=16\pi\Gamma(\tfrac{1}{2})/\Gamma^4(\tfrac{1}{4}). \tag{A18}$$

Of course, for our special values of μ, λ, and β all the restrictions in Eqs. (A3) and (A14) are satisfied and the interchange of summation and integration is also justified.[10]

[1]R. G. Newton, J. Math. Phys. **3**, 75 (1962).
[2]T. Regge, Nuovo Cimento **14**, 951 (1959).
[3]P. C. Sabatier, J. Math. Phys. **9**, 1241 (1968).
[4]M. A. Hooshyar, J. Math. Phys. **16**, 257 (1975).
[5]$b_\lambda^\pm$, which are denoted by $\gamma_\lambda(W^\pm,1)$ in Ref. 3 are referred to as the potential coefficients. For solving the inverse scattering problem one cannot work directly with $b_\lambda^\pm$ and instead has to consider $d_\lambda^\pm$ which are related to $b_\lambda^\pm$ through Eq. (2.10). Because of this relation, we also call $d_\lambda^\pm$ the potential coefficients.
[6]By $f(x)=o(g(x))$ for large values of x, we mean that $f(x)/g(x)$ tends to zero.
[7]N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Clarendon, Oxford, 1965).
[8]Please note the misprint in the corresponding equation of Ref. 4.
[9]By the symbol A_j^i we mean the element (ij) of a matrix A. In the customary notation it would have been written as A_{ij}.
[10]P. C. Sabatier, J. Math. Phys. **7**, 1515 (1966).
[11]In the evaluation of $\tilde{\Lambda}_{m\pm}^{l\pm}$ the signs are decided in accordance with the associated index l or m. For example, in the evaluation of $\tilde{\Lambda}_{m+}^{l-}$, the upper sign is used in any bracket that m appears and the lower sign is used in any bracket that l appears.
[12]P. J. Redmond, J. Math. Phys. **5**, 1547 (1964).
[13]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953).
[14]Bateman Manuscript Projects, *Higher Transcendental Functions*, edited by A. Ederlyi (McGraw-Hill, New York, 1953).

The spinor Helmholtz equation

Pierre Hillion

Institut Henri Poincaré, Paris, France
(Received 8 February 1977)

In this paper, we introduce the spinor Helmholtz equation, a first order partial differential system, which, on the one hand, describes the behavior of a polarized electromagnetic beam at the Fresnel approximation and which, on the other hand, can also be used, from a numerical viewpoint, instead of the scalar Helmholtz equation for solving the Dirichlet problem. We give the main properties of this equation and of its solutions.

INTRODUCTION

The Helmholtz equations for a scalar field $u(r)$ and a vector field $A(r)$ in the usual Euclidean space $\mathbb{R}^3$ are

$$\Delta u(r) + K^2 u(r) = 0,$$

$$\nabla \wedge \nabla \wedge A(r) - K^2 A(r) = 0, \quad \nabla \cdot A(r) = 0.$$

Δ, ∇, are the Laplacian and nabla operators, K the wavenumber, r denotes an arbitrary point in $\mathbb{R}^3$, $r = x\mathbf{i} + y\mathbf{j} + z\mathbf{l}$ with $\mathbf{i}, \mathbf{j}, \mathbf{l}$, an orthonormal basis of $\mathbb{R}^3$.

In this paper, we introduce the spinor Helmholtz equation which is a partial differential equation of the first order (contrary to the two previous ones) for the spinor field $\Psi(r)$. A spinor is a geometrical object defined over a two-dimensional complex space which is the representation space of the two-dimensional unitary unimodular group SU(2).[1]

There exist two reasons for considering such an equation:

1. First, from a numerical viewpoint, since the approximation of derivatives by differences leads to unstable computations, it is often better to use a system of first order partial differential equations than a partial differential equation of the second order provided that in both cases one obtains the same solutions. We will prove that any solution of the spinor Helmholtz equation is also a solution of the scalar Helmholtz equation.

2. Physics provides the second reason; it is well known[2] that for an unpolarized electromagnetic field, according to the Fresnel approximation, that is, broadly speaking, when one is only interested in the energy of the electromagnetic field, one can use the scalar Helmholtz equation instead of the vector equation; but, it was recently shown[3] that, at the same level of approximation, and for a polarized electromagnetic field, one has to use the spinor Helmholtz equation.

In this paper, we give the main properties of the spinor Helmholtz equation and of its solutions, and in a second paper, we will discuss some numerical methods to find these solutions.

In order to simplify calculations, we use the tensorial formalism; the latin indices $i, j, k, \cdots$ take the values 1, 2, 3,

$$x_1 = x, \quad x_2 = y, \quad x_3 = z; \quad \partial_1 = \frac{\partial}{\partial x}, \quad \partial_2 = \frac{\partial}{\partial y}, \quad \partial_3 = \frac{\partial}{\partial z}.$$

We use the Einstein summation convention $A^i B_i = A^1 B_1 + A^2 E_2 + A^3 B_3$, so that, for instance, the previous Helmholtz equations become

$$(\partial^j \partial_j + K^2) u(r) = 0,$$

$$\epsilon^{lmi} \epsilon_{ijk} \partial^j A^k(r) = 0, \quad \partial^j A_j(r) = 0.$$

ϵ_{ijl} is the permutation symbol; ϵ_{ijk} is zero if two or more indices are equal. $\epsilon_{ijk} = 1$ if the ordered set i, j, k is obtained by an even permutation of 1, 2, 3 and $\epsilon_{ijk} = -1$ for an odd permutation. δ_{ij} is the Kronecker symbol.

1. THE SCALAR HELMHOLTZ EQUATIONS

We first give some results which will be useful later, about the scalar Helmholtz equation and its solutions. We begin with the Gauss–Ostrogradski theorem which asserts that if V is a closed region and δV is its bounding surface,

$$\int_V \partial^j A_j(r)\, dv = \int_{\delta V} n^j A_j(r)\, ds$$

provided that the triple integral through V exists and no straight line parallel to an axis meets δV more than a fixed number of times; n^j denotes the outward normal to δV (see Ref. 4).

This theorem holds for unbounded regions V as well as for bounded regions if the integrands of the surface integrals are $O(r^{-3})$ in absolute value as $r \to \infty$. In the following, when we use unbounded regions, we always assume this condition is fulfilled. In particular, we will consider unbounded regions of the Ω type whose boundary $\delta\Omega = S \cup \Sigma$ is made of a surface S and of half a sphere Σ with center on S and with infinite radius. For these regions, the Gauss–Ostrogradski theorem gives

$$\int_\Omega \partial^j A_j(r)\, dr = \int_S n^j A_j(r^s)\, dr^s, \tag{1}$$

where the notation r^s means that r belongs to the surface S.

To mathematically describe the behavior of a complex scalar field $u(r)$, we need in addition to the Helmholtz equation,

$$(\partial^j \partial_j + K^2) u(r) = 0, \quad r \in \Omega, \tag{2}$$

some boundary conditions such as (3a) for the Dirichlet problem or (3b) for the Neumann problem,

$$u(r^s) = f(r^s), \tag{3a}$$

$$\partial_n u(r^s) = g(r^s), \quad r^s \in S \tag{3b}$$

 0022-2488/78/1901-0264$1.00

where $\partial_n u(r^s)$ is the normal derivative to S with $\partial_n = n^j \partial_j$. Of course, we could also consider more general boundary conditions as $\alpha u(r^s) + \beta \partial_n u(r^s) = h(r^s)$ where α and β, are arbitrary complex scalars.

From (1), we easily obtain Green's identities provided that the derivatives exist for the scalar fields $u(r)$, $u'(r)$,

$$\int_\Omega (u'(r)\partial^j \partial_j u(r) + \partial^j u'(r)\, \partial_j u(r))\, dr = \int_S u'(r^s)\, \partial_n u(r^s)\, dr^s, \tag{4}$$

$$\int_\Omega \{u'(r)\, \partial^j \partial_j u(r) - u(r)\, \partial^j \partial_j u'(r)\}\, dr = \int_S \{u'(r^s)\, \partial_n u(r^s) - u(r^s)\, \partial_n u'(r^s)\}\, dr^s. \tag{5}$$

If in (4), we take $u'(r) = \overline{u(r)}$, $u(r)$ being a solution of (2), it becomes

$$\int_S \overline{u(r^s)}\, \partial_n u(r^s)\, dr^s = \int_\Omega \{\partial^j \overline{u(r)}\, \partial_j u(r) - K^2 \overline{u(r)} u(r)\}\, dr$$

and, if we exchange the roles of u and u' and substract,

$$\int_S \partial_n |u(r^s)|^2\, dr^s = 2 \int_\Omega \{\partial^j \overline{u(r)}\, \partial_j u(r) - K^2 \overline{u(r)} u(r)\}\, dr = 2 \int_\Omega L(u, \partial_j u)\, dr \tag{4'}$$

where $L(u, \partial_j u)$ is the Lagrangian density.

Let $G(r, r') = G(r', r)$ be the Green function of the scalar Helmholtz equation,

$$(\partial^j \partial_j + K^2) G(r, r') = -\delta(r - r'), \tag{6}$$

where $\delta(r - r')$ is the Dirac distribution introduced here formally as a shorthand definition that $G(r, r')$ is a fundamental solution of (2) (see Ref. 5). Then, if in (5), we take for $u(r)$ a solution of (2) and for $u'(r)$ the Green function $G(r, r')$, it becomes

$$u(r') = \int_S n^j \{u(r^s) \partial_j G(r^s, r') - G(r^s, r')\, \partial_j u(r^s)\}\, dr^s, \qquad r' \in \Omega, \tag{7}$$

an expression which in theory makes possible the computation of $u(r)$, $r \in \Omega$ from the boundary data on S.

Remark 1: It is easy to check that $u(r)$ defined by (7) is a solution of (2); we have just to use (6) with the remark that $r' \neq r^s$ since (2) is valid in Ω not on S.

Remark 2: To solve the Dirichlet or Neumann problems we must use Green's functions $G_1(r, r')$, $G_2(r, r')$, with suitable boundary data

$$G_1(r^s, r') = 0, \tag{7a}$$

$$\partial_n G_2(r^s, r') = 0. \tag{7b}$$

Then, the solution of (2), (3a) is

$$u(r') = \int_S u(r^s)\, \partial_n G_1(r^s, r')\, dr^s, \quad r' \in \Omega, \tag{8a}$$

while, for (2), (3b), we have

$$u(r') = -\int_S \partial_n u(r^s) G_2(r^s, r')\, dr^s, \quad r' \in \Omega. \tag{8b}$$

Since these problems have a unique solution, (8a) enables us to calculate $\partial_n u(r^s)$ and (8b) enables us to calculate $u(r^s)$.

The Green function for a three-dimensional infinite domain has the simple form[6]

$$G(r, r') = \frac{\exp(iK|r - r'|)}{4\pi |r - r'|}. \tag{9}$$

Remark 3: In the half-space $z \geq 0$ where the surface S is the plane π_0, $z = 0$, we have

$$G_1(r, r') = \frac{1}{4\pi}\left(\frac{\exp(iK|r - r'|)}{|r - r'|} - \frac{\exp(iK|\rho - r'|)}{|\rho - r'|}\right), \tag{9a}$$

$$G_2(r, r') = \frac{1}{4\pi}\left(\frac{\exp(iK|r - r'|)}{|r - r'|} + \frac{\exp(iK|\rho - r'|)}{|\rho - r'|}\right), \tag{9b}$$

$r = (x, y, z)$, $\rho = (x, y, -z)$, $r' = (x', y', z')$,

while relations (8a) and (8b) become, with $R^2 = (x - x')^{1/2} + (y - y'^2) + z^2$,

$$u(r) = -\frac{1}{2\pi} \frac{\partial}{\partial z} \int_{\pi_0} f(x', y') \frac{\exp(iKR)}{R}\, dx'\, dy', \tag{10a}$$

$$u(r) = -\frac{1}{2\pi} \int_{\pi_0} g(x', y') \frac{\exp(iKR)}{R}\, dx'\, dy'. \tag{10b}$$

2. THE SPINOR HELMHOLTZ EQUATION

A. Definition and properties

Let $\Psi(r)$ be a two-component complex spinor $\Psi(r) = \begin{pmatrix} \Psi_1(r) \\ \Psi_2(r) \end{pmatrix}$ and let σ_i, $i = 1, 2, 3$, be the Pauli matrices

$$\sigma_1 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \quad \sigma_2 = \begin{vmatrix} 0 & -i \\ i & 0 \end{vmatrix}, \quad \sigma_3 = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} \tag{11}$$

satisfying the following easy to prove relations

$$\sigma_i^\dagger = \sigma_i, \tag{12a}$$

$$\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij}, \tag{12b}$$

$$\sigma_i \sigma_j = i \epsilon_{ijk} \sigma^k, \tag{12c}$$

where $\sigma_i^\dagger$ is the Hermitian conjugate matrix of σ_i, δ_{ij} is the Kronecker symbol, and ϵ_{ijk} is the permutation symbol.

In the same way that the scalar field $u(r)$ is invariant with respect to the rotation group while the vector field $A(r)$ transforms according to the relation $A_i(r) \to A_i'(r) = a_{ij} A^j(r)$ where a_{ij} is a 3×3 orthogonal matrix, the spinor field $\Psi(r)$ transforms according to the relation $\Psi(r) \to \Psi'(r) = S\Psi(r)$ where S is a 2×2 unimodular unitary matrix $S^\dagger = S^{-1}$, so that for the Hermitian field $\Psi^\dagger(r)$ one has $\Psi^\dagger(r) \to \Psi^\dagger(r) = \Psi^\dagger(r) S^{-1}$, and one can prove the following equality:

$$S \sigma_i S^{-1} = a_{ij} \sigma^j. \tag{12d}$$

The spinor Helmholtz equation [(13)] and its adjoint equation (Hermitian conjugate equation) are then

$$(\sigma^j \overrightarrow{\partial}_j + iK)\Psi(r) = 0, \tag{13}$$

$$\Psi^\dagger_{(r)}(\sigma^j \overleftarrow{\partial}_j - iK) = 0, \tag{13'}$$

where $\overrightarrow{\partial}$ is a derivative acting on the right and $\overleftarrow{\partial}$ is a derivative acting on the left (when no confusion is possible, one writes ∂).

Equations (13) and (13') are convariant under the rotation group in $\mathbb{R}^3$. This follows at once from (12d), from the transformation laws of $\Psi(r)$, $\Psi^\dagger(r)$, and from the fact that ∂_j is a vector operator, $\partial_j \to \partial_j' = a_{jk} \partial^k$.

Now, we will show that Eqs. (13) and (13') describe the propagation of two coupled complex scalar fields; of course to have a well-posed problem we must still give some boundary data. The boundary data will be discussed in the next section.

Theorem 1: Let $\Psi(r)$ be a solution of (13), then each component $\Psi_i(r)$, $i=1,2$ is a solution of (2).

Let us multiply (13) on the left with the operator $(\sigma^l\vec{\partial}_l - iK)$; it becomes

$$(\sigma^l\vec{\partial}_l - iK)(\sigma^j\vec{\partial}_j + iK)\Psi(r)$$
$$= 0$$
$$= (\sigma^l\sigma^j\vec{\partial}_l\vec{\partial}_j + K^2)\Psi(r)$$
$$= \{\tfrac{1}{2}(\sigma^l\sigma^j + \sigma^j\sigma^l)\vec{\partial}_l\vec{\partial}_j + K^2\}\Psi(r).$$

Then, the result follows from (12b).

Of course, this theorem still holds for the components $\psi_i(r)$ $(i=1,2)$ of a solution $\Psi^\dagger(r)$ of (13′) where the bar denotes the complex conjugation

In some sense, the following trivial result can be considered as an inverse of Theorem 1: If, $u_1(r)$, $u_2(r)$, are two arbitrary solutions of (2), then $\Psi(r) = (\sigma^j\vec{\partial}_j - iK) \times \binom{u_1(r)}{u_2(r)}$ is a solution of (13). The proof follows at once from (12b).

Remark: As noted in Ref. 3, $\Psi(r)$ can be considered as the Stokes spinor used to describe the polarization of the electromagnetic field.

B. Green's functions for the spinor Helmholtz equation

Let $\Psi(r)$, $\Phi(r)$, be spinors with the following transformation law under the SU(2) group:

$$\Psi(r) \mapsto \Psi'(r) = S\Psi(r), \quad \Phi(r) \mapsto \Phi'(r) = \Phi(r)S^{-1}.$$

Then, using (12d) and the fact that ∂_j is a vector operator, it is easy to show that the following expressions are vectors:

$$\Phi(r)\sigma_j\Psi(r), \quad \Phi(r)\vec{\partial}_j\Psi(r), \quad \Phi(r)\vec{\partial}_j\Psi(r),$$
$$\Phi(r)(\sigma\wedge\vec{\partial})_j\Psi(r), \quad \Phi(r)(\sigma\wedge\vec{\partial})_j\Psi(r), \tag{14}$$

where the symbol $\wedge$ denotes the vector product so that $(\sigma\wedge\partial)_i = \epsilon_{ijk}\sigma^j\partial^k$; as a consequence, one can use the Gauss–Ostrogradski theorem and relation (1) gives

$$\int_\Omega \vec{\partial}^j(\Phi(r)\sigma_j\Psi(r))\,dr = \int_S n^j(\Phi(r^s)\sigma_j\Psi(r^s))\,dr^s, \tag{14a}$$
$$\int_\Omega \vec{\partial}^j(\Phi(r)\overleftarrow{\partial}_j\Psi(r))\,dr = \int_S n^j(\Phi(r^s)\overleftarrow{\partial}_j\Psi(r^s))\,dr^s, \tag{14b}$$
$$\int_\Omega \vec{\partial}\{\Phi(r)(\sigma\wedge\overleftarrow{\partial})_j\Psi(r)\}\,dr = \int_S n^j\{\Phi(r^s)(\sigma\wedge\overleftarrow{\partial})_j\Psi(r^s)\}\,dr^s. \tag{14c}$$

We shall see later that these three relations are not independent and that both of them imply the third one so that it is just sufficient to consider (14a) and (14b), but we need a slight generalization of the Gauss–Ostrogradski theorem. Let $\Gamma(r,r')$ be a 2×2 matrix with the following transformation law under the SU(2) group: $\Gamma(r,r') \mapsto \Gamma'(r,r') = S\Gamma(r,r')S^{-1}$. Then let us substitute $\Gamma(r,r')$ for $\Phi(r)$ in (14), the corresponding expressions become spin-vectors such as $\Gamma(r,r')\sigma_j\Psi(r)$, that is, geometrical objects with the following transformation law under SU(2) [we still use (12d) or the vector character of ∂_j]:

$$\Gamma(r,r')\sigma_j\Psi(r) \mapsto \Gamma'(r,r')\sigma_j\Psi'(r) = a_{jk}S\Gamma(r,r')\sigma^k\Psi(r).$$

It is easy to generalize the Gauss–Ostrogradski theorem to spin-vectors and the proof in Ref. 4 can be repeated word for word so that it is not reproduced here. Similarly to (14a)–(14c) one obtains

$$\int_\Omega \vec{\partial}^j\{\Gamma(r,r')\sigma_j\Psi(r)\}\,dr = \int_S n^j\{\Gamma(r^s,r')\sigma_j\Psi(r^s)\}\,dr^s, \tag{14′a}$$
$$\int_\Omega \vec{\partial}^j\{\Gamma(r,r')\vec{\partial}_j\Psi(r)\}\,dr = \int_S n^j\{\Gamma(r^s,r')\vec{\partial}_j\Psi(r^s)\}\,dr^s, \tag{14′b}$$
$$\int_\Omega \vec{\partial}^j\{\Gamma(r,r')(\sigma\wedge\vec{\partial})_j\Psi(r)\}\,dr = \int_S n^j\{\Gamma(r^s,r')(\sigma\wedge\vec{\partial})_j\Psi(r^s)\}\,dr^s. \tag{14′c}$$

From (14b) and (14′b) one deduces at once two Green's identities for spinor fields,

$$\int_\Omega\{\Phi(r)\vec{\partial}^j\vec{\partial}_j\Psi(r) - \Phi(r)\vec{\partial}^j\vec{\partial}_j\Psi(r)\}\,dr$$
$$= \int_S n^j\{\Phi(r^s)\vec{\partial}_j\Psi(r^s) - \Phi(r^s)\vec{\partial}_j\Psi(r^s)\}\,dr^s, \tag{15}$$
$$\int_\Omega\{\Gamma(r,r')\vec{\partial}^j\vec{\partial}_j\Psi(r) - \Gamma(r,r')\vec{\partial}^j\vec{\partial}_j\Psi(r)\}\,dr$$
$$= \int_S n^j\{\Gamma(r^s,r')\vec{\partial}_j\Psi(r^s) - \Gamma(r^s,r')\vec{\partial}_j\Psi(r^s)\}\,dr^s. \tag{15′}$$

Let us now introduce the 2×2 Green matrices $\gamma_1(r,r')$, $\gamma_2(r,r')$, for the Eqs. (13) and (13′),

$$(\sigma^j\vec{\partial}_j + iK)\gamma_1(r,r') = -\sigma_0\delta(r-r'), \tag{16a}$$
$$\gamma_2(r,r')(\sigma^j\vec{\partial}_j - iK) = -\sigma_0\delta(r-r'), \tag{16b}$$

where σ_0 is the 2×2 identity matrix.

Using the same proof as in Theorem 1, it is easy to check that $\gamma_1(r,r')$, $\gamma_2(r,r')$, are given by the following expressions where $G(r,r')$ is the Green function of the scalar Helmholtz equation:

$$\gamma_1(r,r') = (\sigma^j\partial_j - iK)G(r,r'), \tag{17a}$$
$$\gamma_2(r,r') = (\sigma^j\partial_j + iK)G(r,r'), \tag{17b}$$

and the matrices $\gamma_i(r,r')$, $i=1,2$, have under the SU(2) group the transformation law assumed for $\Gamma(r,r')$ since $\sigma^j\partial_j \mapsto \sigma^j\partial'_j = a_{kj}\sigma^j\partial^k = S\sigma^j\partial_jS^{-1}$.

Let us remark that $\gamma_2(r,r') \neq \gamma_1^\dagger(r,r')$ since we use the same (in general complex) Green function in the right-hand side of (17a) and (17b) but if one exchanges r and r', it becomes

$$\gamma_1(r',r) = (\sigma^j\partial'_j - iK)G(r',r), \tag{17′a}$$
$$\gamma_2(r',r) = (\sigma^j\partial'_j + iK)G(r',r) \tag{17′b}$$

but $G(r,r') = G(r',r) = G(|r-r'|)$ which implies $\partial_jG(r,r') = -\partial'_jG(r,r')$ so that by comparison of (17) and (17′), we obtain

$$\gamma_1(r,r') = -\gamma_2(r',r), \quad \gamma_2(r,r') = -\gamma_1(r',r). \tag{18}$$

One could expect these relations, taking into account the well-known anticommuting properties of the quantized spinor fields.

We now deduce some important results from these preliminaries. Let us take in, (15′), for $\Psi(r)$ a solution of (13) so that $\partial^j\partial_j\Psi(r) = -K^2\Psi(r)$, and for $\Gamma(r,r')$ let us take the Green matrix $\gamma_1(r,r')$ for which the premultiplication of (16a) on the left by the operator $(\sigma^i\partial_i - iK)$ gives

$$(\partial^j\partial_j + K^2)\gamma_1(r,r') = -(\sigma^j\partial_j - iK)\delta(r-r').$$

As a result, we obtain from (15'),

$$\int_\Omega \{(\sigma^j\partial_j - iK)\delta(r-r')\}\Psi(r)\,dr = \int_S n^j\{\gamma_1(r^s,r')\overrightarrow{\partial}_j\Psi(r^s) - \gamma_1(r^s,r')\overleftarrow{\partial}_j\Psi(r^s)\}\,dr^s$$

and using the properties of the Dirac distribution and of its derivatives,

$$(\sigma^j\partial_j + iK)\Psi(r') = -\int_S n^j\{\gamma_1(r^s,r')\overrightarrow{\partial}_j\Psi(r^s) - \gamma_1(r^s,r')\overleftarrow{\partial}_j\Psi(r^s)\}\,dr^s$$

that is, finally, since $\Psi(r)$ is a solution of (13),

$$\int_S n^j\{\gamma_1(r^s,r')\overrightarrow{\partial}_j\Psi(r^s) - \gamma_1(r^s,r')\partial_j\Psi(r^s)\}\,dr^s = 0. \tag{19}$$

In particular, if $\gamma_{1,1}(r^s,r')$ is the Green matrix (17a) defined with the Green function (7a), $G_1(r,r')$, relation (19) becomes

$$\int_S \{\partial_n\gamma_{1,1}(r^s,r')\}\Psi(r^s)\,dr^s = 0. \tag{19'}$$

Let us compute this expression for r' on S, that is for $r' = r'^s$. From (17a), one has

$$\partial_n\gamma_1(r^s,r') = (\sigma^j\partial_j - iK)\partial_n G_1(r^s,r')$$

and from (8a), $\partial_n G_1(r^s,r'^s) = \delta(r^s - r'^s)$ so that

$$\partial_n\gamma_1(r^s,r'^s) = (\sigma^j\partial_j - iK)\delta(r^s - r'^s).$$

Using this last result in (19') gives

$$(\sigma^j\partial_j + iK)\Psi(r^s) = 0 \tag{20}$$

so, we have proved the following theorem.

Theorem 2: If $\Psi(r)$ is a solution of (13) in Ω, then $\Psi(r^s)$ is a solution of (13) on S.

Let us now consider (14'a) with a solution of (13) for $\Psi(r)$ and with the Green matrix $\gamma_2(r,r')$ for $\Gamma(r,r')$. Then taking into account Eqs. (13) and (16b), it becomes

$$\Psi(r') = \int_S n^j\gamma_2(r^s,r')\sigma_j\Psi(r^s)\,dr^s, \quad r'\in\Omega \tag{21}$$

a relation similar to (7). It is easy to show that $\Psi(r)$ defined by (21) is a solution of (13). Indeed, using the first equation of (18), we have

$$(\sigma^j\partial'_j + iK)\Psi(r') = \int_S n^i(\sigma^j\partial'_j + iK)\gamma_2(r^s,r')\sigma_i\Psi(r^s)\,dr^s = -\int_S n^i(\sigma^j\partial'_j + iK)\gamma_1(r',r^s)\sigma_i\Psi(r^s)\,ds^s,$$

but from (17'a), $(\sigma^j\partial'_j + iK)\gamma_1(r',r^s) = (\Delta' + K^2)G(r',r^s) = -\delta(r'-r^s)$ and so for $r'\in\Omega$, $(\sigma^j\partial'_j + iK)\Psi(r') = 0$.

We will now discuss (21) in order to obtain relations similar to (8a) and (8b).

Theorem 3: If on S, $\Psi(r^s)$ satisfies (20), then

$$\Psi(r') = \int_S \partial_n G_1(r^s,r')\Psi(r^s)\,dr^s, \tag{22a}$$

$$\Psi(r') = -\int_S G_2(r^s,r')\partial_n\Psi(r^s)\,dr^s, \tag{22b}$$

where G_1, G_2 are the Green functions satisfying (7a) and (7b) respectively.

Using definition (17b) of $\gamma_2(r^s,r')$ in (21), leads to

$$\begin{aligned}\Psi(r') &= iK\int_S n^j\sigma_j G(r^s,r')\Psi(r^s)\,dr^s \\ &\quad + \int_S n^j(\sigma^i\partial_i G(r^s,r'))\sigma_j\Psi(r^s)\,dr^s \\ &= iK\int_S n^j\sigma_j G(r^s,r')\Psi(r^s)\,dr^s \\ &\quad + \int_S n^j\{\partial_j G(r^s,r')\}\Psi(r^s)\,dr^s \\ &\quad + i\int_S n^j\{(\sigma\wedge\overrightarrow{\partial})_j G(r^s,r')\}\Psi(r^s)\,dr^s.\end{aligned} \tag{21'}$$

If in (21') we take $G(r^s,r') = G_1(r^s,r')$, the first term is zero from (7a), the third term is also zero for $\partial_k G_1(r^s,r')$ is normal to the surface $G_1(r^s,r') = 0$ and so $\partial_k G_1(r^s,r')$ is parallel to n_k since $G_1(r^s,r') = 0$ on S, but, from the definition of the scalar triple product $n^j(\sigma\wedge\partial)_j = -\sigma^j(n\wedge\partial)_j$ and thus $\sigma^j(n\wedge\partial)_j G_1(r^s,r') = 0$. So, we are left with (22a), but since we proved that $\Psi(r)$ defined by (21) is a solution of (13), it follows from Theorem 2 that (22a) is only valid if (20) is fulfilled. In a way, Theorem 3 is a reciprocal to Theorem 2.

Before proving (22b), let us note that (13) can be written

$$\partial_j\Psi(r) - i(\sigma\wedge\partial)_j\Psi(r) + iK\sigma_j\Psi(r) = 0, \tag{23}$$

a relation which shows, as noticed before, that the three relations (14a)–(14c) [and (14'a)–(14'c)] are not independent. If (23) is multiplied on the left by n^j and if one defines

$$\partial_n = n^j\partial_j, \quad (\sigma\wedge\partial)_n = n^j(\sigma\wedge\partial)_j, \quad \sigma_n = n^j\sigma_j$$

another form for (20) is

$$\partial_n\Psi(r^s) - i(\sigma\wedge\partial)_n\Psi(r^s) + iK\sigma_n\Psi(r^s) = 0. \tag{20'}$$

Let us now take $G(r^s,r') = G_2(r^s,r')$ in (21'); from (7b), the second term is zero,

$$\begin{aligned}\Psi(r') &= iK\int_S n^j\sigma_j G_2(r^s,r')\Psi(r^s)\,dr^s \\ &\quad + i\int_S n^j\{(\sigma\wedge\partial)_j G(r^s,r')\}\Psi(r^s)\,dr^s.\end{aligned} \tag{21''}$$

The expression makes possible the calculation of the normal derivative $\partial'_n\Psi(r'^s)$,

$$\begin{aligned}\partial'_n\Psi(r'^s) &= iK\int_S n^j\sigma_j\partial'_n G_2(r^s,r'^s)\Psi(r^s)\,dr^s \\ &\quad + i\int_S n^j\{(\sigma\wedge\partial)_j\partial'_n G_2(r^s,r'^s)\}\Psi(r^s)\,dr^s\end{aligned}$$

but (8b) implies $\partial'_n G_2(r^s,r'^s) = -\delta(r^s - r'^s)$, using this result and the properties of the Dirac distribution and of its derivatives, the previous expression becomes

$$\partial'_n\Psi(r'^s) = -iK\sigma_n\Psi(r'^s) + i(\sigma\wedge\partial')_n\Psi(r'^s),$$

that is (20').

Let us now introduce this relation in (21''), we have

$$\begin{aligned}\Psi(r') &= -\int_S G_2(r^s,r')\partial_n\Psi(r^s)\,dr^s \\ &\quad + i\int_S n^j\{G_2(r^s,r')(\sigma\wedge\overrightarrow{\partial})_j\Psi(r^s) \\ &\quad + G_2(r^s,r')(\sigma\wedge\overrightarrow{\partial})_j\Psi(r^s)\}\,dr^s.\end{aligned}$$

Let us consider the second integral which becomes with the Gauss–Ostrogradski theorem,

$$\int_S n^j\{G_2(r^s,r')(\sigma\wedge\overrightarrow{\partial})_j\Psi(r^s) + G_2(r^s,r')(\sigma\wedge\overleftarrow{\partial})_j\Psi(r^s)\}\,dr^s = \int_\Omega \partial^j\{G_2(r,r')(\sigma\wedge\overrightarrow{\partial})_j\Psi(r) + G_2(r,r')(\sigma\wedge\overleftarrow{\partial})_j\Psi(r)\}\,dr$$

and it is easy to show that the integrand in the right-hand side is zero. This completes the proof.

We are now in a position to discuss the boundary data to add to (13). From the previous results, it is obvious that one has to consider two kinds of problems:

1. Third boundary value problems in which one gives a solution $\Psi(r^s)$ of (20) on S and then in theory, (22a) makes possible the computation of $\Psi(r)$. This is a third boundary value problem since it appears that (20) written in the form (20') implies a relation between $\Psi(r^s)$,

the normal derivative $\Psi_n(r^s)$, and the tangential derivative $(\sigma\wedge\partial)_n\Psi(r^s)$.

One can easily imagine boundary data satisfying (20): for instance, with two arbitrary solutions $u_1(r)$, $u_2(r)$ of (2), one introduces the spinor $\chi(r)=(\sigma^j\partial_j - iK)\binom{u_1(r)}{u_2(r)}$; $\chi(r)$ is a solution of (13) and its restriction $\chi(r^s)$ on S is a solution of (20).

2. First boundary value problems (Dirichlet problems) in which $\Psi(r^s)$ is given which permits the computation of $(\sigma\wedge\partial)_n\Psi(r^s)$, then $\partial_n\Psi(r^s)$ (and also the higher order normal derivatives) is obtained from (20′). Using (22b), one can in theory compute $\Psi(r)$.

There is a further question, for Dirichlet problem: Can $\Psi(r^s)$ be chosen arbitrarily on S? The following results bring a negative answer.

Indeed, we have not yet used the relations (14a)–(14c). Let us take, in (14a)–(14c), respectively for $\Psi(r)$ and $\Phi(r)$ some solution $\Psi(r)$, $\Psi^\dagger(r)$, of (13) and (13′); then we have

$$\int_S n^j\Psi^\dagger(r^s)\sigma_j\Psi(r^s)\,dr^s=0, \tag{24a}$$

$$\int_S n^j\{\Psi^\dagger(r^s)\vec{\partial}_j\Psi(r^s)-\Psi^\dagger(r^s)\partial_j\Psi(r^s)\}\,dr^s=0, \tag{24b}$$

$$\int_S n^j\vec{\partial}_j(\Psi^\dagger(r^s)\Psi(r^s))ds^s=2\int_\Omega\{\partial^j\Psi^\dagger(r)\partial_j\Psi(r)-K^2\Psi^\dagger(r)\Psi(r)\}dr. \tag{24c}$$

Using (20′) and the Hermitian conjugate equation, one can also deduce from (24a) and (24b)

$$\int_S n^j\{\Psi^\dagger(r^s)(\sigma\wedge\vec{\partial})_j\Psi(r^s)+\Psi^\dagger(r^s)(\sigma\wedge\overleftarrow{\partial})_j\Psi(r^s)\}\,dr^s=0. \tag{24′a}$$

The relations (24b) which put some restriction on the data $\Psi(r^s)$ for the Dirichlet problem, assert that $\Psi^\dagger(r^s)\sigma_j\Psi(r^s)$, $\Psi^\dagger(r^s)\vec{\partial}_j\Psi(r^s)-\Psi^\dagger(r^s)\overleftarrow{\partial}_j\Psi(r^s)$ are tangential vectors while (24c) is similar to (4′). If $\Psi(r)$ is considered as the Stokes spinor, $\Psi^\dagger(r)\sigma_j\Psi(r)$ is the polarization of the electromagnetic field and (24a) means that one has to use a transverse polarization.

Let us now come back to the numerical problem considered in the Introduction, that is for the Dirichlet problem, how to use (13), (20′) instead of (2), (3a). From the uniqueness of the solution in the case of the Dirichlet problem for the scalar Helmholtz equation and from Theorem 1, one knows that if $\psi_1(r^s)=u(r^s)$ on S, then $\psi_1(r)=u(r)$ in Ω where $\psi_1(r)$ is one of the two components of the spinor $\Psi(r)$. So, to solve the Dirichlet problem with the spinor Helmholtz equation one need only give another data $\psi_2(r^s)$ so that (24a) and (24b) are fulfilled, without forgetting that (20′) supplies $\partial_n\Psi(r^s)$. (This is an economic way from a computational viewpoint.)

C. Case of a spinor field in the half-space $z\geqslant 0$

Let us consider here a spinor field in the half-space $z\geqslant 0$ with boundary data on the plane $z=0$. The Eqs. (20), (22a), and (22b) hold valid for these last ones with $G_1(r,r')$, $G_2(r,r')$ given by (9a) and (9b) while the equations (20′) and (24a) become

$$\partial_3\Psi(x,y,0)=i(\sigma_1\partial_y-\sigma_2\partial_x)\Psi(x,y,0)-iK\sigma_3\Psi(x,y,0), \tag{25}$$

$$\int_S|\psi_1(x,y,0)|^2\,dx\,dy=\int_S|\psi_2(x,y,0)|^2\,dx\,dy, \tag{26a}$$

$$\int_S\{\Psi^\dagger(x,y,0)\vec{\partial}_3\Psi(x,y,0)-\Psi^\dagger(x,y,0)\overleftarrow{\partial}_3\Psi(x,y,0)\}dx\,dy=0, \tag{26b}$$

where $\partial_3\Psi(x,y,0)$ denotes the normal derivative on the plane $z=0$. One sees at once that for the Dirichlet problem there exists a very attractive spinor which satisfies (26a) and (26b), identically $\Psi(r^s)=\binom{\Psi(r^s)}{\bar{\Psi}(r^s)}$, where the bar denotes the complex conjugation.

It is interesting to investigate the solutions of the type $\Psi(r)=\Phi(r)\exp(iK_0z)$ where from (13) $\Phi(r)$ satisfies the equations

$$(\sigma^j\partial_j+iK_0\sigma_3+iK)\Phi(r)=0, \tag{27a}$$

$$(\partial^j\partial_j+2iK_0\partial_3+K^3-K_0^3)\Phi(r)=0. \tag{27b}$$

The solutions $\Psi(r)=\Phi(r)\exp(iK_0z)$ have the following interesting property:

Lemma: On every surface S_1 parallel to S (that is every plane $z=z_1$), one has

$$\frac{\partial}{\partial z}\int_{S_1}\Phi^\dagger(r^{s_1})\Phi(r^{s_1})\,dr^s=0. \tag{28}$$

To prove (28), we first note that (27b) and the Hermitian conjugate equation $(\partial^j\partial_j-2iK_0\partial_3+K^2-K_0^2)\Phi(r)=0$ give

$$2iK_0\partial_3(\Phi^\dagger(r)\Phi(r))+\Phi^\dagger(r)\vec{\partial}^3\vec{\partial}_3\Phi(r)-\Phi^\dagger(r)\overleftarrow{\partial}^3\overleftarrow{\partial}_3\Phi(r)$$
$$=-\{\Phi^\dagger(r)\vec{\partial}^\alpha\vec{\partial}_\alpha\Phi(r)-\Phi^\dagger(r)\overleftarrow{\partial}^\alpha\overleftarrow{\partial}_\alpha\Phi(r)\},\quad\alpha=1,2.$$

That is,

$$2iK_0\vec{\partial}_3(\Phi^\dagger(r)\Phi(r))+\vec{\partial}_3(\Phi^\dagger(r)\vec{\partial}_3\Phi(r)-\Phi^\dagger(r)\vec{\partial}_3\Phi(r))$$
$$=-\vec{\partial}^\alpha\{\Phi^\dagger(r)\vec{\partial}_\alpha\Phi(r)-\Phi^\dagger(r)\overleftarrow{\partial}_\alpha\Phi(r)\},$$

so that, if D is a disc in the plane S_1,

$$2iK_0\vec{\partial}_3\int_D\Phi^\dagger(r^s)\Phi(r^s)\,dr^s+\vec{\partial}_3\int_D\{\Phi^\dagger(r^s)\vec{\partial}_3\Phi(r^s)-\Phi^\dagger(r^s)\overleftarrow{\partial}_3\Phi(r^s)\}\,dr^s$$
$$=-\int_D\vec{\partial}^\alpha\{\Phi^\dagger(r^s)\vec{\partial}_\alpha\Phi(r^s)-\Phi^\dagger(r^s)\overleftarrow{\partial}_\alpha\Phi(r^s)\}\,dr^s. \tag{29}$$

Let us apply the Gauss–Ostrogradski theorem to the right-hand side of this relation. It becomes

$$2iK_0\vec{\partial}_3\int_D\Phi^\dagger(r^s)\Phi(r^s)\,dr^s+\vec{\partial}_3\int_D\{\Phi^\dagger(r^s)\vec{\partial}_3\Phi(r^s)-\Phi^\dagger(r^s)\overleftarrow{\partial}_3\Phi(r^s)\}\,dr^s$$
$$=-\int_C\nu^\alpha\{\Phi^\dagger(r^c)\vec{\partial}_\alpha\Phi(r^c)-\Phi^\dagger(r^c)\overleftarrow{\partial}_\alpha\Phi(r^c)\}\,dr^c$$

where C is the boundary of D, ν^α is the outward normal to C, and r^c is an arbitrary point on C.

Now, if the radius of the circle goes to infinity, the first term in the left-hand side becomes $2iK_0\partial_3\times\int_{S_1}\Phi^\dagger(r^s)\Phi(r^s)\,dr^s$ and the second term is zero. By once more using the Gauss–Ostrogradski theorem, one has

$$\vec{\partial}_3\int_{S_1}\{\Phi^\dagger(r^s)\vec{\partial}_3\Phi(r^s)-\Phi^\dagger(r^s)\overleftarrow{\partial}_3\Phi(r^s)\}\,dr^s$$
$$=\partial_3\int_\Omega\vec{\partial}^j\{\Phi^\dagger(r)\vec{\partial}_j\Phi(r)-\Phi^\dagger(r)\overleftarrow{\partial}_j\Phi(r)\}\,dr$$
$$=-4iK_0\vec{\partial}_3\int_\Omega\Phi^\dagger(r)\Phi(r)\,dr$$
$$=0.$$

Now, since for $|r^c|\to\infty$, $\Phi(r^c)\to 0$ (see the Introduction),

the right-hand side of (29) tends to zero and this completes the proof.

From a physical viewpoint, $\Phi^\dagger(r)\Phi(r)$ is a scalar proportional to the energy density of the spinor field, so (28) implies that the energy in a plane normal to the propagation direction is a motion constant. In other words, a transverse wave remains transverse.

D. Helmholtz equations with variable index

The discussion of the solutions of Helmholtz equations when the wavenumber K is a constant is a rather academic problem. We consider here a field in a medium with a slightly variable index. In a more precise way, we consider the scalar Helmholtz equation,

$$(\partial^j\partial_j + K^2n^2(r))u(r) = 0, \tag{30}$$

with

$$n(r) = 1 + \epsilon\mu(r), \quad \epsilon \sup_{r\in\Omega} |\mu(r)| \ll 1. \tag{30'}$$

The quantity ϵ measures the deviation of the index from unity. For the spinor fields, we have

$$(\sigma^j\partial_j + iKn(r))\Psi(r) = 0, \tag{31}$$

$$\Psi^\dagger(r)(\sigma^j\overleftarrow{\partial}_j - iKn(r)) = 0, \tag{31'}$$

but it is clear that Theorem 1 is no more valid. To obtain a weakened form of this theorem, we have to put on $n(r)$ more restrictive conditions than (30'). Among other possibilities, let us consider the two following ones:

(a) Let us assume that, besides (30') one has

$$|\partial_j\mu(r)| < \epsilon K|\mu(r)|. \tag{30''}$$

Then, at the ϵ order, Theorem 1 is valid since from (31), one deduces

$$\{\partial^j\partial_j + K^2n^2(r) + iK\sigma^j(\partial_j n(r))\}\Psi(r) = 0$$

and, using (30') and (30'') at the ϵ order,

$$\{\partial^j\partial_j + K^2(1 + 2\epsilon\mu(r))\}\Psi(r) = 0.$$

(b) One can also assume that $\mu(r)$ is a step function constant on blocks P_{ijl} in $\mathbb{R}^3$, that is, there exist P_{ijl} blocks,

$$P_{ijl} = \{x, y, z; x_i \le x < x_{i+1}, y_j \le y < y_{j+1}, z_l \le z < z_{l+1}\} \tag{30'''}$$

such that

$$n(r) = n(x_i, y_j, z_l) = n_{ijl} \quad \text{for} \quad r \in P_{ijl}.$$

With condition (30'''), Theorem 1 becomes a local theorem only valid on every block P_{ijl}. This condition is very attractive for numerical computations.

From a physical view point, it is interesting to notice that if $\epsilon\mu(r)$ is not a little perturbation, then, at the Fresnel approximation, a polarized electromagnetic beam will have a very different behavior than an unpolarized electromagnetic beam since the first is a solution of (31) and the second is a solution of (30), but now, unlike the case of n constant, the solutions of (31) are not solutions of (30).

3. CONCLUSIONS

In this paper, we gave the main properties of the spinor field equation and its solutions and we proved that it is possible for the Dirichlet problem to use this equation instead of the scalar Helmholtz equation. In complex problems, in particular when $n(r)$ is not a constant, it is conjectured that the spinor Helmholtz equation is easier to solve than the Helmholtz scalar equation and that the corresponding finite difference algorithms have a better stability. We will discuss this point in a second paper.[7]

[1(a)] E. M. Corson, *Introduction to Tensors, Spinors, and Relativistic Waves Equations* (Blackie, London, 1953).

[1(b)] E. P. Wigner, *Group Theory* (Academic, New York, 1959). For a more advanced text, see C. C. Chevalley, *The Algebraic Theory of Spinors* (Columbia, U. P., New York, 1954).

[2] E. Wolf, in *Electromagnetic Waves Propagation*, edited by M. Desirant and J. L. Michiels (Academic, New York, 1960), p. 119.

[3] P. Hillion, J. Opt. Soc. Am. **66**, 865 (1976).

[4] H. Jeffreys and B. Jeffreys, *Methods of Mathematical Physics* (Cambridge, U. P., Cambridge, 1956).

[5] F. John, *Partial Differential Equations* (Springer, Berlin, 1975).

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[7] Submitted to J. Comp. Phys.

Transformation brackets for generalized Bogolyubov–boson transformations[a)]

Yoel Tikochinsky[b)]

Center for Theoretical Physics, Department of Physics and Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 28 July 1977)

We consider the transformation $b_i = \Sigma_{j=1}^{s} (\lambda_{ij} a_j + \mu_{ij} a_j^\dagger)$ between two sets of boson operators, with λ_{ij} and μ_{ij} complex. Closed formulas are derived for the transformation brackets connecting base states of the two sets. As an application, the general quadratic Hamiltonian in one dimension is diagonalized and two examples, involving time-dependent real transformation and time-independent complex transformation, are worked out.

1. INTRODUCTION

It is well known that the problem of finding the eigenvalues and eigenfunctions for a set of s coupled harmonic oscillators can be solved, in analogy with the classical case, by performing the principle axis transformation. Thus, given the Hamiltonian $H = (a^\dagger a)\mathcal{H}\binom{a}{a^\dagger}$ with $(a^\dagger a) = (a_1^\dagger \cdots a_s^\dagger a_1 \cdots a_s)$, a complex transformation

$$b_i = \sum (\lambda_{ij} a_j + \mu_{ij} a_j^\dagger), \quad b_i^\dagger = \sum (\mu_{ij}^* a_j + \lambda_{ij}^* a_j^\dagger) \tag{1}$$

can be found, which diagonalizes the $2s$ by $2s$ Hermitian matrix $\mathcal{H}$. The diagonalizing matrix $A = \binom{\lambda}{\mu^*}\ \binom{\mu}{\lambda^*}$ however, is in general *not* unitary. Once the transformation to the new boson operators has been performed, the eigenvalues of the Hamiltonian can be read from the uncoupled form $H = \sum h_i b_i^\dagger b_i + \text{const}$. To complete the solution, one needs the transformation brackets from base states of the old operators to those of the new operators. It is precisely the calculation of these brackets to which we shall attend ourselves.

Particular cases of the general transformation (1) (both for bosons and fermions), of which the Bogolyubov—Valatin transformation[1] is a notable example, have been found extremely useful in the past. Baranger,[2] has considered the full transformation (1) for the case of fermions. Solutions for the transformation brackets (in the boson case) for particular forms of (1), have been recently published.[3–6] Common to these works, is the assumption of a transformation with real coefficients. As we shall later see, this assumption amounts to a pure point transformation (transforming coordinates among themselves), whereas the general transformation (1) is a full canonical transformation mixing coordinates and momenta.

In deriving the transformation brackets, it is convenient to employ the so-called coherent states (eigenstates of the annihilation operators). Using these states, cumbersome recurrence relations are replaced by simple differential equations whose solutions can be readily found. The relevant properties of the coherent states are summerized in Sec. 2. Section 3 treats the one-dimensional case. The general transformation brackets are derived both in the coherent state representation and in the occupation number representation. As an application, the general one-dimensional quadratic Hamiltonian is diagonalized in Sec. 4, and two specific examples, one with time-dependent real coefficients and another, with time-independent complex coefficient are presented. In Sec. 5, the s-dimensional transformation is worked out. Conditions for the $2s$ by $2s$ matrix $A = \binom{\lambda}{\mu^*}\ \binom{\mu}{\lambda^*}$ to be regular and to connect one set of boson operators with another set, are spelled out, and the most general form of such a matrix, subject to an additional reality condition, is found. Finally, the transformation brackets are derived in the coherent state representation. Our main results are Eqs. (23) and (25) for the one-dimensional case, and Eq. (84) with definitions (63) and (83) and conditions (64) for the s-dimensional case.

2. PROPERTIES OF THE COHERENT STATES[7]

In this section we summarize those properties of coherent states which pertain to our discussion. The coherent states are defined as eigenstates of the annihilation operator a

$$a|\alpha\rangle = \alpha|\alpha\rangle, \tag{2}$$

where α is any complex number. In terms of the occupation number states, these states are given by

$$|\alpha\rangle = \exp(-|\alpha|^2/2) \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle, \tag{3}$$

or

$$|\alpha\rangle = \exp(-|\alpha|^2/2) \exp(\alpha a^\dagger) |0\rangle \tag{4}$$

Note that the vacuum state in both representations coincide. The coherent states form a complete set

$$\int \frac{d^2\alpha}{\pi} |\alpha\rangle\langle\alpha| = 1, \tag{5}$$

where

$$d^2\alpha \equiv d(\mathrm{Re}\alpha)\, d(\mathrm{Im}\alpha). \tag{6}$$

The scalar product of two different states does not vanish. Instead, we have

$$\langle\beta|\alpha\rangle = \exp[-\tfrac{1}{2}|\beta|^2 - \tfrac{1}{2}|\alpha|^2 + \beta^*\alpha]. \tag{7a}$$

In particular, the normalization of the states is

$$\langle\alpha|\alpha\rangle = 1. \tag{7b}$$

[a)] This work is supported in part through funds provided by ERDA under Contract EY-76-C-02-3069.*000.
[b)] Address as of August 1977, Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem, Israel.

 0022-2488/78/1901-0270$1.00

Any state $|\psi\rangle$ can be expanded in terms of the coherent states $|\alpha\rangle$,

$$|\psi\rangle = \int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2/2)\ \psi(\alpha^*)|\alpha\rangle, \tag{8}$$

where

$$\psi(\alpha^*) \equiv \exp(|\alpha|^2/2)\langle\alpha|\psi\rangle \tag{9}$$

is an analytic function of α^*. In particular, the representation inverse to (3) is given by

$$|n\rangle = \frac{1}{\sqrt{n!}} \int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2/2)(\alpha^*)^n|\alpha\rangle. \tag{10}$$

Hence,

$$\langle n|n'\rangle = \delta_{nn'} = \int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2) \frac{1}{\sqrt{n!}}(\alpha)^n \frac{1}{\sqrt{n'!}}(\alpha^*)^{n'}. \tag{11}$$

Finally, for any analytic function $\psi(z)$, the following relation holds,

$$\int \frac{d^2\alpha}{\pi} \exp(\beta^*\alpha - |\alpha|^2)\,\psi(\alpha^*) = \psi(\beta^*). \tag{12}$$

3. THE ONE-DIMENSIONAL CASE

Consider the transformation

$$b = \lambda a + \mu a^\dagger, \quad b^\dagger = \mu^* a + \lambda^* a^\dagger. \tag{13}$$

Using the boson commutation relations we have

$$|\lambda|^2 - |\mu|^2 = 1. \tag{14}$$

The inverse transformation

$$a = \lambda^* b - \mu b^\dagger, \quad a^\dagger = -\mu^* b + \lambda b^\dagger \tag{15}$$

automatically satisfies (14). We shall use the notation

$$a|\alpha\rangle_a = \alpha|\alpha\rangle_a, \quad b|\beta\rangle_b = \beta|\beta\rangle_b \tag{16}$$

and expand the new coherent states in terms of the old coherent states

$$|\beta\rangle_b = \int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2/2)\psi_\beta(\alpha^*)|\alpha\rangle_a, \tag{17}$$

where

$$\psi_\beta(\alpha^*) = \exp(|\alpha|^2/2)\ {}_a\langle\alpha|\beta\rangle_b. \tag{18}$$

Consider the matrix element

$$\begin{aligned} {}_a\langle\gamma|b|\beta\rangle_b &= \beta\ {}_a\langle\gamma|\beta\rangle_b = {}_a\langle\gamma|\lambda a + \mu a^\dagger|\beta\rangle_b \\ &= \lambda\int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2/2)\psi_\beta(\alpha^*)\alpha\ {}_a\langle\gamma|\alpha\rangle_a \\ &\quad + \mu\gamma^*\int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2/2)\psi_\beta(\alpha^*)\ {}_a\langle\gamma|\alpha\rangle_a, \end{aligned}$$

where we have employed Eqs. (13), (16), (17), and (18). Using the value of the scalar product (7a), and Eqs. (12) and (18), we obtain

$$\lambda\frac{\partial}{\partial\gamma^*}\psi_\beta(\gamma^*) + (\mu\gamma^* - \beta)\,\psi_\beta(\gamma^*) = 0.$$

Hence,

$$\psi_\beta(\gamma^*) = \psi_\beta(0)\exp\left[\frac{\beta}{\lambda}\gamma^* - \frac{1}{2}\frac{\mu}{\lambda}\gamma^{*2}\right]. \tag{19}$$

In order to determine the bracket

$$\psi_\beta(0) = {}_a\langle 0|\beta\rangle_b, \tag{20}$$

we shall first calculate the expansion coefficients of the new vacuum in terms of the old states. Specializing Eq. (19) to the case $\beta = 0$, we have

$$\psi_0(\alpha^*) = \psi_0(0)\exp[-\tfrac{1}{2}\tau\alpha^{*2}]$$

with $\tau = \mu/\lambda$. The factor $\psi_0(0) = {}_a\langle 0|0\rangle_b$ can now be determined by use of the normalization condition (7b). Indeed,

$$\begin{aligned} 1 = {}_b\langle 0|0\rangle_b &= \int \frac{d^2\alpha}{\pi}\ {}_b\langle 0|\alpha\rangle_a\ {}_a\langle\alpha|0\rangle_b \\ &= \int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2)|\psi_0(\alpha^*)|^2 \\ &= |\psi_0(0)|^2 \int \frac{d^2\alpha}{\pi} \exp(-|\alpha|^2)\exp(-\tfrac{1}{2}\tau^*\alpha^2 - \tfrac{1}{2}\tau\alpha^{*2}). \end{aligned}$$

The last integral is a simple two-dimensional Gaussian integral. Immediate evaluation gives

$$\text{Integral} = (1 - |\tau|^2)^{-1/2}. \tag{21}$$

Simplifying with the aid of (14), we have

$$1 = |\psi_0(0)|^2\,|\lambda|.$$

Since the phase of the new vacuum relative to the old vacuum can be chosen at will, we secure

$${}_a\langle\alpha|0\rangle_b = |\lambda|^{-1/2}\exp\left(-\tfrac{1}{2}|\alpha|^2 - \frac{\mu}{2\lambda}\alpha^{*2}\right). \tag{22}$$

The needed bracket ${}_a\langle 0|\beta\rangle_b = {}_b\langle\beta|0\rangle_a^*$ [Eq. (20)], can now be obtained from Eq. (22) by invoking the inverse transformation. Indeed, Eq. (15) tells us that we have to replace λ by λ^* and μ by $-\mu$ in Eq. (22). Thus,

$${}_a\langle 0|\beta\rangle_b = |\lambda|^{-1/2}\exp\left(-\tfrac{1}{2}|\beta|^2 + \frac{\mu^*}{2\lambda}\beta^2\right).$$

Substituting this result in Eq. (19) we obtain, with the aid of Eq. (14),

$$\begin{aligned} {}_a\langle\alpha|\beta\rangle_b = |\lambda|^{-1/2}\exp\Big(&-\tfrac{1}{2}|\alpha|^2 - \tfrac{1}{2}|\beta|^2 + \frac{\alpha^*\beta}{\lambda} \\ &- \frac{\mu\alpha^{*2}}{2\lambda} + \frac{\mu^*\beta^2}{2\lambda}\Big). \end{aligned} \tag{23}$$

In the special case of the identity transformation ($\lambda = 1$, $\mu = 0$), the overlap integral (23) reduces to Eq. (7b), as it should.

An alternative way for obtaining Eq. (23) is to employ representation (4) for the new states together with expression (22) for the new vacuum. Since[8]

$$\begin{aligned} \exp(\beta b^\dagger) &= \exp[\beta(\lambda^* a^\dagger + \mu^* a)] \\ &= \exp(\tfrac{1}{2}\beta^2\lambda^*\mu^*)\exp(\beta\lambda^* a^\dagger)\exp(\beta\mu^* a), \end{aligned} \tag{24}$$

we have

$$\begin{aligned} {}_a\langle\gamma|\beta\rangle_b &= {}_a\langle\gamma|\exp(-|\beta|^2/2)\exp(\beta b^\dagger)|0\rangle_b \\ &= |\lambda|^{-1/2}\exp(-|\beta|^2/2)\exp(\beta^2\lambda^*\mu^*/2) \\ &\quad \times\int \frac{d^2\alpha}{\pi}\exp(-|\alpha|^2/2)\exp[-(\mu/2\lambda)\alpha^{*2}] \\ &\quad \times {}_a\langle\gamma|\exp(\beta\lambda^* a^\dagger)\exp(\beta\mu^* a)|\alpha\rangle_a \\ &= |\lambda|^{-1/2}\exp(-|\beta|^2/2)\exp(-|\gamma|^2/2) \end{aligned}$$

$$\times\exp(\beta^2\lambda^*\mu^*/2)\exp(\beta\lambda^*\gamma^*)$$
$$\times\int\frac{d^2\alpha}{\pi}\exp(-|\alpha|^2)\exp[(\gamma^*+\beta\mu^*)\alpha]$$
$$\times\exp[-(\mu/2\lambda)\alpha^{*2}].$$

The last integral is equal by (12) to

$$\exp\left(-\frac{\mu}{2\lambda}(\gamma^*+\beta\mu^*)^2\right).$$

Simplifying, with the aid of (14), we regain Eq. (23).

To obtain the transformation brackets in the occupation number representation, we expand the last three exponentials in Eq. (23) in powers of α^* and β and use Eqs. (10) and (11). The result is

$${}_a\langle n|N\rangle_b=0 \quad \text{if } (N-n) \text{ is odd}, \tag{25a}$$

$${}_a\langle n|N\rangle_b=\left(\frac{n!N!}{|\lambda|}\right)^{1/2}\left(\frac{\mu^*}{2}\right)^{(N-n)/2}\lambda^{-(N+n)/2}$$
$$\times\sum_k\frac{(-1)^k}{k!(n-2k)!\,[k+\frac{1}{2}(N-n)]!}\left(\frac{|\mu|}{2}\right)^{2k}, \tag{25b}$$

where k is an integer satisfying

$$0\le k, \quad -\tfrac{1}{2}(N-n)\le k\le\tfrac{1}{2}n. \tag{25c}$$

For λ and μ real, we set $\lambda=\cosh y$ and $\mu=\sinh y$ [to satisfy (14)]. Equation (25) reduces in this case to well known results. [See, for example, Ref. 6, Eq. (2.22).]

To see what is implied by limiting λ and μ to real values, let us go back to coordinates and momenta. Introduce

$$x=[\hbar/(2m\omega)]^{1/2}\ (a^\dagger+a)$$

and

$$P_x=i[m\omega\hbar/2]^{1/2}\ (a^\dagger-a), \tag{26a}$$

where m and ω are constants having the dimensions of mass and frequency. Similarly, let

$$y=[\hbar/(2m\omega)]^{1/2}\ (b^\dagger+b)$$
$$P_y=i(m\omega\hbar/2)^{1/2}\ (b^\dagger-b). \tag{26b}$$

With these definitions, transformation (13) takes the form

$$y=(\mathrm{Re}\lambda+\mathrm{Re}\mu)x-(\mathrm{Im}\lambda-\mathrm{Im}\mu)/(m\omega)P_x$$
$$P_y=m\omega(\mathrm{Im}\lambda+\mathrm{Im}\mu)x+(\mathrm{Re}\lambda-\mathrm{Re}\mu)P_x. \tag{27}$$

Thus, in general, transformation (27) is a full canonical transformation mixing coordinates and momenta. Equation (27) reduces to a point transformation if and only if $\mathrm{Im}\lambda=\mathrm{Im}\mu=0$.

4. APPLICATION TO QUADRATIC HAMILTONIAN IN ONE DIMENSION

Consider the Hamiltonian

$$H=\alpha_0+\alpha_1a+\alpha_1^*a^\dagger+\alpha_2a^2+\alpha_2^*a^{\dagger2}+\alpha_3a^\dagger a \tag{28}$$

with $\alpha_0=\alpha_0^*$ and $\alpha_3=\alpha_3^*$. The coefficients α_i may be time dependent. Using matrix notation, we rewrite H in the form

$$H=(a^\dagger a)\begin{pmatrix}\alpha_3-\alpha_0 & \alpha_2^*\\ \alpha_2 & \alpha_0\end{pmatrix}\begin{pmatrix}a\\ a^\dagger\end{pmatrix}+(\alpha_1\alpha_1^*)\begin{pmatrix}a\\ a^\dagger\end{pmatrix}. \tag{29}$$

We now seek a transformation

$$A=\begin{pmatrix}\lambda & \mu\\ \mu^* & \lambda^*\end{pmatrix} \tag{30}$$

with $\det A=|\lambda|^2-|\mu|^2=1$, such that A diagonalizes the Hermitian matrix

$$\mathcal{H}=\begin{pmatrix}\alpha_4 & \alpha_2^*\\ \alpha_2 & \alpha_0\end{pmatrix},\quad \alpha_4=\alpha_3-\alpha_0. \tag{31}$$

That is,

$$\mathcal{H}'=(A^{-1})^\dagger\mathcal{H}A^{-1}=\begin{pmatrix}\alpha_4' & 0\\ 0 & \alpha_0'\end{pmatrix}. \tag{32}$$

Since $\mathcal{H}'$ is also Hermitian, $\alpha_4'=\alpha_3'-\alpha_0'$ and α_0' are real. With

$$\begin{pmatrix}b\\ b^\dagger\end{pmatrix}=A\begin{pmatrix}a\\ a^\dagger\end{pmatrix} \tag{33}$$

and

$$(\alpha_1'\alpha_1'^*)=(\alpha_1\alpha_1^*)A^{-1}, \tag{34}$$

we have

$$H=(b^\dagger b)\mathcal{H}'\begin{pmatrix}b\\ b^\dagger\end{pmatrix}+(\alpha_1'\alpha_1'^*)\begin{pmatrix}b\\ b^\dagger\end{pmatrix}$$
$$=\alpha_3'b^\dagger b+\alpha_0'+\alpha_1'b+\alpha_1'^*b^\dagger. \tag{35}$$

Equation (15) gives

$$A^{-1}=\begin{pmatrix}\lambda^* & -\mu\\ -\mu^* & \lambda\end{pmatrix}. \tag{36}$$

Thus, α_1' is explicitly given by

$$\alpha_1'=\alpha_1\lambda^*-\alpha_1^*\mu^*. \tag{37}$$

The eigenvalue equation (32), is equivalent to two pairs of homogeneous equations for the unknowns λ^*, μ^* and λ, μ, namely,

$$\begin{aligned}&(\alpha_4-\alpha_4')\lambda^*-\alpha_2^*\mu^*=0,\\ &\alpha_2\lambda^*-(\alpha_0+\alpha_4')\mu^*=0,\end{aligned} \tag{38a}$$

and

$$\begin{aligned}&(\alpha_0-\alpha_0')\lambda-\alpha_2\mu=0,\\ &\alpha_2^*\lambda-(\alpha_0'+\alpha_4)\mu=0.\end{aligned} \tag{38b}$$

A necessary and sufficient condition for the existence of nontrivial solutions to these equations is the vanishing of the corresponding determinants, i.e.,

$$-(\alpha_4-\alpha_4')(\alpha_0+\alpha_4')+|\alpha_2|^2=0$$

and

$$-(\alpha_0-\alpha_0')(\alpha_0'+\alpha_4)+|\alpha_2|^2=0.$$

Hence,

$$\alpha_0'=\tfrac{1}{2}\{(\alpha_0-\alpha_4)+[(\alpha_0+\alpha_4)^2-4|\alpha_2|^2]^{1/2}\} \tag{39a}$$

and

$$\alpha_4'=\tfrac{1}{2}\{-(\alpha_0-\alpha_4)+[(\alpha_0+\alpha_4)^2-4|\alpha_2|^2]^{1/2}\}. \tag{39b}$$

We have used the plus sign in front of the square root to ensure $\alpha_0'\to\alpha_0$ and $\alpha_4'\to\alpha_4$ in the limit $\alpha_2\to0$. The eigenvalues remain real as long as the perturbation satisfies

$$|\alpha_2|^2\le(\alpha_0+\alpha_4)^2/4. \tag{40}$$

From Eqs. (39) we have

$$\alpha_4 - \alpha_4' = \alpha_0 - \alpha_0'; \tag{41}$$

hence, the two pairs of Eq. (38) reduce to a single pair, the second pair being the complex conjugate of the first. Solving for μ we have

$$\mu = \lambda\alpha_2^*/(\alpha_0' + \alpha_4). \tag{42}$$

Substituting this expression in Eq. (14), the absolute value of λ is determined. Since λ can be chosen real and positive, we secure

$$\lambda = [(\alpha_0' + \alpha_4)/(\alpha_0' + \alpha_4')]^{1/2}. \tag{43}$$

By way of illustration, consider the following examples. Let

$$H(t) = \frac{p^2}{2m}\exp(-2\gamma t) + \tfrac{1}{2}m\omega^2 x^2 \exp(2\gamma t). \tag{44}$$

Using Eq. (26a), the time-dependent Hamiltonian for the damped harmonic oscillator takes the form

$$H(t) = \tfrac{1}{2}\hbar\omega[\sinh 2\gamma t(a^{\dagger 2} + a^2) + 2\cosh 2\gamma t(a^\dagger a + \tfrac{1}{2})]. \tag{45}$$

Thus, $\alpha_0 = \frac{1}{2}\alpha_3 = \alpha_4 = \frac{1}{2}\hbar\omega\cosh 2\gamma t$, $\alpha_1 = 0$, and $\alpha_2 = \frac{1}{2}\hbar\omega\sinh 2\gamma t$. Equations (39), (42), (43), and (35) give

$$\alpha_0' = \tfrac{1}{2}\hbar\omega, \quad \alpha_3' = \hbar\omega,$$

$$\lambda = \cosh\gamma t, \quad \mu = \sinh\gamma t$$

and

$$H = \hbar\omega(b^\dagger b + \tfrac{1}{2}).$$

Thus, the eigenvalues of this time-dependent Hamiltonian are time-independent

$$E_n = \hbar\omega(n + \tfrac{1}{2}). \tag{46}$$

On the other hand, the eigenfunctions are time dependent. Since the transformation in this case is a real point transformation, we expect that [see Eq. (27)]

$$\psi_n = \psi_n(y = \exp(\gamma t)x), \tag{47}$$

where $\psi_n(x)$ is the normalized free harmonic oscillator wavefunction. Inspection of the Schrödinger equation $H\psi_n = E_n\psi_n$, with H given by Eq. (44), shows that this is indeed the case. Thus, the transformation $y = \exp(\gamma t)x$ which carries the classical damped harmonic oscillator equation of motion

$$\ddot{x} + 2\gamma\dot{x} + \omega^2 x = 0$$

into the free oscillator equation

$$\ddot{y} + (\omega^2 - \gamma^2)y = 0,$$

plays a similar role in the quantum-mechanical case. Furthermore, from Eq. (47) we see that the total probability for finding the particle anywhere decays as $\exp(-\gamma t)$. The decay of the probability for finding the particle at a given point $x \neq 0$, follows a more complicated rule. The last decay agrees qualitatively with the classical prediction. Finally, the expansion of the harmonic oscillator wavefunction $\psi_N(\exp(\gamma t)x)$ in terms of the functions $\psi_n(x)$ is given by (25) with λ and μ as deduced above.

Our next example treats the time-independent Hamiltonian

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 x^2 + \gamma(xp + px) = \hbar\omega(a^\dagger a + \tfrac{1}{2}) + i\hbar\gamma(a^{\dagger 2} - a^2). \tag{48}$$

This time we have $\alpha_0 = \frac{1}{2}\alpha_3 = \alpha_4 = \frac{1}{2}\hbar\omega$, $\alpha_1 = 0$, and $\alpha_2 = -i\hbar\gamma$. Equations (39), (42), (43), and (35) give

$$\alpha_0' = \tfrac{1}{2}\hbar\Omega, \quad \alpha_3' = \hbar\Omega,$$

$$\lambda = [(\Omega + \omega)/(2\Omega)]^{1/2}, \quad \mu = i\gamma\{2/[\Omega(\Omega + \omega)]\}^{1/2},$$

and

$$H = \hbar\Omega(b^\dagger b + \tfrac{1}{2})$$

with

$$\Omega = (\omega^2 - 4\gamma^2)^{1/2}.$$

The transformation in this case is a full complex canonical transformation. The energies of the perturbed Hamiltonian (48) are $E_n = \hbar\Omega(n + \frac{1}{2})$.

5. THE s-DIMENSIONAL CASE

A. Conditions on the transformation

Consider the transformation

$$b_i = \sum_{j=1}^{s}(\lambda_{ij}a_j + \mu_{ij}a_j^\dagger),$$

$$b_i^\dagger = \sum_{j=1}^{s}(\mu_{ij}^* a_j + \lambda_{ij}^* a_j^\dagger), \quad i = 1, \ldots, s. \tag{49a}$$

Using matrix notation we have

$$\begin{pmatrix} b \\ b^\dagger \end{pmatrix} = A\begin{pmatrix} a \\ a^\dagger \end{pmatrix}, \quad A = \begin{pmatrix} \lambda & \mu \\ \mu^* & \lambda^* \end{pmatrix}. \tag{49b}$$

The commutation relations satisfied by the boson operators and the requirement that the matrix A shall have an inverse, impose some conditions on the s by s matrices λ and μ. Thus,

$$[b_i, b_j^\dagger] = \delta_{ij} \tag{50}$$

imply

$$\lambda\lambda^\dagger - \mu\mu^\dagger = 1, \tag{51}$$

while $[b_i, b_j] = 0$ leads to

$$\lambda\tilde{\mu} = \mu\tilde{\lambda}. \tag{52}$$

Similarly, the inverse transformation

$$A^{-1} = \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix}$$

should satisfy

$$\alpha\alpha^\dagger - \beta\beta^\dagger = 1 \tag{53}$$

and

$$\alpha\tilde{\beta} = \beta\tilde{\alpha}, \tag{54}$$

while $A^{-1}A = AA^{-1}$ imply

$$\alpha\lambda + \beta\mu^* = 1, \tag{55}$$

$$\alpha\mu + \beta\lambda^* = 0, \tag{56}$$

and

$$\lambda\alpha + \mu\beta^* = 1, \tag{57}$$

$$\lambda\beta + \mu\alpha^* = 0. \tag{58}$$

It is shown in Appendix A that Eqs. (51)—(54), (57), and (58) alone determine α and β uniquely, namely,

$$\alpha = \lambda^\dagger, \quad \beta = -\tilde{\mu}. \tag{59}$$

Substituting these solutions into Eqs. (55) and (56) we have

$$\lambda^\dagger\lambda - \tilde{\mu}\mu^* = 1, \tag{60}$$

and

$$\lambda^\dagger\mu = \tilde{\mu}\lambda^*. \tag{61}$$

From Eq. (51) it is clear that λ must have an inverse. Furthermore, the four conditions (51), (52), (60), and (61) are not independent. In fact, using λ^{-1}, it is easy to derive Eq. (60), say, from the remaining three conditions. In what follows, we will find it *convenient* to impose an additional reality condition on the matrix λ, namely,

$$\lambda^\dagger\lambda = (\lambda^\dagger\lambda)^*. \tag{62}$$

In terms of the matrices

$$\sigma = \lambda^{-1} \text{ and } \tau = \lambda^{-1}\mu, \tag{63}$$

all conditions can be summarized as follows:

$$\tau = \tilde{\tau} \text{ (symmetric)}, \tag{64a}$$

$$\sigma\sigma^\dagger + \tau\tau^\dagger = 1, \tag{64b}$$

$$\sigma\sigma^\dagger\tau = \widetilde{(\sigma\sigma^*\tau)} \text{ (symmetric)}, \tag{64c}$$

and

$$\sigma\sigma^\dagger = (\sigma\sigma^\dagger)^* \text{ (reality condition)}. \tag{64d}$$

The reality condition ensures that the symmetric matrices $\mathrm{Re}\tau$ and $\mathrm{Im}\tau$ commute. Indeed, by Eqs. (64b) and (64d)

$$\tau\tau^\dagger = (\mathrm{Re}\tau)^2 + (\mathrm{Im}\tau)^2 + i(\mathrm{Im}\tau\cdot\mathrm{Re}\tau - \mathrm{Re}\tau\cdot\mathrm{Im}\tau)$$
$$= \text{real matrix}.$$

Hence,

$$[\mathrm{Re}\tau,\ \mathrm{Im}\tau] = 0. \tag{65}$$

It is easy to demonstrate that matrices σ and τ satisfying Eqs. (64) do exist. Indeed, let C be any real diagonal matrix with elements C_i satisfying $0 < C_i \leq 1$, and let τ' be a complex symmetric matrix commuting with C and satisfying $\tau'\tau'^* = 1 - C$. Thus, τ' consists of square symmetric matrices along the main diagonal in accordance with the multiplicity of the C's. Choose an arbitrary real orthogonal matrix Q and an arbitrary complex unitary matrix U. Then

$$\tau = \tilde{Q}\tau' Q \tag{66a}$$

and

$$\sigma = \tilde{Q}C^{1/2}U \tag{66b}$$

is a solution of Eqs. (64). Indeed, as shown in Appendix B, any solution of Eqs. (64) is necessarily of the form (66).

As in the one-dimensional case, the condition for the transformation (49) to be a point transformation, is

$$\mathrm{Im}\lambda = \mathrm{Im}\mu = 0. \tag{67}$$

B. The transformation brackets

We shall use the following condensed notation:

$$|\alpha\rangle = |\alpha_1, \ldots, \alpha_s\rangle,$$
$$d^2\alpha = d^2\alpha_1 \cdots d^2\alpha_s,$$
$$\psi_\beta(\alpha) = \psi_{\beta_1,\ldots,\beta_s}(\alpha_1, \ldots, \alpha_s), \text{ etc.}$$

Thus

$$a_i|\alpha\rangle_a = \alpha_i|\alpha\rangle_a, \quad b_i|\beta\rangle_b = \beta_i|\beta\rangle_b. \tag{68}$$

Starting with the matrix element ${}_a\langle\gamma|b_i|\beta\rangle_b$ and proceeding exactly as in the one-dimensional case, we obtain the following system of differential equations,

$$\sum_{j=1}^{s}\left(\lambda_{ij}\frac{\partial}{\partial\alpha_j^*}\psi_\beta + \mu_{ij}\alpha_j^*\psi_\beta\right) = \beta_i\psi_\beta, \quad i = 1, \ldots, s \tag{69}$$

for the function

$$\psi_\beta(\alpha^*) = \exp(\tfrac{1}{2}\textstyle\sum_j|\alpha_j|^2)\ {}_a\langle\alpha|\beta\rangle_b. \tag{70}$$

Since λ is regular, system (69) can be brought to canonical form, namely,

$$\frac{\partial}{\partial x_i}\psi_\beta = \sum_j(\sigma_{ij}\beta_j - \tau_{ij}x_j)\psi_\beta, \quad x_i \equiv \alpha_i^*, \tag{71}$$

where we have used definitions (63). From the fact that ψ_β is an analytic function of its arguments and from the structure of Eq. (71), it is clear that a single number, e.g., $\psi_\beta(\mathbf{0}) = {}_a\langle\mathbf{0}|\beta\rangle_b$ determines $\psi_\beta(\mathbf{x})$ completely. Indeed, solving the $i = 1$ equations we obtain (dropping temporarily the index β)

$$\psi(x_1, \ldots, x_s) = \psi_1(x_2, \ldots, x_s) \times\exp[\textstyle\sum\sigma_{1j}\beta_j x_1 - \sum_{j=2}\tau_{1j}x_j x_1 - \frac{1}{2}\tau_{11}x_1^2] \tag{72}$$

with

$$\psi_1(x_2, \ldots, x_s) \equiv \psi(0, x_2, \ldots, x_s).$$

We now use the $i = 2$ equation to drive an equation for ψ_1,

$$\frac{\partial\psi_1}{\partial x_2} = (\textstyle\sum\sigma_{2j}\beta_j - \sum_{j=2}\tau_{2j}x_j)\,\psi_1.$$

The solution for ψ_1, namely,

$$\psi_1(x_2, \cdots, x_s) = \psi_2(x_3, \cdots, x_s)\exp[\textstyle\sum\sigma_{2j}\beta_j x_2 - \sum_{j=3}\tau_{2j}x_j x_2 - \frac{1}{2}\tau_{22}x_2^2]$$

with

$$\psi_2(x_3, \cdots, x_s) = \psi_1(0, x_3, \cdots, x_s)$$

is substituted in Eq. (72) and the $i = 3$ equation is used to derive an equation for $\psi_2(x_3, \ldots, x_s)$. Proceeding in this way, we finally obtain

$$\psi_\beta(\mathbf{x}) = \psi_\beta(\mathbf{0})\exp[\textstyle\sum_{ij}\sigma_{ij}\beta_j x_i - \frac{1}{2}\sum_{ij}\tau_{ij}x_i x_j]. \tag{73}$$

Now the bracket

$$\psi_\beta(\mathbf{0}) = {}_a\langle\mathbf{0}|\beta\rangle_b \tag{74}$$

has to be determined. Specializing (73) to the case

$\beta=0$, we have

$$\psi_0(\mathbf{x})=\psi_0(0)\exp(-\tfrac{1}{2}\sum_{ij}\tau_{ij}x_ix_j) \tag{75a}$$

or, using Eq. (70)

$${}_a\langle\boldsymbol{\alpha}|\mathbf{0}\rangle_b=\psi_0(0)\exp[-\tfrac{1}{2}\sum_j|\alpha_j|^2-\tfrac{1}{2}\sum_{ij}\tau_{ij}\alpha_i^*\alpha_j^*]. \tag{75b}$$

Again, the normalization of the new vacuum state will be used to determine $\psi_0(0)$, namely

$$\begin{aligned}1={}_b\langle\mathbf{0}|\mathbf{0}\rangle_b&=\int\frac{d^2\boldsymbol{\alpha}}{\pi^s}\,{}_b\langle\mathbf{0}|\boldsymbol{\alpha}\rangle_a\,{}_a\langle\boldsymbol{\alpha}|\mathbf{0}\rangle_b\\&=|\psi_0(0)|^2\int\frac{d^2\boldsymbol{\alpha}}{\pi^s}\exp[-\sum|\alpha_j|^2\\&\quad-\tfrac{1}{2}\sum(\tau_{ij}^*\alpha_i\alpha_j+\tau_{ij}\alpha_i^*\alpha_j^*)]\\&=|\psi_0(0)|^2\int\frac{d^2\boldsymbol{\alpha}}{\pi^s}\exp\{-(\tilde{\alpha}^*)(\alpha)\\&\quad-\mathrm{Re}[(\tilde{\alpha}^*)\tau(\alpha^*)]\}.\end{aligned} \tag{76}$$

The evaluation of the last integral is greatly simplified if $\mathrm{Re}\tau$ and $\mathrm{Im}\tau$ commute [Eq. (65)]. Indeed, this was our *only* reason for introducing the reality condition (64d). Since $\mathrm{Re}\tau$ and $\mathrm{Im}\tau$ are two commuting, symmetric, and real matrices, there exists a real orthogonal matrix Q, which diagonalizes both matrices. That is,

$$Q\,\mathrm{Re}\tau\tilde{Q}=\tau_R',\quad Q\,\mathrm{Im}\tau\tilde{Q}=\tau_I' \tag{77}$$

with τ_R' and τ_I' diagonal. Introducing the new integration variables

$$(x)=Q(\mathrm{Re}\alpha) \tag{78a}$$

and

$$(y)=Q(\mathrm{Im}\alpha) \tag{78b}$$

we have

$$\begin{aligned}\widetilde{(\alpha^*)}(\alpha)&=\widetilde{(\mathrm{Re}\alpha)}(\mathrm{Re}\alpha)+\widetilde{(\mathrm{Im}\alpha)}(\mathrm{Im}\alpha)\\&=(\tilde{x})(x)+(\tilde{y})(y)\end{aligned}$$

and

$$\begin{aligned}\mathrm{Re}[\widetilde{(\alpha^*)}\tau(\alpha^*)]&=\widetilde{(\mathrm{Re}\alpha)}\,\mathrm{Re}\tau(\mathrm{Re}\alpha)-\widetilde{(\mathrm{Im}\alpha)}\,\mathrm{Re}\tau(\mathrm{Im}\alpha)\\&\quad+2(\mathrm{Re}\alpha)\,\mathrm{Im}\tau(\mathrm{Im}\alpha)\\&=(\tilde{x})\,\tau_R'(x)-(\tilde{y})\tau_R'(y)+2(\tilde{x})\tau_I'(y).\end{aligned}$$

Moreover, since (78) is a real orthogonal transformation, the Jacobian of the transformation is equal to 1. Thus, the multiple integral in Eq. (76) reduces to a product of s two-dimensional integrals,

$$\begin{aligned}\mathrm{Integral}=\prod_{i=1}^{s}\int\frac{dx_i\,dy_i}{\pi}\exp[&-x_i^2-y_i^2\\&-\tau_{Ri}'x_i^2+\tau_{Ri}'y_i^2-2\tau_{Ii}'x_iy_i].\end{aligned}$$

The two-dimensional integrals are precisely the ones evaluated in the one-dimensional case. Using Eq. (21), we obtain

$$\mathrm{Integral}=\prod_{i=1}^{s}[1-|\tau_i'|^2]^{-1/2}.$$

The last result can be further simplified with the aid of Eqs. (64b) and (77). Since

$$1-\tau\tau^\dagger=\tilde{Q}(1-\tau'\tau'^*)Q=\sigma\sigma^\dagger,$$

we have

$$(1-|\tau_i'|^2)^{-1/2}=L_i$$

where

$$L^{-2}=Q\sigma\sigma^\dagger\tilde{Q} \tag{79}$$

is a diagonal matrix with positive elments L_i^{-2}. Thus,

$$\mathrm{Integral}=\prod_{i=1}^{s}L_i=[\det(\sigma\sigma^\dagger)]^{-1/2}, \tag{80}$$

where the last equality follows from Eq. (79). Collecting the results and using the fact that the phase of the new vacuum relative to the old vacuum is arbitrary, we have

$$\begin{aligned}{}_a\langle\boldsymbol{\alpha}|\mathbf{0}\rangle_b&=[\det(\sigma\sigma^\dagger)]^{1/4}\\&\quad\times\exp[-\tfrac{1}{2}\sum_j|\alpha_j|^2-\tfrac{1}{2}\sum_{ij}\tau_{ij}\alpha_i^*\alpha_j^*].\end{aligned} \tag{81}$$

The bracket ${}_a\langle\mathbf{0}|\boldsymbol{\beta}\rangle_b={}_b\langle\boldsymbol{\beta}|\mathbf{0}\rangle_a^*$ [Eq. (74)], can now be obtained from Eq. (81) by invoking the inverse transformation (59). Thus, replacing λ by $\lambda^\dagger$ and μ by $-\tilde{\mu}$, we obtain

$$\begin{aligned}{}_a\langle\mathbf{0}|\boldsymbol{\beta}\rangle_b&=[\det(\sigma\sigma^\dagger)]^{1/4}\\&\quad\times\exp(-\tfrac{1}{2}\sum_j|\boldsymbol{\beta}_j|^2-\tfrac{1}{2}\sum_{ij}\hat{\tau}_{ij}^*\beta_i\beta_j)\end{aligned} \tag{82}$$

where

$$\hat{\tau}=(\lambda^\dagger)^{-1}\tilde{\mu}=\sigma^\dagger\tau\tilde{\sigma}^{-1}. \tag{83}$$

In deriving Eq. (82), the relation

$$\det(\sigma^\dagger\sigma)=\det(\sigma\sigma^\dagger)$$

has been used. Substituting the result for ${}_a\langle\mathbf{0}|\boldsymbol{\beta}\rangle_b$ in Eq. (73) and using Eq. (70), we finally secure

$$\begin{aligned}{}_a\langle\boldsymbol{\alpha}|\boldsymbol{\beta}\rangle_b&=[\det(\sigma\sigma^\dagger)]^{1/4}\exp[-\tfrac{1}{2}\sum_j|\alpha_j|^2\\&\quad-\tfrac{1}{2}\sum_j|\boldsymbol{\beta}_j|^2+\sum_{ij}\sigma_{ij}\alpha_i^*\boldsymbol{\beta}_j\\&\quad-\tfrac{1}{2}\sum_{ij}\tau_{ij}\alpha_i^*\alpha_j^*+\tfrac{1}{2}\sum_{ij}\hat{\tau}_{ij}^*\beta_i\beta_j].\end{aligned} \tag{84}$$

The general expression for the transformation brackets in the occupation number representation is too cumbersome to evaluate. However, following the steps leading to Eq. (25), the quadratures for some special transformation, e.g., the two-dimensional Bogolyubov—Valatin transformation,[1,6] can be easily performed.

APPENDIX A

In this appendix we solve for the inverse transformation

$$A^{-1}=\begin{pmatrix}\alpha & \beta\\ \beta^* & \alpha^*\end{pmatrix}.$$

Starting with Eq. (58) we have

$$\beta=-\lambda^{-1}\mu\alpha^*. \tag{A1}$$

Inserting this expression into Eq. (57), we secure

$$\alpha^{-1}=\lambda-\mu\lambda^{*-1}\mu^*. \tag{A2}$$

Using (A1) again in Eq. (54) we find, with the aid of Eq. (52),

$$\alpha\alpha^\dagger\lambda^{-1}\mu=\lambda^{-1}\mu\alpha^*\tilde{\alpha}. \tag{A3}$$

Employing the same expression in Eq. (53), we obtain

$$\alpha\alpha^\dagger-\lambda^{-1}\mu\alpha^*\tilde{\alpha}\mu^\dagger(\lambda^\dagger)^{-1}=1.$$

Thus, with the aid of Eqs. (A3) and (51) we find

$$\alpha\alpha^\dagger = \lambda^\dagger\lambda. \tag{A4}$$

Let

$$\alpha = \lambda^\dagger\gamma \tag{A5}$$

be the solution of (A4), that is, define $\gamma = (\lambda^\dagger)^{-1}\alpha$. Then

$$\gamma\gamma^\dagger = (\lambda^\dagger)^{-1}\alpha\alpha^\dagger\lambda^{-1} = 1.$$

Since γ is regular, it follows that $\gamma^{-1} = \gamma^\dagger$. Substituting $\alpha = \lambda^\dagger\gamma$ in Eq. (A2), we have

$$\alpha^{-1} = \gamma^\dagger(\lambda^\dagger)^{-1} = \lambda - \mu\lambda^{*-1}\mu^* = \lambda - \mu\mu^\dagger(\lambda^\dagger)^{-1},$$

where the last equality follows from Eq. (52). Hence, by Eq. (51),

$$\gamma^\dagger = \lambda\lambda^\dagger - \mu\mu^\dagger = 1$$

and

$$\alpha = \lambda^\dagger. \tag{A6}$$

Finally, using $\alpha = \lambda^\dagger$ in (A1), we have

$$\beta = -\lambda^{-1}\mu\tilde{\lambda} = -\lambda^{-1}\lambda\tilde{\mu} = -\tilde{\mu} \tag{A7}$$

where the second equality follows from Eq. (52).

APPENDIX B

Let σ and τ satisfy Eqs. (64). We want to show that these matrices are necessarily of the form (66) with τ' and C as specified in the main text. Since $S \equiv \sigma\sigma^\dagger$ is a real symmetric matrix, there exists a real orthogonal matrix Q such that

$$QS\tilde{Q} = C, \tag{B1}$$

where C is a diagonal matrix with elements C_i. The matrix S is regular, hence, $C_i > 0$. Moreover, since $S = 1 - \tau\tau^\dagger$, $C_i \leq 1$. Thus, the diagonal elements C_i satisfy

$$0 < C_i \leq 1. \tag{B2}$$

Now

$$\sigma\sigma^\dagger = S = \tilde{Q}CQ = \tilde{Q}C^{1/2}C^{1/2}Q = (\tilde{Q}C^{1/2})(\tilde{Q}C^{1/2})^\dagger.$$

Let $\sigma = \tilde{Q}C^{1/2}U$ be a solution of the last equation, that is, define

$$U = C^{-1/2}Q\sigma.$$

Then

$$UU^\dagger = C^{-1/2}Q\sigma\sigma^\dagger\tilde{Q}C^{-1/2} = C^{-1/2}CC^{-1/2} = 1,$$

where the second equality follows from (B1). Since U is regular, $U^{-1} = U^\dagger$. Thus,

$$\sigma = \tilde{Q}C^{1/2}U \tag{B3}$$

where U is unitary. Now define

$$\tau' = Q\tau\tilde{Q}. \tag{B4}$$

Since τ is symmetric [Eq. (64a)], τ' is also symmetric. Moreover, since $[\tau, S] = 0$ [Eq. (64c)], we have $[\tau', C] = 0$. Thus τ' consists of square matrices along the main diagonal in accordance with the multiplicity of the C's. Finally, since $\tau\tau^\dagger = 1 - S$ [Eq. (64b)] we have

$$\tau'\tau'^* = 1 - C. \tag{B5}$$

[1] N. N. Bogolyubov, J. Phys. USSR **11**, 23 (1947); Zh. Eksp. Teor. Fiz. **34**, 58 (1958) [Sov. Phys. JETP **7**, 41 (1958)]; Nuovo Cimento 7, 794 (1958); J. G. Valatin, Nuovo Cimento 7, 843 (1958).
[2] M. Baranger, Phys. Rev. **122**, 992 (1961); Phys. Rev. **130**, 1244 (1963).
[3] K. Tanabe, J. Math. Phys. **14**, 618 (1973).
[4] M. A. Rashid, J. Math. Phys. **16**, 378 (1975).
[5] A. Keleman, Z. Physik A **274**, 109 (1975).
[6] A. A. Raduta *et al.*, J. Math. Phys. **18**, 648 (1977).
[7] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[8] A. Messiah, *Quantum Mechanics* (Interscience, New York, 1961), Vol. 1.

Diagonalization methods for the general bilinear Hamiltonian of an assembly of bosons

Constantino Tsallis

Centro Brasileiro de Pesquisas Físicas/CNPq, Rio de Janeiro, Brazil
(Received 27 June 1977)

The problem of the exact diagonalization of the Hamiltonian of an assembly of N bilinearly interacting bosons is discussed in what concerns the eigenvalues as well as for the expression of the new boson operators in terms of the old ones. The method is equivalent to the standard equation-of-motion approach, nevertheless sensibly more concise. Three sets of operational rules are indicated, and their use is exhibited in some examples. In some particular cases of practical importance (for example, when all the coefficients of the Hamiltonian are real), the research of the eigenvalues has been compacted as much as possible.

1. INTRODUCTION

It is well known from long date (at least from the date of Bogolyubov's paper[1] on superfluidity in 1947), that the Hamiltonian of an assembly of N bilinearly interacting bosons (or fermions) is susceptible of exact diagonalization, in terms of new noninteracting bosons (or fermions). The standard method used to perform such a diagonalization is the so-called "equation-of-motion approach," proposed by Bogolyubov and Tyablikov[2-4] in the years 1947-49 and by Bohm and Pines[5] in 1953. This approach is formally presented (see for example Refs. 6 and 7) and discussed[8,9] by several authors. It is equally useful for fermion problems[10-13] (see Refs. 14 and 15 for superconductivity) and boson problems[1,16-18] (phonon–phonon,[16] photon–optical phonon,[16] magnon–magnon,[16,17] phonon–pseudomagnon[18] interactions, etc.). Because of the wideness of the applications of this diagonalization problem, we thought it was worthwhile trying to put it in compact operational rules, and this is the purpose of the present work. However, only the boson case is extensively examined, as in the fermion case, the canonical transformation between old and new fermions is given by a unitary matrix with no further complications. This is not so for the boson case, where the canonical transformation is governed by a matrix related to a not necessarily positive metric, a fact which introduces a certain amount of "pathology" in the case.

In Sec. 2 we present the Hamiltonian we are going to deal with; in Sec. 3 the basic ideas of the diagonalization appear, which lead to the three sets of operational rules of Sec. 6; in Secs. 4 and 5 a particular canonical transformation and the treatment of particular Hamiltonians respectively appear; we conclude in Sec. 7 with a practical comparison between the three diagonalizing methods exposed in this paper; finally the Appendix treats the cases $N=1, 2, 3$ ($N=1$ corresponds to the historical form of Bogolyubov's transformation).

2. HAMILTONIAN

Let us consider an assembly of N bilinearly interacting bosons, which might be particles or quasiparticles. The most general[19] quadratic Hamiltonian (which doesn't need to conserve the number of bosons) might be written as follows:

$$H=\sum_{i=1}^{N}\sum_{j=1}^{N}\left(2\omega_{ij}b_i^{\dagger}b_j+\nu_{ij}^{1}b_i^{\dagger}b_j^{\dagger}+\nu_{ij}^{2}\,b_ib_j\right), \tag{1}$$

where the factor 2 has been introduced for future use; ω_{ij}, ν_{ij}^{1}, and ν_{ij}^{2} are complex numbers, and the creation and annihilation operators satisfy

$$[b_i, b_j]=[b_i^{\dagger}, b_j^{\dagger}]=0 \quad \forall\ (i,j), \tag{2a}$$

$$[b_i, b_j^{\dagger}]=\delta_{ij}\equiv\text{Kroenecker's delta}\ \forall\ (i,j). \tag{2b}$$

Our final purpose is of course to present this Hamiltonian in the form

$$H=\sum_{j=1}^{N}2\Omega_jB_j^{\dagger}B_j, \tag{3}$$

where Ω_j should be known real positive functions of the previous parameters, and the new boson operators are known linear combinations of the old ones.

Let us use the notation ω, ν^1, and ν^2 for denoting the matrix $\{\omega_{ij}\}$, $\{\nu_{ij}^{1}\}$, and $\{\nu_{ij}^{2}\}$, respectively. Because of commutation rules (2a) we may always consider ν^1 and ν^2 as symmetric matrices. Furthermore, hermiticity of H implies hermiticity of ω as well as $\nu^{2*}=\nu^1\equiv\nu$, where * denotes the complex conjugate. Hence (1) may be rewritten as follows:

$$H=\sum_{i,j}\left\{\omega_{ij}b_i^{\dagger}b_j+\omega_{ij}^{*}b_ib_j^{\dagger}+\nu_{ij}b_i^{\dagger}b_j^{\dagger}+\nu_{ij}^{*}b_{ij}\right\}, \tag{1'}$$

where $\omega=\omega^{\dagger}$ and $\nu=\nu^{T}$ (+ and T denote the adjoint and the transposed matrix respectively.) Let us now introduce the nomenclature

$$|b\rangle\equiv\begin{bmatrix}b_1\\ \cdot\\ \cdot\\ \cdot\\ b_N\\ b_1^{\dagger}\\ \cdot\\ \cdot\\ \cdot\\ b_N^{\dagger}\end{bmatrix},\qquad \langle b|\equiv|b\rangle^{\dagger}=(b_1^{\dagger},\ldots,b_N^{\dagger},b_1,\ldots,b_N),$$

$$H\equiv\left[\begin{array}{c|c}\omega & \nu\\ \hline \nu^{*} & \omega^{*}\end{array}\right]\qquad (H^{\dagger}=H).$$

 0022-2488/78/1901-0277$1.00

We note that if H conserves the number of b bosons,[20] then $\nu = 0$. The Hamiltonian (1′) and the commutation rules (2a), (2b) may be written as follows:

$$H = \langle b | H | b \rangle, \tag{1″}$$

$$|b\rangle\langle b| - (|b^*\rangle\langle b^*|)^T = J \equiv \begin{bmatrix} 1_N & 0_N \\ 0_N & -1_N \end{bmatrix}, \tag{2″}$$

where $|\cdots\rangle\langle\cdots|$ means the matrix direct product, and 1_N and 0_N denote the $N\times N$ unity and zero matrix respectively. We note that $|b^*\rangle \neq |b\rangle^*$.

3. DIAGONALIZING METHOD

Let us first of all state a basic property: The Hamiltonian given by (1′) will be diagonal in b's operators (this is to say $\nu = 0$ and ω is diagonal) if and only if

$$[H, b_i] = -2\omega_{ii} b_i \quad \forall\ i.$$

The proof is straightforward once we have remarked that in general

$$[H, b_i] = -2\sum_j \{\omega_{ij} b_j + \nu_{ij} b_j^*\} \quad \forall\ i.$$

This is the property we shall use to find the new boson operators B's which put H into diagonal form, this is to say

$$H = \langle B | H_D | B \rangle, \tag{3′}$$

where

$$H_D \equiv \begin{bmatrix} \Omega & 0_N \\ 0_N & \Omega \end{bmatrix}$$

and

$$\Omega \equiv \begin{bmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & \Omega_N \end{bmatrix}.$$

To perform the diagonalization let us propose

$$\langle B| = \langle b| T \quad \text{and} \quad |B\rangle = T^*|b\rangle,$$

where T is a $2N\times 2N$ matrix to be found. In order to have that B_j^* be the adjoint of B_j, T must have a particular form:

$$T = \begin{bmatrix} T_1 & T_2 \\ T_2^* & T_1^* \end{bmatrix}. \tag{4}$$

As we want the B's to be boson operators, they must also satisfy the commutation rules

$$|B\rangle\langle B| - (|B^*\rangle\langle B^*|)^T = J \tag{5}$$

which implies (once we have remarked that $|B^*\rangle = T^*|b^*\rangle$) that

$$T^* J T J = 1_{2N} \quad \text{hence} \quad T^{-1} = J T^* J. \tag{5′}$$

We see as a corollary that the modulus of the determinant of T equals one. It is also easily verified that the ensemble of the matrix T satisfying (4) and (5′) constitutes a Lie group (in general non-Abelian). Relation (5′) may be written

$$T_1^* T_1 - T_2^T T_2^* = 1_N, \tag{5″a}$$

$$T_2^* T_1 - T_1^T T_2^* = 0_N. \tag{5″b}$$

To be sure that H is diagonal we impose

$$[H, B_j] = -2\Omega_j B_j \quad \forall\ j,$$

$$[H, B_j^*] = 2\Omega_j B_j^* \quad \forall\ j$$

or more compactly

$$[H, |B\rangle] = -2JH_D|B\rangle. \tag{6}$$

Taking into account that

$$[H, |B\rangle] = [H, T^*|b\rangle] = T^*[H, |b\rangle] = -2T^* JH|b\rangle,$$

relation (6) immediately implies that

$$T^* JH = JH_D T^*. \tag{6′}$$

Hence

$$T^{-1} HJT = H_D J, \tag{6″}$$

where we have used relation (5′). And taking into account the particular form of T, (6″) may be rewritten as follows:

$$\begin{bmatrix} P_H & R_H \\ -R_H^* & -P_H^T \end{bmatrix} = \begin{bmatrix} \Omega & 0_N \\ 0_N & -\Omega \end{bmatrix},$$

where

$$P_H \equiv T_1^* \omega T_1 + T_2^T \omega^* T_2^* - T_1^* \nu T_2^* - T_2^T \nu^* T_1 = P_H^*, \tag{6‴a}$$

$$R_H \equiv T_1^* \omega T_2 + T_2^T \omega^* T_1^* - T_1^* \nu T_1^* - T_2^T \nu^* T_2 = R_H^T. \tag{6‴b}$$

Let us formulate in another way what we are doing,

$$\begin{aligned} H &= \langle b | H | b \rangle \\ &= \langle b | (TT^{-1}) H (J(T(JJ)T^{-1})J) | b \rangle \\ &= (\langle b | T)(T^{-1} HJTJ)(JT^{-1} J | b\rangle) \\ &= \langle B | H_D | B \rangle, \end{aligned}$$

where we have used relation (5′) in the last step.

Before going on, a few words about a frequent particular case, namely when $\nu = 0$. In this (and only this) case the solution is given by $T_2 = 0_N$, and we have to deal with a standard $N\times N$ diagonalization problem,

$$T_1^* \omega T_1 = \Omega \quad \text{with } T_1^* T_1 = 1_N.$$

Let us now turn back to the general situation. The secular equation of our diagonalization problem is given by

$$\det(HJ - \Omega_j 1_{2N}) \equiv N\text{th degree polynomial in } \Omega_j^2 = 0 \quad \forall\ j, \tag{7}$$

where the fact that only even powers of Ω_j appear, will soon become clear. So our problem will be practically solved if we find a matrix T which simultaneously diagonalizes the matrix HJ and satisfies restrictions (5″a), (5″b). The discussion of the existence and unicity of such a matrix T is beyond the scope of this paper. However let us point out a very suggestive fact:

The number of unknown real quantities is *exactly* the same as the number of real relations between them.[21] We have indeed $(4N^2+N)$ real unknown quantities: $2N^2$ for the complex matrix T_1, $2N^2$ for the complex matrix T_2, and N for the real diagonal matrix Ω. On the other hand, we have $(4N^2+N)$ real equations to solve: N^2 for (5″a) (notice that the concerned matrix is Hermitian), $N(N-1)$ for (5″b) (notice that the concerned matrix is antisymmetric), N^2 for (6‴a) (notice that the concerned matrix is Hermitian), $N(N+1)$ for (6‴b) (notice that the concerned matrix is symmetric), and finally N for (7).

Let us now prove that in the secular equation (7), only even powers of Ω_j appear. Relation (6′) may be rewritten as follows:

$$HJT = TH_DJ$$

or, more explicitly,

$$\left[\begin{array}{c|c} \omega & -\nu \\ \hline \nu^* & -\omega^* \end{array}\right]\left[\begin{array}{c|c} T_1 & T_2 \\ \hline T_2^* & T_1^* \end{array}\right]=\left[\begin{array}{c|c} T_1 & T_2 \\ \hline T_2^* & T_1^* \end{array}\right]\left[\begin{array}{c|c} \Omega & 0_N \\ \hline 0_N & -\Omega \end{array}\right],$$

in other words, the jth column of the left half of T is nothing but the eigenvector associated to the jth eigenvalue of Ω (namely Ω_j), while the jth column of the right half of T constitutes the eigenvector associated to $(-\Omega_j)$. Then we see that the eigenvectors associated to Ω_j and to $(-\Omega_j)$ are intimately related, and that the secular equation (7) contains only powers of Ω_j^2.

Let us assume we found a particular solution[22] (noted $\bar{T}$) of Eq. (6‴b). If we now write

$$T \equiv \bar{T}S \quad \text{with } S^{-1}=JS^*J,$$

relation (6″) may be rewritten as follows:

$$S^{-1}(\bar{T}^{-1}HJ\bar{T})S = S^{-1}\left[\begin{array}{c|c} Q_H & 0_N \\ \hline 0_N & -Q_H^T \end{array}\right]S=\left[\begin{array}{c|c} \Omega & 0_N \\ \hline 0_N & -\Omega \end{array}\right],$$

where

$$Q_H \equiv \bar{T}_1^\dagger\omega\bar{T}_1+\bar{T}_2^T\omega^*\bar{T}_2^* - \bar{T}_1^\dagger\nu\bar{T}_2^* - \bar{T}_2^T\nu^*\bar{T}_1 = Q_H^\dagger.$$

The solution S may be written as follows,

$$S=\left[\begin{array}{c|c} S_1 & 0_N \\ \hline 0_N & S_1^* \end{array}\right]. \tag{8}$$

Therefore,

$$S_1^{-1}Q_HS_1=\Omega$$

with

$$S_1^{-1}=S_1^\dagger.$$

In this way our problem, as in the case $\nu=0$, has been reduced to a standard diagonalization problem of the $N\times X$ hermitic matrix Q_H. The matrix T will be given by

$$T_1=\bar{T}_1S_1 \quad \text{and} \quad T_2=\bar{T}_2S_1^*.$$

In all usual[23] physical Hamiltonians, we want the $\{\Omega_j\}$ to be real (and positive) numbers, therefore

$$\det(H)=\prod_{j=1}^{N}\Omega_j^2>0. \tag{9}$$

We can also see that

$$F\equiv(HJ)^2=\left[\begin{array}{c|c} F_1 & F_2 \\ \hline F_2^* & F_1^* \end{array}\right]$$

where

$$F_1\equiv\omega^2-\nu\nu^*=F_1^\dagger,$$

$$F_2\equiv\nu\omega^*-\omega\nu=-F_2^T.$$

If a matrix T diagonalizes HJ necessarily it also diagonalizes F (*the opposite is not true*[24]), therefore

$$T^{-1}FT=(H_DJ)^2=\left[\begin{array}{c|c} \Omega^2 & 0_N \\ \hline 0_N & \Omega^2 \end{array}\right]. \tag{10}$$

It follows then that all diagonal elements of F_1 are positive, this is to say

$$\sum_{j=1}^{N}\{|\omega_{ij}|^2-|\nu_{ij}|^2\}>0 \quad \forall\ i. \tag{11}$$

It is clear that conditions (9) and (11) are necessary but in general not sufficient.

The general form (4) for T leads to

$$T^{-1}FT=\left[\begin{array}{c|c} P_F & R_F \\ \hline -R_F^* & P_F^T \end{array}\right]$$

with

$$P_F\equiv T_1^\dagger F_1T_1-T_2^TF_1^*T_2^*+T_1^\dagger F_2T_2^*-T_2^TF_2^*T_1=P_F^\dagger,$$

$$R_F\equiv T_1^\dagger F_2T_1^*-T_2^TF_2^*T_2+T_1^\dagger F_1T_2-T_2^TF_1^*T_1^*=-R_F^T,$$

and relation (10) may be rewritten as follows:

$$P_F=\Omega^2 \tag{10′a}$$

$$R_F=0_N \tag{10′b}$$

If we assume we found a particular solution[25] (noted $\bar{\bar{T}}$) of Eq. (10′b), the matrix T may be written as

$$T=\bar{\bar{T}}S$$

with S given by (8) and satisfying

$$S_1^{-1}Q_FS_1=\Omega^2,$$

$$S_1^{-1}=S_1^\dagger,$$

where

$$Q_F\equiv\bar{\bar{T}}_1^\dagger F_1\bar{\bar{T}}_1-\bar{\bar{T}}_2^TF_1^*\bar{\bar{T}}_2^*+\bar{\bar{T}}_1^\dagger F_2\bar{\bar{T}}_2^*-\bar{\bar{T}}_2^TF_2^*\bar{\bar{T}}_1=Q_F^\dagger.$$

Two immediate corollaries are

$$Q_F=Q_H^2$$

and

$$\det(F)=[\det(Q_F)]^2=[\det(\Omega)]^4=\prod_{j=1}^{N}\Omega_j^4.$$

The preliminary research of $\bar{\bar{T}}$ might be of practical importance: $\bar{\bar{T}}^{-1}HJ\bar{\bar{T}}$ might not be diagonal, but it is expected to be much easier to diagonalize than HJ.

4. PARTICULAR SOLUTION $\bar{\bar{T}}$

In this section let us discuss a general way to construct a particular solution $\bar{\bar{T}}$ of Eq. (10′b). We shall first of all treat the general case $N=2$. Let us use the notation

$$F_1 \equiv \begin{bmatrix} f_{11} & f_{12} \\ f_{12}^* & f_{22} \end{bmatrix} \text{ and } F_2 \equiv \begin{bmatrix} 0 & \tilde{f}_{12}\exp(i\varphi_{12}) \\ -\tilde{f}_{12}\exp(i\varphi_{12}) & 0 \end{bmatrix},$$

where $f_{11}, f_{22}, \tilde{f}_{12}$, and φ_{12} are real numbers, and let us propose the following form for $\bar{\bar{T}}$,

$$\bar{\bar{T}}_1 = \cosh\psi_{12}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \bar{\bar{T}}_2 = \exp(i\chi_{12})\sinh\psi_{12}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

with ψ_{12} and χ_{12} also being real numbers. We verify that restriction (5′) is automatically satisfied, therefore it is enough to impose (10′b), which leads to the solution

$$\chi_{12} = \varphi_{12} \text{ and } \tanh 2\psi_{12} = \frac{2\tilde{f}_{12}}{f_{22} - f_{11}}.$$

We see however that our proposal for $\bar{\bar{T}}$ is not satisfactory if $f_{11} = f_{22}$, so let us make another one for this particular situation,

$$\bar{\bar{T}}_1 = \cosh\psi_{12}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \bar{\bar{T}}_2 = \exp(i\chi_{12})\sinh\psi_{12}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

where ψ_{12} and χ_{12} are again real numbers. Restriction (5′) is again satisfied and relation (10′b) leads to the solution

$$\chi_{12} = \varphi_{12} \text{ and } \tanh 2\psi_{12} = \frac{2\tilde{f}_{12}}{f_{12} + f_{12}^*}.$$

We see that we again have troubles in the particular case of f_{12} being a pure imaginary number (let us note $f_{12} = if_{12}''$) simultaneously with $f_{11} = f_{22}$. In order to be complete let us treat this case by making a new proposal,

$$\bar{\bar{T}}_1 = \cosh\psi_{12}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\bar{\bar{T}}_2 = i\exp(i\chi_{12})\sinh\psi_{12}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where ψ_{12} and χ_{12} are once more real numbers. Once more restriction (5′) is satisfied and (10′b) leads to the solution

$$\chi_{12} = \varphi_{12}$$

and

$$\tanh 2\psi_{12} = \frac{\tilde{f}_{12}}{f_{12}''}.$$

Let us finally say that if f_{12}'' vanishes also, then $\tilde{f}_{12}$ must vanish, otherwise Ω_j^2 should become a complex number.[26] So we may choose $\bar{\bar{T}} = 1_4$, as F will be diagonal by hypothesis.

We are able now to make a proposal for $\bar{\bar{T}}$ for any value of N.

Let us use the notation

$$F_1 \equiv \begin{vmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{12}^* & f_{22} & \cdots & f_{2N} \\ \cdots & & & \cdots \\ \cdots & & & \cdots \\ f_{1N}^* & f_{2N}^* & \cdots & f_{NN} \end{vmatrix}$$

and

$$F_2 \equiv \begin{bmatrix} 0 & \tilde{f}_{12}\exp(i\varphi_{12}) & \cdots & \tilde{f}_{1N}\exp(i\varphi_{1N}) \\ -\tilde{f}_{12}\exp(i\varphi_{12}) & 0 & \cdots & \tilde{f}_{2N}\exp(i\varphi_{2N}) \\ \cdots & & & \cdots \\ \cdots & & & \cdots \\ -\tilde{f}_{1N}\exp(i\varphi_{1N}) & -\tilde{f}_{2N}\exp(i\varphi_{2N}) & \cdots & 0 \end{bmatrix},$$

where $\{f_{ij}\}$, $\{\tilde{f}_{ij}\}$, and $\{\varphi_{ij}\}$ are real numbers. Our proposal will be

$$\bar{\bar{T}} = \bar{\bar{T}}^{12}\bar{\bar{T}}^{13}\cdots\bar{\bar{T}}^{1N}\bar{\bar{T}}^{23}\bar{\bar{T}}^{24}\cdots\bar{\bar{T}}^{2N}\cdots\bar{\bar{T}}^{N-1,N}, \tag{12}$$

where we have $\frac{1}{2}N(N-1)$ factors and each $\bar{\bar{T}}^{ij}$ is a $2N \times 2N$ matrix depending on two real numbers (namely ψ_{ij} and χ_{ij}) and its form is given by

$$\bar{\bar{T}}_1^{ij} = \begin{bmatrix} 1 & & & & & & & & \\ & 1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & \ddots & & & & & \\ & & & & \cosh\psi_{ij} & & 0 & & \\ & & & & & \ddots & & & \\ & & & & & & \ddots & & \\ & & & & 0 & & & \cosh\psi_{ij} & \\ & & & & & & & & \ddots \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 & 0 \\ & & & & & & & & & & 0 & 1 \end{bmatrix} \tag{13}$$

(ith column, jth column)

and

$$\bar{\bar{T}}_2^{ij} = \exp(i\chi_{ij})\,\sinh\psi_{ij}$$

$$\times \begin{bmatrix} 0 & 0 & \cdots & & & & & & & \\ 0 & 0 & \cdots & & & & & & & \\ \vdots & & & & & & & & & \\ \vdots & & & & & & & & & \\ \vdots & & & & & & & & & \\ & & & \tau_{11} & \cdots & \tau_{12} & & & & \\ & & & \vdots & & \vdots & & & & \\ & & & \vdots & & \vdots & & & & \\ & & & \vdots & & \vdots & & & & \\ & & & \tau_{12} & \cdots & \tau_{22} & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & \ddots & & \\ & & & & & & & & \ddots & \\ & & & & & & & & 0 & 0 \\ & & & & & & & & 0 & 0 \end{bmatrix} \qquad (14)$$

(ith column, jth column)

where the central matrix τ is given by

$$\tau_{11} = \tau_{22} = 0 \quad \text{and} \quad \tau_{12} = \exp(i\chi_{ij})\sinh\psi_{ij} \qquad (14a)$$

if $f_{ii} \neq f_{jj}$, hence

$$\sum_{k=1}^{N}\{|\omega_{ik}|^2 - |\omega_{jk}|^2 - |\nu_{ik}|^2 + |\nu_{jk}|^2\} \neq 0,$$

$$\tau_{12} = 0 \quad \text{and} \quad \tau_{11} = -\tau_{22} = \exp(i\chi_{ij})\sinh\psi_{ij} \qquad (14b)$$

if $f_{ii} = f_{jj}$ and f_{ij} is not a pure imaginary number, hence

$$\sum_{k=1}^{N}\{\omega_{ik}\omega_{kj} + \omega_{ki}\omega_{jk} - \nu_{ik}\nu_{kj}^* - \nu_{jk}\nu_{ki}^*\} \neq 0,$$

$$\tau_{12} = 0 \quad \text{and} \quad \tau_{11} = \tau_{22} = i\exp(i\chi_{ij})\sinh\psi_{ij} \qquad (14c)$$

if $f_{ii} = f_{jj}$ and f_{ij} is a pure imaginary number.

It is easy to see that in all cases, $\bar{\bar{T}}$ automatically satisfies restriction (5′). On the other hand, relation (10′b) leads to $N(N-1)$ real equations which in principle enable us to find the $N(N-1)$ unknown quantities $\{\psi_{ij}\}$ and $\{\chi_{ij}\}$.

5. PARTICULAR CASES

We intend to expose here a few particular situations which deserve attention because of their practical importance. Let us express the matrix T which diagonalizes HJ in the form

$$T = LX$$

with

$$L \equiv \frac{1}{\sqrt{2}}\begin{bmatrix} 1_N & -1_N \\ 1_N & 1_N \end{bmatrix}.$$

Therefore,

$$T^{-1}HJT = X^{-1}M_L X$$

where

$$M_L \equiv \frac{1}{2}\begin{bmatrix} (\omega-\omega^*)+(\nu-\nu^*) & (\omega+\omega^*)-(\nu+\nu^*) \\ (\omega+\omega^*)+(\nu+\nu^*) & (\omega-\omega^*)-(\nu-\nu^*) \end{bmatrix}.$$

It follows that *in what concerns the eigenvalues* our diagonalization is equivalent to a standard diagonalization of M_L, whose secular equation is given by

$$\det(M_L - \lambda 1_{2N}) = 0. \qquad (15)$$

Now, if $\omega + \omega^* = \pm(\nu + \nu^*)$ this leads to

$$\det[(\omega - \omega^* + \nu - \nu^*)^2 - 4\lambda^2 1_N] = 0,$$

where we have used that the determinant of the matrix is invariant through transposition. Our work is now simplified as we have to deal with an $N\times N$ matrix.

Another relatively simple situation occurs when $\omega = \omega^*$ and $\nu = \nu^*$, as in this case Eq. (15) leads to

$$\begin{vmatrix} \lambda 1_N & -\omega+\nu \\ -\omega-\nu & \lambda 1_N \end{vmatrix} = 0.$$

A simplified method for calculating such a determinant is given in Appendix D.

Let us now turn to another situation. We shall now express T as follows,

$$T = KY,$$

with

$$K \equiv \frac{1}{\sqrt{2}}\begin{bmatrix} 1_N & -i1_N \\ -i1_N & 1_N \end{bmatrix}.$$

Therefore,

$$T^{-1}HJT = Y^{-1}M_K Y$$

with

$$M_K \equiv \frac{1}{2}\begin{bmatrix} (\omega-\omega^*)-i(\nu+\nu^*) & (\nu^*-\nu)+i(\omega+\omega^*) \\ (\nu^*-\nu)-i(\omega+\omega^*) & (\omega-\omega^*)+i(\nu+\nu^*) \end{bmatrix}.$$

As before, and *only* for the research of the eigenvalues, we may use the secular equation

$$\det(M_k - \lambda 1_{2N}) = 0. \qquad (16)$$

In the case $\nu^* - \nu = \pm i(\omega + \omega^*)$ this leads to

$$\det\{[(\omega-\omega^*)-i(\nu+\nu^*)]^2 - 4\lambda^2 1_N\} = 0$$

which again is a simpler problem.

Finally, if $\omega - \omega^* = \nu + \nu^* = 0_N$, Eq. (16) leads to

$$\begin{vmatrix} \lambda 1_N & \nu - i\omega \\ \nu + i\omega & \lambda 1_N \end{vmatrix} = 0.$$

This determinant can be calculated with the method indicated in Appendix D. Let us recall that the cases $\omega - \omega^* = \nu \pm \nu^* = 0_N$ are very frequent in physics.

6. METHODS

From the ideas developed in previous sections, operational methods emerge, which are exposed here. In all of them we must, first of all, present the Hamiltonian to be diagonalized, into the form

$$H = \langle b | H | b \rangle$$

which defines the $2N \times 2N$ matrix

$$H = \begin{bmatrix} \omega & \nu \\ \nu^* & \omega^* \end{bmatrix}.$$

The problem will be considered completely solved if we attain the knowledge (as functions of ω and ν) of the N real eigenvalues $\{\Omega_j\}$ (which define the diagonalized Hamiltonian H_D), and of the $2N \times 2N$ complex matrix T (which defines the new boson operators $\langle B | = \langle b | T$ in terms of the old ones). We recall that T has the form

$$T = \begin{bmatrix} T_1 & T_2 \\ T_2^* & T_1^* \end{bmatrix}$$

which also gives the N first eigenvalues $\{\vec{\mathbf{T}}_j\}$ by the $N \times 2N$ matrical relation

$$\begin{bmatrix} T_1 \\ T_2^* \end{bmatrix} \equiv \left[\vec{\mathbf{T}}_1\right]\left[\vec{\mathbf{T}}_2\right]\cdots\left[\vec{\mathbf{T}}_N\right],$$

where

$$\vec{\mathbf{T}}_j \equiv \begin{bmatrix} t_j^1 \\ \cdot \\ \cdot \\ \cdot \\ t_j^N \\ t_j^{N+1} \\ \cdot \\ \cdot \\ \cdot \\ t_j^{2N} \end{bmatrix}.$$

So the knowledge of T implies the knowledge of $4N^2$ real numbers (only $2N^2$ if T is real). We recall that

$$J \equiv \begin{bmatrix} 1_N & 0_N \\ 0_N & -1_N \end{bmatrix}.$$

Method I:

(1) Find the roots of the secular equation

$$\det(HJ - \lambda 1_{2N}) = N\text{th degree polynomial in } \lambda^2 = 0.$$

Then $\Omega_j = |\lambda_j| \quad (j = 1, 2, \ldots, N)$.

(2) Write, *for each value of* j, the set of $4N$ real equations (only $2N$ if T is real)

$$[(HJ - \Omega_j 1_{2N})\vec{\mathbf{T}}_j]_k = 0 \quad (k = 1, 2, \ldots, 2N),$$

where by $[\cdots]_k$ we are noting the kth component of the vector. Then eliminate an arbitrary one between them and replace it by the real one

$$\sum_{k=1}^{N} |t_j^k|^2 - \sum_{k=N+1}^{2N} |t_j^k|^2 = 1. \tag{17}$$

We have in this way a set of $4N$ independent real equations (only $2N$ if T is real) which in principle leads to the knowledge of the $2N$ complex numbers $\{t_j^k\}$ associated to the chosen value of j. An example of how to use this method is given in Appendix A.

(2′) Alternative possibility for step (2): Find a particular solution $\bar{T}$ of the equations

$$(\bar{T}_1^\dagger \omega \bar{T}_2 + \bar{T}_2^T \omega^* \bar{T}_1^* - \bar{T}_1^\dagger \nu \bar{T}_1^* - \bar{T}_2^T \nu^* \bar{T}_2)_{ij} = 0 \quad (i \geq j),$$

where by $(\cdots)_{ij}$ we note the ijth element of the matrix, and where the norm relation (17) must also be satisfied.

(3′) Calculate the matrix

$$Q_H = \bar{T}_1^\dagger \omega \bar{T}_1 + \bar{T}_2^T \omega^* \bar{T}_2^* - \bar{T}_1^\dagger \nu \bar{T}_2^* - \bar{T}_2^T \nu^* \bar{T}_1$$

and solve the standard diagonalization problem

$$(S_1^{-1} Q_H S_1)_{ij} = \Omega_j \delta_{ij}$$

with $S_1^{-1} = S_1^\dagger$ and $\delta_{ij} \equiv$ Kroenecker's delta.

(4′) T is given by

$$T_1 = \bar{T}_1 S_1 \quad \text{and} \quad T_2 = \bar{T}_2 S_1^*.$$

Method II:

(1) The same as step (1) of Method I.

(2) Calculate the matrix

$$F_1 \equiv \begin{bmatrix} f_{11} & \cdots & f_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ f_{1N}^* & \cdots & f_{NN} \end{bmatrix} = \omega^2 - \nu\nu^*,$$

$$F_2 \equiv \begin{bmatrix} 0 & \tilde{f}_{12}\exp(i\varphi_{12}) & \cdots & \tilde{f}_{1N}\exp(i\varphi_{1N}) \\ -\tilde{f}_{12}\exp(i\varphi_{12}) & 0 & \cdots & \tilde{f}_{2N}\exp(i\varphi_{2N}) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ -\tilde{f}_{1N}\exp(i\varphi_{1N}) & -\tilde{f}_{2N}\exp(i\varphi_{2N}) & \cdots & 0 \end{bmatrix}$$

$$= \nu\omega^* - \omega\nu.$$

(3) Write, for each one of the $\frac{1}{2}N(N-1)$ values of (ij) $[ij = 12, 13, \ldots, 1N, 23, 24, \ldots, 2N, \ldots, (N-1)N]$, the $2N \times 2N$ matrix

$$\bar{\bar{T}}^{ij} \equiv \begin{bmatrix} \bar{\bar{T}}_1^{ij} & \bar{\bar{T}}_2^{ij} \\ (\bar{\bar{T}}_2^{ij})^* & (\bar{\bar{T}}_1^{ij})^* \end{bmatrix},$$

where

$\bar{\bar{T}}_1^{ij}$ is given by expression (13),

and

$\bar{\bar{T}}_2^{ij}$ is given:

by expressions (14) and (14a) if $f_{ii} \neq f_{jj}$;

by expressions (14) and (14b) if $f_{ii} = f_{jj}$ and $\mathcal{R}(f_{ij}) \neq 0$;

by expressions (14) and (14c) if $f_{ii} = f_{jj}$ and $\mathcal{R}(f_{ij}) = 0$.

(4) Calculate the $2N \times 2N$ matrix

$$\bar{\bar{T}} = \bar{\bar{T}}^{12}\bar{\bar{T}}^{13}\ldots\bar{\bar{T}}^{1N}\bar{\bar{T}}^{23}\bar{\bar{T}}^{24}\ldots\bar{\bar{T}}^{2N}\ldots\bar{\bar{T}}^{N-1,N}$$

which will now be expressed in terms of the $N(N-1)$ real numbers $\{\psi_{ij}\}$ and $\{\chi_{ij}\}$. Then present $\bar{\bar{T}}$ in the form

$$\bar{\bar{T}} \equiv \begin{bmatrix} \bar{\bar{T}}_1 & \bar{\bar{T}}_2 \\ \bar{\bar{T}}_2^* & \bar{\bar{T}}_1^* \end{bmatrix}$$

which leads to the knowledge of $\bar{\bar{T}}_1$ and $\bar{\bar{T}}_2$ separately.

(5) Determine $\{\psi_{ij}\}$ and $\{\chi_{ij}\}$ by solving the $N(N-1)$ real equations given by the matrical relation

$$\bar{\bar{T}}_1^\dagger F_2 \bar{\bar{T}}_1^* - \bar{\bar{T}}_2^T F_2^* \bar{\bar{T}}_2 + \bar{\bar{T}}_1^\dagger F_1 \bar{\bar{T}}_2 - \bar{\bar{T}}_2^T F_1^* \bar{\bar{T}}_1^* = 0_N .$$

Substitute the solutions in the expression of $\bar{\bar{T}}$ obtained in step (4), which will now be a function of $\{f_{ij}\}$, $\{\tilde{f}_{ij}\}$, and $\{\varphi_{ij}\}$.

(6) Calculate the $N \times N$ matrix

$$Q_F = \bar{\bar{T}}_1^\dagger F_1 \bar{\bar{T}}_1 - \bar{\bar{T}}_2^T F_1^* \bar{\bar{T}}_2^* + \bar{\bar{T}}_1^\dagger F_2 \bar{\bar{T}}_2^* - \bar{\bar{T}}_2^T F_2^* \bar{\bar{T}}_1 .$$

(7) Proceed to a standard diagonalization of the hermitic matrix Q_F by a unitary matrix S_1 presented in the following form:

$$S_1 \equiv (\vec{\mathbf{S}}_1)(\vec{\mathbf{S}}_2)\cdots(\vec{\mathbf{S}}_N).$$

To perform this, write, *for each value of* j, the set of $2N$ real equations

$$[(Q_F - \Omega_j^2 1_N)\vec{\mathbf{S}}_j]_k = 0 \quad (k = 1, 2, \ldots, N).$$

Then eliminate an arbitrary one between them and replace it by the real equation

$$\|\vec{\mathbf{S}}_j\| = 1.$$

The solution of this set of $2N$ real equations (only N if S_1 is real) gives the vector $\vec{\mathbf{S}}_j$.

(8) Calculate the matrix

$$T_1' = \bar{\bar{T}}_1 S_1 \quad \text{and} \quad T_2' = \bar{\bar{T}}_2 S_1^*$$

and then

$$T' = \begin{bmatrix} T_1' & T_2' \\ (T_2')^* & (T_1')^* \end{bmatrix}.$$

(9) Calculate the matrix

$$H' = (T')^{-1} H J T' J$$

and enter in step (2) or step (2′) of Method I.

An example of use of this method is given in Appendix B.

Method III: This method is applicable only for the research of the eigenvalues $\{\Omega_j\}$ and only for some particular cases:

1st case: $\omega + \omega^* = \pm(\nu + \nu^*)$

Find the roots of the secular equations

$$\det[(\omega - \omega^* + \nu - \nu^*)^2 - \mu 1_N] = 0,$$

then

$$\Omega_j = \tfrac{1}{2}\sqrt{\mu_j} \quad (j = 1, 2, \ldots, N).$$

2nd case: $\omega + \omega^* = \pm i(\nu - \nu^*)$

Find the roots of the secular equation

$$\det\{[(\omega - \omega^*) - i(\nu + \nu^*)]^2 - \mu 1_N\} = 0,$$

then

$$\Omega_j = \tfrac{1}{2}\sqrt{\mu_j} \quad (j = 1, 2, \ldots, N).$$

3rd case: $\omega - \omega^* = \nu - \nu^* = 0_N$

Find the roots of the equation

$$\sum_{n=0}^{N} (-1)^n C_n \mu^{N-n} = 0,$$

where C_n is given in Appendix D with

$$A \equiv -\omega + \nu \quad \text{and} \quad B \equiv -\omega - \nu$$

and then

$$\Omega_j = \sqrt{\mu_j} \quad (j = 1, 2, \ldots, N).$$

4th case: $\omega - \omega^* = \nu + \nu^* = 0_N$

Find the roots of the equation

$$\sum_{n=0}^{N} (-1)^n C_n \mu^{N-n} = 0,$$

where C_n is given in Appendix D with

$$A \equiv \nu - i\omega \quad \text{and} \quad B \equiv \nu + i\omega$$

and then

$$\Omega_j = \sqrt{\mu_j} \quad (j = 1, 2, \ldots, N).$$

Examples of the use of this method are given in Appendix C.

7. CONCLUSION

Let us conclude by saying that the exposed method for diagonalizing any Hamiltonian of N bilinearly interacting bosons is absolutely equivalent to the so-called "equation-of-motion approach." However systematic exploitation of the peculiar boson properties had led to a concise mathematical formulation which allows for the establishment of operational rules. We have only spoken of N bosons; nevertheless the method is equally applicable to N families (or branches) of bosons, by simple identification of the boson operators ($b_1 \equiv b_q$, $b_2 \equiv b_{-q}$, etc.) as it was done, for example, in Refs. 1, 16, 17, and 18.

Finally let us compare the different methods presented in this paper. Method I [steps (1) and (2)] should be considered the most standard way of performing the diagonalization, however if the matrix H is rather complicated (low symmetry, no zeros), the more delayed procedure indicated in Method I [steps (1), (2′), (3′), and (4′)] could be preferable. Furthermore, if H is very complicated, the highly delayed procedure indicated in Method II could be worth while. If we are interested only in the eigenvalues (as it is frequently the case in statistical mechanics), there is no doubt that Method III should be adopted if we are faced with one of its four cases; if not, the problem will be solved by Method I [step (1)].

ACKNOWLEDGMENTS

We acknowledge with pleasure early and fruitful criticism from J. Tavernier as well as useful discussions with R. Tabensky, J.W. Furtado Valle, R. Feijóo, and R. Lobo.

APPENDIX A

Let us treat, by Method I, the cases $N=1$ and $N=2$. For $N=1$ we have $\omega \in \mathbb{R}$ and $\nu = |\nu| \exp(i\varphi) \in \mathbb{C}$. The secular equation is

$$\begin{vmatrix} \omega - \lambda & -|\nu|\exp(i\varphi) \\ |\nu|\exp(-i\varphi) & -\omega - \lambda \end{vmatrix} = 0$$

hence $\lambda = \pm(\omega^2 - |\nu|^2)^{1/2}$ hence $\Omega = (\omega^2 - |\nu|^2)^{1/2}$. We see that it must be $\omega > |\nu|$. Let us propose

$$\bar{T}_1 = \cosh\psi \quad \text{and} \quad \bar{T}_2 = \exp(i\chi)\sinh\psi.$$

Therefore [performing step (2′) of Method I],

$$\chi = \varphi \quad \text{and} \quad \tanh 2\psi = |\nu|/\omega,$$

hence

$$\bar{T}_1 = \frac{1}{\sqrt{2}} \frac{[\omega + (\omega^2 - |\nu|^2)^{1/2}]^{1/2}}{(\omega^2 - |\nu|^2)^{1/4}}$$

and

$$\bar{T}_2 = \frac{\exp(i\varphi)}{\sqrt{2}} \frac{[\omega - (\omega^2 - |\nu|^2)^{1/2}]^{1/2}}{(\omega^2 - |\nu|^2)^{1/4}}.$$

We may then verify that $Q_H = (\omega^2 - |\nu|^2)^{1/2}$, as it is natural. In this case, obviously $T = \bar{T}$.

For $N=2$ we shall only find the eigenvalues. The most general situation is given by

$$\omega \equiv \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12}^* & \omega_{22} \end{bmatrix} \text{ and } \nu \equiv \begin{bmatrix} \nu_{11} & \nu_{12} \\ \nu_{12} & \nu_{22} \end{bmatrix}$$

with ω_{11}, ω_{22} being real numbers and the rest being complex. The secular equation is given by

$$\begin{vmatrix} \omega_{11}-\lambda & \omega_{12} & -\nu_{11} & -\nu_{12} \\ \omega_{12}^* & \omega_{22}-\lambda & -\nu_{12} & -\nu_{22} \\ \nu_{11}^* & \nu_{12}^* & -\omega_{11}-\lambda & -\omega_{12}^* \\ \nu_{12}^* & \nu_{22}^* & -\omega_{12} & -\omega_{22}-\lambda \end{vmatrix} = \lambda^4 - C_1\lambda^2 + C_2 = 0,$$

where

$$C_1 \equiv \omega_{11}^2 + \omega_{22}^2 + 2|\omega_{12}|^2 - |\nu_{11}|^2 - |\nu_{22}|^2 - 2|\nu_{12}|^2$$

and

$$\begin{aligned} C_2 \equiv{}& \omega_{11}^2\omega_{22}^2 + |\omega_{12}|^4 - 2\omega_{11}\omega_{22}|\omega_{12}|^2 - \omega_{11}^2|\nu_{22}|^2 - \omega_{22}^2|\nu_{11}|^2 \\ & - 2(\omega_{11}\omega_{22} + |\omega_{12}|^2)|\nu_{12}|^2 + |\nu_{11}|^2|\nu_{22}|^2 + |\nu_{12}|^4 \\ & - 2\mathcal{R}(\omega_{12}^2\nu_{11}^*\nu_{22}) - 2\mathcal{R}(\nu_{11}\nu_{22}\nu_{12}^{*2}) \\ & + 4\omega_{11}\mathcal{R}(\omega_{12}\nu_{12}^*\nu_{22}) + 4\omega_{22}\mathcal{R}(\omega_{12}\nu_{12}\nu_{11}^*). \end{aligned}$$

Therefore,

$$\Omega_{1,2} = +\left[\frac{C_1}{2} \pm \left(\frac{C_1^2}{4} - C_2\right)^{1/2}\right]^{1/2}.$$

We see that it must be

$$C_1 \geq 0 \quad \text{and} \quad C_1^2 \geq 4C_2 \geq 0.$$

The particular case $\nu_{11} = \nu_{22} = 0$ and $\omega_{12} = \nu_{12} \in \mathbb{R}$ appears in Ref. 18. On the other hand, if we assume that $\omega_{22} = \omega_{12} = \nu_{22} = \nu_{12} = 0$, we reobtain the case $N=1$.

APPENDIX B

We shall treat here the case $N=2$ in the particular case $\omega_{11} = \omega_{22} = 1$, $\nu_{11} = \nu_{22} = 0$, and $\omega_{12} = \nu_{12} \in \mathbb{R}$. The eigenvalues have already been obtained in Appendix A:

$$\Omega_{1,2} = \sqrt{1 \pm 2\omega_{12}}.$$

Therefore, it must be $|\omega_{12}| < \frac{1}{2}$. We verify immediately that $F_2 = 0_2$, hence $\bar{\bar{T}} = 1_4$, therefore

$$Q_F = F_1 = \begin{bmatrix} 1 & 2\omega_{12} \\ 2\omega_{12} & 1 \end{bmatrix} \quad \text{and} \quad H' = H.$$

Now we enter into step (2) of Method I. The equations to determine T are

$$[(HJ - \sqrt{1+2\omega_{12}}\,1_4)\,\vec{\mathbf{T}}_1]_k = 0 \quad (k=1,2,3),$$

$$(t_1^1)^2 + (t_1^2)^2 - (t_1^3)^2 - (t_1^4)^2 = 1,$$

$$[(HJ - \sqrt{1-2\omega_{12}}\,1_4)\vec{\mathbf{T}}_2]_{k=0} \quad (k=1,2,3),$$

$$(t_2^1)^2 + (t_2^2)^2 - (t_2^3)^2 - (t_2^4)^2 = 1.$$

The solution (attained through very boring calculations!) is given by

$$T_1 = \begin{bmatrix} t_1^1 & t_2^1 \\ t_1^1 & -t_2^1 \end{bmatrix} \quad \text{and} \quad T_2 = \begin{bmatrix} t_1^3 & t_2^3 \\ t_1^3 & -t_2^3 \end{bmatrix}$$

with

$$t_1^1 = \frac{|\omega_{12}|}{D_1}, \quad t_1^3 = \frac{\omega_{12}}{|\omega_{12}|}\frac{1 + \omega_{12} - \sqrt{1+2\omega_{12}}}{D_1},$$

$$t_2^1 = \frac{|\omega_{12}|}{D_2}, \quad t_2^3 = \frac{\omega_{12}}{|\omega_{12}|}\frac{-1 + \omega_{12} + \sqrt{1-2\omega_{12}}}{D_2},$$

$$D_1 \equiv \sqrt{2}\,[\omega_{12}^2 - (1 + \omega_{12} - \sqrt{1+2\omega_{12}})^2]^{1/2},$$

$$D_2 \equiv \sqrt{2}\,[\omega_{12}^2 - (1 - \omega_{12} - \sqrt{1-2\omega_{12}})^2]^{1/2}.$$

APPENDIX C

We shall treat here, by Method III, the case $N=3$ for ω and ν a real matrix. The secular equation can be written as follows

$$\mu^3 - C_1\mu^2 + C_2\mu - C_3 = 0,$$

where

$$\begin{aligned} C_1 ={}& \omega_{11}^2 + \omega_{22}^2 + \omega_{33}^2 - \nu_{11}^2 - \nu_{22}^2 - \nu_{33}^2 \\ & + 2(\omega_{12}^2 + \omega_{13}^2 + \omega_{23}^2 - \nu_{12}^2 - \nu_{13}^2 - \nu_{23}^2), \end{aligned}$$

$C_2 =$

$$\begin{aligned} & (\omega_{11}\omega_{22} + \nu_{11}\nu_{22} - \omega_{12}^2 - \nu_{12}^2)^2 - (\omega_{11}\nu_{22} + \omega_{22}\nu_{11} - 2\omega_{12}\nu_{12})^2 \\ & + (\omega_{22}\omega_{33} + \nu_{22}\nu_{33} - \omega_{23}^2 - \nu_{23}^2)^2 - (\omega_{22}\nu_{33} + \omega_{33}\nu_{22} - 2\omega_{23}\nu_{23})^2 \\ & + (\omega_{11}\omega_{33} + \nu_{11}\nu_{33} - \omega_{13}^2 - \nu_{13}^2)^2 - (\omega_{11}\nu_{33} + \omega_{33}\nu_{11} - 2\omega_{13}\nu_{13})^2 \\ & + 2[(\omega_{11}\omega_{23} + \nu_{11}\nu_{23} - \omega_{12}\omega_{13} - \nu_{12}\nu_{13})^2 \\ & - (\omega_{11}\nu_{23} + \nu_{11}\omega_{23} - \omega_{12}\nu_{13} - \omega_{13}\nu_{12})^2 \\ & + (\omega_{22}\omega_{13} + \nu_{22}\nu_{13} - \omega_{12}\omega_{23} - \nu_{12}\nu_{23})^2 \\ & - (\omega_{22}\nu_{13} + \nu_{22}\omega_{13} - \omega_{12}\nu_{23} - \omega_{23}\nu_{12})^2 \\ & + (\omega_{33}\omega_{12} + \nu_{33}\nu_{12} - \omega_{13}\omega_{23} - \nu_{13}\nu_{23})^2 \\ & - (\omega_{33}\nu_{12} + \nu_{33}\omega_{12} - \omega_{13}\nu_{23} - \omega_{23}\nu_{13})^2], \end{aligned}$$

$$C_3 = \begin{vmatrix} \omega_{11}-\nu_{11} & \omega_{12}-\nu_{12} & \omega_{13}-\nu_{13} \\ \omega_{12}-\nu_{12} & \omega_{22}-\nu_{22} & \omega_{23}-\nu_{23} \\ \omega_{13}-\nu_{13} & \omega_{23}-\nu_{23} & \omega_{33}-\nu_{33} \end{vmatrix} \times \begin{vmatrix} \omega_{11}+\nu_{11} & \omega_{12}+\nu_{12} & \omega_{13}+\nu_{13} \\ \omega_{12}+\nu_{12} & \omega_{22}+\nu_{22} & \omega_{23}+\nu_{23} \\ \omega_{13}+\nu_{13} & \omega_{23}+\nu_{23} & \omega_{33}+\nu_{33} \end{vmatrix}.$$

The eigenvalues are given by

$$\Omega_j = +\sqrt{\mu_j} \quad (j=1,2,3).$$

If in the present secular equation we take the particular case $\omega_{33}=\nu_{33}=\omega_{13}=\omega_{23}=\nu_{12}=\nu_{23}=0$, we easily verify the consistence with the secular equation obtained in Appendix A for $N=2$.

APPENDIX D

We want to calculate the determinant

$$\Delta \equiv \left| \begin{array}{c|c} \lambda 1_N & A \\ \hline B & \lambda 1_N \end{array} \right|,$$

where

$$A \equiv \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{1N} & \cdots & a_{NN} \end{bmatrix} = A^T,$$

$$B \equiv \begin{bmatrix} b_{11} & \cdots & b_{1N} \\ \vdots & & \vdots \\ b_{1N} & \cdots & b_{NN} \end{bmatrix} = B^T$$

with $\{a_{ij}\}$ and $\{b_{ij}\}$ being complex numbers. A long, but not complicated, inductive process leads to

$$\Delta = \sum_{n=0}^{N} (-1)^n C_n \lambda^{2(N-n)},$$

where

$$C_0 = 1$$

$$C_1 = \sum_{i=1}^{N} \partial_{ii} b_{ii} + 2\sum_{i<j} \partial_{ij} b_{ij},$$

$$C_n = \sum_{\{\text{all minors}\}} (\alpha^{(n)}\beta^{(n)}) \quad \left(\left[\frac{N!}{n!\,(N-n)!} \right]^2 \text{terms} \right),$$

$$C_N = |A|\,|B|,$$

$\alpha^{(n)} \equiv$ determinant of an $n \times n$ minor of matrix A, constructed without touching the positions of the elements a_{ij}.

$\beta^{(n)} \equiv$ determinant of an $n \times n$ minor of matrix B, which is obtained by making $a_{ij} \to b_{ij}$ in $\alpha^{(n)}$.

In order to clarify the use of this method, we present here the results for $N=2$ and $N=3$;

$$\begin{vmatrix} \lambda & 0 & a_{11} & a_{12} \\ 0 & \lambda & a_{12} & a_{22} \\ b_{11} & b_{12} & \lambda & 0 \\ b_{12} & b_{22} & 0 & \lambda \end{vmatrix} = \lambda^4 - (a_{11}b_{11} + a_{22}b_{22} + 2a_{12}b_{12})\lambda^2 + \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{vmatrix},$$

$$\begin{vmatrix} \lambda & 0 & 0 & a_{11} & a_{12} & a_{13} \\ 0 & \lambda & 0 & a_{12} & a_{22} & a_{23} \\ 0 & 0 & \lambda & a_{13} & a_{23} & a_{33} \\ b_{11} & b_{12} & b_{13} & \lambda & 0 & 0 \\ b_{12} & b_{22} & b_{23} & 0 & \lambda & 0 \\ b_{13} & b_{23} & b_{33} & 0 & 0 & \lambda \end{vmatrix} = \lambda^6 - \lambda^4(a_{11}b_{11} + a_{22}b_{22} + a_{33}b_{33} + 2a_{12}b_{12} + 2a_{13}b_{13} + 2a_{23}b_{23})$$

$$+ \lambda^2 \left\{ \begin{vmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{vmatrix} \begin{vmatrix} b_{22} & b_{23} \\ b_{23} & b_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{13} & a_{33} \end{vmatrix} \begin{vmatrix} b_{11} & b_{13} \\ b_{13} & b_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{vmatrix} + 2 \begin{vmatrix} a_{12} & a_{23} \\ a_{13} & a_{33} \end{vmatrix} \begin{vmatrix} b_{12} & b_{23} \\ b_{13} & b_{33} \end{vmatrix} + 2 \begin{vmatrix} a_{12} & a_{22} \\ a_{13} & a_{23} \end{vmatrix} \begin{vmatrix} b_{12} & b_{22} \\ b_{13} & b_{23} \end{vmatrix} + 2 \begin{vmatrix} a_{11} & a_{12} \\ a_{13} & a_{23} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} \\ b_{13} & b_{23} \end{vmatrix} \right\}$$

$$- \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{12} & b_{22} & b_{23} \\ b_{13} & b_{23} & b_{33} \end{vmatrix}.$$

[1] N. N. Bogolyubov, J. Phys. **11**, 23 (1947).
[2] S. V. Tyablikov, dissertation for Candidate's Degree, Physics Faculty, Moscow State University (1947).
[3] N. N. Bogolyubov, *Lectures on Quantum Statistics* (in Ukrainian) (Radyanska Shkola, Kiev, 1949).
[4] N. N. Bogolyubov and S. V. Tyablikov, Zh. Eksp. Teor. Fiz. **19**, 256 (1949).
[5] D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953).
[6] D. Pines, *The Many-Body Problem* (Benjamin, New York, 1962), Chap. 2.
[7] S. V. Tyablikov, *Methods in the Quantum Theory of Magnetism* (Plenum, New York, 1967), Chap. 4.
[8] H. Suhl and N. R. Werthamer, Phys. Rev. **122**, 359 (1961).
[9] K. Sawada and N. Fukuda, Progr. Theor. Phys. (Kyoto) **25**, 653 (1961).
[10] K. Sawada, K. A. Brueckner, N. Fukuda, and R. Brout, Phys. Rev. **108**, 507 (1957).
[11] R. Brout, Phys. Rev. **108**, 515 (1957).
[12] P. W. Anderson, Phys. Rev. **112**, 1900 (1958).
[13] G. Rickayzen, Phys. Rev. **115**, 795 (1959).
[14] N. N. Bogolyubov, Sov. Phys. JETP **34**, 41 (1958).
[15] J. G. Valatin, Nuovo Cimento 7, 843 (1958).
[16] C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963), Chap. 2 (phonon—phonon), Chap. 3 (photon—optical phonon), Chap. 4 (antiferromagnetic magnons spectrum).

[17]A. Herpin, *Théorie du Magnétisme* (Presses Universitaires de France, France, 1968), Chap. 16.
[18]C. Tsallis, J. Phys. (Paris) **33**, 1121 (1972).
[19]Eventual terms linear in boson operators can be easily removed by defining new boson operators related to the old ones by $a = b + \mu$ and $a^\dagger = b^\dagger + \mu^*$, where $\mu \in \mathbb{C}$. Additive constants in Hamiltonians are not going to be explicitly written in this paper, because of their simpleness and quite frequent irrelevance.
[20]In any case $\mathcal{H}$ is going to conserve the number of B bosons, which are to be introduced.
[21]This is not a *sufficient* condition for the existence of the solution; therefore strictly speaking it guarantees nothing beyond a strong suspicion.
[22]Equation (6‴b) of course admits a great number of solutions, from which only one (independent) simultaneously satisfies (6‴a).
[23]Hamiltonians adapted to describe displacive phase transitions might constitute an exception.
[24]The reason is that the eigenvalues $|\Omega_j|$ and $-|\Omega_j|$ lead, in F, to a degenerate bidimensional subspace related to the eigenvalue Ω_j^2. It is clear that $\nu \neq 0_N$ and $\nu\omega^* - \omega\nu = 0_N$ are compatible (an example is given in Appendix B).
[25]Every $\overline{T}$ is also a $\overline{\overline{T}}$, but the opposite is not true.
[26]If $f_{11} = f_{22}$ and $f_{12} = 0$ the secular equation for F leads to the roots $\Omega_{1,2}^2 = f_{11} \pm i\widetilde{f}_{12}$.

Remarks on the asymptotic behavior of the 't Hooft's magnetic monopole

D. S. Chernavskii

The Lebedev Institute of Physics (FIAN), Leninskii Prospect 53, Moscow, USSR

R. Kerner

Département de Mécanique, Université P. & M. Curie, 4 Place Jussieu, Paris, France
(Received 22 March 1977)

We investigate the asymptotic behavior of the 't Hooft's solution at $r\to 0$ and $r\to\infty$. Using the methods of the qualitative theory of ordinary differential systems, we are able to derive the asymptotic form of the solution up to a higher approximation than in 't Hooft's original paper [G. 't Hooft, Nucl. Phys. B **79**, 276 (1974], as well as show that the Prasad–Sommerfield's exact solution can be considered as an analytical limit of the 't Hooft's solution for $\mu^2\to 0$, $\lambda\to 0$. Some remarks on the stability of the monopole close the paper.

1. INTRODUCTION

The aim of this paper is to analyze some questions arising on the classical level in the theory of the stationary Yang–Mills field coupled with the Higgs–Kibble scalar field. The main questions concern the existence, the asymptotic behavior, and the stability of the solutions of the type proposed by 't Hooft.[1] The methods adapted here are used widely in the qualitative theory of the nonlinear ordinary differential equations; to our knowledge these methods have not yet been utilized in the study of 't Hooft's magnetic monopole. The general properties of 't Hooft's or Prasad–Sommerfield solutions are quite well known now; still, the problem of existence and stability remains open.

The notation hereafter closely follows the notation used in Refs. 1 and 2.

The starting point of the theory is given by the following gauge-invariant equations:

$$D^\mu F^a_{\mu\nu} + eC^a_{bc} D_\nu \Phi^b \Phi^c = 0 \tag{1}$$

and

$$D^\mu D_\mu \Phi^a - \mu^2 \Phi^a + \frac{\lambda}{e^2}(\Phi^b \Phi_b)\Phi^a = 0. \tag{2}$$

Here $a,b,c = 1,2,3$; $\mu,\nu = 0,1,2,3$; $C^a_{bc} = \epsilon_{abc}$, $\mu^2 > 0$;

$$F^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + eC^a_{bc} W^b_\mu W^c_\nu \tag{3}$$

and

$$D_\mu \Phi^a = \partial_\mu \Phi^a + eC^a_{bc} W^b_\mu \Phi^c. \tag{4}$$

In accordance with the Higgs–Kibble postulate we define the stable vacuum state by

$$\langle 0 \,|\, \Phi^a \Phi_a \,|\, 0\rangle = F^2 = \text{const}, \tag{5}$$

where

$$\mu^2 = \lambda F^2/e^2. \tag{6}$$

Following 't Hooft, we seek the solution of the special form satisfying Refs. 1, 3, and 4,

$$W^a_0 \equiv 0, \quad \frac{\partial W^a_i}{\partial t} = 0, \quad \frac{\partial \Phi^a}{\partial t} = 0 \quad (i,j = 1,2,3), \tag{7}$$

where

$$W^a_i = \epsilon_{aik} x^k W(r), \quad \Phi^a = x^a Q(r), \quad r = (x^k x_k)^{1/2}. \tag{8}$$

Inserting expression (8) into Eqs. (1) and (2) we obtain the following system:

$$\frac{d^2W}{dr^2} + \frac{4}{r}\frac{dW}{dr} - 3eW^2 - e^2r^2W^3 - eQ^2 - e^2r^2WQ^2 = 0, \tag{9}$$

$$\frac{d^2Q}{dr^2} + \frac{4}{r}\frac{dQ}{dr} - 4eWQ - 2e^2r^2W^2Q + \mu^2Q - \lambda r^2Q^3 = 0. \tag{10}$$

These equations can be quite simplified by the following substitution:

$$er^2W = 1 - K(r), \quad er^2Q = H(r), \tag{11}$$

and then by the change of the variable r into t

$$\begin{aligned} &r = e^t, \quad t = \log r, \\ &r \to 0 \Longleftrightarrow t \to -\infty, \quad r \to \infty \Longleftrightarrow t \to \infty. \end{aligned} \tag{12}$$

Then we get the following equations for K and H:

$$\begin{aligned} &\ddot{K} = \dot{K} + K^3 - K + KH^2, \\ &\ddot{H} = \dot{H} + 2HK^2 - \mu^2 e^{2t} H + \frac{\lambda}{e^2} H^3, \end{aligned} \tag{13}$$

where $\dot{K} = dK/dt$, etc.

System (13) is still nonautonomous (i.e., it contains the variable t) because of the factor e^{2t} in the last equation. We can formally give it an autonomous and first-order form by putting

$$X = e^{2t}, \quad \dot{K} = P, \quad \dot{H} = R. \tag{14}$$

Then we get:

$$\begin{aligned} &\dot{X} = 2X, \quad \dot{K} = P, \quad \dot{H} = R, \\ &\dot{P} = P + K^3 - K + KH^2, \\ &\dot{R} = R + 2HK^2 - \mu^2 XH + \frac{\lambda}{e^2} H^3. \end{aligned} \tag{15}$$

The set of equations (15) can be interpreted as the differential equation of a curve in the phase-space of five variables (X,K,H,P,R), the parameter of the curve being t, $-\infty < t < +\infty$.

2. ASYMPTOTIC BEHAVIOR AT $r \to 0$

The energy of our system is given by the Hamiltonian

 0022-2488/78/1901-0287$1.00

$$H=\frac{4\pi}{e^2}\int_0^\infty\left[dr\left(\frac{dK}{dr}\right)^2+\frac{(K^2+1)^2}{2r^2}+\frac{H^2K^2}{r^2}\right.$$

$$\left.+\left(r\frac{dH}{dr}-H\right)^2\Big/2r^2-\frac{\mu^2}{2}H^2+\frac{\lambda H^4}{4e^2r^2}\right]. \tag{16}$$

The finite energy condition imposed on this expression will give us the acceptable asymptotic behavior of the functions K and H at $r\to 0$ and $r\to\infty$. The integral (16) will not diverge at $r\to 0$ if the following conditions are satisfied:

$$K=1+O(r^\alpha),\quad \alpha>\tfrac{1}{2}\text{ at }r\to 0,$$
$$H=O(r^\beta),\quad \beta>\tfrac{1}{2}\text{ at }r\to 0. \tag{17}$$

Now, when $r\to\infty$, the function H has to approach the value $H\to Fr$, in order to get the vacuum-expectation value at the limit, according to (5). Then the integral (16) will behave at infinity as

$$\int_R^\infty F^2r^2\,dr\quad\text{for}\quad R\to\infty \tag{18}$$

and can be renormalized by subtracting a constant from the Hamiltonian density. Then the difference will remain finite as long as $\lim_{r\to\infty}K(r)=0$.

However, the condition on the behavior of $H(r)$ at $r\to\infty$ can be made more precise. Namely, if we assume

$$H=Fr+C+\tilde{H}(r) \tag{19}$$

with

$$\tilde{H}(r)\to 0\quad\text{as}\quad r\to\infty, \tag{20}$$

then inserting this expression into (16) we will get the cross term of the type

$$\frac{2\pi\mu^2}{e^2}\int_R^\infty FC\,dr\quad(R\text{ large compared to }1/\mu) \tag{21}$$

which can never be removed by subtraction of a constant energy density of the vacuum. Therefore, we ought to have

$$H\underset{r\to\infty}{\to}Fr+\tilde{H}(r) \tag{22}$$

and $C=0$. Of course this argument fails to be valid if also $\mu^2\to 0$, $\lambda\to 0$.

As it has been already pointed out, the solution of system (15) can be regarded as a curve in our phase space. It is convenient to have a look at the singular points of our system and to classify them. The singular points are the points in phase space in which all the right-hand terms of (15) vanish. It is clear that this is possible if

$$X=0,\quad P=0,\quad R=0,$$
$$K(K^2+H^2-1)=0\quad\text{and}\quad H\left(2K^2+\frac{\lambda}{e^2}H^2\right)=0. \tag{23}$$

The singular points and the isoclines (the lines along which the direction of "velocities" is constant) are displayed in Fig. 1. The only singular points are $K=\pm 1$, $H=0$, and $K=0$, $H=0$. Along the circle $K^2+H^2=1$ and the H axis the "velocities" are horizontal; along the K axis they are vertical. We note that this image is valid for $X=0$, i.e., $t=-\infty$ or $r=0$; there can still exist some singular points at $r\to\infty$, which we shall investigate later.

It is clear that the trajectory satisfying conditions (17) at $r\to 0$ has to start at the singular point $K=+1$, $H=0$. In order to see what the asymptotic behavior is of such a solution at $r\to 0$ we have to linearize our system in the vicinity of $K=1$, $H=0$. Then we get (we use the same letters for the infinitesimal variations of K, H, P, etc.):

$$\dot{X}=2X,\quad \dot{K}=P,\quad \dot{H}=R,\quad \dot{P}=P+2K,\quad \dot{R}=R+2H. \tag{24}$$

The characteristic exponents are given by the equation

$$\det\begin{pmatrix}2-\beta & 0 & 0 & 0 & 0\\ 0 & -\beta & 0 & 1 & 0\\ 0 & 0 & -\beta & 0 & 1\\ 0 & 2 & 0 & 1-\beta & 0\\ 0 & 0 & 2 & 0 & 1-\beta\end{pmatrix}=0, \tag{25}$$

which gives the following solutions:

$$\beta_1=\beta_2=\beta_3=2,\quad \beta_4=\beta_5=-1. \tag{26}$$

The point $K=1$, $H=0$ is an unstable *saddle point* (with both positive and negative exponents). The asymptotic solutions compatible with condition (17) are

$$H\approx A_1e^{2t},\quad K\cong 1+A_2e^{2t},\quad\text{and}\quad X=e^{2t}, \tag{27}$$

other solutions being incompatible with the finiteness of energy at $r\to 0$. The corresponding eigenvectors are directed at the angles $\tan\varphi_{\beta=2}=2$ in the RH plane and $\tan\varphi_{\beta=2}=2$ in the PK plane. Of course, the eigenvectors directed at the angle $\tan\varphi_{\beta=-1}=-1$ in the RH or in the PK planes correspond formally to asymptotic solutions, but they have to be *eliminated* because of condition (17).

Let us make this point clear. In general, the solutions of the system (9), (10), or the equivalent one, (15), form a four-parameter continuum. If two boundary conditions are fixed, like $K=1$, $H=0$ at $r=0$, we are still left with a two-parameter continuum. The conditions of type (17) should not, in general, restrain this continuum. This would be the case if all the characteristic values (26) satisfied (17), or if the point $K=1$, $H=0$ were a *nodal* point, i.e., a point through which there pass multiple trajectories under the same angle. However this is not the case, and as we have seen, condition (17) is equivalent to a strict equality, $\beta=2$. Therefore we have at our disposal *one and only one* trajectory starting at $K=1$, $H=0$ and corresponding to a finite energy.

3. ASYMPTOTIC BEHAVIOR AT $r\to\infty$

Now we want to see what happens once our trajectory leaves the KH plane and $r\to\infty$. In order to do this let us assume that $K\to 0$ as $r\to\infty$ and

$$H=Fr+\tilde{H}(r),\quad \tilde{H}(r)\underset{r\to\infty}{\to}0. \tag{28}$$

It is convenient to write down our system as ($C=0$)

$$r^2K''=K^3-K+K(Fr+\tilde{H})^2,$$
$$r^2\tilde{H}''=2K^2(Fr+\tilde{H})-\mu^2(Fr+\tilde{H})r^2+\frac{\lambda}{e^2}(Fr+\tilde{H})^3. \tag{29}$$

As $\mu^2=(\lambda/e^2)F^2$, we get

$$r^2K''=K^3-K+KF^2r^2+2\tilde{H}KFr+K\tilde{H}^2,$$
$$r^2\tilde{H}''=2K^2\tilde{H}+2K^2Fr+2\mu^2r^2\tilde{H}+\frac{3\mu^2}{F}r\tilde{H}^2+\frac{\lambda}{e^2}\tilde{H}^3. \tag{30}$$

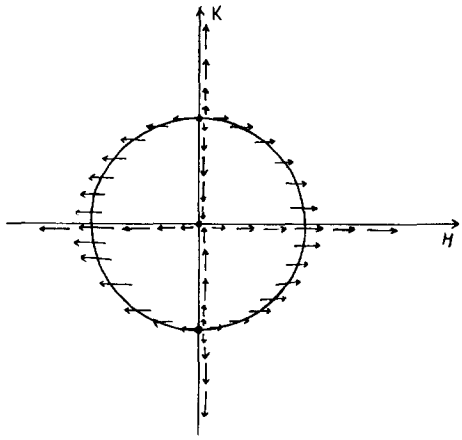

Fig. 1

If we change the variable r into $\xi = r^{-1}$, we get the following system (now $r \to \infty \Leftrightarrow \xi \to 0$):

$$\xi' = -\xi^2, \quad K' = P, \quad \tilde{H}' = R,$$
$$P' = (K^3 - K)\xi^2 + F^2 K + 2FK\tilde{H}\xi + K\tilde{H}^2\xi^2, \tag{31}$$
$$R' = 2K^2\tilde{H}\xi^2 + 2FK^2\xi + 2\mu^2\tilde{H} + \frac{3\mu^2}{F}\tilde{H}^2\xi + \frac{\lambda}{e^2}\tilde{H}^3\xi^2.$$

Clearly there is only one singular point at $r \to \infty$, i. e., $\xi \to 0$. This singular point is given by

$$\xi = 0, \quad P = 0, \quad R = 0, \quad K = 0, \quad \text{and} \quad \tilde{H} = 0. \tag{32}$$

The linearized system in the neighborhood of this point is

$$\xi' = 0, \quad K' = P, \quad \tilde{H}' = R, \quad P' = F^2 K, \quad R' = 2\mu^2\tilde{H}. \tag{33}$$

This system is split up (independent systems for K, P and $\tilde{H}$, R), and the characteristic exponents are

$$\beta_1 = F, \quad \beta_2 = -F, \quad \beta_3 = \mu\sqrt{2}, \quad \beta_4 = -\mu\sqrt{2}. \tag{34}$$

Only the values $\beta_2 = -F$ and $\beta_4 = -\mu\sqrt{2}$ are compatible with the asymptotic conditions (20). Our singular point at $r \to \infty$ is also a *saddle* point. Once more the condition of the finite energy turns out to be equivalent to choosing *one and only one* trajectory out of a four-parameter continuum.

4. REMARKS ON THE UNIQUENESS

Let us consider carefully the number of restrictions following from the above-mentioned behavior of the trajectories.

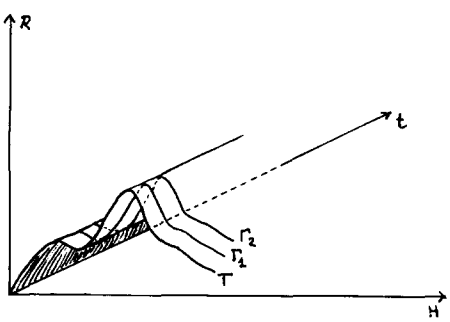

Fig. 3

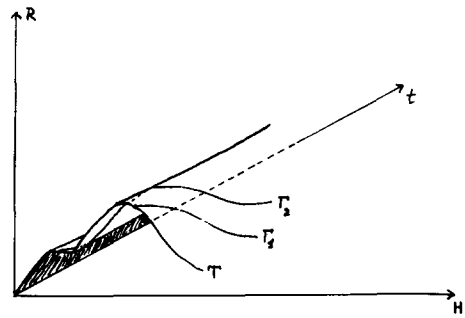

Fig. 2

(a) If our system was autonomous, conditions (17) and (34) would cut out of the parameter-parameter continuum of trajectories *one and only one* at $r = 0$ and *one and only one* trajectory at $r \to \infty$. This would mean that our system is over defined. Indeed, we have only six degrees of freedom: four initial conditions and two free parameters μ^2 and F. For the autonomous systems the change in quantities A_1 and A_2 in (27) do not change the corresponding trajectory, because this can always be reduced to a change $t \to t + \Delta t$ (or $r \to Cr$).

(b) For the nonautonomous systems this is not true in general. The corresponding situation is illustrated in Fig. 2 and Fig. 3. As Fig. 2 shows, for an autonomous system the generalized trajectories in the space containing the t variable all project on the original trajectory T in the phase space. In Fig. 3 and Fig. 4 we see how for a nonautonomous system the generalized trajectories, corresponding to different shifts of t, project onto different trajectories on the phase space.[5]

This means that in our case the changes of A_1, A_2, etc., will really change the trajectories in the phase space. That means, in turn, that our system is not necessarily over defined. The same kind of figures can be drawn at $r \to \infty$. The problem is to find one smooth trajectory starting in the original bundle at $r = 0$ and ending up in the final bundle of trajectories at $r = \infty$.

(c) However, in general, between $r = 0$ and $r = \infty$ both bundles can reduce themselves to the very narrow beams. Such a phenomenon is known under the name of phase autosynchronization in the theory of nonlinear parametric excitation. In such a case a smooth trajectory may not exist (see Fig. 5). That is why our linear approximation is still not conclusive and we have to analyze the situation in more detail.

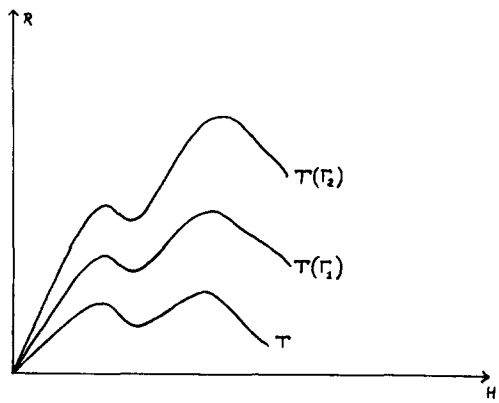

Fig. 4

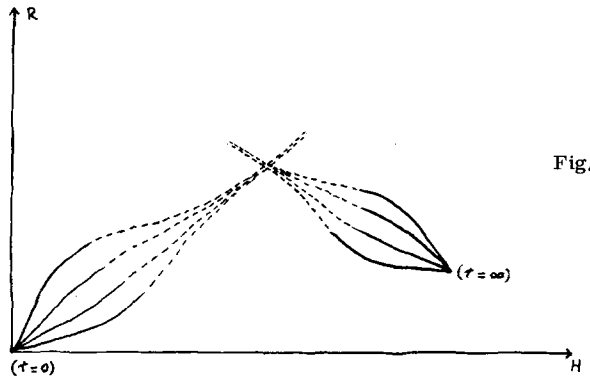

Fig. 5

5. THE PRASAD-SOMMERFIELD LIMIT

We want to see now what happens in the limit $\mu \to 0$, $\lambda \to 0$, where an exact solution is known to exist.

The analysis of the asymptotic behavior at $r \to \infty$ can be continued by taking into account the terms corresponding to the next orders of magnitude. As $\tilde{H} = H - Fr$, let us call

$$\varphi = \tilde{H}/r = H/r - F. \tag{35}$$

Then we get ($\xi = 1/r$)

$$\begin{aligned} K'' &= (K^3 - K)\xi^2 + KF^2 + 2KF\varphi + K\varphi^2, \\ \varphi'' &= 2FK^2\xi^2 + 2\varphi K^2\xi^2 - 2\varphi'\xi + 2\mu^2\varphi + \frac{\lambda}{e^2}\varphi^3 + \frac{3F\lambda}{e^2}\varphi^2. \end{aligned} \tag{36}$$

The approximation we propose now consists in keeping the following terms:

$$K'' = KF^2 + 2KF\varphi, \quad \varphi'' + \frac{2\varphi'}{r} = 2\mu^2\varphi. \tag{37}$$

The second equation gives the Yukawa potential

$$\varphi = A\frac{\exp(\mu\sqrt{2}r)}{r} + B\frac{\exp(-\mu\sqrt{2}r)}{r}. \tag{38}$$

If μ is finite, we can keep only the second term; however, if $\mu \to 0$ (and μr is very small) the function φ tends to

$$\varphi = \frac{A+B}{r} + \mu\sqrt{2}(A-B). \tag{39}$$

In the limit $\mu \to 0$, $A - B$ is arbitrary, and $A + B = D$ is some constant; φ approaches $(A+B)/r = D/r$. At this limit the first equation in (37) becomes

$$K'' = KF^2 + \frac{2KFD}{r} = K\left(F^2 + \frac{2FD}{r}\right). \tag{40}$$

F being just a scaling parameter here, we can put

$$\rho = Fr \tag{41}$$

and write

$$\frac{d^2K}{d\rho^2} = K\left(1 + \frac{2D}{\rho}\right). \tag{42}$$

Now we seek our solution in the form

$$K = \alpha(\rho)e^{\rho} + \beta(\rho)e^{-\rho} \tag{43}$$

with

$$\alpha = \sum_{n=0}^{\infty} a_n\rho^n, \quad \beta = \sum_{n=0}^{\infty} b_n\rho^n. \tag{44}$$

Then we get (e^{ρ} and $e^{-\rho}$ being linearly independent)

$$\alpha'' + 2\alpha' - \frac{2D}{\rho}\beta = 0, \quad \alpha \tag{45}$$

which gives

$$\begin{aligned} &\sum_{n=0}^{\infty} n(n-1)a_n\rho^{n-2} + 2\sum_{n=0}^{\infty}(n-D)a_n\rho^{n-1} = 0, \\ &\sum_{n=0}^{\infty} n(n-1)b_n\rho^{n-2} - 2\sum_{n=0}^{\infty}(n+D)b_n\rho^{n-1} = 0. \end{aligned} \tag{46}$$

The recurrence formulas are then ($a_0 = 0$, $b_0 = 0$)

$$\begin{aligned} a_{n+1} &= (-2)\frac{n-D}{(n+1)n}a_n = \cdots \\ &= (-2)^n\frac{(n-D)(n-D-1)\cdots(1-D)}{(n+1)!n!}a_1, \\ b_{n+1} &= 2\frac{n+D}{(n+1)n}b_n = \cdots \\ &= 2^n\frac{(n+D)(n+D-1)\cdots(1+D)}{(n+1)!n!}b_1, \end{aligned} \tag{47}$$

and we have still the freedom to choose two parameters, a_1 and b_1, at will.

Had we started the same development at another singular point, i.e., $\varphi = H/r + F$, we would get

$$\begin{aligned} a_{n+1} &= (-2)^n\frac{(n+D)(n+D-1)\cdots(1+D)}{(n+1)!n!}a_1, \\ b_{n+1} &= 2^n\frac{(n-D)(n-D-1)\cdots(1-D)}{(n+1)!n!}b_1, \end{aligned} \tag{48}$$

α and β being interchanged. The series are particularly simple if D takes on the positive or negative integer values. Namely, in the first case (47) for $D = m$ = positive integer, α becomes a polynomial and β becomes

$$\beta(\rho) = \sum_{n=0}^{\infty} C_m^{n+m}\frac{1}{2(n+1)}\frac{(2\rho)^{n+1}}{n!}b_1, \tag{49}$$

e.g., for $m = 1$

$$\beta(\rho) = b_1\rho\sum_{n=0}^{\infty}\frac{(2\rho)^n}{n!} = b_1\rho e^{2\rho} \tag{50}$$

and K is badly divergent. If D is a *negative integer*, $D = -m$, then $\beta(\rho)$ reduces to a polynomial of order m, whereas for $\alpha(\rho)$ we get

$$\alpha(\rho) = \sum_{n=0}^{\infty} C_m^{n+m}\frac{1}{2(n+1)}\frac{(-2\rho)^{n+1}}{n!}. \tag{51}$$

E.g., for $D = -1$

$$\beta(\rho) = b_1\rho = b_1Fr \tag{52}$$

and

$$\alpha(\rho) = -a_1\rho e^{-2\rho} = -a_1Fr\exp(-2Fr). \tag{53}$$

Now $K(r)$ is equal to

$$K(r) = Fr(b_1 - a_1)\exp(-Fr) \tag{54}$$

which exactly gives the asymptotic behavior of Prasad–Sommerfield's solution. It seems at this stage that we have lost one degree of freedom (both independent solutions become coincident for $D=-1$), and instead of two independent constants a_1, b_1 there is only one left: $b_1 - a_1$. It remains true for any $D=$ negative integer. However, for an arbitrary negative D we still get two independent solutions. This last statement can be illustrated as follows: Suppose $D=-1+\epsilon$, ϵ very small. Then we get

$$a_{n+1}=(-2)^n\frac{(n+1-\epsilon)(n-\epsilon)\cdots(2-\epsilon)}{(n+1)!\,n!}a_1 \tag{55}$$

and

$$b_{n+1}=2^n\frac{(n-1+\epsilon)(n-2+\epsilon)\cdots(1-\epsilon)\epsilon}{(n+1)!\,n!}b_1. \tag{56}$$

Developing these expressions in the powers of ϵ and keeping only the linear term, we get

$$a_{n+1}=\frac{(-2)^n}{n!}a_1-\epsilon(-2)^n\frac{\sum_{k=1}^{n+1}(1/K)}{2n!\,(n-1)!}a_1, \tag{57}$$

$$b_{n+1}=\epsilon 2^n\frac{1}{2n!\,(n-1)!}b_1. \tag{58}$$

Therefore, we find *two* independent solutions

$$K=Fr(b_1-a_1)\exp(-Fr)+\epsilon[b_1f_2(r)\exp(-Fr) - a_1f_1(r)\exp(Fr)], \tag{59}$$

where

$$f_1(r)=\sum_{n=0}^{\infty}\frac{(-2Fr)^{n+1}\sum_{k=1}^{n+1}(1/K)}{(-4)n!\,(n-1)!} \tag{60}$$

and

$$f_2(r)=\sum_{n=0}^{\infty}\frac{(2Fr)^{n+1}}{4n!\,(n-1)!}\,. \tag{61}$$

All these series are absolutely convergent. The Prasad–Sommerfield solution corresponds to the case $D=-1$. Since series (60) behaves asymptotically as a polynomial, we have to put $a_1=0$, and we are left exactly with two degrees of freedom: D and b_1.

6. STABILITY

The problem of stability can be approached by a method proposed by Derrick,[7] in which the knowledge of the exact solutions is not necessary. We choose the variations of the functions $K(r)$ and $\tilde{H}(r)$ in the form

$$\delta K=\epsilon r\frac{dK}{dr}\,,\quad \delta\tilde{H}(r)=\epsilon r\frac{d\tilde{H}}{dr}\,,\quad \text{with } \epsilon\ll 1. \tag{62}$$

This is equivalent to the transform $K\to K(\chi r)$, $\chi=1+\epsilon$.

The Hamiltonian is then equal to

$$H=\frac{4\pi}{e^2}\int_0^{\infty}\Big[dr\ (K')^2\chi+\frac{(K^2-1)^2}{2r^2}\chi+\frac{F^2K^2}{\chi} +\frac{\tilde{H}^2K^2}{r^2}\chi+\frac{(r\tilde{H}'-\tilde{H})^2}{2r^2}\chi+\frac{\lambda}{e^2}\Big(F^2\frac{\tilde{H}^2}{\chi}+F\frac{\tilde{H}^3}{r}+\frac{\tilde{H}}{r^2}\chi\Big)\Big]. \tag{63}$$

The second variation $\delta^2H/\delta\chi^2$ over χ at $\chi=1$ is equal to

$$\frac{\delta^2H}{\delta\chi^2}\bigg|_{\chi=1}=4\pi\mu^2\int_0^{\infty}dr\Big[\frac{K^2}{\lambda}+\frac{\lambda}{e^4}\tilde{H}^2\Big] \tag{64}$$

and is positive. This means that the solution is stable with respect to the variations of the type in Eq. (62). We point out that in the case $F=0$, $\tilde{H}=0$, any autolocalized solution would be unstable under this kind of variation, and is hence absolutely unstable. In the 't Hooft model the stability follows from the minus sign of the "mass term" of the Higgs–Kibble field.[6]

ACKNOWLEDGMENTS

This work has been carried out during the stay of both authors at the Instituut voor Theoretische Fysica of the Utrecht University. We wish to express our thanks to Professor T. Ruijgrok and Professor M. Veltman, for their marvelous hospitality. The stimulating discussions and remarks due also to Professor J. A. Tjon, Dr. D. Darby, Dr. W. Troost, and Dr. G. Lemoine are gratefully acknowledged.

[1] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).
[2] M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 151 (1975).
[3] R. Kerner, C. R. Acad. Sci. Paris **270**, 1467 (1970).
[4] A. M. Polyakov, Zh. Ekps. Teor. Fiz. Red. **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)]; Yu. S. Tupkin, V. A. Fateev, and A. S. Shvarts, Zh. Eksp. Teor. Fiz. Pis. Red. **21**, 91 (1975) [JETP Lett. **21**, 42 (1975)].
[5] A. Andronov, A. Khaikin, S. Vitt, *The Theory of Oscillations* (Pergamon, New York, 1962).
[6] T. T. Wu and C. N. Yang in H. Mark and S. Fernbach, *Properties of Matter under Unusual Conditions* (Wiley, New York, 1969).
[7] A. Derrick, Math. Phys. 5, 1252 (1964).

WKB approximation for bound states by Heisenberg matrix mechanics

Abraham Klein[a)]

Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19174
(Received 28 April 1977)

The WKB approximation for bound states is derived from Heisenberg matrix mechanics. It is shown that the result can be obtained from the Bohr-Sommerfeld quantization of a suitably chosen solution of the classical equations of motion. It is also derived from the commutation relation and from a quantum form of Hamilton's variational principle.

I. INTRODUCTION

The traditional approach to the WKB approximation has been through the study of the Schrödinger equation in configuration space, giving rise to the apparatus of connection formulas. This approach has perhaps reached its most satisfactory form in the monograph of Fröman and Fröman.[1,2] Quite independently, recent times have witnessed an expanding study of semiclassical approximations by means of path integral methods[3,4]; this study has served as a prime stimulus in the development of quantization procedures for nonlinear systems by Dashen, Hasslacher, and Neveu.[5,6]

Other contributions in this new field have revealed alternative techniques which may provide increased understanding of semiclassical approximation methods. At least two additional distinct approaches may be cited, the method of canonical quantization about a classical solution,[7-9], and the method of Heisenberg matrix mechanics.[10,11] The purpose of this note is to show how Heisenberg's method may be used to give a novel and completely elementary discussion of the WKB approximation which makes contact with the earliest ideas about the quantization of bound, periodic motions.

It is noteworthy that Heisenberg's matrix mechanics has remained almost dormant as a practical tool for most of the history of quantum mechanics. Recent efforts suggest that this has been a matter of taste rather than necessity.[12-14] Even these limited efforts suggest the availability of powerful nonperturbative methods for some strong coupling situations.

In the approach to WKB, it is natural within the Heisenberg framework to define semiclassical as the domain of a large quantum number, $n \gg 1$. We look for an expansion of the form

$$E(n) = E_{\text{class}}(n)\left(1 + \frac{a}{n} + \frac{b}{n^2} + \cdots\right), \tag{1.1}$$

together with a similar expansion for various matrix elements. Here $E_{\text{class}}(n)$ is the result emerging from the Bohr quantization condition; we intend the implication that it can be obtained from a suitable classical solution. Within this framework, the WKB result means that we can rewrite (1.1) as

$$E(n) = E_{\text{class}}(n + \tfrac{1}{2})[1 + (b/n^2)]. \tag{1.2}$$

Viewed thus, it suggests that the full WKB can be derived from a *suitably defined* classical solution.

The proof of (1.2) is given in Sec. III after a careful account of the Bohr—Sommerfeld quantization in Sec. II is used to lay most of the necessary groundwork for our proof. Section IV contains some observations supplemental to the main proof. In Sec. V, we describe several alternative, but equivalent proofs, the most useful for further work utilizing a quantum version of Hamilton's variational principle in classical mechanics.

II. PRELIMINARIES AND REVIEW OF THE BOHR-SOMMERFELD QUANTIZATION

Consider a Hamiltonian for a system with one degree of freedom, x ($\hbar = 1$),

$$H = \tfrac{1}{2}p^2 + V(x), \tag{2.1}$$

$$[x, p] = i. \tag{2.2}$$

We find it convenient to eliminate the operator p from the problem by utilizing the equation

$$p = \dot{x} = -i[x, H]. \tag{2.3}$$

The equations of motion and commutation relations which follow from (2.1) and (2.2) are then taken in the form

$$[[x, H], H] = \frac{\partial V}{\partial x}, \tag{2.4}$$

$$[x, [x, H]] = -1, \tag{2.5}$$

and the energy may be calculated as the expectation value of

$$H = \tfrac{1}{2}[x, H][H, x] + V(x) \tag{2.6}$$

in the bound eigenstate n,

$$\langle n | H | n' \rangle = E(n)\delta_{nn'}. \tag{2.7}$$

We are interested in the limiting value of $E(n)$ for large n. The usual method of obtaining this result is to apply the Bohr—Sommerfeld condition

$$\oint p\, dx = 2\pi n, \tag{2.8}$$

taken over a complete classical period, and p is the solution of

$$E_{\text{class}}(n) \equiv E = \tfrac{1}{2}p^2 + V(x) \tag{2.9}$$

with the appropriate sign. In the remainder of this section we shall describe some not-quite-so-well-known aspects of this procedure which will be helpful in under-

a)Supported in part by the Energy Research and Development Authority through contract AT(E11-1) 3071 Theoretical.

 0022-2488/78/1901-0292$1.00

standing how to carry out the extension from Bohr–Sommerfeld to WKB.

One well-known alternative to (2.8) makes use of the correspondence meaning of the classical frequency, namely

$$\omega \equiv \omega(n) = \frac{dE(n)}{dn}. \tag{2.10}$$

If the classical frequency is known as a function of E, $\omega = \omega(E)$, then (2.10) can be integrated

$$n = \int \frac{dE}{\omega(E)}. \tag{2.11}$$

The equivalence of (2.8) and (2.11) is established by utilizing the formula for the classical period

$$T = \frac{2\pi}{\omega} = \frac{d}{dE} \oint p\, dx \tag{2.12}$$

which follows from the energy integral (2.9). Indeed, combining (2.10) and (2.12), we see that

$$\frac{dn}{dE} = \frac{1}{\omega} = \frac{T}{2\pi} = \frac{1}{2\pi}\frac{d}{dE} \oint p\, dx. \tag{2.13}$$

Although superfluous in practice in the one-dimensional case, it is important in principle for this case and essential also in practice in the multidimensional case to recognize that to find $\omega(E)$ we need an explicit periodic solution of the equation of motion

$$\ddot{x}(t) = -\frac{\partial V(x)}{\partial x} \tag{2.14}$$

in the form $x = x(t, \omega) = x(t + T, \omega)$. We then compute

$$E = E(\omega) = \frac{1}{T}\int_0^T H(\dot{x}(t, \omega), x(t, \omega))\, dt. \tag{2.15}$$

(Though computation of the time average is redundant here, it plays a role later.) The inversion of (2.15) gives $\omega(E)$ to be inserted into (2.11).

So far we have considered only the energy. We can obtain information about matrix elements by re-expressing (2.15) in terms of the Fourier coefficients of $x(t, \omega)$,

$$x(t, \omega) = \sum_{p=-\infty}^{\infty} x_p(\omega) \exp(ip\omega t), \tag{2.16}$$

where we choose the $x_p(\omega) = x_{-p}(\omega)$ and real. If we think of $V(x)$ as a power series in x, it has the Fourier representation

$$V(x(t, \omega)) = \sum_{-\infty}^{\infty} V_p(\omega) \exp(ip\omega t), \tag{2.17}$$

and a similar expansion may be written for $H = \frac{1}{2}\dot{x}^2 + V$. In terms of these definitions, (2.15) is of course equivalent to

$$E(\omega) = H_0(\omega) = \sum_{p=-\infty}^{\infty} \tfrac{1}{2}(p\omega)^2 x_{-p}(\omega)x_p(\omega) + V_0(\omega). \tag{2.18}$$

Progress now consists in comparing (2.18) with the diagonal matrix $\langle n|H|n\rangle$, computed from (2.6) in the large n limit. Let us work with $V(x) = \lambda x^4$ as an illustration. We then have

$$E(n) = \langle n|H|n\rangle$$

$$= \sum_p \tfrac{1}{2}(E_{n+p} - E_n)^2 \langle n|x|n+p\rangle\langle n+p|x|n\rangle$$

$$+ \sum_{pp'p''} \lambda\langle n|x|n+p\rangle\langle n+p|x|n+p'\rangle$$

$$\times\langle n+p'|x|n+p''\rangle\langle n+p''|x|n\rangle. \tag{2.19}$$

To reduce (2.19) to the form (2.18), several points must be considered. First we must restrict the expansion of $(E_{n+p} - E_n)$ to linear terms,

$$E_{n+p} - E_n \approx \frac{dE(n)}{dn} p = \omega(n)p. \tag{2.20}$$

Second, since the $x_p(\omega)$ are real and even in p, we must equate the upward and downward-going transitions, for comparison of classical and quantum expressions leads us to require ($p > 0$)

$$\langle n|x|n+p\rangle\langle n+p|x|n\rangle \equiv x_p x_{-p}, \tag{2.21}$$

$$\langle n|x|n-p\rangle\langle n-p|x|n\rangle \equiv x_{-p}x_p = x_p x_{-p}. \tag{2.22}$$

There are many ways of satisfying (2.21) and (2.22). One of these will correspond to the Bohr–Sommerfeld quantization and another to WKB. In this section we shall consider the following identification, which, as argued below, corresponds to Bohr quantization,

$$x_1(n) = x_1(\omega(n)) = \langle n-1|x|n\rangle, \tag{2.23}$$

$$x_p(n) = x_p(\omega(n)) = \langle n_p - p|x|n_p\rangle, \quad p \neq 1. \tag{2.24}$$

Here n_p is a number which depends on p, which we are *not free to prescribe*. The reason is that, with the reality conditions imposed, the $x_p(\omega)$ in (2.16) are uniquely determined functions of ω and thus the single assignment (2.23) determines $\omega(n)$. Therefore, (2.24) is merely a definition of n_p. It is obvious that n_p is a number differing little from n for sufficiently small p.

All matrix elements (2.19) are now expanded about the values (2.23), (2.24) as a leading approximation. To the leading order the interaction is treated in the cavalier fashion indicated by the approximation

$$\langle n+p|x|n+p'\rangle \approx x_{p-p'} = x_{p'-p}. \tag{2.25}$$

Putting (2.20)–(2.25) together, we have for (2.19)

$$E(n) \approx \sum_p \tfrac{1}{2}(p\omega)^2 x_{-p}(n)x_p(n)$$

$$+ \sum_{pp'p''} \lambda x_p x_{p'-p}x_{p''-p'}x_{-p''}, \tag{2.26}$$

which is precisely of the form (2.18) for the chosen potential energy. The corrections, as will be clear in the next section, are of relative order n^{-1} compared to the term exhibited, for the definition (2.23). Only for the different choice of $x_p(n)$ are the corrections as small as the relative order n^{-2}.

We must now discuss the most crucial point of our considerations: Why does the choice (2.23), (2.24) and no other correspond to the Bohr–Sommerfeld quantization (2.8) or (2.11)? The transition from (2.10) to (2.11), in particular, makes it clear that a constant of integration has been dropped. In particular these formulas give the leading large n dependence for $E(n)$ as a function of n, in general, of the form

$$E(n) \sim n^a, \quad a > 0, \tag{2.27}$$

i.e., the Bohr–Sommerfeld quantization must yield an energy which vanishes when $n = 0$. This means that the proper identification of $x_p(\omega)$ with an approximate matrix element $x_p(n)$, which is an integral part of the correspondence principle procedure, must in this case be such as to yield the stated property $E(0) = 0$.

Of course the quantum expression for $E(0)$,

$$E(0) = \sum_{p>0} \tfrac{1}{2}(E_p - E_0)^2 |\langle 0|x|p\rangle|^2 + \langle 0|V(x)|0\rangle, \tag{2.28}$$

does not vanish, but we are here instructed to replace $\langle 0|x|1\rangle$ by $\langle -1|x|0\rangle$ and $\langle 0|x|p\rangle$, $p \neq 0$ by $\langle 0_p - p|x|0_p\rangle$ according to (2.23) and (2.24). But it is part of the definition of the physical Hilbert space that

$$\langle -1|x|0\rangle = 0. \tag{2.29}$$

This further implies that for $n = 0$ we are dealing with the solution $x(t) = 0$ of Newton's law (assuming that we have chosen this point as minimum of the potential energy) and therefore also the $x_p(\omega(0)) = 0$, $p \neq 1$. Thus $E(0) = 0$ as required.

In understanding the previous discussion it may be helpful to emphasize the distinction between the definition of the Bohr–Sommerfeld approximation and its domain of validity. The approximation is, after all, defined for all n; this definition carries with it the implications attendant upon (2.29). The approximation is valid for n sufficiently large so that the quantum sum (2.26) in which p takes on both positive and negative values may for all numerical purposes be extended to $|p| = \infty$ and therefore identified with the Fourier series (2.18).

We have not yet exhausted our interest in the Bohr–Sommerfeld approximation, but shall return to the discussion in Sec. IV, since we have now laid the framework for the derivation of WKB, to which we therefore proceed directly.

III. PROOF OF WKB

As another statement of the results of the previous section we have shown that if in the classical expression (2.18) for the energy, we replace the classical Fourier coefficients $x_p(\omega)$ by the matrix elements of (2.23) and (2.24), then the resulting expression $E_{\text{class}}(n)$ is the same as obtained from the formula

$$\oint p(x, E_{\text{class}}(n))\, dx = 2\pi n. \tag{3.1}$$

[Actually to the leading order in n we could have been less careful and simply assigned $x_p(n)$ to $\langle n - p|x|n\rangle$, since the error is again $O(n^{-1})$.] In this section we shall show that with the symmetrical identification

$$x_p(n) = x_p(\omega(n)) = \langle n - p + \tfrac{1}{2}p|x|n + \tfrac{1}{2}p\rangle \tag{3.2}$$

$$= \langle n - \tfrac{1}{2}p|x|n + \tfrac{1}{2}p\rangle \tag{3.3}$$

the resulting classical expression for $E(n)$ (and therefore the solution to Newton's equation) reproduces the quantum theory except for errors which are $O(n^{-2})$.

Considering the real value $p = 1$, we see that (3.3) is obtained from the Bohr–Sommerfeld classical solution (2.23), (2.24) by the replacement $n \to n + \frac{1}{2}$, i.e., the formula

$$\oint p(x, E)\, dx = 2\pi(n + \tfrac{1}{2}) \tag{3.4}$$

is derived. We therefore turn to the proof of the assertions made concerning Eqs. (3.2) and (3.3).

We start with as precise a statement as we can make on the definition of semiclassical limit: Let A and B be monomials in x and consider the diagonal element

$$\langle n|AB|n\rangle \approx \sum_{\nu>0}^{\nu_{\max}} \{\langle n|A|n+\nu\rangle\langle n+\nu|B|n\rangle + \langle n|A|n-\nu\rangle\langle n-\nu|B|n\rangle\}. \tag{3.5}$$

We assume that n is both large and strategically located so that: (i) The maximum value $\nu_{\max}$ needed for numerical convergence satisfies $\nu_{\max} \ll n$. (ii) All relevant matrix elements in (3.5) are defined. (iii) Most vitally, the sum is symmetrical with respect to upgoing and downgoing transitions.

Treating n as a continuous variable, we define the symmetrical matrix element ($\bar{n} = n + \frac{1}{2}$)

$$\langle n - \tfrac{1}{2}p|x|n + \tfrac{1}{2}p\rangle \equiv x_p(\bar{n}) = x_{-p}(\bar{n}). \tag{3.6}$$

We then prove the theorem, α an integer,

$$\langle n|x^{\alpha+1}|n\rangle = \sum_{p_1 \cdots p_\alpha} x_{p_1}(\bar{n})\, x_{p_2 - p_1}(\bar{n}) x_{p_3 - p_2}(\bar{n}) \times \cdots x_{-p_\alpha}(\bar{n})\,[1 + O(n^{-2})]. \tag{3.7}$$

As emphasized in the previous section, this expression can be rewritten as a time average if we view $x_p(\bar{n})$ as the Fourier coefficients of a classical dynamical variable $x(t, \bar{n})$ given by the Fourier series

$$x(t, \bar{n}) = \sum_{p=-\infty}^{\infty} x_p(\bar{n}) \exp(ip\omega t). \tag{3.8}$$

We already know that

$$(2\pi/T) = \omega = \omega(n) = dE(n)/dn. \tag{3.9}$$

With the definition (3.8), (3.7) becomes

$$\langle n|x^{\alpha+1}|n\rangle = \frac{1}{T(\bar{n})} \int_0^{T(\bar{n})} dt\,[x(t, \bar{n})]^{\alpha+1} \times [1 + O(n^{-2})], \tag{3.10}$$

We turn to the proof of (3.7). We write

$$\langle n|x^{\alpha+1}|n\rangle = \sum_{p>0} \{\langle n|x|n+p\rangle\langle n+p|x^\alpha|n\rangle + \langle n|x|n-p\rangle\langle n-p|x^\alpha|n\rangle\}. \tag{3.11}$$

Let us expand

$$\langle n|x|n \pm p\rangle = x_{\pm p}(\bar{n}) \pm \tfrac{1}{2}p\partial_{\bar{n}} x_p(\bar{n}) + \cdots, \tag{3.12}$$

and note that a derivative with respect to n implies an additional factor of n^{-1}. For example, then, in the consideration of (3.11) and (3.12), we encounter

$$\sum_{p>0} \tfrac{1}{2} p \partial_n x_p(\bar{n})(\langle n+p|x^\alpha|n\rangle - \langle n-p|x^\alpha|n\rangle)$$
$$\approx -\sum_p \tfrac{1}{2} p \partial_n x_p(\bar{n})[p \partial_n x_p^\alpha(\bar{n})], \tag{3.13}$$

and this is smaller by a factor of $O(n^{-2})$ compared to the leading term of (3.11).

We thus have

$$\begin{aligned}
&\langle n|x^{\alpha+1}|n\rangle \\
&\approx \sum_{p>0}\{x_p(\bar{n})\langle n+p|x^\alpha|n\rangle + x_{-p}(\bar{n})\langle n-p|x^\alpha|n\rangle\} \\
&= \sum_{p>0}\sum_{p_1=-\infty}^{\infty}\{x_p(\bar{n})\langle n+p|x|n+p_1\rangle\langle n+p_1|x^{\alpha-1}|n\rangle \\
&\quad + x_{-p}(\bar{n})\langle n-p|x|n-p_1\rangle\langle n-p_1|x^{\alpha-1}|p\rangle\} \\
&= \sum_{p>0}\sum_{p_1}\{x_p(\bar{n})x_{p_1-p}(\bar{n})\langle n+p_1|x^{\alpha-1}|n\rangle \\
&\quad + x_{-p}(\bar{n})x_{p-p_1}(\bar{n})\langle n-p_1|x^{\alpha-1}|n\rangle\} \\
&\quad \times[1+O(n^{-2})],
\end{aligned} \tag{3.14}$$

the argument being a repetition of (3.13). The continued repetition of this argument results in (3.7). A similar theorem may easily be developed for monomials containing the momentum to an even power.

Insofar as $V(x)$ can be represented as a formal power series in x, (3.7) or (3.10) gives its expectation value to the order required. To calculate the energy we also need the expansion

$$E(n+p) - E(n) = p\omega(n) + \tfrac{1}{2}p^2(d\omega(n)/dn) + \cdots. \tag{3.15}$$

The kinetic energy term is then easily seen to share the accuracy of the potential energy term. Altogether, then, up to terms of relative order n^{-2}, we find

$$\begin{aligned}
\langle n|H|n\rangle &= E_{\text{class}}(\bar{n}) \\
&= \sum_{p>0}[p\omega(\bar{n})]^2 x_p^2(\bar{n}) + V(x(\bar{n}))_0 \\
&= \frac{1}{T(\bar{n})}\int_0^{T(\bar{n})} dt\,[\tfrac{1}{2}\dot{x}^2(t,\bar{n}) + V(x(t,\bar{n}))] \\
&= \frac{1}{T}\int_0^T dt\,H(\dot{x}(t,\bar{n}), x(t,\bar{n})).
\end{aligned} \tag{3.16}$$

According to the introductory remarks of this section, this completes a proof of the WKB bound state formula. Alternative proofs are given in Sec. V. It may be objected that our derivation is missing some cases. For example, it does not cover the "one-sided" potential

$$\begin{aligned}
&V = V(x) \quad \text{(analytic)}, \quad x > 0, \\
&V = \infty, \quad x < 0,
\end{aligned} \tag{3.17}$$

for which the WKB result is known to be

$$\oint p\,dx = (n+\tfrac{3}{4})h. \tag{3.18}$$

We easily obtain this result, however, by reflecting $V(x)$ with respect to the origin and considering the symmetric problem in the full space. Let P and X be the canonical coordinates for the latter problem. Then we write

$$\oint P\,dX = (n'+\tfrac{1}{2})h, \quad n' = 0, 1, \cdots. \tag{3.19}$$

Two remarks transform (3.19) into (3.18), namely

$$\oint P\,dX = 2\oint p\,dx, \tag{3.20}$$

and

$$n' = 2n+1, \quad n = 0, 1, 2, \cdots, \tag{3.21}$$

i.e., the odd solutions of (3.19) are the solutions of (3.18).

The adaptations of WKB to a radial coordinate come out in a standard way and need not be discussed separately. It is also clear that the present approach via the correspondence principle and Fourier series can be extended to systems which are multiply periodic.

IV. ADDITIONAL REMARKS

In the usual discussion of semiclassical quantization, the emphasis is on the energy quantization (3.4). In this paper we have reminded the reader that if one has an explicit classical solution $x(t, \omega)$, one also has information about matrix elements. For example, in the Bohr–Sommerfeld quantization the quantity $x_p(\omega(n))$, which can be computed by substituting $\omega = \omega(n) = [dE(n)/dn]$ after the energy quantization has been carried out, has been identified as the matrix element $\langle n-p|x|n\rangle$. The corresponding identification in the case of WKB was made in the previous section.

Confining ourselves in this section to the Bohr approximation, we remark that there is a well-known method of obtaining the energy and the matrix elements in a joint procedure.

In the discussion of the previous sections, we made no apparent use of the commutation relation (2.5). Treated in the same manner as the equations of motion, the diagonal element yields in leading order the "quantization condition"

$$\frac{d}{dn}\sum_{p>0} 2p^2\omega(n)x_p^2(n) = 1. \tag{4.1}$$

In conjunction with the solution of the classical equations of motion which we utilize in the form

$$x_p = x_p(\omega), \tag{4.2}$$

this equation will determine asymptotic solutions

$$\omega(n) \sim n^\alpha, \quad x_p(n) \sim n^\beta. \tag{4.3}$$

The procedure for solving (2.18) (after quantum transcription) and (4.1) together is a more exact version of the often-quoted approximate argument which determines $E_{\text{class}}(n)$ by minimizing the classical energy $E(p, x)$ subject to the "uncertainty relation"

$$px \sim \tfrac{1}{2}n\hbar. \tag{4.4}$$

Indeed minimization of (2.18) with respect to the x_p for fixed ω, subject to (4.1) is the precise version of this calculation and is once more equivalent to Bohr quantization.

One may wonder finally that the quantization condition (4.1) did not enter explicitly in the previous sections. This conundrum is resolved by noting that (4.1) and the correspondence principle (2.10) are equivalent. This can be seen by differentiating (2.18) with respect to n. We find

$$\omega(n) = dE(n)/dn$$
$$= \sum_{p>0} \frac{d}{dn}[(p\omega(n))^2 x_p^2] + \sum_p \left(\frac{\partial V}{\partial x}\right)_p \frac{\partial x_p}{\partial n}$$
$$= \omega(n)\frac{d}{dn}\sum_p 2p^2\omega(n)x_p(n)^2 = \omega(n). \tag{4.5}$$

The essential, not previously noted ingredient in this derivation is the recognition of the Fourier transform of the equation of motion

$$(p\omega)^2 x_p(\omega) = \left(\frac{\partial V}{\partial x}\right)_p. \tag{4.6}$$

Thus the quantization condition (4.1) which seemed to play no overt role in the previous considerations was present all along in the guise of (2.10).

V. OTHER DERIVATIONS

In the light of the concluding observation of the previous section, one may ask if it is possible to transform the commutator (2.2) or (2.5) more directly into the phase quantization condition. We now show how this can be done.

The first step in the procedure is repetition of the derivation of (4.1), but now expanding all matrix elements about the symmetrically defined matrix element $x_p(\bar{n})$. We again find (4.1) with $n \to \bar{n}$. Integrating, we have

$$S(\bar{n}) \equiv \sum_{p>0} 2p^2\omega(\bar{n})x_p^2(\bar{n}) = n + c, \tag{5.1}$$

where c is a constant of integration. The constant c must be chosen to have the value one-half for the reason explained in Sec. II: The corresponding Bohr–Sommerfeld sum is a sum over the squares of the elements $x_p(n)$ which, as we have explained, vanish when $n = 0$, i.e., $S(n) = n$ with no additive constant; consequently

$$S(\bar{n}) = S(n + \tfrac{1}{2}) = n + \tfrac{1}{2}. \tag{5.2}$$

We may now easily transform (5.2) into the phase integral condition, since

$$S(\bar{n}) = \frac{T(\bar{n})}{2\pi} \sum_{\text{all } p} [p\omega(\bar{n})]^2 x_p(\bar{n})x_{-p}(\bar{n})$$
$$= \frac{1}{2\pi}\int_0^{T(\bar{n})} dt\, \dot{x}^2(t, \bar{n})$$
$$= \frac{1}{2\pi}\oint p\, dx, \tag{5.3}$$

since $p = \dot{x}$.

Though this derivation directly from the commutation relation is more succinct than that given previously, at least some of this terseness comes from taking for granted the matters discussed in detail in Sec. II. It appears moreover that the direct approach through the commutation relation does not readily generalize to problems in field theory, whereas the approach through the energy and the correspondence principle does.

It is therefore of interest to describe yet another approach to the derivation of (4.1) or (5.1) which can be carried through, not only in a field theory context, but can be applied as well to continuum states. We shall describe this method quite briefly here since we plan to develop it in future publications to field theory bound state and scattering problems and to particle scattering and reactions. The essential steps are:

(i) Expand the matrix elements of the Heisenberg equations of motion in powers of n^{-1} to the desired accuracy. In the problem defined by (2.4) we have

$$[E(n \pm p) - E(n)]^2 \langle n | x | n \pm p\rangle = \left\langle n \left| \frac{dV}{dx} \right| n \pm p \right\rangle. \tag{5.4}$$

Using the method described in connection with the derivation of (3.7), we find that both of Eqs. (5.4) lead to the equation

$$D_p(\bar{n}) \equiv \left\{[p\omega(\bar{n})]^2 x_p(\bar{n}) - \left(\frac{dV(x)}{dx}\right)_p\right\} = 0, \tag{5.5}$$

and that corrections to this equation are $O(n^{-2})$. [This can be seen from the fact that each of (5.4) is corrected to $O(n^{-1})$ by the same absolute quantity appearing with opposite sign.] Equation (5.5) is, of course, a Fourier component of the classical equations of motion.

(ii) Now recognize that (5.5) can be derived from the variational principle

$$\delta_{x_{-p}(\bar{n})} L(\bar{n}) = 0, \tag{5.6}$$

$$L(\bar{n}) = \sum_p \tfrac{1}{2}[p\omega(\bar{n})]^2 x_p(\bar{n})x_{-p}(\bar{n}) - (V(x))_0$$
$$= \sum_p [p\omega(\bar{n})]^2 x_p(\bar{n})x_{-p}(\bar{n}) - E(\bar{n}), \tag{5.7}$$

or

$$L(\bar{n}) \equiv C(\bar{n}) - E(\bar{n}), \tag{5.8}$$

where $E(\bar{n})$ is the energy given in (3.16). In this variation $\omega(\bar{n})$ is *held fixed*. Since $\omega(\bar{n})$ is a semiclassical approximation to an energy difference, this constancy can be related to the Rayleigh–Ritz principle.

(iii) The essential final set of steps emerge from the observation that because of (5.6) the derivative of $L(\bar{n})$ with respect to n can be computed in at least two different ways since

$$\frac{dL(\bar{n})}{dn} = \frac{\partial L(\bar{n})}{\partial n}. \tag{5.9}$$

To compute the left-hand side we use the second form of (5.7), namely

$$\frac{dL(\bar{n})}{dn} = \frac{dC(\bar{n})}{dn} - \omega(\bar{n}), \tag{5.10}$$

which doesn't use the variational principle. On the other hand,

$$\frac{\partial L(\bar{n})}{\partial n} = \frac{\partial C(\bar{n})}{\partial n} - \frac{\partial E(\bar{n})}{\partial n} \tag{5.11}$$

in which we differentiate $\omega(\bar{n})$ only. One sees then that the equation

$$1=\frac{1}{\omega(n)}\left(\frac{dC(\bar{n})}{dn}-\frac{\partial C(\bar{n})}{dn}+\frac{\partial E(\bar{n})}{\partial n}\right) \tag{5.12}$$

is precisely (4.1), from which the derivation of WKB proceeds as before.

The reader may be surprised that we have been able to derive a quantization condition from the variational principle, which is basically Hamilton's variational principle in Fourier form. It follows from previous work,[11,14] however, that this variational principle can be derived from a Rayleigh–Ritz principle (for the energy) with constraints which *build in* the commutation relations. It is therefore not that astonishing that this information can be extracted by a suitably chosen variation.

It is perhaps not *at all* surprising that time-dependent versions of the variational principle can also be constructed. We postpone discussion of such principles to a future occasion, however, as they are most useful for the continuum state problem.

[1]N. Fröman and P.O. Fröman, *JWKB Approximation: Contributions to the Theory* (North-Holland, Amsterdam, 1965).
[2]For an interesting application of the methods of Ref. 1, see C. Rosenzweig and J.B. Krieger, J. Math. Phys. **9**, 849 (1965).
[3]See for instance, M.C. Gutwiller, J. Math. Phys. **8**, 1979 (1967); **10**, 1004 (1969); **11**, 1791 (1970); **12**, 343 (1971).
[4]The most extensive compilation of references that we have seen is in S. Levit and U. Smilansky, Report WIS-76/57-Ph, Weizmann Institute of Science, Rehovoth.
[5]R.F. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D. **10**, 1414 (1974).
[6]J.L. Gervais and B. Sakita, Phys. Rev. D **11**, 2943 (1975).
[7]E. Tomboulis, Phys. Rev. D **12**, 1678 (1975).
[8]N. Christ and T.D. Lee, Phys. Rev. D **12**, 1606 (1975).
[9]M. Creutz, Phys. Rev. D **12**, 3126 (1975).
[10]J. Goldstone and R. Jackiw, Phys. Rev. D **11**, 1486 (1975).
[11]A. Klein and F. Krejs, Phys. Rev. D **12**, 3112 (1975); D **13**, 3282, 3295 (1976).
[12]Francis R. Halpern, J. Math. Phys. **14**, 219 (1973); **15**, 733 (1974); **16**, 2469 (1975).
[13]J.H. Henkel and C.A. Uzes, Phys. Rev. D **8**, 4430 (1973); C.A. Uzes and J.H. Henkel, Phys. Rev. D **8**, 3487 (1973); C.A. Uzes, Phys. Rev. D **14**, 3362 (1976).
[14]C.T. Li, A. Klein, and F. Krejs, Phys. Rev. D **12**, 2311 (1975).

Phase space path integrals, without limiting procedure[a)]

Maurice M. Mizrahi

Center for Naval Analyses of The University of Rochester, 1401 Wilson Boulevard, Arlington, Virginia 22209
(Received 12 May 1977)

This paper defines path integrals in phase space without using a time-division approach followed by a limiting process, thereby generalizing a similar procedure used in configuration space. This is useful since the path integral approach cannot always be formulated in configuration space (e.g., when the Hamiltonian is arbitrary) but can always be formulated in phase space. The most general Gaussian measure, absorbing the quadratic portion of the functional to be integrated, is constructed, and large classes of path integrals are evaluated with respect to it. Applications are given to the perturbation expansion and the semiclassical (WKB) expansion for arbitrary Hamiltonians.

I. INTRODUCTION

The quantum-mechanical propagator $\langle q_b, t_b | q_a, t_a\rangle$, or probability amplitude that a particle at position q_a at time t_a will be at position q_b at time t_b, can be written as a phase space path integral:

$$\langle q_b, t_b | q_a, t_a\rangle \equiv K \doteq \int_{\rho} \left[\frac{dp\,dq}{2\pi\hbar}\right] \exp\frac{i}{\hbar}\int_{t_a}^{t_b} \{p(t)\dot{q}(t) - H[p(t), q(t), t]\}\,dt, \tag{1}$$

where H is the classical Hamiltonian of the system, or a quantity suitably related to it,[1] and ρ is the space of phase space paths (q,p) satisfying $q(t_a)=q_a$ and $q(t_b)=q_b$, with $p(t)$ unrestricted. The integral is usually defined by the time division procedure,[2] i.e.,

$$K \equiv \lim_{\substack{m\to\infty \\ \max_j (t_{j+1}-t_j)\to 0}} \int_{\mathbb{R}^{2m+1}} \frac{dq_1\cdots dq_m\,dp_0\,dp_1\cdots dp_m}{(2\pi\hbar)^{m+1}} \times \exp\frac{i}{\hbar}\left\{\sum_{j=0}^{m}\left[p_j\left(\frac{q_{j+1}-q_j}{t_{j+1}-t_j}\right) - H\left(p_j, \frac{q_{j+1}+q_j}{2}, t_j\right)\right](t_{j+1}-t_j)\right\} \tag{2}$$

with $q_{m+1}\equiv q_b$, $q_0 = q_a$, $t_0 \equiv t_a$, and $t_{m+1}\equiv t_b$. We work in one dimension to simplify the discussion. The results can be easily generalized to n dimensions. The convention of summing over repeated indices is used throughout.

The limiting process makes the scheme difficult to use for computational purposes, not to mention questions of mathematical legality. It has been done away with in the case of the Wiener functional integral,[3] and the method was later extended to Feynman path integrals in the configuration space of quantum mechanics.[4–9] The new formalism rests on defining what plays the role of a measure in path space by its Fourier transform, which is a simple closed-form expression. This is all that is needed to completely define the object and reduce many path integrals to ordinary definite integrals.[10] We do not treat the mathematical problems here, as we are mainly concerned with developing computational techniques.

The purpose of this paper is to extend this limiting-procedure-free formalism to phase space. This is necessary not only from the point of view of completeness, but also because phase space path integrals are more basic than configuration space path integrals. Indeed, the latter provide a solution to the Schrödinger equation only for Hamiltonian operators quadratic in the momenta, whereas the former apply to arbitrary Hamiltonian operators,[6,11] a useful generalization. Further, the principle of democracy, which states that all paths contribute with the same amplitude (although with different phases) is realized only in phase space: There is no infinite "normalization factor" preceding the exponential in (1), only the natural volume element in phase space, $dp\,dq/h$. Finally, once the problem is cast in phase space, it is easy to go from there to configuration space (which is a subspace of phase space) if need be, but not vice-versa.

After constructing the most general Gaussian measure in phase space, we evaluate large classes of path integrals with respect to it, and present applications to the perturbation expansion and the semiclassical expansion for arbitrary Hamiltonians.

II. CONSTRUCTION OF THE PHASE SPACE MEASURE

We wish to construct the most general Gaussian measure $w(p,q)$ in phase space, the one which will absorb the entire quadratic term in the functional to be integrated. To be more specific, this measure will be equivalent to

$$dw(p,q) \sim \frac{1}{K_0}\left[\frac{dp\,dq}{2\pi\hbar}\right] \exp\frac{i}{\hbar} \times \int_{t_a}^{t_b} \{p(t)\dot{q}(t) - H_0[p(t), q(t), t]\}\,dt, \tag{3}$$

where

$$H_0(p,q,t) = g(t)\frac{p^2}{2m} + \tfrac{1}{2}f(t)q^2 + k(t)pq \tag{4}$$

and K_0 is the normalization factor, ensuring that

$$\int_{\rho} dw(p,q) = 1. \tag{5}$$

It is readily observed that K_0 must be the propagator associated with the Hamiltonian H_0. The functions $g(t)$, $f(t)$, and $k(t)$ depend on the problem investigated. If one wishes to write a path integral for a Hamiltonian of the form $H_0 + \alpha H_1$, where H_1 contains the terms beyond quadratic, then the measure w enables one to obtain the propagator as a perturbation expansion in

[a)] Paper presented at the NATO Advanced Study Institute on Path Integrals and Their Applications in Quantum, Statistical, and Solid State Physics, Antwerp, Belgium, July 17—30, 1977.

 0022-2488/78/1901-0298$1.00

powers of α. If, in a more useful application, one first expands the action functional about the classical position and classical momentum, $q_c(t)$ and $p_c(t)$, then the measure w yields the terms of a semiclassical (WKB) expansion of the propagator. The functions g, f, and k then contain $q_c(t)$ and $p_c(t)$. All this will be further examined below.

A proper way to define (and use) w without the time-slicing procedure that (3) entails is to build its Fourier transform. The Fourier transform of w can be written as

$$\mathcal{F}w(\mu,\nu)=\int_{\rho}\exp(-i\langle\mu,q\rangle-i\langle\nu,p\rangle)\,dw(p,q), \tag{6}$$

where μ and ν are elements of $\mathcal{M}$, the space of bounded measures on the time interval $T\equiv[t_a,t_b]$. Notation: If μ is induced by a function, i. e., $d\mu(t)=f(t)\,dt$, then

$$\langle\mu,q\rangle\equiv\int_T q(t)\,d\mu(t); \tag{7}$$

if μ is δ_s, the delta function at s, then

$$\langle\delta_s,q\rangle\equiv q(s). \tag{8}$$

The fundamental observation is that if we put $d\mu(t)=B(t)\,dt/\hbar$ and $d\nu(t)=A(t)\,dt/\hbar$, then the Fourier transform (6) is nothing other than K/K_0, where K is the propagator corresponding to the auxiliary Hamiltonian

$$H(p,q,t)=g(t)p^2/2m+\tfrac{1}{2}f(t)q^2+k(t)pq+A(t)p+B(t)q. \tag{9}$$

Both K and K_0 can be calculated exactly given the associated classical paths. Indeed, since both correspond to quadratic Hamiltonians, their semiclassical (WKB) approximations are exact. The latter are given by

$$K_{\mathrm{WKB}}=(M/2\pi i\hbar)^{1/2}\exp(iS_c/\hbar), \tag{10}$$

where S_c is the action functional evaluated at the classical position and momentum q_c and p_c, and M is the Van Vleck–Morette function $-\partial^2S_c/\partial q_a\,\partial q_b$. Thus, the problem of determining the phase space measure w reduces to solving the classical problem for H and H_0. Note that the quantum operators corresponding to the pq terms in H and H_0 are the symmetrized[12] $\frac{1}{2}(\mathbf{PQ}+\mathbf{QP})$.

We first state the main theorem, then we prove it. An intuitive justification for it is found in the Appendix.

Theorem 1: The normalized Gaussian measure $w(p,q)$ in phase space $\mathcal{P}$ corresponding to

$$dw(p,q)\sim\frac{1}{K_0}\left[\frac{dp\,dq}{2\pi\hbar}\right]\times\exp\left\{\frac{i}{\hbar}\int_{t_a}^{t_b}[p(t)\dot q(t)-\frac{1}{2m}g(t)p^2(t)-\tfrac{1}{2}f(t)q^2(t)-k(t)p(t)q(t)]\,dt\right\} \tag{11}$$

has the following Fourier transform[13]:

$$\mathcal{F}w(\mu,\nu)=\exp\Big\{-i\langle\mu,\bar q\rangle-i\langle\nu,\bar p\rangle-\frac{i\hbar}{2}\int_T\int_T G_{ab}(t,t')\,d\mu(t)\,d\mu(t')-i\hbar\int_T\int_T\bar G(t,t')\,d\mu(t)\,d\nu(t')-\frac{i\hbar}{2}\int_T\int_T G_p(t,t')\,d\nu(t)\,d\nu(t')\Big\} \tag{12}$$

$$\equiv\exp\Big\{-i\int_T\bar r(t)\,d\tilde\alpha(t)-\frac{i\hbar}{2}\int_T\int_T d\alpha(t)\mathcal{G}(t,t')\,d\tilde\alpha(t')\Big\}, \tag{13}$$

where:

(1) $\mathcal{P}\equiv\{[p(t),q(t)]$ on $T\equiv[t_a,t_b]\,|\,q(t_a)=q_a,\ q(t_b)=q_b,\ p(t)$ unrestricted$\}$. (14)

(2) The normalization factor K_0 is the propagator corresponding to the Hamiltonian

$$\mathbf{H}_0=[g(t)/2m]\,\mathbf{P}^2+\tfrac{1}{2}f(t)\mathbf{Q}^2+\tfrac{1}{2}k(t)(\mathbf{PQ}+\mathbf{QP}), \tag{15}$$

for which the WKB approximation is exact.

(3) $d\alpha(t)\equiv(d\mu(t),d\nu(t))$, $d\tilde\alpha(t)\equiv\begin{pmatrix}d\mu(t)\\ d\nu(t)\end{pmatrix}$. (16)

(4) $\bar r(t)\equiv(\bar q(t),\bar p(t))$,

the average path in $\mathcal{P}$ with respect to the measure w,

$$=(q_{c0}(t),p_{c0}(t)), \tag{17}$$

where q_{c0} and p_{c0} are the classical position and momentum corresponding to H_0. They are related by

$$p_{c0}(t)=\frac{m}{g(t)}\left[\frac{d}{dt}-k(t)\right]q_{c0}(t). \tag{18}$$

(5) $\mathcal{G}(t,t')\equiv\begin{pmatrix}G_{ab}(t,t') & \bar G(t,t')\\ \bar G(t',t) & G_p(t,t')\end{pmatrix}$ (19)

is a Green function of the small disturbance operator in phase space corresponding to H_0:

$$\mathcal{O}=\begin{pmatrix}-f(t) & -k(t)-d/dt\\ -k(t)+d/dt & -(1/m)g(t)\end{pmatrix}, \tag{20}$$

i. e.,

$$\mathcal{O}\mathcal{G}(t,t')=\delta(t-t')\begin{pmatrix}1&0\\0&1\end{pmatrix}. \tag{21}$$

$\mathcal{G}(t,t')$ is independent of q_a and q_b.

(6) $G_{ab}(t,t')$ is the (symmetric) Green function of the small disturbance operator in configuration space which vanishes at both end points:

$$\mathcal{S}\equiv\frac{-m}{g(t)}\left[\frac{d^2}{dt^2}-\frac{\dot g(t)}{g(t)}\frac{d}{dt}-\dot k(t)+\frac{1}{m}f(t)g(t)-k^2(t)+\frac{\dot g(t)k(t)}{g(t)}\right]. \tag{22}$$

(Note that q_{c0} satisfies $\mathcal{S}q_{c0}=0$.)

I. e.,

$$\mathcal{S}G_{ab}(t,t')=\delta(t-t'),\quad G_{ab}(t,t')=G_{ab}(t',t),\quad G_{ab}(t_a,t)=G_{ab}(t_b,t)=0. \tag{23}$$

(Note that q_{c0} satisfies $\mathcal{S}q_{c0}=0$.)

(7) $\bar G(t,t')=\frac{m}{g(t')}\left[\frac{\partial}{\partial t'}-k(t')\right]G_{ab}(t,t')$. (24)

(8) $G_p(t,t')=\frac{m^2}{g(t)g(t')}\left[\frac{\partial}{\partial t}-k(t)\right]\left[\frac{\partial}{\partial t'}-k(t')\right]G_{ab}(t,t')-mg^{-1}(t)\,\delta(t-t')$. (25)

The δ function term in (25) is always cancelled by a similar term. When $t=t'$, G_{ab} and G_p are continuous, but $\bar{G}$ has a jump of magnitude 1:

$$\left(\lim_{t\to t'} - \lim_{t'\to t}\right)\bar{G}(t,t') = 1. \tag{26}$$

Note that the measure $w(p,q)$ does not split the path integral into an integral over momentum space followed by an integral over configuration space, each with its own measure. Thus one truly has a "phase space" path integral. However, the measure w induces in a natural manner measures w_{ab} and w_p on configuration space alone and momentum space alone by

$$\mathcal{J}w_{ab}(\mu) \equiv \mathcal{J}w(\mu, 0) \quad \text{and} \quad \mathcal{J}w_p(\nu) = \mathcal{J}w(0,\nu). \tag{27}$$

The measure w_{ab} in the configuration space of paths such that $q(t_a)=q_a$ and $q(t_b)=q_b$ is studied in Refs. 7 and 9.

Proof of Theorem 1

The Lagrangian L_0 corresponding to H_0 in (4) and the Lagrangian L corresponding to the auxiliary H in (9) are

$$L(q,\dot{q},t) = [m/2g(t)][\dot{q} - A(t) - k(t)q]^2 - \tfrac{1}{2}f(t)q^2 - B(t)q, \tag{28}$$

$$L_0(q,\dot{q},t) = [m/2g(t)][\dot{q} - k(t)q]^2 - \tfrac{1}{2}f(t)q^2. \tag{29}$$

The classical paths q_c and q_{c0} satisfy the Euler–Lagrange equations

$$\mathcal{Q}q_c(t) = u(t), \tag{30}$$

$$\mathcal{Q}q_{c0}(t) = 0, \tag{31}$$

where $\mathcal{Q}$ is a second-order linear differential operator:

$$\mathcal{Q} \equiv \frac{d^2}{dt^2} - \frac{\dot{g}(t)}{g(t)}\frac{d}{dt} - \dot{k}(t) + \frac{1}{m}f(t)g(t) - k^2(t) + \frac{\dot{g}(t)}{g(t)}k(t) \tag{32}$$

and $u(t)$ depends on $A(t)$ and $B(t)$:

$$u(t) \equiv \dot{A}(t) - (1/m)B(t)g(t) + k(t)A(t) - A(t)\dot{g}(t)/g(t). \tag{33}$$

Both classical paths go through q_a at t_a and q_b at t_b. The substitutions

$$q_c(t) = D_c(t)[g(t)/g(t_a)]^{1/2} \quad \text{and} \quad q_{c0}(t) = D_{c0}(t)[g(t)/g(t_a)]^{1/2} \tag{34}$$

eliminate the d/dt term in $\mathcal{Q}$, and replace (30) and (31) by

$$\mathcal{D}D_c(t) = -u(t)[g(t_a)/g(t)]^{1/2} \tag{35}$$

$$\mathcal{D}D_{c0}(t) = 0, \tag{36}$$

where

$$\mathcal{D} \equiv -\frac{d^2}{dt^2} - w(t), \tag{37}$$

with

$$w(t) = \frac{1}{2}\frac{\ddot{g}(t)}{g(t)} - \frac{3}{4}\frac{\dot{g}^2(t)}{g^2(t)} + \frac{\dot{g}(t)k(t)}{g(t)} - k^2(t) + \frac{1}{m}f(t)g(t) - \dot{k}(t). \tag{38}$$

Note that, for an arbitrary function F,

$$\mathcal{Q}[\sqrt{g(t)}\ f(t)] = -\sqrt{g(t)}\ \mathcal{D}f(t). \tag{39}$$

Let $D_1(t)$ and $D_2(t)$ be two solutions of (36), subject to the boundary conditions

$$D_1(t_b) = 1, \quad \dot{D}_1(t_b) = 0, \qquad D_2(t_b) = 0, \quad \dot{D}_2(t_b) = -1. \tag{40}$$

The Wronskian $W = \dot{D}_1D_2 - D_1\dot{D}_2$ is constant for equations of the form $\mathcal{D}D(t) = 0$. In this case the boundary conditions indicate that W is equal to 1. Since W is different from 0, D_1 and D_2 are linearly independent, and the general solution of (36) is a linear combination of D_1 and D_2. If we define the antisymmetric kernel $J(t,t')$ by

$$J(t,t') \equiv D_1(t')D_2(t) - D_1(t)D_2(t'), \tag{41}$$

then the classical path q_{c0} can be written as

$$q_{c0}(t) = \frac{\sqrt{g(t)}}{J(t_a,t_b)}\left[q_a\frac{J(t,t_b)}{\sqrt{g(t_a)}} + q_b\frac{J(t_a,t)}{\sqrt{g(t_b)}}\right]. \tag{42}$$

The classical path q_c can be easily shown to be

$$q_c(t) = q_{c0}(t) - \sqrt{g(t)}\int_T \frac{u(s)}{\sqrt{g(s)}}\,G(s,t)\,ds, \tag{43}$$

where G is the (symmetric) Green function of $\mathcal{D}$ which vanishes at both end points:

$$\mathcal{D}G(t,t') = \delta(t-t'), \quad G(t,t') = G(t',t), \tag{44}$$

$$G(t_a,t) = G(t_b,t) = 0. \tag{45}$$

This Green function can be built from the solutions D_1 and D_2 of $\mathcal{D}D = 0$. It is[6,8,9]

$$G(t,t') = \frac{J(t_a,t)J(t',t_b)Y(t'-t) + J(t_a,t')J(t,t_b)Y(t-t')}{J(t_a,t_b)}, \tag{46}$$

$Y(t)$ being the Heaviside step function, equal to 1 for $t>0$ and 0 otherwise. This can be verified by direct substitution. If $u(s)$ is replaced by its expression (33) in terms of A and B and the $\dot{A}$ term is integrated by parts (the integrated term vanishes), then the difference $\xi(t)$ of the classical paths depends linearly on A and B as follows:

$$\xi(t) \equiv q_c(t) - q_{c0}(t) = \int_T A(s)\,\omega(s,t)\,ds + \int_T B(s)\,\sigma(s,t)\,ds, \tag{47}$$

where

$$\omega(s,t) \equiv \left(\frac{g(t)}{g(s)}\right)^{1/2}\left[\frac{1}{2}\frac{\dot{g}(s)}{g(s)} - k(s) + \frac{\partial}{\partial s}\right]G(s,t), \tag{48}$$

$$\sigma(s,t) \equiv \frac{1}{m}\sqrt{g(t)g(s)}\,G(s,t). \tag{49}$$

As we established earlier, the Fourier transform of the measure w is the ratio K/K_0 of the propagators corresponding to H and H_0, which in turn happened to be exactly equal to their WKB approximants. If $d\mu(t) \equiv B(t)\,dt/\hbar$ and $d\nu(t) \equiv A(t)\,dt/\hbar$, then

$$\mathcal{J}w(B,A) = K/K_0 = \sqrt{M/M_0}\,\exp\{(i/\hbar)\int_T L(q_c,\dot{q}_c,t)\,dt - (i/\hbar)\int_T L_0(q_{c0},\dot{q}_{c0},t)\,dt\}. \tag{50}$$

The Van Vleck–Morette functions M and M_0 are equal since H and H_0 differ only by terms linear in p and q. We give their value for completeness. It is

$$M = M_0 = m/J(t_a, t_b)\sqrt{g(t_a)\,g(t_b)}\,. \tag{51}$$

This can be easily proved. Indeed,

$$M \equiv -\frac{\partial^2 S_c}{\partial q_a\,\partial q_b} = \frac{\partial p_c(t_a)}{\partial q_b} \tag{52}$$

since $p_c(t_a) = -\partial S_c/\partial q_a$. The momentum corresponding to the Lagrangian L in (28) is

$$p = \frac{\partial L}{\partial \dot{q}} = \frac{m}{g(t)}\,[\dot{q} - A(t) - k(t)\,q], \tag{53}$$

and hence $\partial p_c(t_a)/\partial q_b = mg^{-1}(t_a)\,\partial \dot{q}_c(t_a)/\partial q_b$. The result can then be easily established by using (42), (41), and (40), along with the fact that the Wronskian of D_1 and D_2 is 1. It is verified by direct differentiation of $S_0[q_{c0}]$, which is shown explicitly in (121b).

Substituting (28) and (29) in (50) yields

$$\begin{aligned}\mathcal{J}w(B,A) = \exp\left(\frac{i}{\hbar}\right)\Bigg\{&\left(\frac{m}{2}\right)\int_T \frac{dt}{g(t)}\,[\dot{\xi}(t) - k(t)\,\xi(t) - A(t)]\\ &\times[\dot{\xi}(t) + 2\dot{q}_{c0}(t) - A(t) - k(t)\,\xi(t) - 2k(t)\,q_{c0}(t)]\\ &- \int_T B(t)\,[q_{c0}(t) + \xi(t)] - \tfrac{1}{2}\int_T f(t)\,\xi(t)\\ &\times[\xi(t) + 2q_{c0}(t)]\Bigg\}\,.\end{aligned} \tag{54}$$

Now substituting for $\xi(t)$ its expression in (47) yields the full explicit dependence of $\mathcal{J}w(B,A)$ on A and B, which is of the form

$$\begin{aligned}\mathcal{J}w(B,A) = \exp\left(\frac{i}{\hbar}\right)\Bigg\{&-\int_T \bar{q}(t)\,B(t)\,dt - \int_T \bar{p}(t)\,A(t)\,dt\\ &- \tfrac{1}{2}\int_T\int_T G_{ab}(t,t')\,B(t)\,B(t')\,dt\,dt'\\ &- \tfrac{1}{2}\int_T\int_T G_p(t,t')\,A(t)\,A(t')\,dt\,dt'\\ &- \int_T\int_T \bar{G}(t,t')\,B(t)\,A(t')\,dt\,dt'\Bigg\}\,.\end{aligned} \tag{55}$$

The various functions entering this expression are calculated below one by one and found to be as given in the statement of the theorem. Since they will involve small disturbance equations, we think it useful to first exhibit these equations.

Equations of small disturbances

The small disturbance equation (or Jacobi equation, stability equation, or equation of geodesic deviation in the language of curved spaces) is that satisfied by a small deviation from the classical path. Thus, since the Euler–Lagrange (or Hamilton) equations yielding the classical path are obtained by setting the first variation of the action functional equal to zero, the small disturbance equation is obtained by setting the second variation of the action equal to zero. For Lagrangian actions, it is

$$S\,\alpha(t) = 0, \tag{56}$$

where S is the small disturbance operator in configuration space:

$$\begin{aligned}S = \Bigg\{&-\frac{\partial^2 L}{\partial \dot{q}^2}\,\frac{d^2}{dt^2} - \left[\frac{d}{dt}\left(\frac{\partial^2 L}{\partial \dot{q}^2}\right)\right]\frac{d}{dt}\\ &+ \frac{\partial^2 L}{\partial q^2} - \left[\frac{d}{dt}\left(\frac{\partial^2 L}{\partial q\,\partial \dot{q}}\right)\right]\Bigg\}_{q=q_c}\,.\end{aligned} \tag{57}$$

For Hamiltonian actions, it is

$$O\begin{pmatrix}\alpha(t)\\ \beta(t)\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}, \tag{58}$$

where O is the small disturbance operator in phase space:

$$O = \begin{pmatrix} -\dfrac{\partial^2 H}{\partial q^2} & -\dfrac{\partial^2 H}{\partial q\,\partial p} - \dfrac{d}{dt}\\[2ex] -\dfrac{\partial^2 H}{\partial p\,\partial q} + \dfrac{d}{dt} & -\dfrac{\partial^2 H}{\partial p^2}\end{pmatrix}_{\substack{q=q_c\\ p=p_c}}\,. \tag{59}$$

In the case of H_0, S and O are given by (22) and (20). An interesting observation: the elements of O can be used to form S as follows:

$$S = -f(t) + \left[k(t) + \frac{d}{dt}\right]\left[\frac{m}{g(t)}\right]\left[k(t) - \frac{d}{dt}\right]\,. \tag{60}$$

Note also that S and Q are related by

$$S = [-m/g(t)]\,Q, \tag{61}$$

i.e., the Euler–Lagrange and small disturbance equations are the same—a characteristic of quadratic Lagrangians. The solutions of the small disturbance equations, i.e., $\alpha(t)$ and $\beta(t)$ in (56) and (58), are obtained by differentiating the classical position and momentum $q_c(t)$ and $p_c(t)$ with respect to q_a, q_b, or any other constant of integration.[6,8,9]

Calculation of the elements of the measure w

All the calculations below involve integrations by parts where the integrated term vanishes due to (45). The comma denotes differentiation with respect to the variable indicated. Thus with reference to (55), we have:

The BB term

$$\begin{aligned}G_{ab}(t,t') &= 2\sigma(t,t') + \int_T \left[f(s) - \frac{mk^2(s)}{g(s)}\right]\sigma(t,s)\,\sigma(t',s)\,ds\\ &\quad - m\int_T \frac{ds}{g(s)}\,\sigma_{,s}(t,s)\,\sigma_{,s}(t',s)\\ &\quad + m\int_T \frac{ds}{g(s)}\,k(s)[\sigma(t,s)\,\sigma(t',s)]_{,s}\\ &= 2\sigma(t,t') + \int_T ds\,\sigma(t,s)\\ &\quad \times\Bigg\{f(s) + m\,\frac{\partial}{\partial s}\left[\frac{1}{g(s)}\,\frac{\partial}{\partial s}\right] - m\left[\frac{d}{ds}\,\frac{k(s)}{g(s)}\right]\\ &\quad - \frac{mk^2(s)}{g(s)}\Bigg\}\,\sigma(t',s).\end{aligned} \tag{62}$$

The operator between curly brackets can be easily shown to be $mg^{-1}(t)\mathcal{Q}$, i.e. $-\mathcal{S}$. From (39), (45), and (49) we can establish the relation:

$$\mathcal{S}_t\,\sigma(u,t)=\delta(u-t). \tag{63}$$

Substituting in (62), we have

$$G_{ab}(t,t')=\sigma(t,t')\equiv m^{-1}\sqrt{g(t)\,g(t')}\;G(t,t'). \tag{64}$$

Therefore, $G_{ab}(t,t')$ is indeed a Green function of $\mathcal{S}$ which vanishes at t_a and t_b. Since it is symmetric, it is continuous across the diagonal $t=t'$. QED

The AB term

$$\begin{aligned}\bar{G}(t,t')&=\omega(t',t)+\frac{m}{g(t')}\left[\frac{\partial}{\partial t'}-k(t')\right]\sigma(t,t')\\&\quad+\int_T\sigma(t,s)\,\omega(t',s)\left[f(s)-\frac{mk^2(s)}{g(s)}\right]ds\\&\quad+m\int_T\frac{ds}{g(s)}\;k(s)[\omega(t',s)\,\sigma(t,s)]_{,s}\\&\quad-m\int_T\frac{ds}{g(s)}\;\omega_{,s}(t',s)\,\sigma_{,s}(t,s)\end{aligned} \tag{65}$$

$$=\omega(t',t)+\frac{m}{g(t')}\left[\frac{\partial}{\partial t'}-k(t')\right]\sigma(t,t')-\int_T\omega(t',s)\,\mathcal{S}_s\,\sigma(t,s) \tag{66}$$

$$=\frac{m}{g(t')}\left[\frac{\partial}{\partial t'}-k(t')\right]G_{ab}(t,t') \tag{67}$$

in view of (64) and (23). QED

Using the specific expression [(64) with (46)] of G_{ab}, we have

$$\begin{aligned}\bar{G}(t,t')&=J^{-1}(t_a,t_b)\left[\frac{g(t)}{g(t')}\right]^{1/2}\left\{\left[\frac{\dot{g}(t')}{2g(t')}-k(t')\right]\right.\\&\quad\times[J(t_a,t)\,J(t',t_b)\,Y(t-t')+J(t_a,t')\,J(t,t_b)\,Y(t'-t)]\\&\quad+J(t_a,t)\,J_{,t'}(t',t_b)\,Y(t-t')\\&\quad\left.+J_{,t'}(t_a,t')\,J(t,t_b)\,Y(t'-t)\right\}.\end{aligned} \tag{68}$$

It is readily verified that $\bar{G}$ has a discontinuity of magnitude 1 across the diagonal $t=t'$: $(\lim_{t\to t'}-\lim_{t'\to t})\bar{G}(t,t')=1$. For this, one only needs (41) and the fact that the Wronskian of D_1 and D_2 is 1. QED

The AA term

$$\begin{aligned}G_p(t,t')&=\int_T dv\,\omega(t,v)\,\omega(t',v)\left[f(v)-\frac{mk^2(v)}{g(v)}\right]\\&\quad-m\int_T\frac{dv}{g(v)}\;\omega_{,v}(t,v)\,\omega_{,v}(t',v)-\frac{m}{g(t)}\;\delta(t-t')\\&\quad+m\int_T\frac{dv}{g(v)}\;k(v)[\omega(t',v)\,\omega(t,v)]_{,v}\\&\quad+\frac{2m}{g(t')}\left[\frac{\partial}{\partial t'}-k(t')\right]\omega(t,t')\\&=\frac{2m}{g(t')}\left[\frac{\partial}{\partial t'}-k(t')\right]\omega(t,t')-\frac{m}{g(t)}\;\delta(t-t')\\&\quad-\int_T\omega(t,v)\,\mathcal{S}_v\,\omega(t',v)\,dv.\end{aligned} \tag{69}$$

Using (63) and (64), we have

$$\mathcal{S}_t\,\omega(t',t)=\frac{m}{g(t)}\left[k(t')-\frac{\dot{g}(t')}{g(t')}-\frac{\partial}{\partial t'}\right]\delta(t-t'), \tag{70}$$

which gives

$$\begin{aligned}G_p(t,t')&=\frac{m}{g(t')}\left[\frac{\partial}{\partial t'}-k(t')\right]\omega(t,t')-\frac{m}{g(t)}\;\delta(t-t')\\&=\frac{m}{\sqrt{g(t)}}\left[\frac{1}{2}\;\frac{\dot{g}(t)}{g(t)}-k(t)+\frac{\partial}{\partial t}\right]\frac{\bar{G}(t,t')}{\sqrt{g(t)}}-\frac{m\delta(t-t')}{g(t)}\\&=\frac{m}{g(t)}\left[\frac{\partial}{\partial t}-k(t)\right]\bar{G}(t,t')-\frac{m\delta(t-t')}{g(t)}\end{aligned} \tag{71}$$

$$=\frac{m^2}{g(t)\,g(t')}\left[\frac{\partial}{\partial t}-k(t)\right]\left[\frac{\partial}{\partial t'}-k(t')\right]G_{ab}(t,t')-\frac{m\delta(t-t')}{g(t)}. \tag{72}$$

Using the specific expression [(64) with (46)] of G_{ab}, we find that the δ function term cancels another similar term (due to the fact that the Wronskian of D_1 and D_2 is 1), and we are left with the following expression for G_p, symmetric and continuous across the diagonal $t=t'$:

$$\begin{aligned}G_p(t,t')&=[mY(t-t')/\sqrt{g(t)\,g(t')}]\,J^{-1}(t_a,t_b)\\&\quad\times[J(t_a,t)\,J(t',t_b)\,\gamma(t)\,\gamma(t')+J(t_a,t)\,J_{,t'}(t',t_b)\,\gamma(t)\\&\quad+J_{,t}(t_a,t)\,J(t',t_b)\,\gamma(t')+J_{,t}(t_a,t)\,J_{,t'}(t',t_b)]\\&\quad+t\sim t',\end{aligned} \tag{73}$$

where $\gamma(t)\equiv-k(t)+\dot{g}(t)/2g(t)$ and $F(t,t')+t\sim t'\equiv F(t,t')+F(t',t)$.

The B term

$$\begin{aligned}\bar{q}(t)&=q_{c0}(t)+\int_T dt'\,q_{c0}(t')\,\sigma(t,t')\left[f(t')-\frac{mk^2(t')}{g(t')}\right]\\&\quad+m\int_T\left[\frac{dt'}{g(t')}\right][k(t')\,q_{c0}(t')\,\sigma_{,t'}(t,t')\\&\quad+k(t')\,\dot{q}_{c0}(t')\,\sigma(t,t')-\dot{q}_{c0}(t')\,\sigma_{,t'}(t,t')]\\&=q_{c0}(t)+m\int_T\left[\frac{dt}{g(t')}\right]\sigma(t,t')\,\mathcal{Q}_{t'}\,q_{c0}(t')=q_{c0}(t)\end{aligned} \tag{74}$$

by virtue of (31). QED

The A term

$$\begin{aligned}\bar{p}(t)&=\frac{m}{g(t)}\left[\frac{d}{dt}-k(t)\right]q_{c0}(t)+\int_T dt'\,\omega(t,t')\,q_{c0}(t')\\&\quad\times\left[f(t')-\frac{mk^2(t')}{g(t')}\right]+m\int_T\frac{dt'}{g(t')}[k(t')\,\omega(t,t')\,\dot{q}_{c0}(t')\\&\quad+k(t')\,q_{c0}(t')\,\omega_{,t'}(t,t')-\dot{q}_{c0}(t')\,\omega_{,t'}(t,t')]\end{aligned}$$

$$= \frac{m}{g(t)}\left[\frac{d}{dt} - k(t)\right] q_{c0}(t) + m\int_T \frac{dt'}{g(t')}\,\omega(t,t')\mathcal{Q}_{t'} q_{c0}(t')$$

$$= \frac{m}{g(t)}\left[\frac{d}{dt} - k(t)\right] q_{c0}(t)$$

$$= p_{c0}(t) \tag{75}$$

by virtue of (31) and the fact that $p_{c0} = (\partial L_0/\partial \dot{q})_{q=q_{c0}}$. QED

The relations we have derived so far make it simple to verify that $\mathcal{G}(t,t')$ in (19) is indeed a Green function of the small disturbance operator (20) in phase space. The "11" term of the resulting matrix is $\delta(t-t')$ because of (67) and (60). The "12" term is 0 because of (25), (60), and the fact that $(\partial/\partial t + \partial/\partial t')\,\delta(t-t') = 0$. The "21" term is 0 because of (67). Finally, the "22" term is $\delta(t-t')$ because of (71). The fact that $(\bar{q},\bar{p})$ is the average path will be proved later in the paper.

Example 1: The free particle

For a free particle, $k(t) = f(t) = 0$, $g(t) = 1$, $H_0 = p^2/2m$, $\mathcal{S} = -m d^2/dt^2$, $D_1(t) = 1$, $D_2(t) = t_b - t$, $J(t,t') = t' - t$, and $M = m/T$. The covariance of the corresponding measure in phase space is (19), where

$$G_{ab}(t,t') = [(t'-t_a)(t_b - t)\,Y(t-t') + (t-t_a)(t_b - t')\,Y(t'-t)]/mT, \tag{76}$$

$$\bar{G}(t,t') = [(t_b - t)\,Y(t-t') - (t-t_a)\,Y(t'-t)]/T, \tag{77}$$

$$G_p(t,t') = -m/T \qquad (T \equiv t_b - t_a). \tag{78}$$

We have $\mathcal{S}\,G_{ab}(t,t') = \delta(t-t')$. The average position and momentum are the classical ones:

$$q_{c0}(t) = [q_b(t-t_a) + q_a(t_b - t)]/T, \tag{79}$$

$$p_{c0}(t) = m(q_b - q_a)/T. \tag{80}$$

The Wiener measure for a free particle in Brownian motion, defined on the configuration space of paths $\mathcal{C}_- \equiv \{q(t) \text{ on } T = [t_a, t_b]\,|\,q(t_a) = 0,\ q(t_b) \text{ unrestricted}\}$ can be readily extended to the phase space $\mathcal{P}_-$ defined by

$$\mathcal{P}_- \equiv \{[p(t), q(t)] \text{ on } T = [t_a, t_b]\,|\,q(t_a) = 0,\ q(t_b) \text{ and } p(t) \text{ unrestricted}\} \tag{81}$$

by letting $t_b \to \infty$. The covariance $\mathcal{G}_-(t,t')$ is then

$$\mathcal{G}_-(t,t') = \begin{pmatrix} (1/m)[(t'-t_a)\,Y(t-t') + (t-t_a)\,Y(t'-t)] & Y(t-t') \\ Y(t'-t) & 0 \end{pmatrix}. \tag{82}$$

The "11" term is a (symmetric) Green function of the small disturbance operator $-m d^2/dt^2$ such that $G(t_a, t) = 0$. It is discussed in Refs. 4–6. Since the Wiener measure is real, the i must be removed from the exponent in (12) and (13).

Example 2: The harmonic oscillator

For a harmonic oscillator, $k(t) = 0$, $g(t) = 1$, $f(t) = m\omega^2$, $H_0 = p^2/2m + m\omega^2 q^2/2$, $\mathcal{S} = -m(d^2/dt^2 + \omega^2)$, $D_1(t) = \cos\omega(t_b - t)$, $D_2(t) = \omega^{-1}\sin\omega(t_b - t)$, $J(t,t') = \omega^{-1}\sin\omega(t'-t)$ and $M = m\omega/\sin\omega T$. The covariance of the corresponding measure in phase space is (19), with:

$$G_{ab}(t,t') = [\sin\omega(t_b - t')\sin\omega(t - t_a)\,Y(t'-t) + \sin\omega(t_b - t)\sin\omega(t' - t_a)\,Y(t-t')]/m\omega\sin\omega T \tag{83}$$

$$\bar{G}(t,t') = [\sin\omega(t_b - t)\cos\omega(t' - t_a)\,Y(t-t') - \cos\omega(t_b - t')\sin\omega(t - t_a)\,Y(t'-t)]/\sin\omega T \tag{84}$$

$$G_p(t,t') = (-m\omega/\sin\omega T)[\cos\omega(t_b - t)\cos\omega(t' - t_a)\,Y(t-t') + \cos\omega(t_b - t')\cos\omega(t - t_a)\,Y(t'-t)]. \tag{85}$$

We have $\mathcal{S}\,G_{ab}(t,t') = \delta(t-t')$. The average position and momentum are the classical ones:

$$q_{c0}(t) = (\sin\omega T)^{-1}[q_a\sin\omega(t_b - t) + q_b\sin\omega(t - t_a)], \tag{86}$$

$$p_{c0}(t) = m\omega(\sin\omega T)^{-1}[q_b\cos\omega(t - t_a) - q_a\cos\omega(t_b - t)]. \tag{87}$$

III. PATH INTEGRATION IN PHASE SPACE

We now show how to carry out the path integral of a cylindrical functional with respect to an arbitrary measure in phase space given by its Fourier transform. A cylindrical functional is one which depends on only a finite number of terms of the form $\langle\mu, q\rangle$ or $\langle\nu, p\rangle$, i.e., $\int_T q(t)\,d\mu(t)$ or $\int_T p(t)\,d\nu(t)$.

Theorem 2: Let w be a measure in phase space $\mathcal{P}$ defined by its Fourier transform $\mathcal{J}w(\mu,\nu)$. A cylindrical functional F on $\mathcal{P}$ can be integrated over $\mathcal{P}$ with respect to the measure w by reducing it to an ordinary integral as follows:

$$\begin{aligned} I &\equiv \int_{\mathcal{P}} F(\langle\mu_1, q\rangle, \ldots, \langle\mu_n, q\rangle, \langle\nu_1, p\rangle, \ldots, \langle\nu_m, p\rangle)\,dw(p,q) \\ &= \int_{\mathbb{R}^{n+m}} F(u_1, \ldots, u_n, v_1, \ldots, v_m)\,du_1\cdots du_n \\ &\quad \times dv_1\cdots dv_m(2\pi)^{-n-m} \\ &\quad \times \int_{\mathbb{R}^{n+m}} \mathcal{J}w(\xi^1\mu_1 + \cdots + \xi^n\mu_n, \eta^1\nu_1 + \cdots + \eta^m\nu_m) \\ &\quad \times \exp i(\xi^1 u_1 + \cdots + \xi^n u_n + \eta^1 v_1 + \cdots + \eta^m v_m) \\ &\quad \times d\xi^1\cdots d\xi^n\,d\eta^1\cdots d\eta^n. \end{aligned} \tag{88}$$

Proof: This proof is similar to the ones used for similar formulas in configuration space path integrals without limiting procedure.[5,7] Consider the linear continuous mapping $P_{n,m}$:

$$P_{n,m}: \mathcal{P} \to \mathbb{R}^{n+m} \text{ by } (q,p) \mapsto (u,v)$$

$$\text{where } \begin{cases} u_i = \langle\mu_i, q\rangle & \text{for } i = 1 \text{ to } n \\ v_j = \langle\nu_j, p\rangle & \text{for } j = 1 \text{ to } m \end{cases} \tag{89}$$

Under this mapping, we have

$$I = \int_{\mathbb{R}^{n+m}} F(u_1, \ldots, u_n, v_1, \ldots, v_m)\,dw_{P_{n,m}}(u,v), \tag{90}$$

where $w_{P_{n,m}}$ is the image of w under $P_{n,m}$. This image is a measure in $\mathbb{R}^{n+m}$. By theorem,[3] $\mathcal{J}w_{P_{n,m}}(\xi,\eta) = \mathcal{J}w[\tilde{P}_{n,m}(\xi,\eta)]$, where $\xi \in \mathbb{R}^n$, $\eta \in \mathbb{R}^m$, and $\tilde{P}_{n,m}$ is the

transpose mapping from $\mathbb{R}^{n+m}$ to $\mathcal{M}$, the space of bounded measures on the time interval $T=[t_a, t_b]$. We have

$$\begin{aligned}&\langle \tilde{P}_{n,m}(\xi,\eta);(q,p)\rangle \\ &\quad \equiv \langle (\xi,\eta); P_{n,m}(q,p)\rangle = \langle (\xi,\eta);(u,v)\rangle \\ &\quad = \xi\cdot u+\eta\cdot v=\xi^i\langle \mu_i,q\rangle+\eta^j\langle \nu_j,p\rangle \\ &\quad = \langle (\xi^i\mu_i,\eta^j\nu_j);(q,p)\rangle \end{aligned} \tag{91}$$

and hence $\tilde{P}_{n,m}(\xi,\eta)=(\xi^i\mu_i,\eta^j\nu_j)$. Therefore,

$$\begin{aligned}dw_{P_{n,m}}(u,v) &= \mathcal{F}^{-1}_{\xi,\eta}[\mathcal{F}w(\xi^i\mu_i,\eta^j\nu_j)] \\ &= (2\pi)^{-n-m}\,du_1\cdots du_n\,dv_1\cdots dv_m \\ &\quad\times\int_{\mathbb{R}^{n+m}}\exp(i\xi^i u_i+i\eta^j v_j) \\ &\quad\times \mathcal{F}w(\xi^i u_i,\eta^j v_j)\,d\xi^1\cdots d\xi^n\,d\eta^1\cdots d\eta^m. \quad \text{QED}\end{aligned} \tag{92}$$

Corollary 1: If F depends only on p (resp. q), the path integral reduces to an integral over momentum (resp. configuration) space. In compressed notation:

$$\begin{aligned}&\int_\rho F(\langle \nu,p\rangle)\,dw(p,q) \\ &\quad=\int_{\mathbb{R}^m}\frac{dv}{(2\pi)^m}\,F(v)\int_{\mathbb{R}^m}\mathcal{F}w(0,\eta\circ\nu) \\ &\qquad\times\exp(i\eta\circ v)\,d\eta,\end{aligned} \tag{93}$$

$$\begin{aligned}&\int_\rho F(\langle \mu,q\rangle)\,dw(p,q) \\ &\quad=\int_{\mathbb{R}^n}\frac{du}{(2\pi)^n}\,F(u)\int_{\mathbb{R}^n}\mathcal{F}w(\xi\circ\mu,0)\exp(i\xi\circ u)\,d\xi.\end{aligned} \tag{94}$$

Thus, in the second case, the measure $w(p,q)$ in phase space has the same effect as the measure $w_{ab}(q)$ in the configuration space C_{ab} of paths such that $q(t_a)=q_a$ and $q(t_b)=q_b$, i.e.,

$$\int_\rho F(\langle \mu,q\rangle)\,dw(p,q)=\int_{C_{ab}}F(\langle \mu,q\rangle)\,dw_{ab}(q). \tag{95}$$

C_{ab} and w_{ab} were introduced and studied in Ref. 7.

Moments formula:

$$\begin{aligned}&\int_\rho\langle \mu_1,q\rangle\cdots\langle \mu_n,q\rangle\langle \nu_1,p\rangle\cdots\langle \nu_m,p\rangle\,dw(p,q) \\ &\quad=i^{m+n}\frac{\partial^{m+n}}{\partial\xi^1\cdots d\xi^n\,d\eta^1\cdots d\eta^m} \\ &\qquad\times\mathcal{F}w(\xi^1\mu_1+\cdots+\xi^n\mu_n,\eta^1\nu_1+\cdots+\eta^m\nu_m)\big|_{\xi=\eta=0}.\end{aligned} \tag{96}$$

Proof: Theorem 2 and the fact that $\int_{\mathbb{R}} x\exp(ikx)\,dx=-2\pi i\delta'(x)$ are needed.

Application to the Gaussian measure

If we apply Theorem 2 to the Gaussian measure defined in (12) in Theorem 1, we obtain

$$\begin{aligned}&\int_\rho F(\langle \mu_1,q\rangle,\ldots,\langle \mu_n,q\rangle,\langle \nu_1,p\rangle,\ldots,\langle \nu_m,p\rangle)\,dw(p,q) \\ &\quad=\int_{\mathbb{R}^{n+m}}\frac{F(u_1,\ldots,u_n,v_1,\ldots,v_m)\,du_1\cdots du_n\,dv_1\cdots dv_m}{[(2\pi i\hbar)^{m+n}\det W\cdot\det S]^{1/2}} \\ &\qquad\times\exp(i/2\hbar)\{(S^{-1})^{ij}(v_i-b_i)(v_j-b_j) \\ &\qquad-2(W^{-1}CS^{-1})^{ij}(u_i-a_i)(v_j-b_j) \\ &\qquad+(W^{-1}+W^{-1}CS^{-1}\tilde{C}W^{-1})^{ij}(u_i-a_i)(u_j-a_j)\},\end{aligned} \tag{97}$$

where

$$a_i\equiv\langle \mu_i,\bar{q}\rangle, \tag{98}$$

$$b_i\equiv\langle \nu_i,\bar{p}\rangle, \tag{99}$$

$$W_{ij}\equiv\int_T\int_T G_{ab}(t,t')\,d\mu_i(t)\,d\mu_j(t') \quad (n\times n), \tag{100}$$

$$C_{ij}\equiv\int_T\int_T\bar{G}(t,t')\,d\mu_i(t)\,d\nu_j(t') \quad (n\times m), \tag{101}$$

$$V_{ij}\equiv\int_T\int_T G_p(t,t')\,d\nu_i(t)\,d\nu_j(t') \quad (m\times m), \tag{102}$$

$$S\equiv V-\tilde{C}W^{-1}C \quad (m\times m), \tag{103}$$

$\tilde{C}$ being the transpose of C. As far as we know, every path integral which has been evaluated exactly in the literature can be shown to be a special case of (97).

Proof: The proof is straightforward with repeated use of the formula[14]

$$\begin{aligned}&\int_{\mathbb{R}^n}\varphi(b^iu_i)\exp(-\tfrac{1}{2}A^{ij}u_iu_j)\,du_1\cdots du_n \\ &\quad=[(\sqrt{2\pi})^{n-1}/|c|\sqrt{\det A}\,]\int_{\mathbb{R}}\varphi(u)\exp(-u^2/2c^2)\,du,\end{aligned}$$

$$\text{if}\quad \mathrm{Re}(A^{ij}u_iu_j)\geq 0\quad\forall u\in\mathbb{R}^n, \tag{104}$$

where $c^2\equiv b^ib^j(A^{-1})_{ij}$. Here $\varphi(u)=\exp(iu)$ and one needs $\int_{\mathbb{R}}\exp(ax^2+bx)\,dx=(-\pi/a)^{1/2}\exp(-b^2/4a)$, $\mathrm{Re}(a)\leq 0$.

(105)

If the functional to be integrated does not have a p dependence, then Corollary 1 gives

$$\begin{aligned}&\int_\rho F(\langle \mu_1,q\rangle,\ldots,\langle \mu_n,q\rangle)\,dw(p,q) \\ &\quad=\int_{C_{ab}}F(\langle \mu_1,q\rangle,\ldots,\langle \mu_n,q\rangle)\,dw_{ab}(q) \\ &\quad=\int_{\mathbb{R}^n}\frac{F(u_1,\ldots,u_n)\,du_1\cdots du_n}{(2\pi i\hbar)^{n/2}(\det W)^{1/2}} \\ &\qquad\times\exp\left\{\frac{i}{2\hbar}(W^{-1})^{ij}(u_i-a_i)(u_j-a_j)\right\},\end{aligned} \tag{106}$$

which is formula (59) in Ref. 7. If F has no q dependence, then

$$\begin{aligned}&\int_\rho F(\langle \nu_1,p\rangle,\ldots,\langle \nu_m,p\rangle)\,dw(p,q) \\ &\quad=\int_{\mathbb{R}^m}\frac{F(v_1,\ldots,v_m)\,dv_1\cdots dv_m}{(2\pi i\hbar)^{m/2}(\det V)^{1/2}} \\ &\qquad\times\exp\left[\frac{i}{2\hbar}(V^{-1})^{ij}(v_i-b_i)(v_j-b_j)\right].\end{aligned} \tag{107}$$

Averages and covariances

The moments formula (96) applied to the Gaussian measure w readily gives the average position and momentum for w:

$$\int_\rho q(t)\,dw(p,q)=\bar{q}(t), \tag{108}$$

$$\int_\rho p(t)\,dw(p,q)=\bar{p}(t), \tag{109}$$

indicating that $\bar{p}$ and $\bar{q}$ were correctly identified in the statement of Theorem 1. The covariances are

$$\int_\rho [q(t)-\bar{q}(t)][q(t')-\bar{q}(t')]\,dw(p,q)=i\hbar G_{ab}(t,t'), \tag{110}$$

$$\int_\rho [p(t)-\bar{p}(t)][p(t')-\bar{p}(t')]\,dw(p,q)=i\hbar G_{p}(t,t'), \tag{111}$$

$$\int_\rho [q(t)-\bar{q}(t)][p(t')-\bar{p}(t')]\,dw(p,q)=i\hbar \bar{G}(t,t'). \tag{112}$$

$\bar{G}(t,t')$ is the only covariance not to be continuous across the diagonal $t=t'$. It has a jump of magnitude 1 there, as established earlier. Thus, the correlation between p and q at a given time t with respect to the measure w can only be established to within $i\hbar$.

The set of "important paths"

The variances $i\hbar G_{ab}(t,t)$ and $i\hbar G_{p}(t,t)$, squares of the "standard deviations" $\Delta q(t)$ and $\Delta p(t)$, provide a measure of the degree of dispersion of the Feyman paths about the average position and momentum. We now calculate Δq and Δp for the free particle and the harmonic oscillator, using the results established earlier for these two systems.

Free particle:

$$\Delta q(t)=\left[i\hbar\,\frac{(t-t_a)(t_b-t)}{mT}\right]^{1/2}, \tag{113}$$

$$\Delta p(t)=\left[-\frac{i\hbar m}{T}\right]^{1/2}, \tag{114}$$

$$(\Delta p\circ\Delta q)(t)=\frac{\hbar}{T}\,[(t-t_a)(t_b-t)]^{1/2}. \tag{115}$$

Harmonic oscillator:

$$\Delta q(t)=\left[\frac{i\hbar\,\sin\omega(t_b-t)\,\sin\omega(t-t_a)}{m\omega\,\sin\omega T}\right]^{1/2}, \tag{116}$$

$$\Delta p(t)=\left[-\frac{i\hbar m\omega\,\cos\omega(t_b-t)\cos\omega(t-t_a)}{\sin\omega T}\right]^{1/2}, \tag{117}$$

$$(\Delta p\circ\Delta q)(t)=\frac{\hbar}{2\,|\sin\omega T\,|}\times[\sin2\omega(t-t_a)\sin2\omega(t_b-t)]^{1/2}. \tag{118}$$

In both instances, we have

$$[(\Delta p\cdot\Delta q)(t)]^2\leq(\hbar/2)^2. \tag{119}$$

A first glance at this relation might give the impression that we have obtained the uncertainty principle backwards. In fact, this relation has nothing to do with the uncertainty principle. If Δq and Δp are calculated with respect to $\psi(q,t)$ and $\phi(p,t)$, the wavefunctions of the particles in configuration and momentum spaces at time t, then they reflect the effect of measurement, and $(\Delta p\circ\Delta q)(t)\geq\hbar/2$. But if Δq and Δp are calculated with respect to the phase space measure $w(p,q)$, then they simply reflect which paths are weighed more heavily (i.e., contribute the most) in the sum over paths. To be more precise, they determine how far one must deviate from the average (here, classical) path to still find paths which contribute appreciably to the sum over paths. In these two cases (as in most cases), these "important" paths are so close to the average path in phase space that $(\Delta p\circ\Delta q)^2$ is never larger than $\hbar^2/4$.

Note that the average square velocity in configuration space is infinite, indicating that the "important" paths in configuration space are the nondifferentiable ones, a well-known result. For example, in the case of the free particle,

$$\begin{aligned}[(\Delta\dot{q})(t)]^2&\equiv\int_\rho[\dot{q}(t)-\dot{\bar{q}}(t)]^2\,dw(p,q)\\&=\lim_{t\to t'}\frac{\partial^2}{\partial t\,\partial t'}\int_\rho[q(t)-\bar{q}(t)][q(t')-\bar{q}(t')]\,dw(p,q)\\&=\lim_{t\to t'}\frac{\partial^2}{\partial t\,\partial t'}\,i\hbar G_{ab}(t,t')\\&=\lim_{t\to t'}\left[\frac{-i\hbar m}{T}+i\hbar m\delta(t-t')\right]\\&\to\infty.\end{aligned} \tag{120}$$

Comparison with (114) shows that we do not have $\Delta p(t)=\Delta m\dot{q}(t)$; nor should we expect it, since no relationship is assumed between p and q in the unrestricted sum over paths in phase space.

IV. APPLICATIONS

A. Perturbation expansion

The propagator corresponding to $\mathbf{H}=\mathbf{H}_0+\alpha\mathbf{H}_1$, where $\mathbf{H}_0$ is given by (15), is

$$\langle q_b,t_b\,|\,q_a,t_a\rangle=K_0\int_\rho\exp\left[\frac{-i\alpha}{\hbar}\int_{t_a}^{t_b}H_1[p(t),q(t),t]\,dt\right]dw(p,q), \tag{121}$$

where H_1 is the classical equivalent[1] of $\mathbf{H}_1$, K_0 is the propagator corresponding to $\mathbf{H}_0$, and w is defined in (12). This is a direct application of Theorem 1. By expanding the exponential and carrying out the resulting path integrals by use of the moments formula (96), one obtains the propagator as a power series in α.

Calculation of K_0

K_0, the propagator corresponding to $\mathbf{H}_0$ in (15), is given exactly by its WKB approximation. Thus, we only need to calculate the classical action. The action functional is

$$\begin{aligned}S_0[q]&=\int_T L_0(q,\dot{q},t)\,dt\\&=\int_T\left\{\frac{m}{2g(t)}\,[\dot{q}(t)-k(t)q(t)]^2-\tfrac{1}{2}f(t)q^2(t)\right\}dt\\&=\tfrac{1}{2}\int_T q(t)\,\mathcal{S}\,q(t)\,dt+\frac{m}{2}\left\{\frac{q_b}{g(t_b)}\,[\dot{q}(t_b)-k(t_b)q_b]\right.\\&\quad\left.-\frac{q_a}{g(t_a)}\,[\dot{q}(t_a)-k(t_a)q_a]\right\},\end{aligned} \tag{121a}$$

where $\mathcal{S}$ is the operator (22). This can be easily established by integrations by parts of the $\dot{q}^2$ and $(q^2)^{\cdot}$ terms. At the classical path q_{c0}, the first term vanishes since

$\mathcal{S}q_{c0}=\mathcal{Q}q_{c0}=0$, and only the integrated term remains. Thus, no integrations are required to obtain the action functional at the classical path. (42) gives q_{c0} in terms of the kernel J, and we get

$$S_0[q_{c0}]=\frac{m}{2}\left\{\frac{q_b^2}{g(t_b)}\left[\frac{J_{,2}(t_a,t_b)}{J(t_a,t_b)}+\frac{\dot g(t_b)}{2g(t_b)}-k(t_b)\right]-\frac{q_a^2}{g(t_a)}\left[\frac{J_{,1}(t_a,t_b)}{J(t_a,t_b)}+\frac{\dot g(t_a)}{2g(t_a)}-k(t_a)\right]-\frac{2q_aq_b}{J(t_a,t_b)\sqrt{g(t_a)g(t_b)}}\right\}, \tag{121b}$$

where we have used the fact that $J_{,2}(t_a,t_a)=-J_{,1}(t_b,t_b)=1$. Note that $J_{,2}(t_a,t_b)=D_1(t_a)$ and that $J_{,1}(t_a,t_b)=\hat D_2(t_a)$ ($J_{,i}$ denotes derivative with respect to ith argument). Finally,

$$K_0=[m/2\pi i\hbar J(t_a,t_b)]^{1/2}g^{-1/4}(t_a)g^{-1/4}(t_b)\times\exp\{(i/\hbar)S_0[q_{c0}]\}. \tag{121c}$$

B. Semiclassical expansion

A more useful application of Theorem 1 is to use it to expand the ratio of the propagator to its WKB approximation in a power series in $\hbar$:

$$K=K_{\rm WKB}(1+\hbar K_1+\hbar^2K_2+\cdots). \tag{122}$$

This is the semiclassical expansion, treated in configuration space in Refs. 6, 8, and 9. The terms K_i are path integrals of cylindrical functionals, which can be evaluated exactly using (96).

For example, in the case of a particle in a time-dependent potential, with Hamiltonian $H=p^2/2m+V(q,t)$, one finds[9]

$$K_1=\frac{-i}{8}\int_T V^{(4)}(t)\,G_{ab}^2(t,t)\,dt+\frac{i}{24}\int_{T^2}dt\,ds\,V^{(3)}(t)\,V^{(3)}(s)\times[3G_{ab}(t,t)\,G_{ab}(t,s)\,G_{ab}(s,s)+2G_{ab}^3(t,s)] \tag{122a}$$

where

$$V^{(n)}(t)\equiv[\partial^nV(q,t)/\partial q^n]_{q=q_c}. \tag{122b}$$

Such an expansion has been applied to the anharmonic oscillator.[6,9] Sometimes, due to the peculiarities of the Hamiltonian, a configuration space path integral scheme is not possible. A phase space path integral scheme is always possible. We now show how to generalize the path integral treatment of the semiclassical expansion to arbitrary Hamiltonians (for the cases when the Hamiltonian lends itself to a WKB approximation for the propagator[15]).

Theorem 3: The propagator corresponding to an arbitrary H [see Eq. (1)] can be expressed as the following path integral:

$$K=K_{\rm WKB}\int_{\rho_0}\exp\{(i/\hbar)\,\Omega(q_c,p_c)(x,y)\}\,dw(y,x) \tag{123}$$

if K admits a WKB approximation $K_{\rm WKB}$. We have:

$$(1)\quad K_{\rm WKB}=\left(\frac{-1}{2\pi i\hbar}\,\frac{\partial^2S(q_c,p_c)}{\partial q_a\,\partial q_b}\right)^{1/2}\exp\left[\frac{i}{\hbar}\,S(q_c,p_c)\right]; \tag{124}$$

(2) $(x,y)\in\rho_0\equiv\{[x(t),y(t)]$ on $T\equiv[t_a,t_b]\,|\,x(t_a)=x(t_b)=0,\ y(t)$ unrestricted$\}$; (125)

(3) $\Omega(q_c,p_c)$ is an operator resulting from the expansion of the action functional about the classical path (q_c,p_c):

$$S(q,p)\equiv S(x+q_c,y+p_c)=S(q_c,p_c)+S'(q_c,p_c)(x,y)+(1/2!)\,S''(q_c,p_c)(x,y)+\Omega(q_c,p_c)(x,y); \tag{126}$$

(4) the Gaussian measure w is as in Theorem 1, with

$$\frac{g(t)}{m}=\left.\frac{\partial^2H}{\partial p^2}\right|_{\substack{q=q_c\\p=p_c}}, \tag{127}$$

$$f(t)=\left.\frac{\partial^2H}{\partial q^2}\right|_{\substack{q=q_c\\p=p_c}}, \tag{128}$$

$$k(t)=\left.\frac{\partial^2H}{\partial q\,\partial p}\right|_{\substack{q=q_c\\p=p_c}}. \tag{129}$$

Thus, the average path $(q_{c0}(t),\ p_{c0}(t))$ of Theorem 1 is, in this context, the solution of the small disturbance equation corresponding to H, which vanishes at t_a and t_b.

The path integral can be evaluated by expanding the exponential in a power series, which can then be rearranged to yield a power series in $\hbar$ where the terms depend only on the classical path (q_c,p_c).

Proof: In the expansion (126) of the action, the term $S'(q_c,p_c)(x,y)$ is 0 by definition of the classical path (q_c,p_c) (it yields Hamilton's equations). The term $S''(q_c,p_c)(x,y)/2$ is

$$\tfrac{1}{2}(x,y)\mathcal{O}\begin{pmatrix}x\\y\end{pmatrix} \tag{130}$$

where $\mathcal{O}$ is the small disturbance operator (59). Expanding this term, integrating the $-\frac{1}{2}\int_T x(t)\dot y(t)\,dt$ term by parts to get $\frac{1}{2}\int_T y(t)\dot x(t)\,dt$, and lumping the resulting expression (quadratic in x and y) into the measure by using Theorem 1 yields the desired result. The expression of the measure given in (A3) and the expression (121c) of K_0 can be used to establish that the normalization is indeed $K_{\rm WKB}$ as given in (124). The $\hbar$ in the denominator will always be cancelled by higher powers of $\hbar$ in the numerator, due to the fact that the moments formula needed to evaluate the various path integrals arising in the expansion of the exponential in (123) yields products of covariances; each of which is multiplied by $\hbar$.[6,8,9] The needed solutions of the small disturbance equation can be obtained by differentiating the classical position and momentum with respect to the constants of integration of the classical equations of motion.[6,8,9] Thus, all the terms of the semiclassical expansion can be expressed as definite integrals of known functions once the classical problem is solved in closed form.

V. CONCLUSION

The generalization of the path integral scheme to arbitrary Hamiltonians, which can only be done in

phase space, is best carried out without the limiting process which makes the integrals ambiguous and difficult to compute. This paper has built Gaussian phase space measures which do not require any reference to such a limiting process, shown how to integrate with respect to them, and given examples of how these measures can be of use in solving problems. It would be useful to find non-Gaussian measures, which would absorb larger parts of the functionals to be integrated.

ACKNOWLEDGMENT

I thank Dr. W. Hurley for a discussion.

APPENDIX: INTUITIVE JUSTIFICATION OF THEOREM 1 AND OF PATH INTEGRATION WITHOUT LIMITING PROCEDURE

A Gaussian measure on $\mathbb{R}^n$ which can be written as

$$d\gamma(y) = (2\pi)^{-n/2}(\det C)^{1/2}\exp(-\tfrac{1}{2}C^{ij}y_iy_j)\,dy_1\cdots dy_n \quad \text{(A1)}$$

has as its Fourier transform the exponential of a quadratic form involving the *inverse* of the matrix C:

$$(\mathcal{F}\gamma)(x) \equiv \int_{\mathbb{R}^n}\exp(-ix\cdot y)\,d\gamma(y) = \exp[-\tfrac{1}{2}(C^{-1})_{ij}x^ix^j]\,. \quad \text{(A2)}$$

How does this carry over to infinite-dimensional spaces? This is the question answered in Theorem 1. The phase space measure $w(p,q)$ in (11), after integrations by parts, can be written as[16]

$$dw(p,q) \sim K_0^{-1}\left[\frac{dp\,dq}{2\pi\hbar}\right]\exp\frac{i}{2\hbar}\left\{\int_T (q(t),p(t))\,\mathcal{O}\begin{pmatrix}q(t)\\p(t)\end{pmatrix}dt + [q_b p(t_b) - q_a p(t_a)]\right\}, \quad \text{(A3)}$$

where $\mathcal{O}$ is the operator defined in (20). $\mathcal{O}$ is seen to play the role of the matrix C above. Therefore, by analogy, one expects the Fourier transform of w to be the exponential of a quadratic form involving the inverse of $\mathcal{O}$, i.e., one of the *Green functions* of $\mathcal{O}$, so that

$$\int_T \mathcal{O}_t\,\mathcal{G}(t,t')\,dt' = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}. \quad \text{(A4)}$$

This is precisely what was proved. Which Green's function is used depends on the path space considered (for example, in the free particle case we saw that consideration of the space $\mathcal{P}_-$ instead of $\mathcal{P}$ led to a Green function $\mathcal{G}_-$ different from $\mathcal{G}$). The terms involving q_a and q_b enter when the average path is non-zero. The reason why $\mathcal{O}$ was called the small disturbance operator in Theorem 1 is that this is what it is when the action is expanded about the classical path, as we have seen.

To illustrate this intuitive justification further, we consider the free-particle measure w_{ab} in configuration space. It is (N is the "normalization" necessary in the time-slicing approach):

$$dw_{ab}(q) \sim [dq/N]\exp[(im/2\hbar)\textstyle\int_T \dot{q}^2(t)\,dt], \quad \text{(A5)}$$

which can be rewritten as

$$dw_{ab}(q) \sim \left[\frac{dq}{N}\right]\exp\left\{\frac{i}{2\hbar}\int_T q(t)\left(-m\frac{d^2}{dt^2}\right)q(t)\,dt + \frac{im}{2\hbar}[q_b\dot{q}(t_b) - q_a\dot{q}(t_a)]\right\}. \quad \text{(A6)}$$

The Fourier transform of w_{ab} indeed has as its covariance an inverse of $-md^2/dt^2$, namely the Green function G_{ab} in (76).[17] The form (A6) can be easily generalized to the more general configuration space measure $w_{ab}(q)$ induced by $w(p,q)$ in (11): $-md^2/dt^2$ is replaced by $\mathcal{S}$ in (22)—as shown in (121a)—and the covariance is G_{ab}, the Green function of $\mathcal{S}$ introduced in Theorem 1.

[1] In most cases of physical interest, e.g., when the quantum Hamiltonian operator is $\mathbf{H} = [\mathbf{P} - (e/c)A(\mathbf{Q})]^2/2m + e\phi(\mathbf{Q})$, H in (1) is the classical Hamiltonian H_c. For stronger couplings of Q and P, e.g., when there is a metric $g(\mathbf{Q})$ to be considered, H in (1) is of the form $H_c + O(\hbar^2)$. In M. M. Mizrahi, J. Math. Phys. **16**, 2201—06 (1975), it was shown that H could be the Weyl transform of $\mathbf{H}$. Other possibilities, all yielding the same propagator, are being investigated. At any rate, we always have $H = H_c + O(\hbar^2)$ for Hermitian Hs. In this paper, we assume that $H = H_c$.

[2] See, e.g., R. P. Feynman, Phys. Rev. **84**, 108—28 (1951). Appendix B; H. Davies, Proc. Camb. Phil. Soc. **59**, 147—55 (1963); C. Garrod, Rev. Mod. Phys. **38**, 483—94 (1966); and M. M. Mizrahi, Ref. 1.

[3] See, e.g., N. Bourbaki, *Elêments de mathématiques* (Hermann, Paris, 1969), Vol. XXXV, Book VI, Chap. IX.

[4] C. DeWitt-Morette, Comm. Math. Phys. **28**, 47—67 (1972).

[5] C. DeWitt-Morette, Comm. Math. Phys. **37**, 63—81 (1974).

[6] M. M. Mizrahi, "An Investigation of the Feynman Path Integral Formulation of Quantum Mechanics," Ph. D. dissertation, the University of Texas at Austin, August 1975, unpublished.

[7] M. M. Mizrahi, J. Math. Phys. **17**, 566—75 (1976).

[8] C. DeWitt-Morette, Ann. Phys. (N.Y.) **97**, 367—99 (1976).

[9] M. M. Mizrahi, "WKB Expansions by Path Integrals, with applications to the Anharmonic Oscillator," preprint, University of Texas at Austin and Center for Naval Analyses of the University of Rochester.

[10] Only in the case of the Wiener integral (no "i's" in the exponent) is a bona fide measure obtained. In the case of the Feynman path integral, the imaginary Gaussian measures on $\mathbb{R}^n$, building blocks of the promeasure one hopes to turn into a measure, are not bounded. This fact makes this attempt at mathematical legalization fall through. However, when one works with the Fourier transforms of the promeasure, the boundedness requirement is no longer needed, and progress can be made for computational purposes. C. DeWitt-Morette calls the resulting objects "pseudomeasures," P. Krée [Bull. Soc. Math. France **46**, 143—62 (1976)] calls them "prodistributions." For simplicity we call them "measures," as they are formally used as such.

[11] M. M. Mizrahi, J. Math. Phys. **16**, 2201—06 (1975).

[12] This is a very simple application of more general restrictions on the use of the given WKB approximation formula to a certain class of correspondence rules between the classical and quantum Hamiltonians, found in M. M. Mizrahi, J. Math. Phys. **18**, 786—90 (1977).

[13] In "Path Integration in Phase Space," by C. DeWitt-Morette, A. Maheshwari, and B. Nelson, preprint (to appear in Gen. Rel. Grav.), a similar measure is presented using a different approach. This paper and the present one, written independently, complement each other and should be read concurrently.

[14] This formula can be proved by path integrals—see Ref. 7.

[15] That some Hamiltonian operators do not admit a WKB approximation of their corresponding propagators is shown in Ref. 12.

[16] A double integral, corresponding to the double summation in (A1), can be easily obtained by replacing $(q(t), p(t))$ in (A3) by $\int_T \delta(t-t')(q(t'), p(t'))\,dt'$.

[17] In the case of the free particle in momentum space, a rare case where a measure in momentum space alone can be used, we have

$$dw(p) \sim [dp/N']\exp[(i/2m\hbar)\textstyle\int_T p^2(t)\,dt].$$

The operator corresponding to C is then simply the constant m^{-1}. Its inverse in the sense of (A4) is the constant m/T. It is the negative of $G_p(t,t')$ for the free particle [Eq. (78)] because $p^2/2m$ appears with a different sign in (11).

The linear potential: A solution in terms of combinatorics functions

Adel F. Antippa [a]

Département de Physique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, G9A 5H7, Canada

Alain J. Phares [b]

Department of Physics and Astronomy, University of Montana, Missoula, Montana 59801
(Received 28 December 1976)

The mathematical formalism recently developed for solving multiterm linear recursion relations is used to obtain a solution of the linear potential problem, for any value of the orbital quantum number l, in terms of combinatorics functions. The wavefunctions are given as power series expansions and the energy eigenvalues as the roots of an infinite order polynomial.

I. INTRODUCTION

For a long time the quark confining potential has been taken as a harmonic potential (mainly because it is exactly soluble in both the relativistic and nonrelativistic cases). But, during the past few years, strong theoretical reasons[1] have emerged favoring a linear central confining potential. This gave the linear potential the same importance in particle physics as the Coulomb potential in atomic physics and the harmonic potential in solid state physics. But, while series solutions of the Schrödinger equation for these latter potentials lead to two-term recursion formulas, that of the linear potential leads to a three-term recursion formula with variable coefficients.[2]

The importance that the linear potential has acquired recently, motivated the development of a general formalism for solving multiterm linear recursion relations with variable coefficients.[3] In this paper, we use this formalism to obtain a solution for the linear potential, with arbitrary values of the orbital quantum number l, in terms of combinatorics functions. Previous to this work, and in the context of the Charmonium model,[4] Harrington, Park, and Yildiz[5] obtained an exact solution in the case $l=0$, and a WKB solution for $l\neq 0$. A numerical solution for the combined Coulomb and linear potentials has been given by Eichten *et al.*[6] Extensive work on confining potentials in general, in both nonrelativistic and relativistic cases, has been done by Müller–Kirsten[7] in the framework of the WKB approximation. Gunion and Li[8] as well as Tryon[9] have obtained analytic asymptotic results in the relativistic case. Finally, Critchfield[10] studied the problem of scalar linear potentials in the Dirac equation and solved the resulting recursion relations using numerical methods.

The problem is set up in Sec. II. In Sec. III, we give the explicit expressions for the coefficients of the power series expansion of the wavefunction. In Sec. IV, we derive the energy eigenvalue equation. The computational details are left to Appendices.

[a] Supported in part by the National Research Council of Canada.

[b] Present address: Department of Physics, Villanova University, Villanova, Pennsylvania 19085.

II. GENERAL SOLUTION AND BOUNDARY CONDITIONS

Consider a quark–antiquark system interacting via a linear potential

$$V(r)=V_0+Kr. \tag{2.1}$$

The sum of the quark masses is $M=m_1+m_2$ and their reduced mass is $\mu=m_1m_2/(m_1+m_2)$. Let x and t_{nl} be the dimensionless distance and energy parameters defined through

$$E_{nl}=\left(\frac{K^2}{2\mu}\right)^{1/3}t_{nl}+(M+V_0) \quad \text{and} \quad x=(2\mu K)^{1/3}r. \tag{2.2}$$

E_{nl} are the energy eigenvalues and are labelled by the radial and orbital quantum numbers n and l, respectively. The eigenfunctions of the Hamiltonian for this system are then given by

$$\psi_{nlm}(\mathbf{r})=R_{nl}(r)Y_l^m(\theta,\phi) \quad \text{with} \quad R_{nl}(r)=\frac{u_{nl}(x)}{x}, \tag{2.3}$$

where $Y_l^m(\theta,\phi)$ are the spherical harmonics, $u_{nl}(x)=u_l(x,t_{nl})$, and $u_l(x,t)$ is a solution of the radial differential equation,

$$\left[\frac{d^2}{dx^2}-\frac{l(l+1)}{x^2}-(x-t)\right]u_l(x,t)=0. \tag{2.4}$$

For $l=0$, the solution[5] of Eq. (2.4) with the correct asymptotic behavior is the Airy function $Ai(x-t_{n0})$. Imposing physical threshold behavior gives the energy eigenvalue condition

$$Ai(-t_{n0})=0. \tag{2.5}$$

For $l\neq 0$, the solution of Eq. (2.4) cannot be related to known tabulated functions. A new class of functions have to be considered. These are related to the special combinatorics functions of the first and second kind which were introduced in Ref. 3.

A. General solution

The general solution of Eq. (2.4) is[11]

$$u_l=Aw_1+Bw_2, \tag{2.6}$$

where w_1 and w_2 are two linearly independent solutions that are given by

$$w_1=x^{l+1}\sum_{m=0}^{\infty}b_mx^m, \tag{2.7a}$$

 0022-2488/78/1901-0308$1.00

$$w_2 = \kappa w_1(x) \ln x + x^{-l} \sum_{m=0}^{\infty} c_m x^m. \tag{2.7b}$$

The coefficients b_m are given by

$$m(m+2l+1)b_m + tb_{m-2} - b_{m-3} = 0 \quad \text{for } m > 0 \tag{2.8a}$$

with the initial conditions

$$b_m = b_0 \delta_{m0} \quad \text{for } m \leq 0 \tag{2.8b}$$

and the coefficients c_m are given by

$$m(m-2l-1)c_m + tc_{m-2} - c_{m-3} = 0 \quad \text{for } 0 \leq m \leq 2l \tag{2.9a}$$

and

$$m(m+2l+1)c_{m+2l+1} + tc_{m+2l-1} - c_{m+2l-2} = (2m+2l+1)\kappa b_m \quad \text{for } m \geq 0 \tag{2.9b}$$

with the initial conditions

$$c_m = c_0 \delta_{m0} \quad \text{for } m \leq 0. \tag{2.9c}$$

$\kappa = \kappa(l, t)$ is given in Appendix A, and is in general different from zero.

B. Boundary conditions

(i) *Condition at the origin*: As can be seen from Eq. (2.3), $u_l(x,t)/x$ has to be regular at the origin.

(ii) *Condition at infinity*: Asymptotically, the differential equation (2.4) goes into the equation

$$u_l'' - (x-t)u_l = 0. \tag{2.10}$$

That is, the term $l(l+1)/x^2$, but not t, can be neglected with respect to x. The reason is that t enters in the leading term of the asymptotic solution of Eq. (2.4), while l does not. The solution of Eq. (2.10) which is regular at infinity is the Airy function[12] $Ai(x-t)$. Therefore, the boundary condition at infinity can be expressed as

$$\lim_{x\to\infty} \frac{u_l(x,t)}{\mathrm{Ai}(x-t)} = \text{const}. \tag{2.11}$$

The constants A and B can be determined from the boundary condition at the origin and the normalization of the wavefunction. The boundary condition at infinity then determines the eigenenergies.

III. WAVE FUNCTIONS

The boundary condition at the origin requires that the coefficient B, in Eq. (2.6), vanish. Hence $u_l(x,t)$ is given by

$$u_l(x,t) = x^{l+1} v_l(x,t), \quad v_l(x,t) = \sum_{m=0}^{\infty} b_m(l,t) x^m, \tag{3.1}$$

where the coefficient A has been absorbed in $v_l(x,t)$.

The recursion relation for the b_m is given, through Eqs. (2.8a) and (2.8b) as

$$b_m = f_2(m) b_{m-2} + f_3(m) b_{m-3} \quad \text{for } m > 0 \tag{3.2a}$$

with the initial conditions

$$b_m = \lambda_0 \delta_{m0} \quad \text{for } m \leq 0, \tag{3.2b}$$

where λ_0 is the normalization constant. The functions $\{f_i(m); i = 2, 3\}$ are given by

$$f_2(m) = -tf(m,l), \quad f_3(m) = f(m,l) \tag{3.3a}$$

with

$$f(m,l) = \frac{1}{m(m+2l+1)}. \tag{3.3b}$$

The theory of Ref. 3 gives explicit expressions for the expansion coefficients $b_m(l,t)$. This in turn provides an explicit solution for $u_l(x,t)$, and, once the eigenenergies are determined, gives the wavefunctions through Eq. (2.3).

According to Ref. 3 the solution of Eq. (3.2) is related to the partitions of the interval $(0, m)$ into subintervals of lengths 2 and 3. We note that for m integer the only interval that cannot be partitioned in parts of lengths 2 and 3 is $(0,1)$ thus leading to the vanishing of the coefficient b_1. This can also be easily seen from the recursion relation (3.2a) and the initial condition (3.2b).

Consider one possible partition of $(0, m)$ into n parts, say $(\delta_1, \delta_2, \ldots, \delta_n)$, with

$$\delta_j \in \{2, 3\}, \quad j = 1, \ldots, n \tag{3.4a}$$

and

$$\sum_{j=1}^{n} \delta_j = m. \tag{3.4b}$$

Let p be the number of 2-subintervals in the partition, then, $(n-p)$ is the number of 3-subintervals involved, and, according to Eq. (3.4)

$$2p + 3(n-p) = m \quad \text{or} \quad p = 3n - m. \tag{3.5}$$

From Eq. (3.5) it is seen that there is a unique number p of 2-subintervals in a partition of $(0, m)$ having n parts. Therefore, the number of distinct partitions of $(0, m)$ into n parts is the binomial

$$\binom{n}{p} = \frac{n!}{p!(n-p)!} = \frac{n!}{(3n-m)!(m-2n)!}. \tag{3.6}$$

Consequently if we label by q the distinct partitions of $(0, m)$ into n parts, q will range from 1 to $q_{\max}(n) = \binom{n}{p}$. The values of n for which a partition is possible are included in the interval.

$$n_{\min} = \left[\frac{m+2}{3}\right] - \delta_{m1} \leq n \leq \left[\frac{m}{2}\right] = n_{\max} \quad \text{for } m > 0, \tag{3.7}$$

where the square brackets refer to integer division. n decreases from $n_{\max}$ to $n_{\min}$ by successively exchanging three 2-subintervals by two 3-subintervals. Thus as n changes by units of one, holding m fixed, p changes by steps of 3 as can easily be seen from Eq. (3.5).

Associated with the qth partition of $(0, m)$ into n parts we introduce the set of integers s_j given by,

$$s_0 = 0, \quad s_j = \sum_{i=1}^{j} \delta_j, \quad s_n = m, \tag{3.8}$$

and define the functions

$$F_n^q(0, m) = \prod_{j=1}^{n} f_{\delta_j}(s_j). \tag{3.9}$$

Combining Eqs. (3.3), (3.9), and the preceding analysis on the number of 2-subintervals in a partition we can rewrite this as

$$F_n^q(0, m) = (-t)^{3n-m} \prod_{j=1}^{n} f(s_j, l). \tag{3.10}$$

The special combinatorics functions of the first and second kind are then given by[3]

$$C_2(0,m)=\sum_{n=n_{\min}}^{n_{\max}} C_1(0,m,n)=\sum_{n=n_{\min}}^{n_{\max}}\sum_{q=1}^{q_{\max}(n)} F_n^q(0,m) \tag{3.11}$$

and according to Theorem 4 of Ref. 3 the solution of the recursion relation (3.2a) subject to the initial conditions (3.2b) is given by

$$b_m=\lambda_0 C_2(0,m). \tag{3.12}$$

This can be rewritten as

$$b_m(l,t)=\lambda_0\sum_{n=n_{\min}}^{n_{\max}}\beta_l(0,m;n)(-t)^{3n-m}, \tag{3.13}$$

where the "structure" functions β_l are given by

$$\beta_l(0,m;n)=\sum_{q=1}^{q_{\max}(n)}\prod_{j=1}^{n} f(s_j,l) \tag{3.14a}$$

and

$$\beta_l(0,0;0)=1. \tag{3.14b}$$

If the interval $(0,m)$ cannot be partitioned into n parts, then $\beta_l(0,m;n)=0$. We thus obtain an explicit expression for the coefficients of the power series expansion in terms of two parameters: the normalization constant λ_0, and the energy parameter t.

Before proceeding any further, it may be instructive to explicitly calculate one of the coefficients b_m. In Fig. 1 we give the various intervals and parameters related to the evaluation of b_8. Using this figure and Eqs. (3.10), (3.11), and (3.12) we find

$$b_8=\lambda_0\sum_{n=3}^{4}\sum_{q=1}^{q_{\max}(n)} F_n^q(0,8) \tag{3.15a}$$

or

$$b_8=\lambda_0\left(\sum_{q=1}^{3} F_3^q(0,8)+F_4^1(0,8)\right) \tag{3.15b}$$

with

$$F_3^1(0,8)=-tf(2,l)f(5,l)f(8,l)=\frac{-t}{2\cdot 5\cdot 8\cdot(2+2l+1)(5+2l+1)(8+2l+1)}, \tag{3.16a}$$

$$F_3^2(0,8)=-tf(3,l)f(5,l)f(8,l)=\frac{-t}{3\cdot 5\cdot 8(3+2l+1)(5+2l+1)(8+2l+1)}, \tag{3.16b}$$

$$F_3^3(0,8)=-tf(3,l)f(6,l)f(8,l)=\frac{-t}{3\cdot 6\cdot 8(3+2l+1)(6+2l+1)(8+2l+1)}, \tag{3.16c}$$

$$F_4^1(0,8)=t^4 f(2,l)f(4,l)f(6,l)f(8,l)=\frac{t^4}{2\cdot 4\cdot 6\cdot 8\cdot(2+2l+1)(4+2l+1)(6+2l+1)(8+2l+1)}. \tag{3.16d}$$

$m=8 \quad n_{\min}=\left[\frac{m+2}{3}\right]=3 \quad n_{\max}=\left[\frac{m}{2}\right]=4 \quad 3\le n\le 4$

n	p=3n-m	$q_{\max}(n)=\binom{n}{p}$
3	1	3
4	4	1

n=3: q=1: 2 3 3 → $\vec{s}_{31}=(2,5,8)$; q=2: 3 2 3 → $\vec{s}_{32}=(3,5,8)$; q=3: 3 3 2 → $\vec{s}_{33}=(3,6,8)$

n=4: q=1: 2 2 2 2 → $\vec{s}_{41}=(2,4,6,8)$

FIG. 1. The partitions and parameters necessary for the evaluation of the coefficient b_8 as given by Eq. (3.15a).

By considering even and odd values of m separately, Eq. (3.13) can be rewritten in a form that brings out the polynomial dependence of b_m on t as follows:

$$b_{2k}=\lambda_0\sum_{j=0}^{[k/2]}\beta_l(0,2k;k-j)(-t)^{k-3j} \quad \text{for } k\ge 1, \tag{3.17a}$$

$$b_{2k+1}=\lambda_0\sum_{j=0}^{[(k-1)/3]}\beta_l(0,2k+1;k-j)(-t)^{k-1-3j} \quad \text{for } k\ge 1. \tag{3.17b}$$

The upper limit on j essentially guarantees that only positive powers of t are included. Thus b_m is a polynomial in positive powers of t of order $m/2$ for m even, and $((m-1)/2-1)$ for m odd, and with sucessive powers differing by three units. In Appendix B we present, and illustrate, a method for obtaining explicit expressions for the structure functions $\beta_l(0,m;n)$.

IV. EIGENVALUE EQUATION

To derive the eigenvalue equation we need to express $b_m(l,t)$ in terms of higher order coefficients. This can be done by an extension of the methods of Ref. 3. We first invert the recursion relation (3.2a) to obtain

$$b_m=g_1(m)b_{m+1}+g_3(m)b_{m+3} \quad \text{for } m\ge -2, \tag{4.1a}$$

where the functions $\{g_i(m);i=1,3\}$ are given by

$$g_1(m)=t, \quad g_3(m)=g(m,l) \tag{4.1b}$$

with

$$g(m,l)=(m+3)(m+2l+4). \tag{4.1c}$$

As shown in Appendix C, the solution of Eq. (4.1) is

$$b_m=\sum_{j=0}^{2}\bar{C}_2^*(N+j,m,j)b_{N+j}, \tag{4.2}$$

where N is greater than m but otherwise arbitrary, and the $\bar{C}_2^*$ are, what we call, the conjugate combinatorics functions. The eigenvalue equation is obtained by relating the initial condition at the origin to the boundary condition at infinity by using Eq. (4.3).

Using Eq. (2.8a), the initial conditions (2.8b) can be expressed as,

$$2(l+1)b_1/(-t)=b_{-1}=0. \tag{4.3}$$

It is not difficult to see that this is related to the boundary condition at the origin. On the other hand, the boundary condition at infinity, as expressed by Eq.

(2.11), determines the coefficients $b_N(l,t)$, for large values of N, in terms of the expansion coefficients, a_n, of the Airy function, Ai, according to

$$b_N(l,t) \underset{N\to\infty}{\to} \sum_{i=0}^{\infty} a_{N+i+l+1}\binom{N+i+l+1}{i}(-t)^i. \tag{4.4}$$

The Airy expansion coefficients, a_n, satisfy the recursion relation[12]

$$a_n = \frac{a_{n-3}}{n(n-1)} \quad \text{for } n \geq 3, \tag{4.5a}$$

subject to the initial conditions

$$\frac{a_1}{a_0} = -\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{1}{3})}\,3^{1/3}, \quad a_2 = 0. \tag{4.5b}$$

The solution of Eq. (4.5) is given by

$$a_{3p+i} = \frac{(-1)^i\Gamma(\frac{2}{3})(1-\delta_{i2})}{3^{2p+2i/3}p!\,\Gamma(p+\frac{2}{3}(1+i))}\,a_0$$

$$\text{for } i=0,1,2, \quad \text{and } p=0,1,\ldots,\infty. \tag{4.6}$$

Setting $m=1$ in Eq. (4.2), using Eq. (4.3) for b_1, taking the limit as $N\to\infty$, and using Eq. (4.4) for the b_{N+j}, we obtain the energy eigenvalue equation as

$$\lim_{N\to\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{2} 2(l+1)\bar{C}_2^*(N+j,1,j)a_{N+j+i+l+1}\binom{N+j+i+l+1}{i} \times(-t)^{i-1} = 0. \tag{4.7}$$

The conjugate combinatorics functions, $\bar{C}_2^*$, appearing in Eq. (4.7) are polynomials in t. Therefore, the left-hand side of this equation is an infinite order polynomial in t, which will be denoted $H_l(t)$, and can be written as

$$H_l(t) = \lim_{N\to\infty} H_l^{(N)}(t) = 0 \tag{4.8a}$$

with

$$H_l^{(N)}(t) = \sum_{n=0}^{\infty} 2(l+1)K_n^{(N)}(l)t^{n-1}. \tag{4.8b}$$

Combining Eqs. (4.8a) and (4.8b), exchanging the limit and summation signs, and defining

$$K_n(l) = \lim_{N\to\infty} K_n^{(N)}(l), \tag{4.8c}$$

we obtain the eigenvalue equation as

$$H_l(t) = \sum_{n=0}^{\infty} 2(l+1)K_n(l)t^{n-1} = 0. \tag{4.8d}$$

Thus, to complete the derivation of the eigenvalue equation, one still has to determine the coefficients $K_n^{(N)}(l)$. This is done by a reduction of $\bar{C}_2^*$; that is factoring out its t dependence.

According to Appendix C, the constrained conjugate combinatorics functions $\bar{C}_2^*(N+j,1,j)$, $j=0,1,2$, are related to the partitions of the interval $(1,N+j)$ into parts of lengths 1 and 3 subject to certain constraints. Consider one possible partition of $(1,N+j)$ into n parts, $(\delta_1,\ldots,\delta_n)$, subject to the three conditions

(i) $\delta_i \in \{1,3\}$, $\quad i=1,\ldots,n$, (4.9a)

(ii) $\delta_1 > j$, $\quad$ for $j=0$, 1, or 2, (4.9b)

(iii) $\sum_{i=1}^{n}\delta_i = N+j-1$, $\quad$ for $j=0$, 1, or 2. (4.9c)

As in the above equation, j will always be restricted to one of the values 0, 1, or 2. Let p be the number of parts of length 1 in the partition, then, $(n-p)$ is a number of parts of length 3, and, according to Eq. (4.9),

$$p+3(n-p) = N+j-1 \quad \text{or} \quad 2p = 3n-N-j+1. \tag{4.10}$$

From Eq. (4.10) it is seen that, in a partition of $(1,N+j)$ into n parts, there is a unique number p, and therefore a unique number $(n-p)$, of parts of length 1 and 3 respectively. Hence, the number of distinct partitions of $(1,N+j)$ into n parts subject to the constraints (4.9) is the binomial

$$\binom{n-1+\delta_{j0}}{p} = \frac{(n-1+\delta_{j0})!}{[(N+j-1-n)/2-1+\delta_{j0}]![(3n+1-N-j)/2]!}, \tag{4.11}$$

where δ_{j0} is the Kronecker delta and $j=0,1,2$. Consequently, if we label by q the distinct partitions of $(1,N+j)$ into n parts subject to the constraints (4.9), q will range from 1 to a $q_{\max}(n)$ given by the binomial of Eq. (4.11). The values of n for which a constrained partition is possible are included in the interval

$$n_{\min} = N+j-1-2\left[\frac{N+j-1}{3}\right] \leq n \leq N-1-\delta_{j1} = n_{\max}$$

$$\text{for } N \geq 3, \tag{4.12}$$

where as usual $[\nu]$ refers to the largest integer less than or equal to ν. n decreases from $n_{\max}$ to $n_{\min}$ by successively exchanging three parts of length 1 by one part of length 3. Therefore, n changes by steps of two units. The corresponding change in p, holding $(N+j)$ fixed, is in steps of three units. In addition, the minimum and maximum number of parts of length one is,

$$p_{\min} = (N+j-1) - 3\left[\frac{N+j-1}{3}\right],$$

$$p_{\max} = N-2+\delta_{j0}-\delta_{j1} \quad \text{for } N \geq 3, \tag{4.13}$$

while the minimum and maximum number of parts of length three is

$$(n-p)_{\min} = \tfrac{1}{3}(N+j-1-p_{\max}) = 1-\delta_{j0} \quad \text{for } N \geq 3, \tag{4.14a}$$

$$(n-p)_{\max} = \tfrac{1}{3}(N+j-1-p_{\min}) = \left[\frac{N+j-1}{3}\right] \quad \text{for } N \geq 3. \tag{4.14b}$$

Due to the above discussion, the possible values of p and n are given by

$$p = p_{\min} + 3n', \tag{4.15a}$$

$$n = n_{\min} + 2n', \tag{4.15b}$$

where

$$n' = 0,1,\ldots,n'_{\max} \tag{4.15c}$$

and

$$n'_{\max}=\left[\frac{N+j-1}{3}\right]+\delta_{j0}-1 \quad \text{for } N\geq 3. \tag{4.15d}$$

Associated with the qth partition of $(1,N+j)$ into n parts subject to conditions (4.9), we introduce the set of integers s_i given by

$$s_0=N+j, \quad s_i=s_0-\sum_{r=1}^{i}\delta_r, \quad s_n=1, \tag{4.16}$$

and define the functions

$$G_n^q(N+j,1)=\prod_{r=1}^{n} g_{\delta_r}(s_r). \tag{4.17}$$

Combining Eqs. (4.2), (4.17), and the preceding analysis on the partitions of $(1,N+j)$ into parts $\delta_i\in\{1,3\}$, we can rewrite this as

$$G_n^q(N+j,1)=t^p\prod_{i=1}^{n}{}' g(s_i,l), \tag{4.18a}$$

where $\prod'_{i=1}^{n}$ stands for the product over only those values of i that satisfy the constraint

$$s_{i-1}-s_i=3. \tag{4.18b}$$

The constrained conjugate combinatorics functions, $\bar{C}_2^*$, are then given by (see Appendix C)

$$\bar{C}_2^*(N+j,1,j)=\sum_{n=n_{\min}}^{n_{\max}}\sum_{q=1}^{q_{\max}(n)} G_n^q(N+j,1), \tag{4.19}$$

where it is understood that n will range from $n_{\min}$ to $n_{\max}$ by steps of two units. To reduce the $\bar{C}_2^*$, we introduce the constrained conjugate structure functions, $\bar{\beta}_l^*$, by

$$\bar{\beta}_l^*(N+j,1;n)=\sum_{q=1}^{q_{\max}(n)}\prod_{i}{}' g(s_i,l). \tag{4.20}$$

Using Eqs. (4.15), (4.18), and (4.20), we can rewrite Eq. (4.19) as

$$\bar{C}_2^*(N+j,1,j)=t^{p_{\min}}\sum_{n'=0}^{n'_{\max}} t^{3n'}\bar{\beta}_l^*(N+j,1;n_{\min}+2n') \quad \text{for } N\geq 3. \tag{4.21}$$

This completes the reduction of the $\bar{C}_2^*$.

Combining Eqs. (4.7), (4.8), and (4.21), we obtain

$$H_l^{(N)}(t)=\sum_{i=0}^{\infty}\sum_{j=0}^{2}\sum_{n'=0}^{n'_{\max}} 2(l+1)(-1)^{i+1}\bar{\beta}_l^*(N+j,1;n_{\min}+2n')\times a_{N+j+i+l+1}\binom{N+j+i+l+1}{i} t^{p_{\min}+3n'+i-1}. \tag{4.22}$$

To proceed further in the reduction of (4.22) it is convenient to choose $N=3k+1$, where k is a positive integer. Equations (4.12), (4.13), and (4.15d), then lead to $n_{\min}=k+j$, $p_{\min}=j$, and $n'_{\max}=k+\delta_{j0}-1$. Thus, setting $i=3r'+j'$, Eq. (4.22) can be rewritten as

$$H_l^{(3k+1)}(t)=\sum_{r'=0}^{\infty}\sum_{j'=0}^{2}\sum_{j=0}^{2}\sum_{n'=0}^{k+\delta_{j0}-1} 2(l+1)(-1)^{j'+r'+1}\times\bar{\beta}_l^*(3k+j+1,1;k+j+2n')a_{3k+j+j'+l+2+3r'}\times\binom{3k+j+j'+l+2+3r'}{3r'+j'} t^{3(r'+n')+j+j'-1}. \tag{4.23}$$

We now make the change of variable $r'\to r=r'+n'$, and follow this by exchanging the resulting summation over r with that over n' to obtain

$$H_l^{(3k+1)}(t)=\sum_{r=0}^{\infty}\sum_{j=0}^{2}\sum_{j'=0}^{2}\sum_{n'=0}^{[r,\,k+\delta_{j0}-1]_<} 2(l+1)(-1)^{r-n'+j'+1}\times\bar{\beta}_l^*(3k+j+1,1;k+j+2n')a_{3(k-n'+r)+j+j'+l+2}\times\binom{3(k-n'+r)+j+j'+l+2}{3r-3n'+j'} t^{3r+j+j'-1}, \tag{4.24}$$

where $[\mu,\nu]_<$ stands for the smaller of μ and ν. It is worthwhile remembering that, the conjugate structure functions $\bar{\beta}_l^*(3k+j+1,1;k+j+2n')$ are evaluated over partitions that are constrained by condition (4.9b). This constraint requires that all partitions of $(3k+1+j,1)$ into $(k+j+2n')$ parts have their first part δ_1 greater than j. For $j=0$, δ_1 can be either 1 or 3; that is, the constraint is automatically satisfied. On the other hand, if $j=1$ or 2, then δ_1 has to be of length 3. Thus,

$$\bar{\beta}_l^*(3k+j+1,1;k+j+2n')\big|_{j=0}=\beta_l^*(3k+1,1;k+2n'), \tag{4.25}$$

$$\bar{\beta}_l^*(3k+j+1,1;k+j+2n')=(3k+j+1)(3k+j+2l+2)\beta_l^*(3k-2+j,1;k+j-1+2n') \quad \text{for } j=1,2, \tag{4.26}$$

where β_l^* is an unconstrained conjugate structure function. A detailed study of these functions will be given in Appendix D.

Finally, by comparing Eqs. (4.8b) and (4.24), and making use of Eqs. (4.25) and (4.26), we obtain the the coefficients $K_n^{(N)}(l)$ as

$$\begin{aligned}K_{3r+i}^{(3k+1)}(l)&=\sum_{n'=0}^{[r,\,k]_<}(-1)^{r-n'+i+1}\beta_l^*(3k+1,1;k+2n')a_{3(k-n'+r)+l+2+i}\times\binom{3(k-n'+r)+l+2+i}{3r-3n'+i}-(3k+2)(3k+2l+3)\\&\times\sum_{n'=0}^{[r-\delta_{i0},\,k-1]_<}(-1)^{r-n'+i+1}\beta_l^*(3k-1,1;k+2n')\\&\times a_{3(k-n'+r)+l+2+i}\binom{3(k-n'+r)+l+2+i}{3r-3n'-1+i}\\&+(3k+3)(3k+2l+4)\sum_{n'=0}^{[r-1+\delta_{i2},\,k-1]_<}(-1)^{r-n'+i+1}\\&\times\beta_l^*(3k,1;k+2n'+1)\\&\times a_{3(k-n'+r)+l+2+i}\binom{3(k-n'+r)+l+2+i}{3r-3n'-2+i}\quad \text{for } i=0,1,2.\end{aligned} \tag{4.27}$$

It is interesting to remark that using Eq. (4.6) one immediately obtains the results

$$K_{3r}^{(3k+1)}(3l')=0, \quad K_{3r+2}^{(3k+1)}(3l'+1)=0, \quad K_{3r+1}^{(3k+1)}(3l'+2)=0, \tag{4.28}$$

where l' is a nonnegative integer. Since Eqs. (4.28) remain evidently valid in the limit, $k\to\infty$, then for every value of l, there is a whole set of coefficients, in the eigenvalue equation (4.8d), that vanish.

This completes the formal development of the eigen-

value equation. A detailed study of the coefficients $K_{3r+l}^{(3k+1)}(l)$ will be given in Appendix E.

V. CONCLUSION

We have given explicit expressions for the nonrelativistic linear potential wavefunctions and eigenenergy equation. The wavefunctions were obtained by a direct application of the general formalism for solving linear recursion relations.[3] As such, they provide a first concrete application of this method, and the techniques involved in the reduction of the combinatorics functions, which form the basis in which the solution is expressed.

To obtain the eigenenergy equation, a general approach to the eigenvalue problem was used. The usual method of quantizing by terminating the power series expansion after a finite number of terms only works when the coefficients in the expansion are related by a two-term recursion relation. The coefficients in the power series expansion arising in the linear potential problem are related by a three-term recursion relation, and hence the series cannot be terminated. Instead we directly connected the initial conditions at the origin to the boundary conditions at infinity, thus obtaining the eigenenergies as the roots of an infinite order polynomial. The boundary condition at infinity in this case requires that the wavefunction goes over asymptotically into the Airy function $\mathrm{Ai}(x - t)$.

To connect the boundary conditions at the origin and at infinity we had to generalize the results of Ref. 3 to the case of inverted recursion relations, and obtained their solution through the conjugate combinatorics functions.

APPENDIX A: GENERAL EXPRESSION FOR THE CONSTANT $\kappa(l, t)$

By requiring that w_2 be a solution of the radial equation (2.6), we arrived at the recursion relations,

$$(2m + 2l + 1)\kappa b_m + m(m + 2l + 1)c_{m+2l+1} + tc_{m+2l-1} - c_{m+2l-2} = 0 \quad \text{for } m \geq 0 \tag{A1}$$

and

$$m(m - 2l - 1)c_m + tc_{m-2} - c_{m-3} = 0 \quad \text{for } 0 \leq m \leq 2l \tag{A2}$$

with the initial conditions

$$c_m = c_0\delta_{0m} \quad \text{for } m \leq 0. \tag{A3}$$

Equation (A2) is more conveniently separated into three equations:

$$m(m - 2l - 1)c_m + tc_{m-2} - c_{m-3} = 0 \quad \text{for } 0 \leq m \leq 2l \text{ and } l \geq 2, \tag{A4a}$$

$$m(m - 3)c_m + tc_{m-2} = 0 \qquad \text{for } 0 \leq m \leq 2 \text{ and } l = 1, \tag{A4b}$$

$$0 = 0 \Rightarrow c_0 = \text{arbitrary} \qquad \text{for } l = 0. \tag{A4c}$$

If we now set $m = 0$ in Eq. (A1) we obtain

$$(2l + 1)\kappa b_0 + tc_{2l-1} - c_{2l-2} = 0. \tag{A5}$$

The explicit expressions for c_{2l-1} and c_{2l-2} can be obtained by solving the recursion relations (A4). For $l = 0$, Eqs. (A3) and (A5) give $\kappa b_0 = 0$, and since $b_0 \neq 0$, this implies that

$$\kappa(0, t) = 0. \tag{A6a}$$

For $l = 1$, Eqs. (A3), (A4b), and (A5) give

$$\kappa(1, t) = \frac{1}{3}\frac{c_0}{b_0}. \tag{A6b}$$

For $l \geq 2$ we note that Eq. (A4a) is of the same type as the one appearing in Eq. (3.2), and can be rewritten as

$$c_m = f_2(-m)c_{m-2} + f_3(-m)c_{m-3} \quad \text{for } 0 < m \leq 2l, \tag{A7}$$

where the functions f_2 and f_3 are the same as the ones defined by Eq. (3.3) and c_0 is arbitrary. Thus the whole discussion given in Sec. III follows through, leading to the solution of Eq. (A7) as

$$c_m = c_0 \sum_{n=n_{\min}}^{n_{\max}} \beta_l'(0, m; n)(-t)^{3n-m} \quad \text{for } 0 < m \leq 2l. \tag{A8}$$

The structure functions $\beta_l'(0, m; n)$ are given by

$$\beta_l'(0, m; n) = \sum_{q=1}^{q_{\max}(n)} \prod_{j=1}^{n} f(-s_j, l), \tag{A9}$$

where the function f is given by Eq. (3.3b), $n_{\min}$ and $n_{\max}$ are given by Eq. (3.7), and $q_{\max}(n)$ by Eq. (3.6). The expression for $\beta_l'(0, m; n)$ is analogous to that for $\beta_l(0, m; n)$ given in Eq. (3.14), and the underlying set of partitions is the same in both cases. Combining Eqs. (A5) and (A8) we have

$$\kappa(l, t) = \frac{c_0}{(2l + 1)b_0} \sum_{n=\lceil 2l/3 \rceil}^{l-1} \{\beta_l'(0, 2l - 1; n) + \beta_l'(0, 2l - 2; n)\} \times (-t)^{3n-2l+2}, \quad l \geq 2 \tag{A10}$$

where $\lceil \nu \rceil$ stands for the largest integer less than or equal to ν. In evaluating the lower limit of the summation it should be noted that when $2l + 1$ is an exact multiple of 3, the interval $(2l - 1)$ cannot be divided into $[2l/3]$ parts belonging to the set $\{2, 3\}$, and hence the corresponding structure function $\beta_l'(0, 2l - 1; [2l/3])$ is zero.

As an example we will evaluate Eq. (A10) for $l = 2$. We find

$$\kappa(2, t) = \frac{c_0}{5b_0}[\beta_2'(0, 3; 1) + \beta_2'(0, 2; 1)](-t).$$

From Eqs. (A9), (3.3b), and (3.6) we find

$$\beta_2'(0, 3; 1) = f(-3, 2) = -\tfrac{1}{6},$$

$$\beta_2'(0, 2; 1) = f(-2, 2) = -\tfrac{1}{6},$$

leading to

$$\kappa(2, t) = \frac{t}{15}\frac{c_0}{b_0}. \tag{A11}$$

APPENDIX B: THE STRUCTURE FUNCTIONS $\beta_l(0, m; n)$

According to Eq. (3.14) the structure functions are given by

$$\beta_l(0, m; n) = \sum_{q=1}^{q_{\max}(n)} \prod_{j=1}^{n} f(s_j, l). \tag{B1}$$

Comparing Eq. (B1) with Eq. (3.11) and making use of

Eq. (3.10) we find that

$$C_1(0, m, n) = (-1)^{3n-m}\beta_1(0, m; n). \tag{B2}$$

Consequently by using Eq. (3.3) and the recursion relation obeyed by the combinatorics functions of the first kind as given in Theorem 1 of Ref. 3, it is easy to derive the following recursion relation for the structure functions,

$$\beta_l(0, m; n) = [\beta_l(0, m-2; n-1) + \beta_l(0, m-3; n-1)]f(m, l), \tag{B3}$$

By construction the structure functions $\beta_l(0, m; n)$ are evaluate over the partitions of the interval $(0, m)$ into n parts of lengths 2 and 3. If p represents the number of parts of length 2, then, according to Eq. (3.5), the only intervals that can be partitioned into a given number, n, of parts are those of length

$$m = 3n - p \quad \text{for } p = 0, 1, \ldots, n. \tag{B4}$$

We will therefore study the structure functions $\beta_l(0, 3n-p; n)$ for the above range of values of p, the other structure functions being zero.

There are two special cases in which the structure functions take a simple form. These correspond to $p=0$ and $p=n$, where $q_{\max}(n)=1$, leading to

$$\beta_l(0, 3n; n) = \prod_{j=0}^{n-1} f(3n-3j, l), \tag{B5a}$$

$$\beta_l(0, 2n; n) = \prod_{j=0}^{n-1} f(2n-2j, l). \tag{B6a}$$

Making use of Eq. (3.3b) we then obtain

$$\beta_l(0, 3n; n) = \frac{1}{3^{2n}} \frac{\Gamma(1 + (2l+1)/3)}{\Gamma(n+1)\Gamma(n+1+(2l+1)/3)}, \tag{B5b}$$

$$\beta_l(0, 2n; n) = \frac{1}{2^{2n}} \frac{\Gamma(1 + (2l+1)/2)}{\Gamma(n+1)\Gamma(n+1+(2l+1)/2)}. \tag{B6b}$$

In the general case we derive a relationship between the structure functions corresponding to partitions with p parts of length 2, of a given interval, and those corresponding to partitions with $(p-1)$ parts of length 2, of its subintervals. Let p_i be the set of distinct partitions of $(0, m)$ into n parts, p of which are of length 2, and subject to the two following conditions: (i) The last i parts are all of length 3, and (ii) the $(n-i)$th part is of length 2, as shown in Fig. 2. Thus if P is the set of all distinct partitions of $(0, m)$ into n parts, p of which

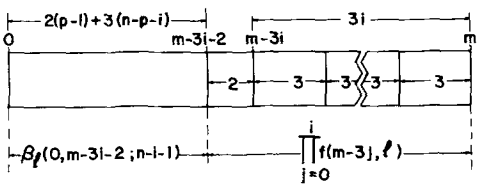

FIG. 2. The partitions belonging to the set P_i and their corresponding representation by structure functions.

are of length 2, we have

$$P = \bigcup_{i=0}^{n-p} P_i \tag{B7a}$$

since the maximum possible number of parts of 3 is $n-p$, and the minimum is 0. Notice that the second condition on the elements of the set P_i, on the one hand requires that $p \geq 1$, and on the other hand guarantees that these sets are disjoint,

$$P_i \cap P_j = 0 \quad \text{for } i \neq j. \tag{B7b}$$

The sum over q in Eq. (B1) is a sum over all distinct partitions of $(0, m)$ into n parts belonging to $\{2, 3\}$. Thus due to Eqs. (B7a) and (B7b), we can rewrite Eq. (B1) as

$$\beta_l(0, m, n) = \sum_{i=0}^{n-p} \sum_{\text{partitions}\in P_i} \prod_{j=1}^{n} f(s_j, l). \tag{B8}$$

But as can be seen from Fig. 2,

$$\sum_{\text{partitions}\in P_i} \prod_{j=1}^{n} f(s_j, l) = \left[\sum_{\text{partitions}\in P_i} \prod_{j=1}^{n-i-1} f(s_j, l)\right] \prod_{j=n-i}^{n} f(s_j, l). \tag{B9}$$

Furthermore

$$\sum_{\text{partitions}\in P_i} \prod_{j=1}^{n-i-1} f(s_j, l) = \beta_l(0, m-3i-2; n-i-1) \tag{B10a}$$

and

$$\prod_{j=n-i}^{n} f(s_j, l) = \prod_{j=0}^{i} f(m-3j, l). \tag{B10b}$$

Combining Eqs. (B8), (B9), (B10a), and (B10b) we obtain

$$\beta_l(0, m; n) = \sum_{i=0}^{n-p} \beta_l(0, m-3i-2; n-i-1) \prod_{j=0}^{i} f(m-3j, l) \qquad 1 \leq p \leq n. \tag{B11}$$

Alternatively we can prove Eq. (B11) by mathematical induction on n, holding p fixed, and remembering that $m = 3n - p$. The proof is straightforward and simple and is based on Eq. (B3).

We now rewrite Eq. (B11) in a form that is more convenient for calculations by using Eq. (B4) and making the changes of variable $i \to n-p-i$ and $j \to n-p-j$, to obtain

$$\beta_l(0, 3n-p; n) = \sum_{i=0}^{n-p} \beta_l(0, 3i+2p-2; i+p-1) \times \prod_{j=i}^{n-p} f(3j+2p, l). \tag{B12}$$

Furthermore

$$\prod_{j=i}^{n-p} f(3j+2p, l) = \frac{3^{2(p+i-1)}\Gamma(i+2p/3)\Gamma(i+(2p+2l+1)/3)}{3^{2n}\Gamma(n+1-p/3)\Gamma(n+1+(2l+1-p)/3)}. \tag{B13}$$

Substituting (B13) in (B12) we finally obtain the desired

expression,

$$\beta_l(0,3n-p;n)=\frac{1}{3^{2n}\Gamma(n+1-p/3)\Gamma(n+1+(2l+1-p)/3)}$$
$$\times\sum_{i=0}^{n-p}2^{2(p+i-1)}\beta_l(0,3i+2p-2;i+p-1)$$
$$\times\Gamma(i+2p/3)\Gamma(i+(2p+2l+1)/3)$$
$$\text{for } p=1,2,\ldots,n. \tag{B14}$$

Since we have an explicit expression for $\beta_l(0,3n;n)$, that given by Eq. (B5b), we are then able, in principle, to calculate explicit expressions for $\beta_l(0,3n-1)$, and so on.

In order to illustrate this method, we will calculate β_l for $p=1$. Making use of Eq. (B5b), we obtain

$$\beta_l(0,3n-1;n)=\frac{\Gamma(1+(2l+1)/2)}{3^{2n}\Gamma(n+\frac{2}{3})\Gamma(n+1+2l/3)}$$
$$\times\sum_{i=0}^{n-1}\frac{\Gamma(i+\frac{2}{3})\Gamma(i+1+2l/3)}{\Gamma(i+1)\Gamma(i+1+(2l+1)/3)}. \tag{B15}$$

In the case $l=0$, this expression can be further simplified by using the identity[13]

$$\sum_{j=1}^{n}\frac{\Gamma(j+b)}{\Gamma(j+a)}=\frac{\Gamma(n+1+b)}{(b+1-a)\Gamma(n+a)}-\frac{\Gamma(b+1)}{(b+1-a)\Gamma(a)} \tag{B16}$$

to obtain

$$\beta_0(0,3n-1;n)=\frac{1}{3^{2n}n!}\left(\frac{\Gamma(\frac{1}{3})}{\Gamma(n+\frac{1}{3})}-\frac{\Gamma(\frac{2}{3})}{\Gamma(n+\frac{2}{3})}\right). \tag{B17}$$

APPENDIX C: INVERTED RECURSION RELATIONS AND CONJUGATE COMBINATORICS

In this Appendix, we would like to extend the proof presented in Ref. 3 to solve the inverted recursion relation,

$$b_m=\sum_{k=1}^{N}g_{a_k}(m)b_{m+a_k},\quad 0<a_1<a_2<\cdots<a_N, \tag{C1}$$

where $g_{a_k}(m)$ are a set of N functions conveniently labeled by a_k. Essentially, we would like to express b_m in terms of higher order coefficients, say, b_{j_0}, $b_{j_0+1},\ldots,b_{j_0+a_N-1}$, for $j_0>m$.

The solution of this problem is related to what we shall call the conjugate combinatorics functions of the first and second kind. The word "conjugate" does not refer to complex conjugation but rather to the relation that the inverted recursion relation and its solution bear to the original recursion relation and its solution in terms of combinatorics functions.

As in Ref. 3, the distinct partitions of (m_1,m_2) into n parts $(\delta_1,\ldots,\delta_n)$ belonging to the set $\mathcal{A}=\{a_k;k=1,\ldots,N\}$, are labeled by the index $q=1,\ldots,q_{\max}(n)$. Corresponding to each partition we introduce the function $G_n^q(m_2,m_1)$ given by

$$G_n^q(m_2,m_1)=\prod_{i=1}^{n}g_{\delta_i}(s_i), \tag{C2}$$

where

$$s_0=m_2,\quad s_i-s_{i-1}=-\delta_i,\quad s_n=m_1 \tag{C3a}$$

or explicitly

$$s_i=m_2-\sum_{k=1}^{i}\delta_k. \tag{C3b}$$

Equations (C2) and (C3) are to be compared with Eqs. (2.11) and (2.13) of Ref. 3.

By summing the functions $G_n^q(m_2,m_1)$ over all distinct partitions of (m_1,m_2) into n parts subject to the constraint that the αth part is greater than d we obtain the constrained conjugate combinatorics function of the first kind,

$$C_1^{\alpha*}(m_2,m_1,n,d)=\sum_{q\in Q_\alpha}G_n^q(m_2,m_1), \tag{C4}$$

where $Q_\alpha\equiv Q_\alpha(m_2-m_1,n,d)$ is the subset of the q's labelling those partitions of (m_1,m_2) whose αth part, δ_α, is greater than d.

On the other hand, by summing the functions $G_n^q(m_2,m_1)$ over all distinct partitions of the interval (m_1,m_2), subject to the constraint that the first part, δ_1, be greater than d, we obtain the constrained conjugate conbinatorics function of the second kind,

$$\bar{C}_2^*(m_2,m_1,d)=\sum_n\sum_{q\in Q_1}G_n^q(m_2,m_1). \tag{C5}$$

Then the analogs of Theorems 1 and 2 of Ref. 3 are:

I. Given an interval (j,m), $j\geq j_0$, and a set $\mathcal{A}$ of partitioning subintervals, then

$$C_1^{1*}(j,m,n,j-j_0)=\sum_{a_k\in\mathcal{A}}g_{a_k}(m)C_1^{1*}(j,m+a_k,n-1,j-j_0)$$
$$\text{for } m<j_0 \text{ and } n\geq 1. \tag{C6}$$

II. Given an interval (j,m), $j\geq j_0$, and a set $\mathcal{A}$ of partitioning subintervals, then

$$\bar{C}_2^*(j,m,j-j_0)=\sum_{a_k\in\mathcal{A}}g_{a_k}(m)\bar{C}_2^*(j,m+a_k,j-j_0)$$
$$\text{for } m<j_0. \tag{C7}$$

The proofs of the above two theorems follow from the Lemmas 1–4 of Ref. 3, and are very similar to the proofs of Theorems 1 and 2, given there. We will

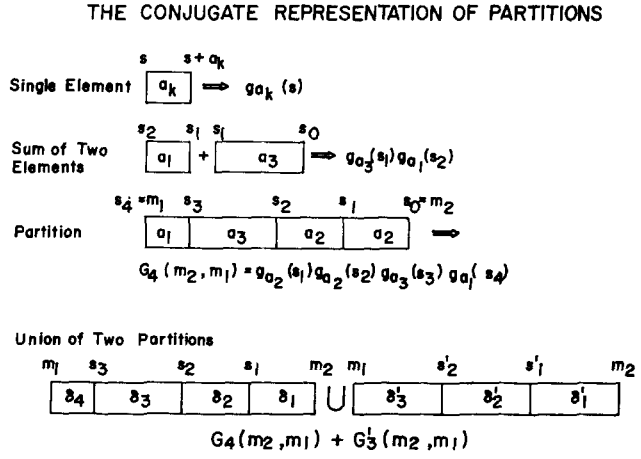

FIG. 3. The partitions of an interval (m_1,m_2) into n parts are represented by the functionals $G_n^q(m_2,m_1)$. The functions $g_{a_k}(X)$ corresponding to the partitioning subintervals a_k, are arbitrary.

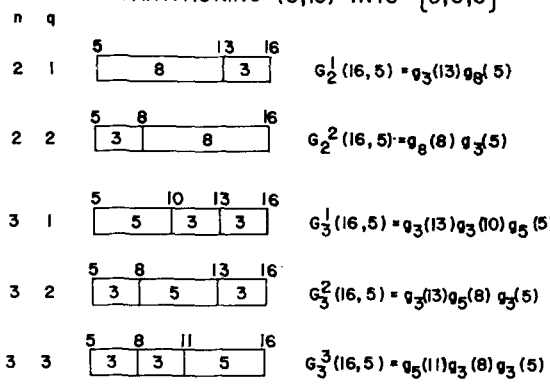

FIG. 4. The partitions of the interval (5,16) into the parts $\{3,5,8\}$, and their corresponding conjugate functionals.

therefore skip the proofs and refer the interested reader to Sec. III of Ref. 3.

Similarly, the analogs of Theorems 3 and 4 of Ref. 3 lead to the following general solution of Eq. (C1):

$$b_m = \sum_{i=0}^{a_N-1} \bar{C}_2^*(j_0+i, m, i) b_{j_0+i}, \tag{C8}$$

where $b_{j_0}, b_{j_0+1}, \ldots, b_{j_0+a_N-1}$ are taken as initial conditions.

In the special case of the linear potential problem, the inverted recursion relation, Eq. (4.1), leads to a set $\mathcal{A}$ containing two elements, $a_1=1$ and $a_2=3$. If, in addition, we set $j_0=N$ (which should not be confounded with the N used above to signify the number of elements in $\mathcal{A}$) in Eq. (C8), we obtain the result given in Eq. (4.2).

The conjugate combinatorics functions provide a representation of partitions leading to the recursion relation (C1). To further clarify the essentials of this representation, we give in Figs. 3, 4, and 5, the analogs of Figs. 1, 2, and 3 of Ref. 3. The differences and similarities between the two types of representations will become clear on comparing the two sets of figures.

APPENDIX D: THE CONJUGATE STRUCTURE FUNCTIONS $\beta_l^*(m, 1; n)$

The unconstrained conjugate structure functions are given by

$$\beta_l^*(m,1;n) = \sum_{q=1}^{q_{\max}(n)} \prod_{i=1}^{n}{}' g(s_i, l), \tag{D1}$$

where the values of s_i associated with the qth partition are given by

$$s_0 = m, \quad s_i = s_0 - \sum_{r=1}^{i} \delta_r, \quad s_n = 1, \tag{D2}$$

and $\delta_r \in \{1,3\}$. $\prod_{i=1}^{n}{}'$ stands for the product over only those values of i which satisfy the condition $s_{i-1} - s_i = 3$. Following the same steps as in the derivation of Eq. (B3) we obtain the following recursion relation,

$$\beta_l^*(m,1;n) = \beta_l^*(m-1,1;n-1) + \beta_l^*(m-3,1;n-1)g(m-3,l). \tag{D3}$$

By construction the conjugate structure functions $\beta_l^*(m,1;n)$ are evaluated over partitions of the interval $(m,1)$ into n parts of lengths 1 and 3. If p represents the number of parts of length 1, then according to Eq. (4.10), the only intervals that can be partitioned into a given number n of parts, are those of length

$$m - 1 = 3n - 2p \quad \text{for } p = 0, 1, \ldots, n. \tag{D4}$$

We will therefore study the conjugate structure functions $\beta_l^*(3n-2p+1, 1; n)$ for the above range of p, all the others being zero.

There are two special cases for which the above functions take a simple form. These correspond to $p=0$ and $p=n$. In both cases $q_{\max}(n)=1$ and we obtain

$$\beta_l^*(3n+1,1;n) = \prod_{j=0}^{n-1} g(3j+1, l) \tag{D5a}$$

and

$$\beta_l^*(n+1,1;n) = 1. \tag{D6}$$

Making use of Eq. (4.2b) we have

$$\beta_l^*(3n+1,1;n) = 32n \frac{\Gamma(n+\frac{4}{3})\Gamma(n+(2l+5)/3)}{\Gamma(\frac{4}{3})\Gamma((2l+5)/3)}. \tag{D5b}$$

In the general case we derive a relationship between the conjugate structure functions corresponding to partitions, having p parts of length 1, of a given interval, and those corresponding to partitions, having $p-1$ parts of length 1, of its subintervals. Let P_l^* be the set of distinct partitions of $(m,1)$ into n parts, p of which are of length 1, and subject to the following two conditions: (i) the last i part are of length 3, and (ii) the $(n-i)$th part is of length 1, as shown in Fig. 6. Thus if P^* is the set of all distinct partitions of $(m,1)$ into n parts, p of which are of length 1, we have

$$P^* = \bigcup_{i=0}^{n-p} P_i^*, \tag{D7a}$$

since the minimum number of parts of length 3 is zero and the maximum is $n-p$. Note that the second condition on the elements of the set P_l^*, on the one hand requires that $p \geq 1$, and on the other hand guarantees that these sets are disjoint,

$$P_i \cap P_j = 0 \quad \text{for } i \neq j. \tag{D7b}$$

The sum over q in Eq. (D1) is a sum over all distinct partitions of $(m,1)$ into n parts belonging to $\{1,3\}$. Thus due to Eqs. (D7a) and (D7b), we can rewrite Eq. (D1) as

$$\beta_l^*(m,1;n) = \sum_{i=0}^{n-p} \sum_{\text{partitions} \in P_i^*} \prod_{j=1}^{n}{}' g(s_j, l). \tag{D8}$$

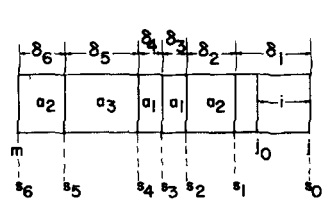

FIG. 5. Notation for the partition of an interval (m,j) into parts, subject to the condition $s_1 < j_0$, or equivalently $\delta_1 > i$.

But as can be seen from Fig. 6

$$\sum_{\text{partitions}\in P_l^*} \prod_{j=1}^{n}{}' g(s_j,l)$$
$$= \left[\sum_{\text{partitions}\in P_l^*} \prod_{j=i+1}^{n}{}' g(s_j,l)\right] \prod_{j=1}^{i} g(s_j,l)$$
$$= \beta_l^*(m-3i-1;n-i-1) \prod_{j=1}^{i} g(s_j,l)$$

and

$$\prod_{j=1}^{i} g(s_j,l) = \prod_{j=1}^{i} g(m-3j,l).$$

Thus

$$\beta_l^*(m,1;n) = \sum_{i=0}^{n-p} \beta_l^*(m-3i-1,1;n-i-1) \times \prod_{j=1}^{i} g(m-3j,l). \tag{D9}$$

Using Eq. (D4) and making the change of variable $i \to n-p-i$ and $j \to n-p-j$ we obtain

$$\beta_l^*(3n-2p+1,1;n) = \sum_{i=0}^{n-p} \beta_l^*(3i+p,1;i+p-1) \times \prod_{j=i}^{n-p-1} g(3j+p+1,l). \tag{D10}$$

Furthermore

$$\prod_{j=i}^{n-p-1} g(3j+p+1,l) = \frac{3^{2n}\Gamma(n-p+(p+4)/3)\Gamma(n-p+(p+2l+5)/3)}{3^{2(i+p)}\Gamma(i+(p+4)/3)\Gamma(i+(p+2l+5)/3)} \tag{D11}$$

and when this is substituted in Eq. (D10) we obtain the desired equation,

$$\beta_l^*(3n-2p+1,1;n) = 3^{2(n-1)}\Gamma(n-(2p-4)/3)\Gamma(n+(2-2p+5)/3) \times \sum_{i=0}^{n-p} \frac{\beta_l^*(3i+p,1;i+p-1)}{3^{2(i+p-1)}\Gamma(i+(p+4)/3)\Gamma(i+(p+2l+5)/3)} \quad \text{for } p=1,2,\ldots,n. \tag{D12}$$

To use Eq. (D12) recursively we make the change of variable $n \to n+p-1$ and $i \to i-1$ to obtain

$$\beta_l^*(3n+p-2,1;n+p-1) = 3^{2n}\Gamma(n+(p+1)/3)\Gamma(n+(p+2l+2)/3) \times \sum_{i=1}^{n} \frac{\beta_l^*(3i+p-3,1;p+i-2)}{3^{2i}\Gamma(i+(p+1)/3)\Gamma(i+(p+2l+2)/3)}. \tag{D13}$$

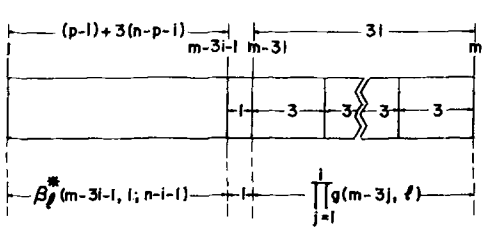

FIG. 6. The partitions belonging to the set P_l^* and their corresponding representation by conjugate structure functions.

We set $i=i_1$ in Eq. (D13) and calculate $\beta_l^*(3i_1+p-3,1;p+i_1-2)$ from Eq. (D13) by making the change of variable $n \to i_1$, $p \to p-1$, $i \to i_2$. This gives

$$\beta_l^*(3i_1+p-3,1;p+i_1-2) = 3^{2i_1}\Gamma(i_1+p/3)\Gamma(i_1+(p+2l+1)/3) \times \sum_{i_2=1}^{i_1} \frac{\beta_l^*(3i_2+p-4,1;p+i_2-3)}{3^{2i_2}\Gamma(i_2+p/3)\Gamma(i_2+(p+2l+1)/3)}. \tag{D14}$$

Substituting Eq. (D14) into Eq. (D13) we obtain

$$\beta_l^*(3n+p-2,1;n+p-1) = 3^{2n}\Gamma(n+(p+1)/3)\Gamma(n+(p+2l+2)/3) \times \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} \frac{\Gamma(i_1+p/3)\Gamma(i_1+(p+2l+1)/3)}{\Gamma(i_1+(p+1)/3)\Gamma(i_1+(p+2l+2)/3)} \times \frac{\beta_l^*(3i_2+p-4,1;p+i_2-3)}{3^{2i_2}\Gamma(i_2+p/3)\Gamma(i_2+(p+2l+1)/3)}. \tag{D15}$$

After p iterations, we obtain the explicit expression

$$\beta_l^*(3n+p-2,1;n+p-1) = \frac{3^{2(n-1)}\Gamma(n+(p+1)/3)\Gamma(n+(p+2l+2)/3)}{\Gamma(\frac{1}{3})\Gamma((2l+5)/3)}\sigma_l(n,p), \quad p=0,1,2,\cdots, \tag{D16}$$

where we have made use of Eq. (D5b) for $\beta_l^*(3i_p-2, 1; i_p-1)$, and $\sigma_l(n,p)$ is defined by

$$\sigma_l(n,p) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} \cdots \sum_{i_p=1}^{i_{p-1}} \prod_{k=1}^{p} \gamma_l(i_k,k,p), \quad p=1,2,\cdots. \tag{D17a}$$

$$\sigma_l(n,0) = 1, \tag{D17b}$$

with

$$\gamma_l(i,k,p) = \frac{\Gamma(i+(p+1-k)/3)\Gamma(i+(p+2l+2-k)/3)}{\Gamma(i+(p+2-k)/3)\Gamma(i+(p+2l+3-k)/3)}. \tag{D18}$$

It is easy to see that

$$\sigma_l(n,p) = \sum_{i=1}^{n} \gamma_l(i,1,p)\sigma_l(i,p-1). \tag{D19}$$

For $l=0$, the expression for $\sigma_l(n,p)$ simplifies considerably and is easily evaluated by using Eq. (B16). As an example

$$\sigma_0(n,1) = \frac{3\Gamma(n+\frac{4}{3})}{n!} - \Gamma(\tfrac{1}{3}). \tag{D20}$$

APPENDIX E: STUDY OF THE COEFFICIENTS

According to Eqs. (4.8c) and (4.8d), the eigenvalue equation for the energy parameter t is given by

$$\sum_{n=0}^{\infty} 2(l+1)K_n(l)t^{n-1} = 0, \tag{E1a}$$

where

$$K_n(l) = \lim_{N\to\infty} K_n^{(N)}(l). \tag{E1b}$$

The general formula from which $K_n^{(N)}(l)$ is obtained is given by Eq. (4.27), and can be rewritten, for $k > r$,

as

$$K_{3r+i}^{(3k+1)}(l)=\sum_{q=0}^{r}(-1)^{r-q+i+1}a_{3(k+r-q)+l+2+i}\binom{3(k+r-q)+l+2+i}{3(r-q)+i}$$
$$\times\Big\{\beta_l^*(3k+1,1;k+2q)-(1-\delta_{qr}\delta_{i0})[3(r-q)+i]$$
$$\times\frac{[3k+2][3(k+1)+2l]}{[3(k+1)+l]}\beta_l^*(3k-1,1;k+2q)$$
$$+[1-\delta_{qr}(1-\delta_{i2})][3(r-q)+i][3(r-q)+i-1]$$
$$\times\frac{3(k+1)[3(k+1)+2l+1]}{[3(k+1)+l][3(k+1)+l+1]}\beta_l^*(3k,1;$$
$$k+2q+1\Big\},\quad i=0,1,2. \tag{E2}$$

Basically, three types of conjugate structure functions need to be evaluated, $\beta_l^*(3k+1,1;k+2q)$, $\beta_l^*(3k-1,1;k+2q)$, and $\beta_l^*(3k,1;k+2q+1)$. Everyone of the above structure functions can be written as $\beta_l^*(3n+p-2,1;n+p-1)$, the first one corresponding to $n=k-q+1$, $p=3q$, the second to $n=k-q$, $p=3q+1$, and the last one to $n=k-q$, $p=3q+2$. Using Eq. (D16), we then find

$$\beta_l^*(3k+1,1;k+2q)$$
$$=\frac{3^{2(k-q)}\Gamma(k+\frac{4}{3})\Gamma(k+(2l+5)/3)}{\Gamma(\frac{4}{3})\Gamma((2l+5)/3)}\sigma_l(k-q+1,3q), \tag{E3a}$$

$$\beta_l^*(3k-1,1;k+2q)$$
$$=\frac{3^{2(k-q-1)}\Gamma(k+\frac{2}{3})\Gamma(k+1+2l/3)}{\Gamma(\frac{4}{3})\Gamma((2l+5)/3)}\sigma_l(k-q,3q+1), \tag{E3b}$$

$$\beta_l^*(3k,1;k+2q+1)$$
$$=\frac{3^{2(k-q-1)}\Gamma(k+1)\Gamma(k+(2l+4)/3)}{\Gamma(\frac{4}{3})\Gamma((2l+5)/3)}\sigma_l(k-q,3q+2). \tag{E3c}$$

By using the triplication formula for the gamma functions we can write the binomial appearing in Eq. (E2) as

$$\binom{3(k+r-q)+l+2+i}{3(r-q)+i}$$
$$=\frac{2\pi}{\sqrt{3}}\,\Gamma\Big(k+r-q+1+\frac{l+i}{3}\Big)\,\Gamma\Big(k+r-q+1+\frac{l+i+1}{3}\Big)$$
$$\times\Gamma\Big(k+r-q+1+\frac{l+i+2}{3}\Big)\Big/\Gamma\Big(r-q+\frac{i+1}{3}\Big)$$
$$\times\Gamma\Big(r-q+\frac{i+2}{3}\Big)\,\Gamma\Big(r-q+\frac{i+3}{3}\Big)\,\Gamma\Big(k+1+\frac{l}{3}\Big)$$
$$\times\Gamma\Big(k+1+\frac{l+1}{3}\Big)\,\Gamma\Big(k+1+\frac{l+2}{3}\Big). \tag{E4}$$

To evaluate the Airy coefficients we set

$$l=3l'+j,\quad j=0,\ 1\text{ or }2,\quad l'=0,1,2,\cdots \tag{E5}$$

and make use of Eq. (4.6) to obtain

$$a_{3(k+r-q+l'+1)+i+j-1}$$
$$=\frac{a_0\Gamma(\frac{2}{3})}{3^{2(k+r-q+l'+1)}\Gamma(k+r-q+l'+2)}$$
$$\times\begin{cases}0, & i+j=0\text{ or }3,\\ \dfrac{1}{\Gamma(k+r-q+l'+\frac{5}{3})}, & i+j=1,\\ \dfrac{1}{3^{2/3}\Gamma(k+r-q+l'+\frac{7}{3})}, & i+j=2,\\ \dfrac{1}{3^2(k+r-q+l'+2)\Gamma(k+r-q+l'+\frac{8}{3})}, & i+j=4.\end{cases} \tag{E6}$$

We note that using Eq. (E6) we have

$$(-1)^{i+1}3^{2(k-q)+1}\Gamma(k+r-q+l'+1+(i+j+1)/3)$$
$$\times\Gamma(k+r-q+l'+1+(i+j+2)/3)a_{3(k+r-q+l'+1)+i+j-1}$$
$$=(1-\delta_{i0}\delta_{j0}-\delta_{i1}\delta_{j2}-\delta_{i2}\delta_{j1})$$
$$\times\frac{(1-2\delta_{i2}\delta_{j2})(-1)^j}{3^{2(i+j-1)/3}}\cdot\frac{a_0\Gamma(\frac{2}{3})}{3^{2(r+l'+1)}}. \tag{E7}$$

Substituting Eqs. (E3), (E4), and (E6) in Eq. (E2) and making use of Eqs. (E5) and (E7) we find

$$K_{3r+i}^{(3k+1)}(3l'+j)$$
$$=(1-\delta_{i0}\delta_{j0}-\delta_{i1}\delta_{j2}-\delta_{i2}\delta_{j1})(1-2\delta_{i2}\delta_{j2})\cdot\frac{(-1)^j}{3^{2(i+j-1)/3}}$$
$$\times\frac{2\pi\Gamma(\frac{2}{3})}{3^{3/2}\Gamma(\frac{1}{3})}a_0\,\frac{(-1)^r}{3^{2(r+l')}\Gamma(2l'+1+(2j+2)/3)}$$
$$\times\sum_{q=0}^{r}\Bigg[\frac{(-1)^q\Gamma(k+r-q+l'+1+(i+j)/3)}{\Gamma(r-q+(i+1)/3)\Gamma(r-q+(i+2)/3)}$$
$$\times\frac{1}{\Gamma(r-q+(i+3)/3)\Gamma(k+l'+1+(j+2)/3)}$$
$$\times\Bigg(\frac{\Gamma(k+\frac{4}{3})\Gamma(k+2l'+1+(2j+2)/3)}{\Gamma(k+l'+1+j/3)\Gamma(k+l'+1+(j+1)/3)}$$
$$\times\sigma_{3l'+j}(k-q+1,3q)-(r-q+i/3)$$
$$\times\frac{\Gamma(k+\frac{5}{3})\Gamma(k+2l'+1+(2j+3)/3)}{\Gamma(k+l'+2+j/3)\Gamma(k+l'+1+(j+1)/3)}$$
$$\times\sigma_{3l'+j}(k-q,3q+1)+(r-q+i/3)$$
$$\times[r-q+(i-1)/3]\,\frac{\Gamma(k+2)\Gamma(k+2l'+1+(2j+4)/3)}{\Gamma(k+l'+2+j/3)(k+l'+2+(j+1)/3)}$$
$$\times\sigma_{3l'+j}(k-q,3q+2)\Bigg)\Bigg]. \tag{E8}$$

From Eq. (E8) it is easily seen that

$$K_{3r}^{(3k+1)}(3l')=0,\quad K_{3r+2}^{(3k+1)}(3l'+1)=0,\quad K_{3r+1}^{(3k+1)}(3l'+2)=0. \tag{E9}$$

Also for $r=0$, $i=0$ the expression simplifies considerably, and we have

$$K_0^{(3k+1)}(3l')=0, \tag{E10a}$$

$$K_0^{(3k+1)}(3l'+1)$$
$$=-\frac{\Gamma(\frac{2}{3})a_0}{3^{2l'+1}\Gamma(\frac{1}{3})\Gamma(3l'+\frac{7}{3})}\,\frac{\Gamma(k+\frac{4}{3})\Gamma(k+2l'+\frac{7}{3})}{\Gamma(k+l'+2)\Gamma(k+l'+\frac{5}{3})}, \tag{E10b}$$

$$K_0^{(3k+1)}(3l'+2)$$
$$=\frac{3^{1/3}\Gamma(\frac{2}{3})a_0}{3^{2l'+2}\Gamma(\frac{2}{3})\Gamma(2l'+3)}\,\frac{\Gamma(k+\frac{4}{3})\Gamma(k+2l'+3)}{\Gamma(k+l'+\frac{7}{3})\Gamma(k+l'+2)}. \tag{E10c}$$

By making use of the identity

$$\lim_{k\to\infty} k^{\beta-\alpha}\frac{\Gamma(k+\alpha)}{\Gamma(k+\beta)}=1 \tag{E11}$$

and Eqs. (E1b) and (E10), we obtain

$$K_0(3l')=0, \tag{E12a}$$

$$K_0(3l'+1)=-\frac{\Gamma(\frac{2}{3})a_0}{3^{2l'+1}\Gamma(\frac{1}{3})\Gamma(2l'+\frac{7}{3})}, \tag{E12b}$$

$$K_0(3l'+2)=+\frac{3^{1/3}\Gamma(\frac{2}{3})a_0}{3^{2l'+2}\Gamma(\frac{1}{3})\Gamma(2l'+3)}. \tag{E12c}$$

For $n>0$, taking the limit as $k\to\infty$ is more complicated due to the summations over the Γ functions. What is needed at this stage are summation theorems for products of quotients of Γ functions; that is generalization of Eq. (B16).

It is worth noting that, in the case $l=0$, Eq. (4.4) is not only valid asymptotically but for every value of N, i.e.,

$$b_N(0,t)=\sum_{i=0}^{\infty} a_{N+i+1}\binom{N+i+1}{i}(-t)^i \quad \text{for } N\geq 0. \tag{E13}$$

The immediate consequence of the above results is that Eq. (4.8c) holds for all values of N at $l=0$, i.e.,

$$K_n(0)=K_n^{(N)}(0) \quad \text{for any } N\geq 0, \tag{E14}$$

or, in other words, $K_n^{(N)}(0)$ has to be N independent. Obviously, a consistency check would be to reproduce this property from the general formula (E8). Setting $l'=0$, $j=0$ in Eq. (E8), we have,

$$K_0(0)=K_0^{(3k+1)}(0)=0, \tag{E15a}$$

$$K_1(0)=K_1^{(3k+1)}(0)=a_0/2, \tag{E15b}$$

$$K_2(0)=K_2^{(3k+1)}(0)=\frac{a_0}{2}\frac{3^{1/3}\Gamma(\frac{2}{3})}{\Gamma(\frac{2}{3})}=-\frac{a_1}{2}. \tag{E15c}$$

Thus for $l=0$, the energy eigenvalue equation becomes

$$H_0(t)=\sum_{n=0}^{\infty} 2K_n(0)t^{n-1}=\{a_0+a_1(-t)+\cdots\}$$
$$=\mathrm{Ai}(-t)=0, \tag{E16}$$

where *Ai* is the Airy function.

[1] E. P. Tryon, Phys. Rev. Lett. **28**, 1605 (1972); A. Casher, J. Kogut, and L. Susskind, *ibid.* **31**, 792 (1973); J. Kogut and L. Susskind, Phys. Rev. D **9**, 3501 (1974); H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973); D. J. Gross and F. Wilczek, *ibid.* **30**, 1343 (1973).

[2] P. M. Morse and H. Feshbach, *Methods of Mathematical Physics* (McGraw-Hill, New York, 1953), Vol. I, pp. 568–576.

[3] A. F. Antippa and A. J. Phares, J. Math. Phys. **18**, 173 (1977).

[4] T. Applequist and H. D. Politzer, Phys. Rev. Lett. **34**, 43 (1975); A. De Rujùla and S. L. Glashow, *ibid.* **34**, 46 (1975); A. De Rujùla, H. D. Politzer, and S. L. Glashow, *ibid.* **34**, 365 (1975).

[5] B. J. Harrington, S. Y. Park, and A. Yildiz, Phys. Rev. Lett. **34**, 168, 706 (1975).

[6] E. Eichten, K. Gottfried, T. Kinoshita, J. Kogut, K. D. Lane, and T. M. Yan, Phys. Rev. Lett. **34**, 370 (1975); E. Eichten, K. Gottfried, T. Kinoshita, K. D. Lane, and T. M. Yan, *ibid.* **36**, 500 (1976).

[7] H. J. W. Müller–Kirsten, Phys. Rev. D **12**, 1110 (1975).

[8] J. F. Gunion and L. F. Li, Phys. Rev. D **12**, 3583 (1975); J. F. Gunion and R. S. Willey, *ibid.* **12**, 174 (1975).

[9] E. P. Tryon, Phys. Rev. Lett. **36**, 455 (1976).

[10] C. L. Critchfield, J. Math. Phys. **17**, 261 (1976).

[11] See for example: B. Spain and M. G. Smith, *Functions of Mathematical Physics* (Van Nostrand Reinhold, London, 1970), Chapter 1; Fred Brauer and John A. Noel, *Ordinary Differential Equations* (Benjamin, New York, 1967), pp. 151–66.

[12] See for example: N. N. Lebedev, *Special Functions and their Applications*, translated by R. A. Silverman (Prentice-Hall, Englewood Cliffs, New Jersey, 1965), Sec. 5.17; *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, D. C., 1964), pp. 446–50.

[13] L. B. W. Jolley, *Summation of Series* (Dover, New York, 1961), p. 10.

Generalization of Dirac's monopole to SU_2 gauge fields

Chen Ning Yang

Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794
(Received 31 May 1977)

Dirac's monopole is generalized to SU_2 gauge fields in five-dimensional flat space or four-dimensional spherical space. The generalized fields have SO_5 symmetry. The method used is related to the concept of orthogonal gauge fields which is developed in an appendix. The angular momenta operators for the SO_5 symmetrical fields are given.

I. INTRODUCTION

The Dirac[1] monopole, which is singular only at the origin in three-dimensional space, satisfies the following properties:

(a) The magnetic flux through *any* closed surface around the origin, is $4\pi g \neq 0$. I. e.,

$$\tfrac{1}{2}\oiint f_{\mu\nu}\,dx^{\mu\nu} = 4\pi g \neq 0, \tag{1}$$

where $dx^{\mu\nu}$ is the *surface element* and is *antisymmetrical* in μ and ν.

(b) It is spherically symmetrical.

The Dirac monopole field is uniquely determined by (a) and (b) for each (allowed) value of g. We remark that if we remove condition (b) then the field is not uniquely determined by (a), since, e.g., the addition of a dipole or any higher order pole to the origin does not change (a).

A further remark is useful. The integral in (1) is independent of any distortion of the closed surface since[2]

$$f_{\mu\nu,\lambda} + f_{\nu\lambda,\mu} + f_{\lambda\mu,\nu} = 0,$$

or

$$\sum^{a} f_{\mu\nu,\lambda} = 0. \tag{2}$$

We want to generalize the Dirac monopole field to SU_2 gauge field. Consider a five-dimensional space with metric

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 + dx_5^2. \tag{3}$$

Consider an SU_2 gauge field which is singular only at the origin. The generalization of (1), which is the first Chern class number, is the second Chern class[3] number,

$$\frac{1}{24}\oiiint f^i_{\mu\nu} f^i_{\alpha\beta}\,dx^{\mu\nu\alpha\beta} = (8\pi^2/3)C_2, \tag{4}$$

where the integral is taken over a closed four-dimensional surface enclosing the origin. This integral is also independent of any distortion of the surface. To see this we use the rules of the gauge Riemannian calculus of Ref. 4 and find in a straightforward manner,

$$\sum^{a} (f^i_{\mu\nu} f^i_{\alpha\beta})_{,\gamma} = \sum^{a} (f^i_{\mu\nu} f^i_{\alpha\beta})_{\|\gamma} = \sum^{a} (f^i_{\mu\nu\|\gamma} f^i_{\alpha\beta} + f^i_{\mu\nu} f^i_{\alpha\beta\|\gamma}) = 0,$$

which is a natural generalization of (2). It follows immediately that (4) is independent of any distortion of the surface provided it always encloses the origin.

We thus search for an SU_2 gauge field satisfying

(a′) $C_2 \neq 0$;

(b′) It is SO_5 symmetrical.

As before, condition (b′) is needed to make the field unique for a given value of C_2.

We shall prove that there are *two and only two* solutions α and β satisfying (a′) and (b′). They are respectively characterized by

$$C_2 = +1 \quad \text{and} \quad C_2 = -1. \tag{5}$$

Furthermore, fields α and β will be defined so that in orthogonal coordinates ξ_1, ξ_2, ξ_3, θ, r, where r is the radial variable and ξ_1, ξ_2, ξ_3, θ are five-dimensional angular coordinates (to be defined later),

$$b^i_r = 0, \quad b^i_\theta = 0, \tag{6}$$

$$b^i_{(\xi_i)} = \text{function only of } \theta,\ \xi. \tag{7}$$

Thus $b^i_r\,dr + b^i_\theta\,d\theta + b^i_{(\xi_i)}\,d\xi^i = $ independent of r and dr.

This means that the radial coordinate r and the angular coordinates can be separated, and the gauge fields α and β are only really dependent on the latter. One can thus view α and β as gauge fields confined to any sphere S_4 with its center at the origin. In this S_4 *viewpoint*, the field α is *self-dual* and *orthogonal* everywhere, and the field β is *self-antidual* and *orthogonal* everywhere. These concepts are defined in Appendix A.

In the five-dimensional viewpoint, α and β are both sourceless and analytic everywhere except at the origin. It is SO_5 symmetrical.

In the S_4 viewpoint, α and β are both sourceless and analytic everywhere, and is SO_5 symmetrical. We define a total "action"

$$L = \oiiint f^i_{\mu\nu} f^{i\mu\nu}\,d(\text{surface}) \tag{8}$$

over S_4. We shall prove that solutions α and β have the least "action" among fields with their respective second Chern class numbers C_2.

The fields α and β will be defined in Secs. II and III in terms of nonintegrable phase factors.[4,6] A reader unfamiliar with this geometrical concept can take Eqs. (34) as the algebraic definition of the fields.

The concepts of orthogonal self-dual and self-antidual fields seem to be very useful. These fields are defined and discussed in Appendix A, where the relationship between these concepts and SO_4 symmetry is also discussed.

[a] Work partially supported by the NSF under Grant PHY 7615328.

 0022-2488/78/1901-0320$1.00

The angular momentum operators in five-dimensional space are exhibited in Sec. X. They contain extra terms to take into account the angular momenta that reside in the field, just as the angular momentum operators for a charged particle in a Dirac monopole field contain extra terms.

II. CONSTRUCTION OF THE SOLUTIONS

We first recall the vector potential **A** for the Dirac monopole. They can be chosen in two[5] regions R_a and R_b which are defined by (in spherical coordinates r, θ, ϕ):

$$R_a\colon\ 0 \leq \theta < (\pi/2) + a,$$

$$R_b\colon\ \pi \geq \theta > (\pi/2) - a \quad (0 < a < \pi/2).$$

In the two regions they are respectively

$$A_r^{(a)} = A_\theta^{(b)} = 0, \quad A_\phi^{(a)} = g(1 - \cos\theta), \tag{9}$$

$$A_r^{(b)} = A_\theta^{(b)} = 0, \quad A_\phi^{(b)} = -g(1 + \cos\theta), \tag{10}$$

where we use tensor notation for the components, so that the expression for A_ϕ in these formulas is $(r \sin\theta)$ times the corresponding A_ϕ of Ref. 5.

It has been emphasized[4,6] that a more intrinsic concept than **A** is the phase factor. For an infinitesimal path from $P: (r, \theta, \phi)$ to $P + dP: (r + dr, \theta + d\theta, \phi + d\phi)$, the phase factor is (in R_a)

$$\begin{aligned}\Phi^{(a)}_{(P+dP)P} &= 1 + ieA_\mu\, dx^\mu \\ &= 1 + ieg(1 - \cos\theta)\, d\phi \\ &\approx [\exp(+2ieg\, d\phi)]^{p(\theta)},\end{aligned} \tag{11}$$

where

$$p(\theta) = \tfrac{1}{2}(1 - \cos\theta). \tag{12}$$

In R_b we obtain

$$\Phi^{(b)}_{(P+dP)P} = [\exp(-2ieg\, d\phi)]^{1-p(\theta)}. \tag{13}$$

Now consider a function, defined everywhere except along the z axis,

$$T(r, \theta, \phi) = \exp(2ige\phi), \tag{14}$$

which is single valued in view of Dirac's quantization condition

$$2ge = \text{integer}. \tag{15}$$

Thus,

$$\Phi^{(a)}_{(P+dP)P} = (T_{P+dP} T_P^{-1})^{p(\theta)}, \tag{16}$$

$$\Phi^{(b)}_{(P+dP)P} = (T^{-1}_{P+dP} T_P)^{1-p(\theta)}. \tag{17}$$

Since $p(0) = 0$, Eq. (16) is applicable near $\theta = 0$. Similarly, Eq. (17) is applicable near $\theta = \pi$ since $1 - p(\pi) = 0$. Equations (16) and (17) define the Dirac field.

We are now ready to generalize to a SU_2 gauge field in five-dimensional space. We shall choose coordinates ξ_i, θ, r $(i = 1, 2, 3)$ such that

$$r = (x_1^2 + \cdots + x_5^2)^{1/2}, \quad x_5 = r\cos\theta \quad (0 \leq \theta \leq \pi). \tag{18}$$

ξ_1, ξ_2, ξ_3 parametrize the three-dimensional sphere (Fig. 1)

$$r = \text{fixed}, \quad \theta = \text{fixed}, \quad \text{or } r = \text{fixed}, \quad x_5 = \text{fixed}. \tag{19}$$

For the time being we shall not specify how to choose ξ_1, ξ_2, ξ_3 to avoid unnecessary distraction at this stage. Consider *any*

$$T(\xi_1, \xi_2, \xi_3, \theta, r)$$

which is an element of SU_2 and is defined and differentiable at all points in the five-dimensional space except on the x_5 axis. Consider any $p(\theta)$ which satisfies

$$p(0) = 1 - p(\pi) = 0. \tag{20}$$

Then (16) and (17) define a gauge field in R_a and R_b respectively. In the region of overlap, we find the following relationship between Eqs. (16) and (17)

$$T^{-1}_{P+dP} \Phi^{(a)}_{(P+dP)P} T_P = \Phi^{(b)}_{(P+dP)P}. \tag{21}$$

To prove this we start from

$$\Phi^{(a)}_{(P+dP)P} = (T_{P+dP} T_P^{-1})^{p(\theta)-1} (T_{P+dP} T_P^{-1}).$$

Thus,

$$\begin{aligned}T^{-1}_{P+dP} \Phi^{(a)}_{(P+dP)P} T_P &= T^{-1}_{P+dP} (T_{P+dP} T_P^{1})^{p-1} T_{P+dP} \\ &= [T^{-1}_{P+dP} (T_{P+dP} T_P^{-1}) T_{P+dP}]^{p-1} \\ &= [T_P^{-1} T_{P+dP}]^{p-1}\end{aligned}$$

which leads to Eq. (21).

Equation (21) shows that $\Phi^{(a)}$ and $\Phi^{(b)}$ define the same gauge field. T_P is thus the "transition function" S for the overlap.[6] *It defines the gauge transformation from region b to region a.*

III. CONSTRUCTION OF THE SOLUTIONS (CONTINUED)

For any T and $p(\theta)$ satisfying Eq. (20) we have a gauge field. It remains to choose an explicit form for T as a function of the coordinates and a $p(\theta)$ that satisfies Eq. (20) so that conditions (a) and (b) are satisfied. For $p(\theta)$ we choose Eq. (12), the same as in Dirac's case. For T, we endeavor to define it as a function independent of r and θ, again imitating (14) for the Dirac case. Thus $T = T(\xi_1, \xi_2, \xi_3)$. Since the sphere (19) has the *same geometry* as the SU_2 group manifold itself, it is natural to define T as the group element represented by the point ξ_1, ξ_2, ξ_3 on the sphere (19). [We observe that this is an exact generalization of the

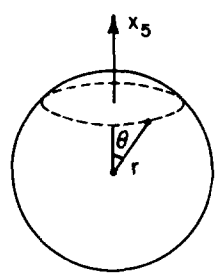

FIG. 1. The coordinates $\xi_1, \xi_2, \xi_3, \theta$, and r in five dimensions. r is the radius $r\cos\theta = x_5$ as illustrated. The equations r = const, θ = const is a three-dimensional sphere symbolized by the dotted curve. It is the generalization of the azimuthal circle in the usual spherical coordination system r, θ, ϕ. ξ_1, ξ_2, ξ_3 parametrizes this S_3, as ϕ parametrizes the usual azimuthal circle.

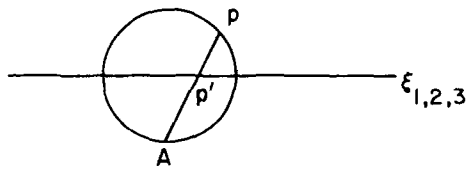

FIG. 2. Projective coordinate for three-dimensional sphere S_3 (in four dimensions). $\xi_{1,2,3}$ is symbolically three-dimensional flat space. A is the "south" pole of the unit sphere S_3. The point p on S_3 is projected to the point p' whose coordinates ξ parametrize the point p. The point A on S_3 corresponds to the point at ∞ in ξ space.

Dirac case where (19) reduces to a circle (Fig. 1). In that case, if $2eg=1$, T as defined by (14) *is* the group element that corresponds geometrically to the point ϕ on the circle.]

The above description gives a geometrical definition of T. To translate it into an explicit formula, we adopt a projective coordinate system ξ_1, ξ_2, ξ_3 for the sphere (19) (Fig. 2):

$$x_i=(r\sin\theta)2\xi_i(1+\xi^2)^{-1} \quad \text{where } \xi=\left(\sum_1^3 \xi_i^2\right)^{1/2}\geqslant 0, \quad i=1,2,3, \tag{22}$$

$$x_4=(r\sin\theta)(1-\xi^2)(1+\xi^2)^{-1}.$$

For fixed $r>0$, and $0<\theta<\pi$, the complete ξ_1, ξ_2, ξ_3 plane plus the point at ∞ maps through Eq. (22) onto the sphere (19) in a one-to-one mapping. The transformation $x_1, \ldots, x_5 \rightleftharpoons \xi_1, \xi_2, \xi_3, \theta, r$ defined by Eqs. (22) and (18) has the metric

$$ds^2=dr^2+r^2\,d\theta^2+r^2(\sin^2\theta)4(1+\xi^2)^{-2}\,d\xi^2. \tag{23}$$

Furthermore, the Jacobian of the transformation is positive,

$$\frac{\partial(x_1x_2x_3x_4x_5)}{\partial(\xi_1\xi_2\xi_3\theta r)}>0.$$

We now define the SU_2 *monopole gauge field* α by Eqs. (12), (16), and (17) together with[7] the following definition of T,

$$R(T)=(1+\xi^2)^{-1}(1-\xi^2+2i\boldsymbol{\xi}\cdot\boldsymbol{\sigma}) \quad [\xi=(\xi_1^2+\xi_2^2+\xi_3^2)^{1/2}], \tag{24α}$$

where σ are the Pauli matrices satisfying $\sigma_1\sigma_2=i\sigma_3$. $R(T)$ means the 2×2 representative of T. Similarly, we define the SU_2 *monopole gauge field* β by Eqs. (12), (16), and (17) together with

$$R(T)=(1+\xi^2)^{-1}(1-\xi^2-2i\boldsymbol{\xi}\cdot\boldsymbol{\sigma}). \tag{24β}$$

IV. POTENTIALS $b^{i(a)}_\mu$ AND $b^{i(b)}_\mu$

Defining the gauge potential b^i_μ by

$$\Phi_{(P+dP)P}=1-b^j_\mu X_j\cdot dx^\mu, \tag{25}$$

we can compute in R_a and R_b respectively $b^{i(a)}_\mu$ and $b^{i(b)}_\mu$ from Eqs. (24), (12), (16), and (17). We shall use tensor notation[8] and write

$$b^i_1, b^i_2, b^i_3, b^i_\theta, b^i_r$$

for

$$b^i_{(\xi_1)}, b^i_{(\xi_2)}, b^i_{(\xi_3)}, b^i_\theta, b^i_r.$$

By putting $d\theta=d\xi^i=0$ in Eqs. (16), (24), and (25) we obtain b^j_r. Since T is independent of r, $\Phi_{(P+dP)P}=$ identity, we have

$$b^j_r=0 \quad \text{in both } R_a \text{ and } R_b. \tag{26}$$

Similarly,

$$b^j_\theta=0 \quad \text{in both } R_a \text{ and } R_b. \tag{27}$$

Putting $d\theta=dr=0=d\xi^2=d\xi^3$, we obtain by Eqs. (25) and (16),

$$-b^j_1\left(-\frac{i}{2}\sigma_j\right)=p(\theta)\left(\frac{\partial T}{\partial\xi^1}\right)T^{-1} \quad \text{in } R_a, \tag{28}$$

where we have substituted $-i\sigma_j/2$ for X_j which it represents, and we write T for $R(T)$.

Substituting Eq. (24) into (28) we can calculate b^i_j. To present the results we define

$$B^i_j\equiv\langle i\,|B|\,j\rangle, \quad D^i_j\equiv\langle i\,|D|\,j\rangle, \tag{29}$$

and

$$B=-8(1+\xi^2)^{-2}[\psi\tilde{\psi}+\tfrac{1}{2}(1-\xi^2)+N], \tag{30}$$

$$D=-\tilde{B}, \tag{31}$$

where ~ means transposed, and

$$\psi=\begin{pmatrix}\xi_1\\ \xi_2\\ \xi_3\end{pmatrix}, \quad N=\begin{pmatrix}0 & -\xi_3 & \xi_2\\ \xi_3 & 0 & -\xi_1\\ -\xi_2 & \xi_1 & 0\end{pmatrix}. \tag{32}$$

The following formulas are useful:

$$N\psi=0, \quad N^2=\psi\tilde{\psi}-\xi^2, \quad B\tilde{B}=16(1+\xi^2)^{-2}. \tag{33}$$

Using these definitions, we find

$$b^i_r=b^i_\theta=0 \quad \text{for both solutions } \alpha \text{ and } \beta, \text{ in both } R_a \text{ and } R_b. \tag{34}$$

For[7] solution α:

$$b^{i(a)}_j=(1-\cos\theta)D^i_j/2, \quad b^{i(b)}_j=(1+\cos\theta)B^i_j/2. \tag{34'α}$$

For solution β:

$$b^{i(a)}_j=(1-\cos\theta)B^i_j/2, \quad b^{i(b)}_j=(1+\cos\theta)D^i_j/2. \tag{34'β}$$

These equations are obtained from Eq. (28). We notice

$$B^j_k\left(-\frac{i}{2}\sigma_j\right)=-R_{,k}R^{-1},$$

where R stands for $R(T)$ of Eq. (24β). Similarly

$$D^i_k\left(-\frac{i}{2}\sigma_j\right)=-R_{,k}R^{-1},$$

where R stands for $R(T)$ of Eq. (24α).

In the overlap region R_{ab}, $b^{i(a)}_j$ and $b^{i(b)}_j$ are related by a gauge transformation, since they were computed from (25) using (16) and (17), which are related by the gauge transformation (21). The gauge transformation (i.e., transition function) from $b^{(b)}$ to $b^{(a)}$ is thus T which is given by (24α) or (24β).

In the rest of the paper, we concentrate on R_a. *Results for R_b are obtained by applying the gauge transformation* (21) *to that for R_a.*

V. FIELD STRENGTHS $f^i_{\mu\nu}$ FOR FIELD β

Applying Eq. (34) to the definitions

$$f^i_{\mu\nu}=b^i_{\mu,\nu}-b^i_{\nu,\mu}-C^i_{jk}b^j_\mu b^k_\nu,$$
$$C^1_{23}=C^2_{31}=C^3_{12}=-C^1_{32}=-C^2_{13}=-C^3_{21}=1,$$

we obtain for field β, in region R_a,

$$f^i_{r\theta}=f^i_{rj}=0, \tag{35}$$
$$f^i_{\theta j}=-(\sin\theta)B^i_j/2, \tag{35'β}$$

and

$$f^i_{jk}=-p(p-1)C^i_{lm}B^l_jB^m_k. \tag{35''β}$$

We have used the relation

$$B^i_{j,k}-B^i_{k,j}-C^i_{lm}B^l_jB^m_k=0 \tag{36}$$

which can be verified directly from definition (30) of B^i_j. It can also be verified by putting $p(\theta)=1$ in Eq. (16). $\Phi^{(a)}_{(P+dP)P}$ is then obviously gauge transformable to unity. Thus the field strength in such a case should vanish. Now when $p(\theta)=1$, Eq. (28) states that $b^j_i=B^j_i$. Equation (36) is then the statement that the field strengths vanish, which we already proved.

Explicit evaluation starting from Eq. (35) gives in R_a for field β:

$$\begin{aligned}
f^1_{\theta 1}&=4(\sin\theta)(1+\xi^2)^{-2}[\xi_1\xi_1+(1-\xi^2)/2],\\
f^1_{\theta 2}&=4(\sin\theta)(1+\xi^2)^{-2}[\xi_1\xi_2-\xi_3],\\
f^1_{\theta 3}&=4(\sin\theta)(1+\xi^2)^{-2}[\xi_1\xi_3+\xi_2],\\
f^1_{23}&=8(\sin^2\theta)(1+\xi^2)^{-3}[\xi_1\xi_1+(1-\xi^2)/2],\\
f^1_{31}&=8(\sin^2\theta)(1+\xi^2)^{-3}[\xi_1\xi_2-\xi_3],\\
f^1_{12}&=8(\sin^2\theta)(1+\xi^2)^{-3}[\xi_1\xi_3+\xi_2].
\end{aligned} \tag{37β}$$

Other components of $f^i_{\theta j}$ and f^i_{jk} can be obtained from Eq. (37β) by cyclic permutation of all indices 1, 2, and 3 (i.e., simultaneously of the SU_2 index and the ξ subscript). The field strengths $f^i_{\mu\nu}$ in R_b are obtained from (37β) by a gauge transformation as discussed in Sec. IV. In the rest of the paper we shall concentrate on R_a.

The field strengths for field α are similar to these. They are discussed in Appendix B.

Since conditions (6) and (7) are satisfied, we can take the S_4 viewpoint mentioned in the Introduction. We apply then the concepts of Appendix A to field β. Obviously, by (A1),

$$\eta_{\mu\nu\alpha\beta}=(8\sin^3\theta)r^4(1+\xi^2)^{-3}\epsilon_{\mu\nu\alpha\beta}. \tag{38}$$

One can then evaluate f^* from Eq. (37β), arriving at

$$f^*=-f. \tag{39β}$$

Similarly for field α one proves this way that

$$f^*=f. \tag{39α}$$

Thus fields α and β are respectively self-dual and self-antidual.

Using Eq. (37β) we can prove (Appendix C) that *field β is self-antidual and orthogonal everywhere*, and

$$H=-\mathcal{E}=-(1+\xi^2)\tilde{B}(4r^2)^{-1}, \tag{40}$$

$$a=r^{-2}. \tag{41}$$

One can similarly show that *field α is self-dual and orthogonal everywhere*, with the same amplitude a given by (41). The inverse square dependence of a on r is the same as in Dirac's monopole.

VI. ANALYTICITY AT $\theta = 0$ AND $\theta = \pi$

For the Dirac monopole, the choice (9) of b_μ in R_a has an apparent singularity at $\theta=0$, since A_ϕ cannot be defined there. However, in Cartesian coordinates at a point on the $+z$ axis,

$$\begin{aligned}
A_x&=-g\frac{1-\cos\theta}{r\sin\theta}\sin\phi=-g\frac{1}{1+\cos\theta}\frac{y}{r^2},\\
A_y&=g\frac{1-\cos\theta}{r\sin\theta}\cos\phi=g\frac{1}{1+\cos\theta}\frac{x}{r^2},\\
A_z&=0.
\end{aligned} \tag{42}$$

Thus $A^{(a)}$ is analytic at $\theta=0$. Similarly we can prove that $A^{(b)}$ is analytic at $\theta=\pi$.

For the fields α and β in five-dimensional space we can, by using Cartesian coordinates at a point on the $+x_5$ axis, in exactly the same manner, prove that $b^{i(a)}_\mu$ in Cartesian coordinates is analytic at $\theta=0$. Similarly we find that $b^{i(b)}_\mu$ in Cartesian coordinates is analytic at $\theta=\pi$. Thus the fields α and β are *both everywhere analytic* except at the origin.

In the S_4 viewpoint, fields α and β are analytic everywhere.

VII. PROOF OF SO_5 SYMMETRY

An SO_5 rotation around the origin generates a new field α' from field α. We shall prove now that field α can be gauge transformed into field α'. This can be done by considering infinitesimal SO_5 rotations. We shall, however, present a different and better proof in the following steps:

(a) Since fields α and α' are both self-dual orthogonal everywhere in the S_4 view, and they have the same amplitude (41), their $f^i_{\mu\nu}$ can both be gauge transformed into the same standard form (A10). [Equation (35) insures that $f^i_{r\mu}=0$ always.] Thus their field strengths are gauge transformable into each other.

(b) Now adopt gauges for α and α' so that $(f^i_{\mu\nu})_\alpha=(f^i_{\mu\nu})_{\alpha'}=$ standard form (A10). It remain to be proved that $(b^i_\mu)_{\alpha'}=(b^i_\mu)_{\alpha'}$. To do this we observe that $(b^i_r)_\alpha=(b^i_r)_{\alpha'}=0$ by definition (since the gauge transformants that we used are independent of r). Next we write down the Bianchi identities for α and α' and subtract the corresponding equations from each other, resulting in

$$C^i_{jk}(\Delta b^j_\mu)f^k_{\nu\lambda}+(\text{cyclic permutations of }\mu\nu\lambda)=0, \tag{43}$$

where

$$\Delta b^i_\mu=(b^i_\mu)_{\alpha'}-(b^i_\mu)_\alpha.$$

Choose μ,ν,λ to be three of the coordinates $\xi_1\xi_2\xi_3\theta$. There are four ways of doing this. Since $i=1$, 2, or 3 in Eq. (43), we thus have 12 equations in the 12 unknowns Δb^i_μ. The determinant of these two 12 equations are known from the standard form (A11). It is simple

to evaluate and it is equal to $16a^{12}=16r^{-24}\neq 0$. Thus

$$\Delta b^i_\mu = 0$$

and we have proved that α and α' are gauge equivalent. The proof for the gauge equivalence of β and β' is similar.

VIII. ADDITIONAL PROPERTIES

(a) Fields α and β are both sourceless, i. e.,

$$f^{i\mu\nu}_{\mu\nu}=0. \tag{44}$$

This is true both when the field is viewed from the five-dimensional viewpoint, if we exclude the origin, or from the S_4 viewpoint. To prove it for the latter view, we use Eqs. (A2), (A3), and $f^*=\pm f$,

$$f^{i\mu\nu}_{\mu\nu}=\eta_{\mu\nu\alpha\beta}f^{*i\alpha\beta\mu\nu}/2=\pm\eta_{\mu\nu\alpha\beta}f^{i\alpha\beta\mu\nu}/2$$

which is zero because of the Bianchi identity. The proof of Eq. (44) for the five-dimensional view follows from this easily.

(b) $f^i_{\mu\nu}f^{i\mu\nu}$ can be evaluated, using Eqs. (A12) and (41),

$$f^i_{\mu\nu}f^{i\mu\nu}=12r^{-4}. \tag{45}$$

(c) In the S_4 viewpoint we can evaluate, via Eq. (A13),

$$f^i_{\mu\nu}f^{*i\mu\nu}=\pm 12r^{-4}, \tag{46}$$

where + is for field α, − for field β. We can now evaluate (4) on a sphere $r=\text{const}$,

$$\tfrac{1}{24}\oint\!\!\!\oint\!\!\!\oint\!\!\!\oint f^i_{\mu\nu}f^i_{\alpha\beta}dx^{\mu\nu\alpha\beta}=\pm r^{-4}\oint\!\!\!\oint\!\!\!\oint\!\!\!\oint d(\text{area})=\pm\tfrac{8}{3}\pi^2, \tag{47}$$

verifying (4) with $C_2=\pm 1$.

(d) Consider the S_4 viewpoint. The sourceless condition (44) which we just proved is[4] the condition that the "action"

$$L=\oint\!\!\!\oint\!\!\!\oint\!\!\!\oint f^i_{\mu\nu}f^{i\mu\nu}d(\text{surface})$$

over the sphere is stationary against changes of the gauge field. But we can prove a stronger statement. Consider any SU_2 gauge field on the sphere. Using the notation of (A6) we see that

$$f^i_{\mu\nu}f^{i\mu\nu}=2(\mathbf{E}^i\cdot\mathbf{E}^i+\mathbf{H}^i\cdot\mathbf{H}^i), \tag{48}$$

$$\tfrac{1}{24}f^i_{\mu\nu}f^i_{\alpha\beta}\eta^{\mu\nu\alpha\beta}=\tfrac{1}{3}\mathbf{E}^i\cdot\mathbf{H}^i. \tag{49}$$

Integrating Eq. (48) over the sphere we get L. Integrating Eq. (49) over the sphere we get by definition (4), $(8\pi^2/3)C_2$. Thus

$$L\geq 12\left|8\pi^2C_2/3\right|=32\pi^2\left|C_2\right|. \tag{50}$$

Since the Chern class number C_2 is always an integer, we find that for all fields for which $C_2\neq 0$, L attains an *absolute minima* $32\pi^2$ for fields α and β. This conclusion is the same as a corresponding one in Ref. 10.

For fields α and β, Eq. (45) leads to $L=32\pi^2$.

IX. FIELDS α AND β AS THE ONLY SO_5 SYMMETRICAL FIELDS

We shall now prove that fields α and β are the only SO_5 symmetrical SU_2 gauge fields other than the trivial case of all $f^i_{\mu\nu}=0$.

To be precise, we assume that field γ, whose strength is not equal to 0, defined on all five-dimensional space except the origin, can be gauge transformed to become any SO_5 rotation of itself. We shall prove that γ is gauge transformable to either field α or field β.

(a) Use the coordinates $\xi_1, \xi_2, \xi_3, \theta, r$. Consider a point P and write the twelve elements f^j_{ri} $(i=1,2,3,\theta;\ j=1,2,3)$ as a 4×3 matrix M. An SO_4 rotation around the r axis at P generates a transformation A on the i index. SO_4 invariance requires that there is a compensating gauge transformation R so that

$$AMR=M.$$

Thus,

$$AM\tilde{M}\tilde{A}=M\tilde{M}.$$

Since A is an irreducible representation of SO_4 and the rank of $M\tilde{M}\leq 3$, we find $M\tilde{M}=0$, i. e., $M=0$. Thus

$$f^j_{ri}=0. \tag{51}$$

(b) Consider a sphere S_4: $r=\text{const}$. The sphere is geometrically SO_5 symmetrical. γ is clearly pointwise SO_4 symmetric at any point P on the sphere. Using Lemma 3 of Appendix A we conclude that γ is orthogonal and self-dual or self-antidual at P. Thus it can be gauge transformed to the standard form (A10) or (A11). SO_5 symmetry implies that a is a function of r alone. Thus $a=a(r)$. Since the gauge transformation can be made independently at every point in five-dimensional space, we conclude that in a proper gauge,

$$(f^i_{\mu\nu})_\gamma=(ar^2)(f^i_{\mu\nu})_{\alpha\text{ or }\beta}. \tag{52}$$

(c) Now we can imitate the arguments of Sec. VII (b) and write down the Bianchi identities for the field γ, and for α or β. Multiplying the latter by (ar^2) and subtracting from the former we obtain for $\mu,\nu,\lambda\neq r$,

$$C^i_{jk}(\Delta b^j_\mu)(ar^2)(f^k_{\nu\lambda})_{\alpha\text{ or }\beta}+(\text{cyclic permutation of }\mu\nu\lambda)=0, \tag{53}$$

where

$$\Delta b^j_\mu=(b^j_\mu)_\gamma-(b^j_\mu)_{\alpha\text{ or }\beta}.$$

Just as in Sec. VII (b), Eq. (53) implies $\Delta b^j_\mu=0$ $(\mu\neq r)$. Substitution into Eq. (52) further leads to

$$ar^2=1. \tag{54}$$

(d) We need only prove now that $b^j_r=0$. To do this we subtract the Bianchi identity again, like in Eq. (53), but in addition use $ar^2=1$ and take one of μ,ν,λ to be r. Because of Eq. (35) we get

$$C^i_{jk}(\Delta b^j_r)(f^k_{\nu\lambda})_{\alpha\text{ or }\beta}=0. \tag{55}$$

It follows trivially that $\Delta b^j_r=0$, i. e., $b^j_r=0$.

X. ANGULAR MOMENTUM OPERATORS

In Dirac's monopole field, the angular momentum of a particle of charge Ze is[5,9]

$$\mathbf{L}=\mathbf{r}\times(\mathbf{p}-Ze\mathbf{A})-Zeg\mathbf{r}r^{-1}. \tag{56}$$

We want to generalize this formula to the field α (or β).

Consider the motion of a particle of isospin I in field α or β in five-dimensional space. Let Y_1, Y_2, Y_3 be the representation of generators X_1, X_2, X_3 for isospin I. Then the generalization of Eq. (56) is

$$L_{\mu\nu} = x_\mu(\partial_\nu + Y_k b_\nu^k) - x_\nu(\partial_\mu + Y_k b_\mu^k) - r^2 f_{\mu\nu}^k Y_k.$$
$$= -L_{\nu\mu} \quad (\mu, \nu = 1, 2, 3, 4, 5). \tag{57}$$

It is important to notice that the wavefunctions are *sections*.[5] The transition function S_{ab} can be read off from Eq. (21), so that a section is defined by

$$\psi^{(a)} = \mathrm{Rep}(T)\psi^{(b)}, \tag{58}$$

where $\mathrm{Rep}(T)$ is the representation of T in the representation generated by the Y's. This formula is exactly the same as the corresponding one in Ref. 5. As in that reference, we consider the *Hilbert space of sections*. Equation (57) is then a Hermitian operator in the Hilbert space. The commutation rules of $L_{\mu\nu}$ can be obtained by direct calculation. After some algebra we obtain

$$[L_{\mu\nu}, L_{\alpha\beta}] = \delta_{\nu\alpha}L_{\mu\beta} - \delta_{\mu\alpha}L_{\nu\beta} - \delta_{\nu\beta}L_{\mu\alpha} + \delta_{\mu\beta}L_{\nu\alpha}, \tag{59}$$

which shows that $L_{\mu\nu}$ are the angular momentum operators.

One can now generalize the monopole harmonies of Ref. 5 to SU_2 monopole harmonies which are harmonic sections on a sphere $r = \text{const}$ in five-dimensional space. We shall return to this problem in a later paper.

One notices that if we make the replacement

$$Y_k \to -iZ, \quad b_\nu^k \to eA_\nu, \quad -iL_{12} \to L_z, \tag{60}$$

then Eq. (57) reduces to Eq. (56).

XI. REMARKS

(a) The fields α and β on a sphere S_4 (in five dimensions) exhibit SO_5 symmetry. The sphere has a nonflat geometry. Does there exist corresponding solutions on a flat four-dimensional space with $ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$? The answer is no if we require maximum symmetry consistent with the geometry, i. e., if we require SO_4 symmetry plus displacement symmetry. (The symmetry group, which we shall call the Poincaré group, has 10 generators and is the natural extension to flat R^4 space of the SO_5 group for S_4 geometry.)

To prove the nonexistence in R^4 of a SU_2 gauge field δ with Poincaré symmetry we proceed exactly as in Sec. IX. If δ exists, it is pointwise SO_4 symmetrical at every point P. Lemma 3 of Appendix A then shows that it is orthogonal and self-dual or self-antidual at every point. Lemmas 1α and 1β then lead to the conclusion that δ can be gauge transformed to the standard form (A10) or (A11), where $a^2 = G$. Now Eq. (A5) and displacement symmetry imply $G = \text{numerical constant}$. Thus $a = \text{const}$ in Eq. (A10) or (A11). i. e., $f_{\mu\nu}^i$ is independent of $x_1x_2x_3x_4$. The Bianchi identity then reads

$$C_{jk}^i b_\mu^j f_{\nu\lambda}^k + (\text{cyclic permutation of } \mu\nu\lambda) = 0. \tag{61}$$

If $a \neq 0$, this is 12 equations in the 12 numbers b_μ^j. The determinant is not equal to 0, as in Eq. (43). Thus $b_\mu^j = 0$. Therefore,

$$f_{\mu\nu}^k = 0. \tag{62}$$

If $a = 0$, then automatically Eq. (62) also holds. Thus in R^4 there is no SU_2 gauge field with strengths not equal to 0 that is SO_4 invariant and displacement invariant.

Belavin, Polyakov, Schwartz, and Tyupkin[10] have exhibited a solution which they call a pseudoparticle solution. It is a sourceless SU_2 gauge field on R^4 which is everywhere analytic. It has a second Chern class number $C_2 = \pm 1$. It does not have displacement invariance, in agreement with the conclusion above. The relationship between this pseudoparticle solution and O(5) symmetry has been discussed by Jackiw and Rebbi,[11] who found that the pseudoparticle when conformally mapped to a sphere S_4 is O(5) symmetrical. According to Sec. IX above, the conformally mapped solution is, exactly, the O(5) symmetrical SU_2 gauge field which is the generalization of Dirac's monopole. Further comments on this relationship will be communicated in a separate paper.

(b) Does there exist a SO_n symmetrical SU_2 gauge field on the n-dimensional flat space (with positive signatures) minus the origin? (We do not consider the trivial case of $f_{\mu\nu}^i \equiv 0$.) We have seen in Sec. IX that for $n = 5$, there are two such fields α and β. We shall now prove that for $n \geq 6$, there are no such fields.

Take the case $n = 6$. Choose *orthogonal* coordinates $\xi_1\xi_2\xi_3\xi_4\xi_5 r$ where r is the radius. We can first easily prove the generalization of Eq. (51),

$$f_{ri}^j = 0 \quad (i = 1, 2, 3, 4, \theta).$$

Next consider a point P and choose the scales of $\xi_1\xi_2\xi_3\xi_4\xi_5$ so that $g_{11} = g_{22} = g_{33} = g_{44} = g_{55} = 1$ at P. Consider $f_{\mu\nu}^j$ for $\mu, \nu = 1, 2, 3, 4$. SO_4 symmetry in the directions of $\xi_1\xi_2\xi_3\xi_4$ leads to

$$f_{12}^i = \pm f_{34}^i,$$

by an argument similar to that in Sec. IX. Similarly, if we consider $\mu, \nu = 1, 2, 3, 5$ we obtain

$$f_{12}^i = \pm f_{35}^i.$$

Now take $\mu, \nu = 2, 3, 4, 5$. SO_4 symmetry in the directions of $\xi_2\xi_3\xi_4\xi_5$ implies orthogonality, so that

$$f_{35}^i f_{34}^i = 0.$$

Thus

$$f_{12}^i f_{12}^i = 0,$$

i. e.,

$$f_{12}^i = 0.$$

We thus find all components of $f = 0$. The proof for the case $n > 6$ is similar.

(c) What happens for $n = 4$? One can find SO_4 symmetric solutions, singular only at the origin, in the following way:

Consider a path $A \to B$ not passing through the origin. Project the path radially onto the unit sphere $r = 1$. Let the projection be called $A'B'$. Let p be a real number. Each point $a, b, c, \ldots, z$ along the path $A'B'$ corresponds to an element of SU_2 which we shall denote by $\underline{a}, \underline{b}, \ldots, \underline{z}$. If the path $A'B'$ is $A'ab \cdots zB'$ we define

$$\Phi_{BA} = \Phi_{B'A'} = (\underline{B'}\underline{z}^{-1})^p(\underline{z}\underline{y}^{-1})^p \cdots (\underline{a}\underline{A'}^{-1})^p.$$

To show that this gauge field is SO_4 invariant, consider any two fixed elements of the group ξ, η. Let $\underline{z}_0 = \xi \underline{z} \eta$, $\underline{B}'_0 = \xi \underline{B}' \eta$, etc.

Then

$$\begin{aligned}\xi \Phi_{BA} \xi^{-1} &= (\xi \underline{B}' z^{-1} \xi^{-1})^p (\xi \underline{z} y^{-1} \xi^{-1})^p \cdots (\xi \underline{a} \underline{A}'^{-1} \xi^{-1})^p \\ &= (\underline{B}'_0 \underline{z}_0^{-1})^p (\underline{z}_0 \underline{y}_0^{-1})^p \cdots (\underline{a}_0 \underline{A}'^{-1}_0)^p \\ &= \Phi_{B'_0 A'_0}.\end{aligned}$$

Now the path $A'_0 \to B'_0$ is an SO_4 rotation of $A' \to B'$, since the transformation $z \to z_0 = \xi \underline{z} \eta$ is an SO_4 rotation. Furthermore, every SO_4 rotation is such a transformation. Thus an SO_4 rotation only produces a gauge transformation.

Notice that p is an arbitrary real number. So we have exhibited a 1 parameter family of SO_4 symmetrical SU_2 gauge fields in R^4 minus origin.

(d) For $n=3$, we are in more familiar geometry. To construct an SO_3 symmetrical SU_2 gauge field ϵ we take a Dirac U(1) monopole b_μ and put $b^1_\mu = b^2_\mu = 0$, $b^3_\mu = b_\mu$. Such gauge fields are, however, not really interesting because the space does not have enough dimensions to develop the full complexity of the group. One consequence of this lack of enough dimensions is the fact, demonstrated in Ref. 6, that field ϵ is of the same gauge type (i. e., same fibre bundle) as the vacuum field $f^i_{\mu\nu} = 0$.

This work was done in April, 1976 during the author's visit to Futan University, China. It is a pleasure to acknowledge the hospitality the author enjoyed during the visit. The work had been reported at the CERN conference of July, 1976.

APPENDIX A: SOME PROPERTIES OF SU_2 GAUGE FIELDS IN FOUR DIMENSIONS

Consider a SU_2 gauge field in four-dimensional space, with signature ++++, flat or otherwise. We define the antisymmetrical tensor η by

$$\eta_{\alpha\beta\mu\nu} = \sqrt{g}\,\epsilon_{\alpha\beta\mu\nu}, \tag{A1}$$

where $\epsilon = \pm 1$ is the antisymmetrical symbol. We define the dual f^* of a field f by

$$f^*_{\alpha\beta} = \tfrac{1}{2}\eta_{\alpha\beta\mu\nu} f_{\mu\nu}. \tag{A2}$$

Clearly,

$$f^{**} = f. \tag{A3}$$

We only consider coordinate choices that leave $\eta_{1234} > 0$. In other words, a reflection in four-dimensional space is not considered a legitimate transformation. We adopt the terminology at any point P,

$$\begin{aligned} &f^* = f \text{ at } P \longleftrightarrow f \text{ is self-dual at } P, \\ &f^* = -f \text{ at } P \longleftrightarrow f \text{ is self-antidual at } P. \end{aligned} \tag{A4}$$

We further call a gauge field "orthogonal" at a point P if at that point

$$f^i_{\mu\nu} f^{j\lambda\nu} = a^2 \delta^{ij} \delta^\lambda_\mu + a \epsilon^{ijk} f^{k\cdot\lambda}_{\mu}. \tag{A5}$$

It is clear that the orthogonality and self duality properties of a field f at a point is independent of the choice of gauge or the choice of the coordinate system. We shall call the scalar a the *amplitude* of the orthogonal field at P. It is independent of the choice of coordinates and can be positive or negative.

Consider any field f at a point P. Adopt a coordinate system so that at P the metric $g_{\mu\nu} = \delta_{\mu\nu}$. Write the field strengths in the following form:

$$f^j_{\mu\nu} = \begin{bmatrix} 0 & H^j_3 & -H^j_2 & E^j_1 \\ -H^j_3 & 0 & H^j_1 & E^j_2 \\ H^j_2 & -H^j_1 & 0 & E^j_3 \\ -E^j_1 & -E^j_2 & -E^j_3 & 0 \end{bmatrix} \quad (f^j_{12} = H^j_3, \text{ etc.}). \tag{A6}$$

We shall consider f^j $(j = 1, 2, 3)$ as three 6-vectors. The matrices

$$\mathcal{H} = \begin{bmatrix} H^1_1 & H^2_1 & H^3_1 \\ H^1_2 & H^2_2 & H^3_2 \\ H^1_3 & H^2_3 & H^3_3 \end{bmatrix}, \quad \mathcal{E} = \begin{bmatrix} E^1_1 & E^2_1 & E^3_1 \\ E^1_2 & E^2_2 & E^3_2 \\ E^1_3 & E^2_3 & E^3_3 \end{bmatrix} \tag{A7}$$

will be called magnetic and electric matrices. It is obvious that

$$\begin{aligned} &\mathcal{E} = \mathcal{H} \longleftrightarrow \text{self-duality}, \\ &\mathcal{E} = -\mathcal{H} \longleftrightarrow \text{self-antiduality}. \end{aligned} \tag{A8}$$

By substituting (A6) into (A5) we find

$$\begin{aligned} &\mathcal{E} = \mathcal{H},\ \mathcal{H} = a\Gamma \longleftrightarrow \text{self-duality + orthogonality}, \\ &\mathcal{E} = -\mathcal{H},\ \mathcal{H} = a\Gamma \longleftrightarrow \text{self-antiduality + orthogonality}. \end{aligned} \tag{A9}$$

In Eq. (A9), Γ is an orthogonal matrix with determinant $+1$.

A gauge transformation multiples $\mathcal{E}$ and $\mathcal{H}$ from behind by an orthogonal 3×3 matrix R of determinant unity. Thus if $\mathcal{H} = a\Gamma$, there always exists a gauge transformation to make $\mathcal{H} \to aI$. Hence, we have

Lemma 1α: Consider a gauge field which is self-dual and orthogonal at a point P. Consider any coordinate system so that at P, $g_{\mu\nu} = \delta_{\mu\nu}$. The field at P can be gauge transformed to a *standard form* for such fields:

$$\begin{aligned} f^1_{\mu\nu} &= a \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}, \quad f^2_{\mu\nu} = a \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \\ f^3_{\mu\nu} &= a \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \end{aligned} \tag{A10}$$

(e. g., $f^1_{14} = a$). Equation (A10) can also be written as

$$\mathcal{E} = \mathcal{H} = aI \quad (I = \text{unit matrix}). \tag{A10'}$$

Lemma 1β: Consider a gauge field which is self-antidual and orthogonal at a point P. Consider any coordinate system so that at P, $g_{\mu\nu} = \delta_{\mu\nu}$. The field at P can be gauge transformed to a *standard form* for such fields,

$$f^1_{\mu\nu}=a\begin{bmatrix}0&0&0&-1\\0&0&1&0\\0&-1&0&0\\1&0&0&0\end{bmatrix},\quad f^2_{\mu\nu}=a\begin{bmatrix}0&0&-1&0\\0&0&0&-1\\1&0&0&0\\0&1&0&0\end{bmatrix},$$

$$f^3_{\mu\nu}=a\begin{bmatrix}0&1&0&0\\-1&0&0&0\\0&0&0&-1\\0&0&1&0\end{bmatrix}. \tag{A11}$$

Equation (A11) can also be written as

$$-\mathcal{E}=\mathcal{H}=aI \quad (I=\text{unit matrix}). \tag{A11'}$$

Lemma 2: A field which is orthogonal and self-dual or self-antidual at P satisfies, at P,

$$f^i_{\mu\nu}f^{i\mu\nu}=12a^2, \tag{A12}$$

$$f^i_{\mu\nu}f^{*i\mu\nu}=\tfrac{1}{2}f^i_{\mu\nu}\eta^{\mu\nu\alpha\beta}f^i_{\alpha\beta}=\pm 12a^2, \tag{A13}$$

where the + sign is for the case of self-dual fields and the − sign is for the case of self-antidual fields, and a is the amplitude of f.

The proof is trivial.

Consider a four-dimensional space that has geometrically SO_4 symmetry at a point P. Examples are the flat space, a S_4 sphere, or the more general space M_4,

$$ds^2=d\rho^2+\rho^2[e(\rho)]^2\frac{4d\xi^2}{1+\xi^2},\quad -\infty<\xi<\infty, \tag{A14}$$

where $e(\rho)$ is any function of ρ. [If $\rho e=\sin\rho$, we get the sphere S_4, of Eq. (23).] M_4 has SO_4 symmetry at the point $\rho=0$.

For a gauge field γ defined on a space that has geometrically SO_4 symmetry at a point P, we can generate another field γ' by rotating the whole potential (and field around P by an SO_4 rotation. Is γ gauge equivalent to γ' as far as the field strength at P is concerned? If it is, we say that the field is *pointwise* SO_4 *symmetrical* at P.

We shall call a field self-dual orthogonal or self-antidual orthogonal at a point P, *regular* at P. We now have a geometrical meaning of regularity (we shall show later that orthogonality is equivalent to regularity).

Lemma 3: Consider a space that has geometrical SO_4 symmetry at a point P. Then for a SU_2 field,

pointwise SO_4 symmetry at P ⟷ regularity at P. (A15)

Proof:

(a) That the right-hand side implies the left-hand side follows from Lemmas 1α and 1β. To prove the converse, we start with the field $f^i_{\mu\nu}$ of Eq. (A6). An SO_3 rotation around P means, for the field strengths at P,

$$\mathcal{H}\to\Gamma\mathcal{H},\quad \mathcal{E}\to\Gamma\mathcal{E},$$

where Γ is an orthogonal matrix with determinant unity. Pointwise SO_4 symmetry at P implies that there exists a compensating gauge transformation which causes a multiplication from the right by the 3×3 representation R of the compensating SU_2 gauge rotation at the point, i.e., for every Γ there exists an R so that

$$\Gamma\mathcal{H}R=\mathcal{H},\quad \Gamma\mathcal{E}R=\mathcal{E}. \tag{A16}$$

Thus

$$\Gamma\mathcal{H}\tilde{\mathcal{H}}\tilde{\Gamma}=\mathcal{H}\tilde{\mathcal{H}}. \tag{A17}$$

Hence

$$\mathcal{H}\tilde{\mathcal{H}}=h^2I. \tag{A18}$$

Similarly we find that $\mathcal{E}\tilde{\mathcal{E}}$ and $\mathcal{E}\tilde{\mathcal{H}}$ are proportional to the unit 3×3 matrix I. Thus $\mathcal{E}$ and $\mathcal{H}$ are proportional to each other.

(b) Now make a transformation at P that mixes the indices 1 and 4. It is easy to see that $\mathcal{E}+\mathcal{H}$ and $\mathcal{E}-\mathcal{H}$ are independently rotated:

$$\mathcal{E}+\mathcal{H}\to\Gamma_1(\mathcal{E}+\mathcal{H}),$$

$$\mathcal{E}-\mathcal{H}\to\Gamma_2(\mathcal{E}-\mathcal{H}).$$

For $\mathcal{E}$ and $\mathcal{H}$ to remain proportional we must have either

$$\mathcal{E}+\mathcal{H}=0 \quad\text{or}\quad \mathcal{E}-\mathcal{H}=0. \tag{A19}$$

Equations (A18) and (A19) show that we can gauge transform the field at P to the standard form Eqs. (A10) or (A11). This completes the proof of the lemma.

Lemma 4: Choose coordinates so that $g_{\mu\nu}=\delta_{\mu\nu}$ at P. An orthogonal field at P satisfies at P

$$f^1f^1=f^2f^2=f^3f^3=-a^2, \tag{A20}$$

$$f^1f^2=-f^2f^1=-af^3, \quad\text{and cyclic permutation}, \tag{A21}$$

$$f^1f^2f^3=a^3, \tag{A22}$$

where f^i is a 4×4 antisymmetrical matrix with elements

$$\langle\mu|f^i|\nu\rangle=f^i_{\mu\nu}.$$

The proof is easy, starting from definition (A5).

Lemma 5: A field orthogonal at P is either self-dual or self-antidual at P. Therefore, it is regular at P.

Proof: If $a=0$, $f^1=f^2=f^3=0$, and the lemma is proved. If $a\neq0$, f^3/a is an antisymmetrical real matrix whose square is -1, according to (A20). By a well-known theorem one can, by an SO_4 rotation of coordinates at P, bring f^3 into the form displayed in (A10) or (A11). Thus f^3 is either self-dual or self-antidual. (i) If f^3 is self-dual, $f^3=ai\sigma_2$ where $\sigma_1,\sigma_2,\sigma_3,\tau_1,\tau_2,\tau_3$, are the standard 4×4 Pauli matrices. $i(f^1)$ is imaginary Hermitian, and anticommutes with σ_2. Thus $i(f^1)$ is a sum of $\sigma_1\tau_2$ and $\sigma_3\tau_2$ with real coefficients. By another SO_4 rotation of coordinates we can make $f^1=a\sigma_1(i\tau_2)$, leaving $f^3=ai\sigma_2$. $i(f^2)$ is imaginary Hermitian and anticommutes with both f^1 and f^3. Thus $f^2=\xi(a^2)^{1/2}(-i\sigma_3\tau_2)$, $\xi=\pm1$. Thus f^1,f^2,f^3 are all self-dual. (ii) If f^3 is self-antidual, we can similarly prove that f^1 and f^2 are also self-antidual.

Lemma 6:

$$[f^i_{\mu\nu}f^{j\lambda\nu}+f^j_{\mu\nu}f^{i\lambda\nu}=2a^2\delta^{ij}\delta^\lambda_\mu] \longleftarrow \text{(A5)}. \tag{A23}$$

Proof: That (A5) implies the left-hand side is obvious. If the left-hand side holds, the sign of a is for us to choose. We choose coordinates at P so that $g_{\mu\nu}=\delta_{\mu\nu}$. Then (A20) holds. In fact

$$f^if^j+f^jf^i=-2a^2\delta^{ij}. \tag{A24}$$

Now the proof of Lemma 5 depends only on this equation. Following that proof we conclude that there are only two possibilities.

(i) We can by an SO_4 rotation at P bring the f's into the following form,

$$f^1 = |a|\sigma_1(i\tau_2), \quad f^2 = \xi|a|(-i\sigma_3\tau_2), \quad f^3 = |a|i\sigma_2.$$

Thus f^1, f^2, f^3 satisfy (A20) and (A21) with $a = \xi|a|$. (A20) and (A21) together imply (A5).

(ii) f^3 is antiselfdual. The proof is similar.

APPENDIX B: FIELD STRENGTHS FOR FIELD α

Using Eq. (31) we find how to obtain $(f^i_{\mu\nu})_\alpha$ from $(f^i_{\mu\nu})_\beta$:

$$[f^1_{\theta i}(\xi, \theta)]_\alpha = -[f^1_{\theta i}(-\xi, \theta)]_\beta \quad (i = 1, 2, 3), \tag{B1}$$

$$[f^1_{ij}(\xi, \theta)]_\alpha = [f^1_{ij}(-\xi, \theta)]_\beta \quad (i, j = 1, 2, 3). \tag{B2}$$

APPENDIX C: PROOF THAT FIELD β IS ORTHOGONAL SELF-ANTIDUAL

One proof consists in evaluating the left-hand side of Eq. (A5) using Eq. (37β). The calculation is long but straightforward.

Another proof follows the steps in Appendix A by starting with a scale change from variable $\xi_1, \xi_2, \xi_3, \xi_4 \to y_1, y_2, y_3, y_4$,

$$\theta = \theta_0 + y_4 r^{-1}, \quad \xi_i = (\xi_i)_0 + y_i(1 + \xi_0^2)(2r_0 \sin\theta)^{-1}. \tag{C1}$$

Then *in the y variables*, at $y_\mu = 0$ the metric is unity. $f^i_{\mu\nu}$ in the y variables are easily obtained from Eq. (37β). We arrange it in the form of Eqs. (A6) and (A7), obtaining

$$\mathcal{H} = -\mathcal{E} = -(1 + \xi^2)\tilde{B}(4r^2)^{-1}, \tag{C2β}$$

where we have dropped the subscript 0 in all variables. Using Eq. (33) we find

$$\mathcal{H}\tilde{\mathcal{H}} = r^{-4}I \quad (I = \text{unit matrix}). \tag{C3}$$

To determine the value of $\det\mathcal{H}$ we put $\xi = 0$. Then $B = -4I$. Thus

$$\det\mathcal{H} > 0 \quad \text{everywhere.} \tag{C4}$$

We can then make a gauge transformation to make

$$-\mathcal{E} = \mathcal{H} \to \mathcal{H}R = r^{-2}I.$$

Thus we have arrived at the standard form Eq. (A11) showing that field β is orthogonal self-antidual with $a = r^{-2}$, in agreement with Eq. (41).

The same calculations can be made for field α. Using Eqs. (B1) and (B2), we see that all formulas are unchanged, except Eq. (C2β) becomes in this case

$$\mathcal{H} = \mathcal{E} = -(1 + \xi^2)B(4r^2)^{-1}. \tag{C2α}$$

[1]P. A. M. Dirac, Proc. Roy. Soc. A 133, 60 (1931).

[2]We use the notation of Chen Ning Yang, *Proceedings of the Sixth Hawaii Topical Conference*, 1975 (University of Hawaii Press, Hawaii, 1976). Specifically, $\sum^a$ means antisymmetrical sum with respect to all Greek indices. $,\lambda$ means $\partial/\partial\chi^\lambda$. Dummy indices are summed over. X_i are the generators of SU_2 satisfying $[X_1, X_2] = X_3$, etc. $C^1_{23} = 1$.

[3]Shiing-shen Chern, *Topics in Differential Geometry* (Inst. for Adv. Study, Princeton, 1951); Ann. Math. (N.Y.) 47, 85 (1946).

[4]Chen Ning Yang, Rev. Lett. 33, 445 (1974), and in *Proceedings of the Sixth Hawaii Topical Conference*, 1975 (University of Hawaii Press, Hawaii, 1976).

[5]Tai Tsun Wu and Chen Ning Yang, Nucl. Phys. B 107, 365 (1976).

[6]Twi Tsun Wu and Chen Ning Yang, Phys. Rev. D 12, 3845 (1975).

[7]We denote by $(xx\alpha)$ an equation that is valid for field α, $(xx\beta)$ denotes one that is valid for field β. An equation number without α or β is valid for both fields.

[8]$\xi^1 = \xi_1$, $\xi^2 = \xi_2$, and $\xi^3 = \xi_3$ are not tensors. $d\xi^1$, $d\xi^2$, and $d\xi^3$ are.

[9]M. Fierz, Helv. Phys. Acta 17, 27 (1944).

[10]A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B 59, 85 (1975). See generalization in E. Witten, Phys. Rev. Lett. 38, 121 (1977).

[11]R. Jackiw and C. Rebbi, Phys. Rev. D 14, 517 (1976).

Orthocomplemented subspaces of nondegenerate partial inner product spaces[a)]

J.-P. Antoine

Institut de Physique Théorique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

A. Grossmann

Centre de Physique Théorique, CNRS, Marseille, Cedex 2, France
(Received 29 July 1977)

A vector subspace of a nondegenerate partial inner product space is orthocomplemented if and only if it satisfies a certain topological regularity property. If a nondegenerate, positive definite partial inner product space has "many" orthocomplemented subspaces, it is a Hilbert space.

1. INTRODUCTION

In Refs. 1 and 2 we have studied nondegenerate partial inner product spaces. Their definition runs as follows:

Call *compatibility* any symmetric binary relation $\#$ in a set V. For any $f \in V$, denote by $\{f\}^{\#}$ the set of all $g \in V$ such that $g \# f$. If $S \subseteq V$, write

$$S^{\#} = \bigcap_{f \in S} \{f\}^{\#}.$$

A compatibility is called *linear* if V is a vector space and if all $\{f\}^{\#}$ are vector subspaces of V.

A *partial inner product space* is a vector space V over $\mathbb{C}$, with a linear compatibility relation and with a Hermitian form $f, g \to \langle f|g\rangle$, defined whenever f and g are compatible, i.e., precisely when $f \# g$. (The partial inner product $\langle f|g\rangle$ is antilinear in its dependence on the argument on the left.)

Two vector subspaces E, F of a partial inner product space V are said to form a *dual pair* if

(i) $f \# g$ for all $f \in E$, $g \in F$

and

(ii) If $f \in E$ and $\langle f|g\rangle = 0$ for all $g \in F$, then $f = 0$. If $g \in F$ and $\langle f|g\rangle = 0$ for all $f \in E$, then $g = 0$.

If V and $V^{\#}$ are a dual pair, we say that V is *nondegenerate*. So, a nondegenerate partial inner product space is a vector space with a linear compatibility relation, and a Hermitian form on compatible vectors, such that V and $V^{\#}$ are a dual pair.

Nondegeneracy is a statement about both the Hermitian form $\langle f|g\rangle$ and about the supply of compatible pairs of vectors. Namely, $V^{\#}$ consists of vectors that are "infinitely good" in the sense of being compatible with all vectors of V; the assumption of nondegeneracy means that no $f \in V$ except 0 is orthogonal to all infinitely good vectors.

Very many spaces in analysis (e.g., distributions or sequence spaces) have a natural partial inner product structures. See Refs. 1 and 2 for examples.

If all vectors in V are mutually compatible, then V is a nondegenerate *inner* product space (Refs. 3 and 4). If, furthermore, $\langle f|f\rangle > 0$ for $f \neq 0$, then V is a *pre-Hilbert* space. Finally, a pre-Hilbert space that is complete for the norm topology is a *Hilbert* space. So, a Hilbert space is a very special nondegenerate partial inner product space.

A vector subspace W of a nondegenerate *inner* product space V is called *orthocomplemented* if every vector in V can be written, in a unique way, as a sum of a vector in W and a vector orthogonal to W.

If V is Hilbert, its orthocomplemented subspaces are easily characterized: $W \subseteq V$ is orthocomplemented if and only if it is closed. This result is one of the main reasons for the usefulness of Hilbert spaces.

For arbitrary (nondegenerate) *inner* product spaces, one has the following result: $W \subseteq V$ is orthocomplemented if and only if the "intrinsic" weak topology $\sigma(W, W)$ on W coincides with the topology $\sigma(V, V)|_W$ induced by V; this happens precisely when the intrinsic and induced Mackey topologies coincide: $\tau(W, W) = \tau(V, V)|_W$. If V is Hilbert, this reduces to the statement above (see Refs. 3 and 4 and the references given there).

The aim of the present paper is to define and characterize orthocomplemented subspaces of arbitrary nondegenerate *partial* inner product spaces. Our main result is Theorem 4.3, which says that a subspace is orthocomplemented if and only if a suitable family of induced topologies coincides with a family of intrinsic topologies. Orthocomplemented subspaces can also be defined as ranges of orthogonal projections, which have a natural intrinsic definition (Sec. 3).

These results are remarkably similar to those obtained for inner product spaces. Stated otherwise, as soon as the inner product may be indefinite, it does not make much difference whether it is defined everywhere or not.

An interesting case is that of a space V, for which the partial inner product is positive definite, in the sense that $\langle f|f\rangle > 0$ for every nonzero $f \in V^{\#}$. If such a space has "too many" orthocomplemented subspaces, then it is necessarily a Hilbert space (Sec. 8). This generalizes a known result about pre-Hilbert spaces (Proposition 7.1[5,6]).

[a)]Dedicated to Professor L. P. Bouckaert on the occasion of his seventieth birthday.

 0022-2488/78/1901-0329$1.00

Notation

We recall here some terminology and results of Refs. 1 and 2.

If E, F is any dual pair in the nondegenerate partial inner product space V, the weak topology (Mackey topology) on E of the dual pair is denoted by $\sigma(E, F)$ ($\tau(E, F)$).

The *assaying subsets* of V are defined as subsets $S \subseteq V$ that satisfy

$$S^{\#\#} = S.$$

Every assaying subset is a vector subspace of V. Both V and $V^{\#}$ are assaying subsets. If V is an inner product space, then $V = V^{\#}$ is the only assaying subset.

$F(V)$ denotes the family of all assaying subsets of V, ordered by inclusion and considered as an abstract partially ordered set. If $r \in F(V)$, the corresponding assaying subset will be denoted by V_r. We write, furthermore,

$$V_{\bar{r}} \equiv (V_r)^{\#}.$$

V_r and $V_{\bar{r}}$ are a dual pair.

2. REDUCTION OF COMPATIBILITY RELATIONS

Let V be a vector space with a linear compatibility relation $\#$.

Given a direct sum decomposition $V = V_1 \oplus V_2$, we can ask whether $\#$ can be reconstructed from its restriction to V_1, V_2 and the knowledge of compatible pairs $f_1 \in V_1$, $f_2 \in V_2$.

2.1 *Definition*: A direct sum decomposition $V = V_1 \oplus V_2$ *reduces* the linear compatibility $\#$ if the condition $f \# g$ ($f \in V$, $g \in V$) is equivalent to the four conditions

$$f_1 \# g_1, \quad f_2 \# g_2, \quad f_1 \# g_2, \quad f_2 \# g_1$$

$$(f_1 \in V_1,\ f_2 \in V_2,\ g_1 \in V_1,\ g_2 \in V_2,\ f_1 + f_2 = f,\ g_1 + g_2 = g),$$

i.e., if $\#$ can be expressed in terms of components. The compatibility relation $\#$ is *absolutely reduced* if every vector in V_1 is compatible with every vector in V_2. The condition $f \# g$ is then equivalent to $f_1 \# g_1$ and $f_2 \# g_2$; the spaces V_1 and V_2 are decoupled from each other.

2.2 *Example:* Starting with two vector spaces V_1, V_2 with linear compatibility relations $\#_1, \#_2$, one can easily endow $V = V_1 \oplus V_2$ with compatibility relations that are reduced by the direct sum decomposition: It is enough to introduce any linear compatibility *between* V_1 and V_2 (i.e., a linear symmetric relation $f_1 \#_{12} g_2$) and declare $f \# g$ if and only if $f_1 \#_1 g_1$, $f_2 \#_2 g_2$, $f_1 \#_{12} g_2$, $f_2 \#_{12} g_1$).

2.3 *Example*: Of direct sum decomposition that does *not* reduce a given compatibility:

Let $V = V_1 \oplus V_2$, V_1 one-dimensional, spanned by a vector f that is not compatible with itself.

2.4 *Proposition*: Let V be a vector space with a linear compatibility relation $\#$. Let

$$V = V_1 \oplus V_2 \tag{1}$$

be a direct sum decomposition of V, and P_1, P_2 the corresponding (algebraic) projections. Then the following four conditions are equivalent:

(i) The compatibility relation $\#$ is reduced by the direct sum decomposition (1);

(ii) $\{P_1 f\}^{\#} \supseteq \{f\}^{\#}$ for every $f \in V$;

(iii) $\{P_2 f\}^{\#} \supseteq \{f\}^{\#}$ for every $f \in V$;

(iv) $\{f\}^{\#} = \{P_1 f\}^{\#} \cap \{P_2 f\}^{\#}$ for every $f \in V$.

The proof is straightforward and can be omitted.

The compatibility relation $\#$ is absolutely reduced by the decomposition (1) if and only if $P_1 f \# P_1 g$ implies $P_1 f \# g$ and $f \# P_1 g$; the same relations follow then for P_2. When this happens, we shall say that P_1 and P_2 are *absolute*.

If, say, $V_1 \subseteq V^{\#}$, then P_1 and P_2 are absolute.

3. ORTHOGONAL PROJECTIONS

3.1 *Definition*: Let V be a nondegenerate partial inner product space. We shall say that a linear map P from all of V into V is an *orthogonal projection* if

(i) $P^2 = P$,

(ii) $\{Pf\}^{\#} \supseteq \{f\}^{\#}$ for every $f \in V$,

(iii) if $f \# g$, then $\langle g | Pf \rangle = \langle Pg | f \rangle$.

The set of all orthogonal projections in V is denoted by $\mathrm{Proj}(V)$.

Conditions (i) and (iii) mimic the familiar Hilbert space definition. Condition (ii) is natural in view of Proposition 2.4. It says that the decomposition $V = PV \oplus (1 - P)V$ reduces the compatibility in V.

By Proposition 3.3 of Ref. 2, P maps every assaying subset V_r into itself continuously for $\tau(V_r, V_{\bar{r}})$. Thus $P \in \mathrm{Op}(V, V)$, the natural domain of P is all of V, and $P^* = P$.

An equivalent definition runs as follows:

Consider in $\mathrm{Op}(V, V)$ the subset[7] $\mathcal{A}(V, V)$ consisting of all operators such that A and A^* improve behavior [i.e., that $\{r, r\} \in J(A)$ or, equivalently, $AV_r \subseteq V_r$ and $A^* V_r \subseteq V_r$ for all $r \in F(V)$]. It is easy to see that $\mathcal{A}(V, V)$ is a $*$-algebra. Then the orthogonal projections are precisely the symmetric (i.e., $P^* = P$) idempotents of that algebra.

3.2 *Proposition*: Let V be a nondegenerate partial inner product space, and P_1, P_2 orthogonal projections in V. Then

(i) 1 (the identity) and 0 (the zero map) belong to $\mathrm{Proj}(V)$,

(ii) if $P_1 P_2 = P_2 P_1$, then $P_1 P_2 \in \mathrm{Proj}(V)$.

(iii) if $P_1 P_2 = P_2 P_1 = 0$, then $P_1 + P_2 \in \mathrm{Proj}(V)$.

(iv) in particular, $1 - P \in \mathrm{Proj}(V)$ for every $P \in \mathrm{Proj}(V)$.

The proof is an immediate verification.

Example: If $f \in V^{\#}$ and $\langle f | f \rangle = 1$, then $P_f = |f\rangle\langle f| \in \mathrm{Proj}(V)$. By (iii) above, this generalizes to a finite

sum of projections corresponding to a family of mutually orthogonal, normalized vectors.

Additional examples of orthogonal projections will be discussed in Sec. 6.

3.3 *Proposition*: Let $P \in \mathrm{Proj}V$, and $Q = 1 - P$. Then, for every $r \in F(V)$,

(i) V_r decomposes into a (topological) direct sum $V_r = PV_r \oplus QV_r$,

(ii) the representatives P_{rr}, Q_{rr} are related by $\ker P_{rr} = \text{range } Q_{rr}$,

(iii) PV_r and $PV_{\bar{r}}$ are a dual pair (and so are QV_r and $QV_{\bar{r}}$),

(iv) one has

$$(PV_r)^{\perp} \cap V_{\bar{r}} = QV_{\bar{r}},$$

(v) if $r \leq s$, then PV_r is dense in PV_s [for the topology $\tau(V_s, V_{\bar{s}})$].

Proof: (i) and (ii) follow immediately from $PQ = 0$ and $PV_r \subseteq V_r$, $QV_r \subseteq V_r$ (continuously). In order to prove (iii) notice first that PV_r and $PV_{\bar{r}}$ are compatible. Now take $f \in PV_r$ and $f \in (PV_{\bar{r}})^{\perp}$. The first relation shows that f is orthogonal to $QV_{\bar{r}}$. Since it is also orthogonal to $PV_{\bar{r}}$, it is orthogonal to $V_{\bar{r}}$, and is consequently zero. In order to prove (iv), notice that

$$(PV_r)^{\perp} \cap V_{\bar{r}} \subseteq [(PV_r)^{\perp} \cap PV_{\bar{r}}] \cup [(PV_r)^{\perp} \cap QV_{\bar{r}}]$$

$$= (PV_r)^{\perp} \cap QV_{\bar{r}} = QV_{\bar{r}}.$$

The inclusion in the other direction is immediate.

To prove (v), we compute the closure of PV_r in V_s [with $\tau(V_s, V_{\bar{s}})$] using Proposition 5.1 of Ref. 1. We have

$$\mathrm{cl}(PV_r) = ((PV_r)^{\perp} \cap V_{\bar{s}})^{\perp} \cap V_s = ((PV_r)^{\perp} \cap V_{\bar{r}} \cap V_{\bar{s}})^{\perp} \cap V_s$$

$$= (QV_{\bar{r}} \cap V_{\bar{s}})^{\perp} \cap V_s = (QV_{\bar{s}})^{\perp} \cap V_s = PV_s.$$

In the special case where V is an inner product space, this proposition reduces to a result stated in Ref. 3, Sec. III.6.

3.4 *Remark*: As a special case, we find the useful result:

$$\mathrm{Ker}P = \mathrm{range}Q = (PV^{\#})^{\perp} = (W \cap V^{\#})^{\perp}$$

3.5 *Corollary*: If W is the range of an orthogonal projection, then it is a nondegenerate partial inner product space with the inherited compatibility and partial inner product.

4. ORTHOCOMPLEMENTED SUBSPACES

4.1 *Definition*: Let V be a nondegenerate partial inner product space and W a vector subspace of V. We shall say that W is *orthocomplemented* if there exists a vector subspace $Z \subseteq V$ such that

(i) $W \cap Z = \{0\}$, $W + Z = V$,

(ii) the direct sum decomposition $V = W \oplus Z$ reduces the compatibility relation in V,

(iii) if $f \in W$, $g \in Z$, and if $f \# g$, then $\langle f | g \rangle = 0$.

4.2 *Definition*: A vector subspace W of a nondegenerate partial inner product space V is said to be *topologically regular* if

(i) for every assaying subset $V_r \subseteq V$, the intersections $W^r = W \cap V_r$ and $W^{\bar{r}} = W \cap V_{\bar{r}}$ are a dual pair in V, and

(ii) the intrinsic Mackey topology $\tau(W^r, W^{\bar{r}})$ coincides with the induced Mackey topology $\tau(V_r, V_{\bar{r}})|_{W^r}$.

Remark: We shall see that in this statement Mackey topologies can be replaced by weak topologies.

4.3 *Theorem*: For a vector subspace W of the nondegenerate partial inner product space V the following conditions are equivalent:

(i) W is orthocomplemented,

(ii) W is the range of an orthogonal projection,

(iii) W is topologically regular.

We shall divide the proof into several lemmas.

4.4 *Lemma*: A vector subspace $W \subseteq V$ is orthocomplemented if and only if it is the range of an orthogonal projection.

Proof: Let W be orthocomplemented, $V = W \oplus Z$. For every $f \in V$, let $f = f_w + f_z$ be its unique decomposition. Write $f_w = Pf$. Then P is an orthogonal projection. Indeed, P is idempotent and improves behavior by assumption. In order to verify that P is symmetric, let $g \# h$, $g = g_w + g_z$, $h = h_w + h_z$. Then

$$\langle Pg | h \rangle = \langle g_w | h_w + h_z \rangle = \langle g_w | h_w \rangle = \langle g_w | Ph \rangle = \langle g | Ph \rangle$$

All the inner products exist since $\#$ is reduced.

Conversely, let $W = PV$ be the range of an orthogonal projection. Then W is clearly orthocomplemented with $Z = (1 - P)V$.

4.5 *Lemma*: If W is the range of an orthogonal projection, then, for every r, W^r, and $W^{\bar{r}}$ are a dual pair.

This is (iii) of Proposition 3.3.

4.6 *Lemma*: If P is an orthogonal projection and T is any subset of $V_{\bar{r}}$, then

$$T^{\circ} \cap W^r = (PT)^{\circ} \cap W^r$$

where T° is the polar of T (see Ref. 1).

Proof: Let $h \in V_{\bar{r}}$, and $f \in W^r$. Then $\langle f | h \rangle = \langle Pf | h \rangle = \langle f | Ph \rangle$. It follows that $\{h\}^{\circ} \cap W^r = \{Ph\}^{\circ} \cap W^r$ for every $h \in V_{\bar{r}}$. Consequently, $T^{\circ} \cap W^r = (PT)^{\circ} \cap W^r$ for any $T \subseteq V_{\bar{r}}$.

4.7 *Lemma*: If W is the range of an orthogonal projection then, for every $r \in F(V)$ the topology $\sigma(W^{\bar{r}}, W^r)$ coincides with $\sigma(V_{\bar{r}}, V_r)|_{W^{\bar{r}}}$.

Proof: In Lemma 4.6, take T a finite subset of V_r, obtaining

$$T^{\circ} \cap W^{\bar{r}} = (PT)^{\circ} \cap W^{\bar{r}}.$$

As T runs over all finite subsets of V_r, the lhs of the above equation goes over a basis of neighbourhoods for the inherited topology, and the rhs over a basis of neighborhoods for $\sigma(W^{\bar{r}}, W^r)$.

4.8 *Lemma*: Let P be an orthogonal projection and V_τ an assaying subspace of V. Let $Z \subseteq V_{\bar\tau}$ be convex and $\sigma(V_{\bar\tau}, V_\tau)$-compact. Then $PZ \subseteq W^{\bar\tau}$ is convex and $\sigma(W^{\bar\tau}, W^\tau)$-compact. Conversely, if $S \subseteq W^{\bar\tau}$ is convex and $\sigma(W^{\bar\tau}, W^\tau)$-compact, then there exists a convex $\sigma(V_{\bar\tau}, V_\tau)$-compact Z such that $S = PZ$.

Proof: The convexity of PZ is clear. The $\sigma(V_{\bar\tau}, V_\tau)$-compacity of PZ follows from $\sigma(V_{\bar\tau}, V_\tau)$-continuity of $P_{\bar\tau\tau}$ (Lemma 3.1 of Ref. 2) and the $\sigma(W^{\bar\tau}, W^\tau)$-compacity of PZ follows then from Lemma 4.7 above. Conversely, if $S \subseteq W^{\bar\tau}$ is $\sigma(W^{\bar\tau}, W^\tau)$-compact, it is also $\sigma(V_{\bar\tau}, V_\tau)$-compact by Lemma 4.7. Since $PS = S$, we can choose $Z = S$.

4.9 *Lemma*: If W is the range of an orthogonal projection, then W is topologically regular.

Proof: We have to show that, for every τ, the Mackey topology $\tau(W^\tau, W^{\bar\tau})$ coincides with $\tau(V_\tau, V_{\bar\tau})|_W$. A basis of neighborhoods for $\tau(W^\tau, W^{\bar\tau})$ is given by the sets $S^\circ \cap W^\tau$, with S ranging over all $\sigma(W^{\bar\tau}, W^\tau)$-compact, convex subsets of $|_{W^\tau}$. A basis of neighborhoods for $\tau(V_\tau, V_{\bar\tau})|_{W^\tau}$ is given by the sets $Z^\circ \cap W^\tau$ with Z ranging over all $\sigma(V_{\bar\tau}, V_\tau)$-compact, convex subsets of $V_{\bar\tau}$. By Lemmas 4.6 and 4.8, the two families of neighborhoods coincide.

Alternate proof of Lemma 4.9: According to (i) of Proposition 3.3, V_τ decomposes into a topological direct sum

$$V_\tau = PV_\tau \oplus QV_\tau$$

with $W^\tau = PV_\tau$, $Q = 1 - P$, and so does $V_{\bar\tau}$:

$$V_{\bar\tau} = PV_{\bar\tau} \oplus QV_{\bar\tau}.$$

By a standard theorem,[8] W^τ, with the topology induced by $\tau(V_\tau, V_{\bar\tau})$, is then topologically isomorphic to the quotient V_τ/QV_τ, the latter being equipped with its quotient topology. By (iv) of the same Proposition 3.3

$$QV_\tau = (PV_{\bar\tau})^{\perp\perp} \cap V_\tau = (W^{\bar\tau})^{\perp\perp} \cap V_\tau.$$

Since V_τ carries its Mackey topology $\tau(V_\tau, V_{\bar\tau})$, the quotient topology on V_τ/QV_τ is $\tau(V_\tau/(W^{\bar\tau})^{\perp\perp} \cap V_\tau, W^{\bar\tau}) = \tau(W^\tau, W^{\bar\tau})$ (Mackey topologies are inherited by quotients; see Ref. 9, IV. 4.1 Coroll. 3). This concludes the alternative proof of Lemma 4.9.

In order to complete the proof of Theorem 4.3, we have to show that every topologically regular subspace is the range of an orthogonal projection. It is convenient to introduce here an auxiliary notion:

If W is a vector subspace of V, one may consider on it the restriction of all linear forms $\langle f|$, with $f \in W$, and ask in analogy to the above whether all such forms can be "internally generated." This gives the following:

We say that W is *self-sufficient in linear forms* if the following holds:

For every $f \in V$, there exists a $f_W \in W$, at least as good as f, such that

$$\langle f_W|_{W \cap \{f\}^\#} = \langle f|_{W \cap \{f\}^\#}.$$

The notation is explained in Sec. 3 of Ref. 1.

4.10 *Proposition*: (i) If W is topologically regular, then it is self-sufficient in linear forms.

(ii) If W is self-sufficient in linear forms, then it is the range of a (unique) orthogonal projection.

Proof of (i): Let $f \in V$. We want to find a $f_W \in W$ such that $\langle f_W|_{W \cap \{f\}^\#} = \langle f|_{W \cap \{f\}^\#}$. Define r by $V_r = \{f\}^\#$, $V_{\bar r} = \{f\}^{\#\#}$, and consider the intersections $W^r = W \cap V_r$, $W^{\bar r} = W \cap V_{\bar r}$. By the assumption of regularity, W^r and $W^{\bar r}$ are a dual pair, and $\tau(W^r, W^{\bar r})$ coincides with the topology inherited from $\tau(V_r, V_{\bar r})$. Consequently, the restriction of $\langle f|$ to $W^r = W \cap \{f\}^\#$ is continuous for $\tau(W^r, W^{\bar r})$. So there exists a unique $f_W \in W^{\bar r}$ such that $\langle f_W|_{W^r} = \langle f|_{W^r}$. Since $W^{\bar r} \subseteq V_{\bar r} = \{f\}^{\#\#}$, one has that $\{f_W\}^\# \supseteq \{f\}^\#$, i.e., f_W is at least as good as f, which proves the statement.

In order to prove part (ii) of Proposition 4.10, we need two small lemmas.

4.11 *Lemma*: Let W be self-sufficient in linear forms. Given $f \in V$, there exists only one $f_W \in W$ such that $\{f_W\}^\# \supseteq \{f\}^\#$ and that $\langle f_W| = \langle f|$ on $W \cap \{f\}^\#$.

Proof: Let f_W^1 and f_W^2 be two such vectors and $g = f_W^1 - f_W^2$. Then $g \in W$ and $\langle g|h\rangle = 0$ for every $h \in W \cap \{f\}^\#$. *A fortiori*, g is orthogonal to every $h \in W \cap V^\#$, hence $g = 0$ by the lemma that follows.

4.12 *Lemma*: Let W be self-sufficient in linear forms. Then

$$(W \cap V^\#)^{\perp\perp} \cap W = \{0\}.$$

Proof: Let $f \in V^\#$, then $\{f\}^\# = V$, $W \cap \{f\}^\# = W$. By self-sufficiency, there exists at least one $f_W \in W \cap V^\#$ such that $\langle f_W|_W = \langle f|_W$. Let $g \in W$ be orthogonal to $W \cap V^\#$. Then

$$0 = \langle f_W|g\rangle = \langle f|g\rangle,$$

that is, g is orthogonal to every $f \in V^\#$. Hence $g = 0$.

Proof of Proposition 4.10 (*ii*): The correspondence $f \to f_W$ is well defined by virtue of Lemma 4.11. Write $f_W = Pf$. Then we claim that P is an orthogonal projection. Since f_W is at least as good as f, P improves behaviour. For any $f \in W$, $f = Pf$; thus W is the range of P, and P is idempotent since for any $f \in V$, $Pf \in W$, and $P^2 f = Pf$. Finally, let $f \# g$. Since P improves behavior, $f \# Pg$, $Pf \# g$ and $Pf \# Pg$. By the definition of P, $\langle Pf|h\rangle = \langle f|h\rangle$ for every $f \in W$. In particular, we have (for any $f \# g$)

$$\langle f|Pg\rangle = \langle Pf|Pg\rangle = \langle Pf|g\rangle.$$

This completes the proof of Theorem 4.3.

This theorem gives a complete characterization of all orthocomplemented subspaces. It is gratifying to notice that exactly the same result holds for the simple case of inner product spaces (Ref. 3, Theorem III. 7.2).

5. THE ORDER STRUCTURE OF PROJ (V)

Since an orthocomplemented W defines uniquely the corresponding projection P_W, we may define a partial order in Proj(V) by

$$P_W \leq P_Y \text{ if and only if } W \subseteq Y.$$

It is easy to see that this is equivalent to

$$P_W P_Y = P_Y P_W = P_W.$$

There are many questions about Proj(V) which we are unable to answer in general. For example: When is Proj(V) a (complete) lattice?

We are able, however, to compare Proj(V) and the partially ordered set Proj($V^\#$) of all the orthogonal projections in the *inner* product space $V^\#$.

5.1 *Theorem*: Let V be an nondegenerate partial inner product space and W an orthocomplemented subspace of V. Then $W \cap V^\#$ is an orthocomplemented subspace of the *inner* product space $V^\#$. The correspondence $W \to W \cap V^\#$ defines an injective order-preserving map α from Proj(V) into Proj($V^\#$). An orthogonal projection $P \in \text{Proj}(V^\#)$ lies in the range of α if and only if it is continuous in every one of the topologies $\tau(V^\#, V_r)$.

Proof: Let W be an orthocomplemented subspace of V and P_W the corresponding projection. One verifies immediately that the restriction of P_W to $V^\#$ is an orthogonal projection in $V^\#$; as such, it is continuous for $\tau(V^\#, V^\#)$, which is the topology induced on $V^\#$ by $\tau(V, V^\#)$. Consequently, $P_W V^\#$ is an orthocomplemented subspace of $V^\#$. Furthermore, $P_W V^\# = W \cap V^\#$, which proves the first assertion.

Notice that, for each $r \in F$, the restriction of P_W to $V^\#$ is continuous in the topology induced on $V^\#$ by V_r, namely $\tau(V^\#, V_{\bar{r}})$.

Conversely, let $P \in \text{Proj}(V^\#)$ be continuous for every one of the topologies $\tau(V^\#, V_q)$. Then P can be extended to a unique $P_W \in \text{Proj}(V)$. Indeed P can be considered as the representative (from $V^\#$ to $V^\#$) of a unique element P_W of Op(V, V); our assumptions guarantee that this element is an orthogonal projection.

5.2 *Proposition*: Let V be a nondegenerate partial inner product space, W a finite-dimensional subspace of V. Then W is orthocomplemented if and only if it is contained in $V^\#$ and nondegenerate, i.e., $W \cap W^\perp = \{0\}$.

Proof: Let W be finite-dimensional and orthocomplemented. By Proposition 3.3(v), $W \cap V^\#$ is dense in W, which means $W \cap V^\# = W$, thus $W \subseteq V^\#$. By (iii) of the same proposition, $PV = W$ and $PV^\# = W$ are a dual pair, i.e., $W \cap W^\perp = \{0\}$.

Conversely, if $W \subseteq V^\#$, we have $W \cap V^\# = W$ and $W \cap V_r = W$ for every $r \in F$. Since $W \cap W^\perp = \{0\}$, $\tau(W, W)$ is well-defined and coincides with the topology induced by $\tau(V_r, V_{\bar{r}})$, for any r, since all separated topologies coincide on a finite-dimensional subspace. Thus W is topologically regular, and, therefore, orthocomplemented.

5.3 *Corollary*: An orthogonal projection of finite-dimensional range is absolute.

Notice that every vector of $W \cap W^\perp$ is necessarily orthogonal to itself. So, if the partial inner product is definite (i.e., $f \# f$ and $\langle f | f \rangle = 0$ imply $f = 0$), the condition of nondegeneracy is superfluous. Anyway it is remarkable that not even all finite-dimensional subspaces of V are orthocomplemented. We have already noted in the Introduction that, in the general case, the situation for orthocomplemented subspaces is essentially the same whether the inner product is defined everywhere or not. We see here that, for finite-dimensional subspaces, there is no difference at all (so all the necessary information may be found in Ref. 3).

6. EXAMPLES AND FURTHER REMARKS

6.1 *Examples*: (1) Let $V = L^1_{\text{loc}}(X; d\mu)$ (Example 5 of Ref. 1, Sec. 4) or, in particular, $V = \omega$ (space of all complex sequences) (Example 2 of Ref. 1). Take any partition of X into two measurable subsets, of nonzero measure, $X = \Omega \cup \Omega'$. Then V is decomposed in two orthocomplemented subspaces:

$$V = L^1_{\text{loc}}(\Omega, d\mu_\Omega) \oplus L^1_{\text{loc}}(\Omega', d\mu_{\Omega'})$$

where $\mu_\Omega, \mu_{\Omega'}$, are the restrictions of μ to Ω, Ω' resp. The orthogonal projection P_Ω is the operator of multiplication by the characteristic function χ_Ω of Ω. Similarly, $P_{\Omega'} = 1 - P_\Omega$ is multiplication by $\chi_{\Omega'}$. They are both absolute.

(2) An example of nonabsolute projection may be obtained as follows. Consider again $V = L^1_{\text{loc}}(X, d\mu)$ with the following compatibility relation: $f \# g$ if at least one of them belongs to $L^\infty_c(X, d\mu)$. Then $V^\# = L^\infty_c(X, d\mu)$ and the only assaying subspaces are V and $V^\#$. Let Ω be a measurable subset of X and P be the multiplication by χ_Ω, the characteristic function of Ω. This P is obviously an orthogonal projection. Take now an element $f \in V$ with support in Ω and another one, g say, with support in the complement $\Omega' = X \setminus \Omega$, but none of them in $V^\#$. Thus f and g are not compatible. By construction, we have $Pf = f$ and $(1 - P)g = g$. Thus Pf and $(1 - P)g$ are not compatible, i.e., P is not absolute.

(3) Take again $V = T(X, \mu)$, the "partial Fock space" discussed in Ref. 2, Sec. 6, Example 6. The operator P_n defined by $(P_n f)_k = \delta_{nk} f_n$ is clearly an absolute projection, with range $0 \oplus \cdots \oplus V^{(n)} \oplus 0 \cdots$. On the other hand, the symmetrization operator S, defined by $(Sf)_n = S_n f_n$, where S_n symmetrizes the n arguments of f_n, is not an orthogonal projection. Indeed $S_n f_n$ and g_n are not necessarily compatible, whenever f_n and g_n are, since the integral $\int |S_n f_n| \, |g_n| \, d\mu$ need not converge, even if $\int |f_n g_n| \, d\mu$ does.

6.2 *Quotients*: Let W be an orthocomplemented subspace of V. Let P be the orthogonal projection corresponding to W: $W = PV$ and Q the projection $1 - P$.

Consider the vector space V/W. (Its elements are W-cosets of the additive group of V). Notice that $g \in W$ is equivalent to $Qg = 0$, so that the vector Qf is independent of the choice of f in a W-coset.

If f and g are two elements of V/W, they will be called compatible (in V/W) iff Qf and Qg are compatible (in V). The scalar product of $\dot{f}$ and $\dot{g}$ is defined as $\langle Qf | Qg \rangle$.

In this way V/W becomes a partial inner product space which clearly may be identified with $W^\perp$). So, as for inner product spaces (in particular, Hilbert spaces),

the notions of orthocomplemented subspaces and quotient spaces coincide.

6.3 *Orthogonal sums*: Let V_1 and V_2 be nondegenerate partial inner product spaces, with compatibilities $\#_1$ and $\#_2$. Let $\#_{12}$ be a compatibility between V_1 and V_2, such that every vector in $V_1^{\#_1}$ is $\#_{12}$-compatible to all of V_2, and every vector in $V_2^{\#_2}$ is $\#_{12}$-compatible to all of V_1. Consider the direct sum $V = V_1 \oplus V_2$, and define in it a compatibility following the procedure of Sec. 2.

For all compatible pairs define the inner product

$$\langle\{v_1, v_2\}|\{v_1', v_2'\}\rangle = \langle v_1|v_1'\rangle + \langle v_2|v_2'\rangle.$$

It is nondegenerate.

In order to extend this definition to infinitely many summands $\{V_j\}_{j\in J}$, one has to further restrict the compatibility in $V = \prod_{j\in J} V_j$ by the requirement

$$\sum_{j\in J} |\langle v_j|w_j\rangle| < \infty.$$

An example of this is the "partial Fock space" $T(x, \mu)$ discussed in Ref. 2, Sec. 6, Ex. 6.

7. ORTHOCOMPLEMENTED SUBSPACES OF PRE-HILBERT SPACES

The aim of this section is to illustrate the above results by specializing them to the familiar case of a pre-Hilbert space E. At the end of this section we shall state a result of Piron, Amemiya, and Araki, which says that a pre-Hilbert space with "many" projections is necessarily a Hilbert space. This result will be generalized in the next section.

Let E be a pre-Hilbert space. Theorem 4.3 applies immediately: A subspace is orthocomplemented if and only if it is the range of an orthogonal projection, P, i.e., an idempotent ($P^2 = P$), symmetric ($P^* = P$), $\tau(E, E)$-continuous map. But now, E carries also the norm topology ($\|f\|^2 = \langle f|f\rangle$) which is in general finer than the Mackey topology $\tau(E, E)$.

If E is norm-complete (that is, a Hilbert space), both topologies coincide, and the orthocomplemented subspaces are exactly the norm-closed ones. However, if E is not norm-complete, this is no longer true. The range of an idempotent, symmetric, *norm*-continuous operator may fail to be orthocomplemented. Also a subspace may be orthocomplemented without being norm-complete (e.g., E itself!).

We have seen in Proposition 3.3 that every orthocomplemented subspace is $\tau(E, E)$-closed. But the converse is not true in general: Not every $\tau(E, E)$-closed subspace is orthocomplemented.

In order to give a counterexample, we remind the reader that a vector subspace $W \subseteq E$ is $\tau(E, E)$-closed [or equivalently $\sigma(E, E)$-closed] if and only if $W^{\perp\perp} = W$. Consider now $E = \mathcal{S}(\mathbb{R})$ (Schwartz testing functions). Let $W = \{\varphi \mid \varphi \in \mathcal{S},\ \varphi(x) = 0 \text{ for } x \leq 0\}$. Then $W^\perp = \{\psi \mid \psi \in \mathcal{S},\ \psi(x) = 0 \text{ for } x \geq 0\}$. So $W^{\perp\perp\perp} = W$. However, W is not orthocomplemented, since every $\chi \in W + W^\perp$ satisfies $\chi(0) = 0$; consequently, $W + W^\perp \neq \mathcal{S}$.

If we require that every $\tau(E, E)$-closed subspace of a pre-Hilbert space be orthocomplemented, we force E to be a Hilbert space. Namely,[5,6]

7.1 *Proposition*: Let E be a pre-Hilbert space. Assume that every $\sigma(E, E)$-closed vector subspace [or, equivalently, every $\tau(E, E)$-closed vector subspace] is orthocomplemented. Then E is a Hilbert space.

8. SPACES WITH MANY PROJECTIONS

In this section we shall extend Proposition 7.1 to nondegenerate partial inner product spaces, such that $V^\#$ is pre-Hilbert. If, in such a space, every $\tau(V, V)$-closed subspace is orthocomplemented, then V is a Hilbert space, i.e., all elements of V are compatible and V is complete in the norm topology.

We shall prove a stronger statement in which the assumption is only that certain $\tau(V, V)$-closed subspaces are orthocomplemented.

8.1 *Theorem*: Let V be a nondegenerate partial inner product space such that $\langle f|f\rangle > 0$ for every nonzero $f \in V^\#$. Assume that the following condition holds:

If a vector subspace $W \subseteq V$ is the $\tau(V, V^\#)$-closure of its "infinitely good core" $W \cap V^\#$, then W is orthocomplemented.

Then V is a Hilbert space.

Proof: Let S be any $\tau(V^\#, V^\#)$-closed subspace of $V^\#$. The $\tau(V, V^\#)$-closure of S is then, according to Proposition 3.3, $W = (S^\perp \cap V^\#)^\perp$ and therefore,

$$\begin{aligned} W \cap V^\# &= (S^\perp \cap V^\#)^\perp \cap V^\# \\ &= \tau(V^\#, V^\#)\text{-closure of } S \\ &= S. \end{aligned}$$

By assumption, W is orthocomplemented in V.

By Theorem 5.1, $W \cap V^\# = S$ is orthocomplemented in V. Since S was arbitrary, it follows from Theorem 7.1, that $V^\#$ is a Hilbert space.

Now the topology induced on $V^\#$ by $\tau(V, V)^\#$ is $\tau(V^\#, V^\#)$. But $\tau(V^\#, V^\#)$ is the norm topology on $V^\#$, since $V^\#$ is a Hilbert space. Thus $V^\#$ is complete in the topology induced by V; hence it is closed in V. Since it is also dense, we have necessarily $V^\# = V$.

Our theorem says, in particular, that, say $\mathcal{S}'(\mathbb{R})$, the space of tempered distributions (which is certainly not a Hilbert space), contains subspaces that are the closure of their infinitely good core, and yet are not orthocomplemented. An example is provided by the set of distributions with support in a closed interval; the nonexistence of a complementary (orthogonal) subspace is easily proved with the help of Theorem 5.1 and the example given in Sec. 7.

[1]J.-P. Antoine and A. Grossmann, "Partial inner product spaces-I. General properties," J. Funct. Anal. **23**, 369–78 (1976).
[2]J.-P. Antoine and A. Grossmann, "Partial inner product spaces-II. Operators," J. Funct. Anal. **23**, 379–91 (1976).
[3]J. Bognar, *Indefinite Inner Product Spaces* (Springer-Verlag, Berlin-Heidelberg-New York, 1974).
[4]A. Z. Jadczyk, "Geometry of indefinite-metric spaces," Rep. Math. Phys. **2**, 263–76 (1971).
[5]C. Piron, "Axiomatique Quantique," Helv. Phys. Acta **37**, 439–68 (1964).
[6]I. Amemiya and H. Araki, "A remark on Piron's paper," Publ. RIMS (Kyoto) A **2**, 423–27 (1967).
[7]This object was introduced in the following for the particular case of a nested Hilbert space: F. Debacker-Mathot, "Some operator algebras in nested Hilbert spaces," Commun. Math. Phys. **42**, 183–93 (1975).
[8]A. P. Robertson and W. Robertson, *Topological Vector Spaces* (Cambridge U. P., Cambridge, 1964).
[9]H. H. Schaefer, *Topological Vector Spaces* (Springer-Verlag, Berlin-Heidelberg, New York, 1970).

Application of a novel interpolative perturbation scheme to the determination of anharmonic oscillator wavefunctions

Charles A. Ginsburg

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

Elliott W. Montroll[a)]

Physical Dynamics, Inc., [b)] *La Jolla, California 92038*
(Received 27 July 1977)

This paper is concerned with a scheme for finding approximate wavefunctions and energy levels of anharmonic oscillators with even power anharmonicities. Generally one can easily find simple asymptotic forms for wavefunctions for small and large oscillator displacements y. In the case of a quartic oscillator the wavefunction with anharmonicity parameter λ,

$$\psi_0(y) = \exp\{-[a_{11}(\lambda)y^4 + (2\lambda/9)y^6]^{1/2}\} \tag{A}$$

with $a_{11}(\lambda) = E_0^2$ and

$$E_0^3 - (1/4)E_0 - (1/3)\lambda = 0, \tag{B}$$

while not exact for all y, does have the correct asymptotic properties. Furthermore, the appropriate root of the cubic equation for E_0 deviates by no more than 4% from the exact values of E_0 in the range $0 \le \lambda < \infty$. In our interpolative method of improving $\psi_0(y)$, the correct asymptotic behavior of $\psi_0(y)$ is preserved by introducing an extra function of λ into the exponential function (A) and changing the power of the polynomial in the exponent:

$$\psi_0(y) = \exp\{-[a_{21}(\lambda)y^8 + a_{22}(\lambda)y^{10} + (2\lambda/9)^2 y^{12}]^{1/4}\}. \tag{C}$$

Through the proper choice of $a_{21}(\lambda)$ and $a_{22}(\lambda)$ one obtains ground state energies which deviate by no more than 0.5% over the full range $0 \le \lambda < \infty$. Futher improvement is achieved by increasing the degree of the polynomial in the exponential. Excited state wavefunctions are obtained by multiplying exponentials such as (A) and (C) by polynomials in y.

I. INTRODUCTION

Many physical processes are characterized by differential equations whose independent variables are position, y, and time, t, and which involve a single dimensionless parameter. In the case of a Navier–Stokes fluid the parameter is the Reynolds number. The Schrödinger equation of an anharmonic oscillator contains a dimensionless anharmonic force constant λ. For certain eigenvalue problems which are associated with the subject models, it is easy to find the asymptotic form of the eigenfunctions for very small y as well as for very large y. The aim of this report is to introduce a scheme for combining the available information on two regimes to obtain formulas for wavefunctions and energy levels which are valid over the full range of y and of the basic parameter of the model. Our strategy here is to discuss our method in considerable detail in terms of the manner in which it is applied to the calculation of wavefunctions of an anharmonic oscillator. In future papers the methods will be applied to other systems.

Consider an anharmonic oscillator with Hamiltonian

$$H_\alpha(\omega, \lambda) = \tfrac{1}{2}(p^2 m^{-1} + m\omega^2 x^2) + \lambda' x^{2\alpha} \quad \text{with } \lambda' > 0. \tag{I. 1}$$

In the Schrödinger representation the energy levels E' are the eigenvalues of

$$[-(\hbar^2/2m)d^2/dx^2 + \tfrac{1}{2}m\omega^2 x^2 + \lambda' x^{2\alpha}]\psi(x) = E'\psi(x). \tag{I. 2}$$

We direct our attention to the presentation of certain approximations to $\psi_n(x)$, especially for the case $\alpha = 2$.

Equation (I. 2) has the simplified form

$$\left(-\frac{1}{2}\frac{d^2}{dy^2} + \tfrac{1}{2}y^2 + \lambda y^{2\alpha}\right)\psi(y) = E\psi(y), \tag{I. 3a}$$

in terms of the dimensionless quantities

$$E \equiv E'/\hbar\omega, \quad y \equiv x(m\omega/\hbar)^{1/2}, \quad \text{and} \quad \lambda \equiv \lambda'\hbar^{\alpha-1}/m^\alpha\omega^{\alpha+1}. \tag{I. 3b}$$

When y is small (in the limit $\lambda \to 0$) the ground state wavefunction has the form

$$\psi_0(y) \sim \exp(-\tfrac{1}{2}y^2). \tag{I. 4}$$

On the other hand, if when $|y|$ is very large, and $\lambda \neq 0$, one may assume that

$$\psi_0(y) \sim \exp(-\gamma|y|^\beta), \tag{I. 5}$$

then substitution of (I. 5) into (I. 3a) and retention of only the highest order terms in y implies that

$$\beta = 1 + \alpha \quad \text{and} \quad \gamma^2 = 2\lambda/(1+\alpha)^2. \tag{I. 6}$$

Now let us consider, in the case $\alpha = 2$, the interpolation function

a)This research was partially supported by the Fluid Dynamics Branch of the Office of Naval Research.
b)Permanent address, Institute for Fundamental Studies, University of Rochester, Rochester, New York 14627.

 0022-2488/78/1901-0336$1.00

$$\psi_0(y) = \exp\{-[a(\lambda)y^4 + (2\lambda/9)y^6]^{1/2}\}. \quad \text{(I. 7)}$$

The function $a(\lambda) \to \frac{1}{2}$ as $\lambda \to 0$. It will be chosen generally in such a manner that as $y \to 0$

$$\left(-\frac{1}{2}\frac{d^2}{dy^2} + \tfrac{1}{4}y^2 - E'\right)\psi_0(y) = o(y^2). \quad \text{(I. 8)}$$

The function $\psi_0(y)$ is the exact ground state wavefunction for a potential

$$V(y) = \left[\left(E_0 + \frac{1}{2}\frac{d^2}{dy^2}\right)\psi_0(y)\right]\Big/\psi_0(y). \quad \text{(I. 9)}$$

It is easily shown from (I. 7) that, with $a(\lambda) \equiv a$,

$$\frac{d^2\psi_0(y)}{\psi_0(y)\,dy^2} = \frac{4ay^2[1 + (\lambda y^2/3a)]^2}{[1 + (2\lambda y^2)/9a]} - \frac{2a^{1/2}[1 + (\lambda y^2/a) + (4\lambda^2 y^4/27a^2)]}{[1 + (2\lambda y^2)/9a]^{3/2}}. \quad \text{(I. 10)}$$

The necessary conditions for $\psi_0(y)$ to satisfy (I. 8) are that

$$(a - \tfrac{1}{4}) = \lambda/3a^{1/2} \quad \text{and} \quad E_0 = a^{1/2}. \quad \text{(I. 11)}$$

Then

$$V(y) = a^{1/2} + \frac{2ay^2[1 + (\lambda/3a)y^2]^2}{[1 + (2\lambda/9a)y^2]} - \frac{a^{1/2}[1 + (\lambda/a)y^2 + (4\lambda^2/27a^2)y^4]}{[1 + (2\lambda/9a)y^2]^{3/2}} \quad \text{(I. 12)}$$

$$\sim \begin{cases} \frac{1}{2}y^2 & \text{as } y \to 0, \\ \lambda y^4 & \text{as } y \to \infty. \end{cases} \quad \text{(I. 13)}$$

Actually $V(y)$ is an excellent approximation to

$$V_0(y) = \tfrac{1}{2}y^2 + \lambda y^4 \quad \text{(I. 14)}$$

over the entire range $-\infty < y < \infty$. For example, when $\lambda = \frac{9}{4}$, so that $a = 1$, the relative deviation of $V(y)$ from $V_0(y)$,

$$r(y) = [V(y) - V_0(y)]/V_0(y), \quad \text{(I. 15)}$$

never exceeds 4%. We have the following values of $r(y)$ for a number of selected values of y:

y	0	1	2	3	4	5	6	10
$r(y)$	0	0.026	0.038	0.024	0.016	0.012	0.0052	0.0035

The fact that $V(y)$ deviates so little from $V_0(y)$ suggests that the ground state energy levels associated with the two potential energy functions should also be very close to each other.

The required values of a and the ground state energy level E_0 follow from Eq. (I. 11), with E_0 being a solution of the cubic equation

$$E_0^3 - \tfrac{1}{4}E_0 - \lambda/3 = 0. \quad \text{(I. 16)}$$

When $\lambda^2 < \frac{1}{48}$ one has the trigonometric form of E_0,

$$E_0 = 3^{-1/2}\cos[\tfrac{1}{3}\cos^{-1}(4\lambda \cdot 3^{1/2})]. \quad \text{(I. 17)}$$

As $\lambda \to 0$, $E_0 \to 3^{-1/2}\cos(\pi/6) = \frac{1}{2}$ as it should. When $\lambda^2 > \frac{1}{48}$ one has the Cardan solution

$$E_0 = [\lambda/6 + \tfrac{1}{6}(\lambda^2 - \tfrac{1}{48})^{1/2}]^{1/3} + [\lambda/6 - \tfrac{1}{6}(\lambda^2 - \tfrac{1}{48})^{1/2}]^{1/3} = a^{1/2}. \quad \text{(I. 18)}$$

As $\lambda \to \infty$ this yields

$$E_0 \sim (\lambda/3)^{1/3} + (1/576\lambda)^{1/3} + \cdots \sim 0.693\lambda^{1/3} + 0.120\lambda^{-1/3} + \cdots \quad \text{(I. 19)}$$

which, when compared to the exact result of Ref. 1,

$$E_0 \sim 0.668\lambda^{1/3} + 0.144\lambda^{-1/3} + \cdots \quad \text{(I. 20)}$$

is shown to be in error by less than 4% as $\lambda \to \infty$. This is the maximum error over the full range $0 \leq \lambda < \infty$. At the intermediate points $\lambda = 1$, 2, and 50 one obtains from (I. 18),

λ	E approx.	E (exact)[1,2]	% error
1	0.813	0.804	1.1%
2	0.969	0.952	1.8%
50	2.568	2.500	2.7%

A cubic equation analogous to (I. 16) has been derived by a completely different method in Ref. 3. Equation (10) of that reference with $N = 2$ corresponds to (I. 16). The mean field theory approach of Ref. 3 leads, in the first approximation, to about the same uniform accuracy as that given above.

With these good results for a first approximation we now develop a systematic procedure for improving them and we show how the scheme can be extended for the calculation of wavefunctions and energy levels associated with excited states.

II. SUCCESSIVE APPROXIMATIONS TO GROUND STATE WAVEFUNCTION

In this section we introduce a set of successive improvements to the interpolation formula (I. 7) for the ground state wavefunction $\psi_0(y)$. In their increasing order of complexity, each will be associated with a potential which agrees with $V_0(y)$ to one higher power in y^2 in the small y regime. All will approach λy^4 as $y \to \infty$. We will find that to each new order there will be an improvement in the precision of the ground state energy level over the full range of $0 \leq \lambda < \infty$. The set of functions to be considered is

$$\psi_0^{(1)}(y) = \exp[-(a_{11}y^4 + a_{12}y^6)^{1/2}], \quad \text{(II. 1)}$$

$$\psi_0^{(2)}(y) = \exp[-(a_{21}y^8 + a_{22}y^{10} + a_{23}y^{12})^{1/4}], \quad \text{(II. 2)}$$

$$\psi_0^{(3)}(y) = \exp[-(a_{31}y^{12} + a_{32}y^{14} + a_{33}y^{16} + a_{34}y^{18})^{1/6}], \quad \text{(II. 3)}$$

etc.

We will outline the procedure for the determination of the $\{a_{jK}\}$ and the ground state energy by using $\psi_0^{(2)}(y)$ as an example. The calculations for higher order $\psi_0^{(j)}(y)$ are longer but proceed in the same manner. As $|y| \to \infty$,

$$\psi_0^{(2)}(y) \sim \exp(-a_{23}^{1/4}|y|^3). \quad \text{(II. 4)}$$

Hence, from (I. 6), $a_{23}^{1/4} = (2\lambda/9)^{1/2}$ so that

$$a_{23} = (2\lambda/9)^2. \quad \text{(II. 5)}$$

As $y \to 0$

$$\psi_0^{(2)}(y) \sim \exp[-a_{21}^{1/4}y^2 - \tfrac{1}{4}a_{22}a_{21}^{-3/4}y^4 - \tfrac{1}{4}a_{21}^{-3/4}(a_{23} - \tfrac{3}{8}a_{22}^2a_{21}^{-1})y^6 - \cdots] \quad \text{(II. 6)}$$

so that

$$\frac{d^2\psi_0^{(2)}}{dy^2} = \{-2a_{21}^{1/4} + (4a_{21}^{1/2} - 3a_{22}a_{21}^{-3/4})y^2 + [4a_{22}a_{21}^{-1/2} - \tfrac{15}{2}a_{23}a_{21}^{-3/4} + \tfrac{45}{16}a_{22}^2 a_{21}^{-7/4}]y^4 + \cdots\} \times \psi_0^{(2)}. \quad \text{(II. 7)}$$

From the Schrödinger equation

$$\frac{1}{2}\frac{d^2\psi_0}{dy^2} = (-E + \tfrac{1}{2}y^2 + \lambda y^4)\psi_0, \quad \text{(II. 8)}$$

we find by equating the constant term on each side of the equation as well as the coefficients of y^2 and y^4,

$$E = a_{21}^{1/4}, \quad 4a_{21}^{1/2} - 3a_{22}a_{21}^{-3/4} = 1, \quad \text{(II. 9a)}$$

$$4a_{22}a_{21}^{-1/2} - \tfrac{15}{2}a_{21}^{-3/2}a_{23} + \tfrac{45}{16}a_{22}^2/a_{21}^{7/4} = 2\lambda. \quad \text{(II. 9b)}$$

Notice that a_{22} can be expressed in terms of a_{21} (and hence in terms of E),

$$a_{22} = \tfrac{1}{3}(4a_{21}^{1/2} - 1)a_{21}^{3/4}. \quad \text{(II. 10)}$$

When one sets $a_{21}^{1/4} = E$ and $a_{23} = (2\lambda/9)^2$ [as given by Eq. (II. 5)] in (II. 9b) he obtains the following characteristic equation for E,

$$E^6 - \tfrac{23}{62}E^4 - \tfrac{6}{31}\lambda E^3 + \tfrac{15}{496}E^2 - \tfrac{10}{279}\lambda^2 = 0, \quad \text{(II. 11)}$$

which has an alternative form

$$E = \frac{1}{2} + \frac{16\lambda(27E^3 + 5\lambda)}{9E^2(2E+1)(124E^2 - 15)}. \quad \text{(II. 12)}$$

When λ is small, Eq. (II. 12) can be solved by iteration. The first iteration yields, if one neglects all resulting terms of $O(\lambda^2)$,

$$E = \tfrac{1}{2} + \tfrac{3}{4}\lambda, \quad \text{(II. 13)}$$

which is the same result that one obtains from first-order perturbation theory. Upon a second iteration, now retaining terms linear in λ in the numerator and denominator of (II. 12) after the λ coefficient is extracted, one finds

$$E = \tfrac{1}{2} + \tfrac{3}{4}\lambda\left(\frac{1 + \frac{323}{54}\lambda}{1 + \frac{153}{16}\lambda}\right), \quad \text{(II. 14)}$$

which is reminiscent of Padé approximant[4,5] forms for E. As will be noted, when we use $\psi_0^{(3)}$ the resulting E agrees with second order perturbation theory to within terms of $O(\lambda^2)$. However, it is known from the work of Bender and Wu[6] that there is no point in developing E as a series expansion in powers of λ since such a series diverges for all values of λ.

When λ is large, it has been shown[1] that

$$E \sim b_1\lambda^{1/3} + b_2\lambda^{-1/3} + b_3\lambda^{-1} + b_4\lambda^{-5/3} + \cdots. \quad \text{(II. 15)}$$

The coefficients b_1, b_2, b_3 can be obtained by substituting (II. 15) into (II. 11) and setting coefficients of $\lambda^{6/3}$, $\lambda^{4/3}$, $\lambda^{2/3}, \cdots$, etc. equal to zero. The first two resulting equations are

$$b_1^6 - \tfrac{6}{31}b_1^3 - \tfrac{10}{279} = 0, \quad \text{(II. 16a)}$$

$$6b_1^5 b_2 - \tfrac{23}{62}b_1^4 - \tfrac{18}{31}b_1^2 b_2 = 0. \quad \text{(II. 16b)}$$

The quadratic equation (II. 16a) in b_1^3 can be solved immediately to yield the positive root

$$b_1^3 = 0.309395 \quad \text{with} \quad b_1 = 0.6763. \quad \text{(II. 17a)}$$

When this is substituted in (II. 16b) it gives the b_2 value

$$b_2 = 0.13302 \quad \text{(II. 17b)}$$

so that for large λ

$$E \sim 0.676\lambda^{1/3} + 0.133\lambda^{-1/3} + \cdots, \quad \text{(II. 18)}$$

which is to be compared with the exact[1] expression (I. 20). Note that it is a considerable improvement over the first order result (I. 19). As $\lambda \to \infty$ one obtains the greatest percentage deviation from comparing the results of (II. 11) with exact ones. That maximum error of 1.2% is thus an improvement over the 4% error cited at the end of Sec. I. We compare the appropriate root of (II. 11) for a wide range of λ to exact ground state energy levels[1,2] in Table I.

Given the ground state energy, the three constants required in the ground state wavefunction (II. 2) are, using (II. 9a) and (II. 5)

$$a_{21} = E^4, \quad a_{23} = (2\lambda/9)^2, \quad \text{(II. 19a)}$$

$$a_{22} = (4E^2 - 1)E^3/3. \quad \text{(II. 19b)}$$

It is easy to proceed to third order using (II. 3). We merely collect a few of the results here. First

$$\frac{d^2\psi_0^{(3)}}{dx^2} = [-2a_{31}^{1/6} + y^2(4a_{31}^{1/3} - 12c_1) + y^4(16c_1a_{31}^{1/6} - 30c_2) + (16c_1^2 + 24c_2a_{31}^{1/6} - 56c_3)y^6 + \cdots]\psi_0^{(3)}, \quad \text{(II. 20)}$$

where

$$c_1 = a_{32}/6a_{31}^{5/6}, \quad \text{(II. 21a)}$$

$$c_2 = \left(a_{33} - \frac{5}{12}\frac{a_{32}^2}{a_{31}}\right)\Big/6a_{31}^{5/6}, \quad \text{(II. 21b)}$$

$$c_3 = \left(a_{34} - \frac{5}{6}\frac{a_{32}a_{33}}{a_{31}} + \frac{55}{216}\frac{a_{32}^3}{a_{31}^2}\right)\Big/6a_{31}^{5/6}. \quad \text{(II. 21c)}$$

By substituting (II. 20) into the Schrödinger equation (II. 8) we find by equating the constant term on each side of the equation as well as the coefficients of y^2, y^4, and y^6,

$$E = a_{31}^{1/6}, \quad 4a_{31}^{1/3} - 12c_1 = 1, \quad \text{(II. 22)}$$

$$8c_1a_{31}^{1/6} - 15c_2 = \lambda, \quad \text{(II. 23)}$$

$$2c_1^2 + 3c_2a_{31}^{1/6} - 7c_3 = 0, \quad \text{(II. 24)}$$

$$a_{34} = (2\lambda/9)^3. \quad \text{(II. 25)}$$

TABLE I. Comparison of approximate ground state energy levels with exact ones.

λ	E (2nd approx.)	E (3rd approx.)	E(exact)[1,2]
0.05	0.53257	0.53264	0.53264
0.10	0.55900	0.55913	0.55915
0.30	0.63808	0.63803	0.63799
0.50	0.69683	0.69637	0.69618
0.70	0.74517	0.74428	0.74390
1.00	0.80593	0.80444	0.80377
2.00	0.95617	0.95312	0.95157
50	2.5278	2.5107	2.49971
200	3.9781	3.9496	3.9209
10^3	6.7768	6.7272	6.6942
8×10^3	13.5336	13.4336	13.3669
20×10^3	18.3638	18.2280	18.1372

In a manner analogous to that in which (II. 11) was derived we find

$$187E^9 - \tfrac{1563}{16}E^7 - \tfrac{99}{2}\lambda E^6 + \tfrac{495}{32}E^5 + \tfrac{315}{32}\lambda E^4 - \tfrac{175}{256}E^3 - \tfrac{35}{54}\lambda^3 = 0. \quad \text{(II. 26)}$$

Once E is known, a_{32} and a_{33} are found from (II. 23) and (II. 24) to be

$$a_{32} = \tfrac{1}{2}E^5(4E^2 - 1), \quad \text{(II. 27a)}$$

$$a_{33} = \tfrac{5}{48}E^4(4E^2 - 1)^2 + \tfrac{4}{15}E^6(4E^2 - 1) - \tfrac{2}{5}E^5. \quad \text{(II. 27b)}$$

Equation (II. 26) has the alternative form

$$E = \frac{1}{2} + \frac{4\lambda}{27}\left(\frac{42768E^6 - 8505E^4 + 560\lambda^2}{E^3(2E+1)(176E^2 - 35)(68E^2 - 5)}\right). \quad \text{(II. 28)}$$

If one substitutes the result that corresponds to first-order perturbation theory (II. 13) into this equation and expands in powers of λ, he retains the traditional second-order perturbation theory result

$$E = \tfrac{1}{2} + \tfrac{3}{4}\lambda - \tfrac{21}{8}\lambda^2. \quad \text{(II. 29)}$$

When this is substituted into (II. 28) and terms to order λ^2 are retained in numerator and denominator, it yields a Padé type formula

$$E = \tfrac{1}{2} + \tfrac{3}{4}\lambda\left(\frac{1 + \frac{62}{3}\lambda + \frac{193609}{4374}\lambda^2}{1 + \frac{145}{6}\lambda + \frac{887}{8}\lambda^2}\right) \quad \text{(II. 30)}$$

which is fairly accurate for $\lambda < \frac{1}{4}$. As has already been noted, while the Padé method converges, it converges only slowly. Even for λ as small as 1, the large λ asymptotic formulas derived below are better. Roots calculated directly from (II. 26) however, over the full range of positive λ, give excellent agreement with the exact results of Refs. 1 and 2.

When λ is large we postulate the form (II. 15) to be introduced into (II. 26). By setting the coefficients of λ^3 and $\lambda^{7/3}$ equal to zero, as they must be, we find the conditions

$$b_1^9 - \tfrac{9}{34}b_1^6 - \tfrac{35}{10098} = 0, \quad \text{(II. 31a)}$$

$$b_2 = \frac{(1042b_1^3 - 105)}{1056b_1(17b_1^2 - 3)}, \quad \text{(II. 31b)}$$

which yield the asymptotic expression

$$E \sim 0.6713\lambda^{1/3} + 0.1384\lambda^{-1/3} + \cdots \quad \text{(II. 32)}$$

which is to be compared with the exact expression[1]

$$E \sim 0.6680\lambda^{1/3} + 0.1437\lambda^{-1/3} + \cdots.$$

As $\lambda \to \infty$, the relative error of the estimate (II. 32) is 0.49%, about half that given by the second-order estimate (II. 18). We compare the appropriate root of (II. 26) for a wide range of λ with the second-order roots and with the exact ground state energy levels in Table I. With knowledge of E the a_{3j} are given by (II. 22), (II. 25), and (II. 27).

We have used the following scheme to calculate the appropriate roots of (II. 11) and (II. 26), which we call respectively second and third order ground state energies, while we refer to the E_0 of Eqs. (I. 17) and (I. 18) as first order ground state energies. The first order energies are used as the first approximation in a standard iterative process to find the roots of (II. 11). The final roots of (II. 11) are the numbers in the second column of Table I. These second order energies are used as the first approximation for an iterative root search of Eq. (II. 26). The results of that search are the third order energies which are tabulated in the third column of Table I.

Were we to proceed to higher order expressions for $\psi_0^{(n)}(y)$ we would obtain recurrence formulas analogous to (II. 22)—(II. 25). Then we would use the exact value of E_0 calculation in Ref. 1 (or apply the computer program developed to produce the tables in Ref. 1) to calculate a_{j1}. From a_{j1}, the analog of the recurrence formulas in (II. 22) would be applied to calculate a_{j2}, etc. This would save us the trouble of calculating the new estimate to the energy levels in each order of approximation. Actually we have developed a computer program for calculating the a_{nj}'s of the generalization of Eq. (II. 3) to higher order. This will be discussed elsewhere.

We now proceed to extend the ideas used in this section to find wavefunctions of excited states of our anharmonic oscillator.

III. WAVEFUNCTION FOR FIRST FEW EXCITED STATES

We follow in the spirit of the last section and construct first order approximations to excited state wavefunctions by multiplying functions of the form chosen for the ground state by polynomials. Thus we postulate that

$$\psi_1^{(1)}(y) = y\exp[-(a_{11}y^4 + a_{12}y^6)^{1/2}], \quad \text{(III. 1a)}$$

$$\psi_2^{(1)}(y) = (1 + b_{12}y^2)\exp[-(a_{11}y^4 + a_{12}y^6)^{1/2}], \quad \text{(III. 1b)}$$

$$\psi_3^{(1)}(y) = y(1 + c_{12}y^2)\exp[-(a_{11}y^4 + a_{12}y^6)^{1/2}], \quad \text{(III. 1c)}$$

etc.

The second order approximations will be written as

$$\psi_1^{(2)}(y) = y\exp[-(a_{21}y^8 + a_{22}y^{10} + a_{23}y^{12})^{1/4}], \text{ etc.} \quad \text{(III. 2)}$$

Higher order approximations follow in a similar pattern.

When $|y|$ is very large

$$\psi_n^{(1)}(y) \sim P_n(y)\exp(-a_{12}^{1/2}|y|^3), \quad \text{(III. 3)}$$

$P_n(y)$ being a polynomial of degree n so that

$$\frac{d^2\psi_n^{(1)}(y)}{dy^2} = \{[P_n''(y) - 3a_{12}^{1/2}y^2P_n'(y) - 6a_{12}^{1/2}yP_n(y)] - 3[P_n'(y) - 3a_{12}^{1/2}y^2P_n(y)]y^2a_{12}^{1/2}\}\exp(-a_{12}^{1/2}y^3) \quad \text{(III. 4a)}$$

$$\sim 9a_{12}y^4P_n(y)\exp(-a_{12}^{1/2}y^3) \text{ as } y \to \infty. \quad \text{(III. 4b)}$$

As in the last section we note that in the limit $|y| \to \infty$ the Schrödinger equation becomes

$$-\frac{1}{2}\frac{d^2\psi_n^{(1)}(y)}{dy^2} \sim \lambda y^4P_n(y)\exp(-a_{12}^{1/2}|y|^3). \quad \text{(III. 5)}$$

When (III. 4b) is substituted into (III. 5), and appropriate coefficients are equated, we again find, as in the ground state

$$a_{12} = 2\lambda/9. \quad \text{(III. 6)}$$

In the regime of small y we approximate the first excited state wavefunction by

$$\psi_1^{(1)}(y) \sim y \exp(-a_{11}^{1/2}y^2 - \tfrac{1}{2}a_{12}a_{11}^{-1}y^4) \tag{III. 7a}$$

so that

$$\frac{1}{2}\frac{d^2\psi_1^{(1)}(y)}{dy^2} = [-3a_{11}^{1/2} + (2a_{11} - 5a_{12}a_{11}^{-1/2})y^2 + \cdots]\psi_1^{(1)}(y). \tag{III. 7b}$$

When one substitutes this expression into the Schrödinger equation

$$\frac{1}{2}\frac{d^2\psi_1^{(1)}(y)}{dy^2} = (\tfrac{1}{2}y^2 + \lambda y^4 - E_1)\psi_1^{(1)}(y)$$

and he equates the constant term on each side of the resulting equation as well as the coefficients of y^2, he finds

$$E_1 = 3a_{11}^{1/2}, \tag{III. 8a}$$

$$4a_{11}^{3/2} - a_{11}^{1/2} - 10a_{12} = 0. \tag{III. 8b}$$

This pair of equations yields the characteristic equation for E_1 [remembering (III. 6)]

$$E_1^3 - \tfrac{9}{4}E_1 - 15\lambda = 0. \tag{III. 9}$$

When $\lambda^2 < \frac{3}{400}$, the trigonometric form of the solution of this cubic is appropriate,

$$E_1 = 3^{1/2}\cos\tfrac{1}{3}\cos^{-1}(20\lambda \cdot 3^{1/2}). \tag{III. 10}$$

When $\lambda^2 > \frac{3}{400}$ the Cardan solution is appropriate,

$$E_1 = (\tfrac{15}{2}\lambda)^{1/3}\left\{\left[1 + \left(1 - \frac{3}{400\lambda^3}\right)^{1/2}\right]^{1/3} + \left[1 - \left(1 - \frac{3}{400\lambda^2}\right)^{1/2}\right]^{1/3}\right\}. \tag{III. 11}$$

As $\lambda \to 0$, (III. 10) reduces to $E_1 \to \frac{3}{2}$ as it should; and as $\lambda \to \infty$, (III. 11) becomes

$$E_1 \sim (15)^{1/3}\lambda^{1/3} + \tfrac{3}{4}(15)^{-1/3}\lambda^{-1/3} + \cdots \tag{III. 12a}$$

$$= 2.466\lambda^{1/3} + 0.304\lambda^{-1/3} + \cdots \tag{III. 12b}$$

which when compared to the exact[1] asymptotic expansion

$$E_1 \sim 2.394\lambda^{1/3} + 0.358\lambda^{-1/3} + \cdots \tag{III. 13}$$

indicates that as $\lambda \to \infty$ the relative error of the estimate is only 3%. The percentage error of (III. 12b) for several intermediate values of λ are tabulated below.

λ	E_1 (approx.)	E_1 (exact)[1,2]	% error
1	2.769	2.738	1.1%
2	3.348	3.293	1.7%
50	9.168	8.915	2.8%

By comparing this with the table below Eq. (I. 20) we observe that the relative errors for the first order approximation of the first excited state are about the same as for the ground state.

We now proceed to develop a scheme for producing the wavefunctions for all excited states. We postulate the small y form of the excited state wavefunctions to be [generalizing (III. 7a)]

$$\psi(y) = P(y)\exp(-a_{11}^{1/2}y^2 - \tfrac{1}{2}a_{12}a_{11}^{-1}y^4) \equiv P(y)\psi_0(y). \tag{III. 14}$$

Then

$$\frac{1}{2}\frac{d^2\psi}{dy^2} = \frac{1}{2}\left[\frac{d^2P}{dy^2} - 4y(a_{11}^{1/2} + a_{12}a_{11}^{-1/2}y^2)\frac{dP}{dy} + 2(2a_{11}y^2 - a_{11}^{1/2} - 3a_{12}a_{11}^{-1/2}y^2)P\right]\psi_0. \tag{III. 15}$$

Since

$$\left[\frac{1}{2}\frac{d^2}{dy^2} - (\tfrac{1}{2}y^2 - E + \lambda y^4)\right]\psi = 0 \tag{III. 16}$$

we find that $P(y)$ must satisfy (neglecting the term λy^4 which was also neglected in the calculation of the coefficient a_{11} in ψ_0 and ψ_1; in second order this term is always included)

$$\frac{d^2P}{dy^2} - 4y(a_{11}^{1/2} + a_{12}a_{11}^{-1/2}y^2)\frac{dP}{dy} + [(4a_{11} - 6a_{12}a_{11}^{-1/2} - 1)y^2 + 2(E - a_{11}^{1/2})]P = 0. \tag{III. 17}$$

The series solution for $P(y)$ now to be constructed is

$$P(y) = \sum_{j=0}^{\infty} a_j y^{j+\rho}, \tag{III. 18}$$

where ρ is to be set equal to zero for the even order wavefunctions ψ_0, ψ_2, $\psi_4, \cdots$ and to one for odd order ψ_1, ψ_3, $\psi_5, \cdots$. When (III. 18) is substituted into (III. 17) the coefficient of each power $y^{j+\rho}$ for $j = 0, 1, 2, \cdots$ vanishes. This yields a two term recurrence formula for the a_j,

$$(j+\rho+1)(j+\rho+2)a_{j+2} + 2a_j(E - a_{11}^{1/2}[1 + 2j + 2\rho]) + a_{j-2}(4a_{11} - 1 - 2a_{12}a_{11}^{1/2}[2j - 1 + 2\rho]) = 0. \tag{III. 19}$$

Since the series (III. 18) starts with a_0, we set $a_{-2} = a_{-4} \equiv 0$.

We first consider in detail this set of equations in the even order wavefunction case where $P = a_0 + a_2y^2 + a_4y^4 + \cdots$ so that $j = 2m$ and $\rho = 0$. Then

$$(2m+2)(2m+1)a_{2m+2} + 2a_{2m}(E - a_{11}^{1/2}[1 + 4m]) + a_{2m-2}(4a_{11} - 1 - 2a_{12}a_{11}^{-1/2}(4m - 1)) = 0. \tag{III. 20}$$

The first few members of this set are, for $m = 0, 1, 2, \cdots$,

$$\begin{aligned} &2(E - a_{11}^{1/2})a_0 + 2 \cdot 1a_2 = 0, \\ &(4a_{11} - 6a_{12}a_{11}^{1/2} - 1)a_0 + 2(E - 5a_{11}^{1/2})a_2 + 4 \cdot 3a_4 = 0, \\ &(4a_{11} - 14a_{12}a_{11}^{-1/2} - 1)a_2 + 2(E - 9a_{11}^{1/2})a_4 + 6 \cdot 5a_6 = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &[4a_{11} - 2(4n-1)a_{12}a_{11}^{-1/2} - 1]a_{2n-2} + 2[E - (4n+1)a_{11}^{1/2}]a_{2n} + 2(n+1)(2n+1)a_{2n+2} = 0. \end{aligned} \tag{III. 21}$$

Now suppose we wish to terminate the series for $P(y)$ so that it will be a polynomial of order $(2n-2)$. When $n = 1$, $P(y)\psi_0(y)$ corresponds to the ground state wavefunction, $n = 2$ the second excited state, $n = 3$ the fourth excited state, $n = 4$ the sixth. This termination would result from setting

$$a_{2n} = a_{2n+2} = 0. \tag{III. 22}$$

Since we wish a_{2n-2} to be nonvanishing, we then require that

$$4a_{11} - 1 = 2(4n-1)a_{12}a_{11}^{-1/2} = 4(4n-1)\lambda/9a_{11}^{1/2}. \tag{III. 23}$$

On this basis the nonvanishing members of the set of equations (III. 21) are

$$(E - a_{11}^{1/2})a_0 + 1\cdot 1a_2 = 0, \tag{III. 24a}$$

$$4(n-1)a_{12}a_{11}^{-1/2}a_0 + (E - 5a_{11}^{1/2})a_2 + 2\cdot 3a_4 = 0, \tag{III. 24b}$$

$$4(n-2)a_{12}a_{11}^{-1/2}a_2 + (E - 9a_{11}^{1/2})a_4 + 3\cdot 5a_6 = 0, \tag{III. 24c}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$4(1)a_{12}a_{11}^{-1/2}a_{2n-4} + (E - [4n-3]a_{11}^{1/2})a_{2n-2} = 0. \tag{III. 24d}$$

The condition that this set of equations has a solution is that the determinant of the coefficients vanish: i. e., if we define

$$w = 4a_{12}a_{11}^{-1/2}, \tag{III. 25}$$

$$\begin{vmatrix} (E-a_{11}^{1/2}) & 1\cdot 1 & 0 & \cdots & 0 & 0 \\ (n-1)w & (E-5a_{11}^{1/2}) & 2\cdot 3 & \cdots & 0 & 0 \\ 0 & (n-2)w & (E-9a_{11}^{1/2}) & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & 0 & \cdots & 1\cdot w & [E-(4n-3)a_{11}^{1/2}] \end{vmatrix} = 0. \tag{III. 26}$$

Since a_{12} is given by (III. 6) and $a_{11}^{1/2}$ is a solution of the cubic equation (III. 23), which we write in the form

$$a_{11}^{3/2} - \tfrac{1}{4}a_{11}^{1/2} - (4n-1)\lambda/9 = 0, \tag{III. 27}$$

the only unknown in the characteristic equation (III. 26) is the energy E. To make contact with the results of Sec. I, let $n=1$. Then (III. 26) and (III. 27) become

$$E = a_{11}^{1/2} \quad \text{and} \quad a_{11}^{3/2} - \tfrac{1}{4}a_{11}^{1/2} - \tfrac{1}{3}\lambda = 0, \tag{III. 28}$$

which is equivalent to (I. 11), and yields the ground state characteristic equation (I. 16).

When $n=2$, which corresponds to the second excited state, Eqs. (III. 26) and (III. 27) yield [with a_{12} given by (III. 6)]

$$E^2 - 6Ea_{11}^{1/2} + 5a_{11} - 4a_{12}a_{11}^{-1/2} = 0, \tag{III. 29a}$$

$$a_{11}^{3/2} - \tfrac{1}{4}a_{11}^{1/2} - \tfrac{7}{9}\lambda = 0. \tag{III. 29b}$$

The form of the solution of the cubic equation for $a_{11}^{1/2}$ depends on the value of λ. If $\lambda < 3^{1/2}/28$, then

$$a_{11}^{1/2} = 3^{-1/2}\cos[\tfrac{1}{3}\cos^{-1}(28.\,3^{-1/2}\lambda)]. \tag{III. 30a}$$

When $\lambda > 3^{1/2}/28$ one uses the Cardan solution

$$a_{11}^{1/2} = (\tfrac{7}{18}\lambda)^{1/3}\left\{\left[1+\left(1-\frac{3}{784\lambda^2}\right)^{1/2}\right]^{1/3} + \left[1-\left(1-\frac{3}{784\lambda^2}\right)^{1/2}\right]^{1/3}\right\}. \tag{III. 30b}$$

With a_{11} known,

$$E_2 = a_{11}^{1/2}[3 + 2(1 + a_{12}a_{11}^{-3/2})^{1/2}]. \tag{III. 31}$$

As $\lambda \to 0$, $a_{11}^{1/2} \to \frac{1}{2}$ and $E \to \frac{5}{2}$ as it should. As $\lambda \to \infty$,

$$a_{11}^{1/2} \sim \left(\frac{7\lambda}{9}\right)^{1/3} + \left(\frac{1}{1344\lambda}\right)^{1/3} + \cdots \tag{III. 32}$$

so that

$$1 + a_{12}a_{11}^{-3/2} \sim \tfrac{7}{9} - [1 - (\tfrac{3}{3136})^{1/3}\lambda^{-2/3} + \cdots] \tag{III. 33}$$

and

$$E_2 \sim 4.\,844\lambda^{1/3} + 0.\,409\lambda^{-1/3} + \cdots \tag{III. 34a}$$

which is to be compared with the exact asymptotic form

$$E_2 \sim 4.\,697\lambda^{1/3} + 0.\,494\lambda^{-1/3} + \cdots. \tag{III. 34b}$$

The relative error in E_2 from (III. 34a), in the limit as $\lambda \to \infty$, is 3. 1%. As in the case of lower energy levels, this is the largest relative error in the positive λ range. At the intermediate points $\lambda = 1$, 2, and 50, one obtains from (III. 34a) and (III. 34b):

λ	E (approx.)	E (exact)[1,2]	% error
1	5. 257	5. 179	1. 5%
2	6. 430	6. 304	2. 0%
50	17. 96	17. 44	3. 0%

Notice that when $n=2$ in the characteristic equation (III. 26), it has the form

$$\begin{vmatrix} (E-a_{11}^{1/2}) & 7 \\ 3w & (E-5a_{11}^{1/2}) \end{vmatrix} = 0. \tag{III. 35}$$

As $\lambda \to 0$, from (III. 25), $w \to 0$ so that one of the resulting energy levels approaches $\frac{5}{2}$ and the other $\frac{1}{2}$. The $\frac{5}{2}$ corresponds to E_2 as given by (III. 31) while that which corresponds to $\frac{1}{2}$ is the ground state. It is the other root of (III. 29a), i. e.,

$$E_0 = a_{11}^{1/2}[3 - 2(1 + a_{12}a_{11}^{-3/2})^{1/2}]. \tag{III. 36}$$

When λ is large this equation combined with (III. 32) yields the following asymptotic expansion for E_0,

$$E_0 \sim 0.\,6734\lambda^{1/3} + 0.\,1348\lambda^{-1/3} + \cdots. \tag{III. 37}$$

which is to be compared with the exact result (I. 20). Notice that this is a slightly better approximation to the exact asymptotic form than was the second order approximation (II. 18).

It is now apparent that the wavefunction (III. 14) with $P(y)$ defined by (III. 18) with $\rho = 0$ can be interpreted as the ground state wavefunction. When the series is terminated at y^{2n-2}, E_0 is the smallest root of the characteristic equation (III. 26). The eigenfunction is an alternative to the generalization of (II. 3). To a given order a terminated form of (III. 18) might give energy levels which are as good or better than a generalization of (II. 3). However, when the series is terminated, the resulting ground state wavefunction does not vanish in the proper manner for large y since it becomes

$$\psi_0(y) \sim cy^{2n-2}\exp[-(2\lambda/9)^{1/2}y^3],$$

rather than $\exp[-(2\lambda/9)^{1/2}y^3]$. When one is not concerned with the precision of $\psi_0(y)$ for large y this is not a serious difficulty.

The higher even order energy levels $E_4, E_6, \cdots$ are easily obtainable from (III. 24), (III. 25), and (III. 27) if n is not too large. The large n energy levels will be discussed in the next section.

We now extend the above discussion to the case of odd numbered energy levels and wavefunctions by letting $\rho = 1$ in Eqs. (III. 18) and (III. 19). Then the analog of (III. 21) is

$$2(E - 3a_{11}^{1/2})a_0 + 2(3\cdot 1)a_2 = 0, \tag{III. 38a}$$

$$(4a_{11} - 10a_{12}a_{11}^{-1/2} - 1)a_0 + 2(E - 7a_{11}^{1/2})a_2 + 2(5\cdot 2)a_4 = 0, \tag{III. 38b}$$

$$(4a_{11} - 18a_{12}a_{11}^{-1/2} - 1)a_2 + 2(E - 11a_{11}^{1/2})a_4 + 2(7\cdot 3)a_6 = 0, \tag{III. 38c}$$

.....

$$(4a_{11} - 2[4n+1]a_{12}a_{11}^{-1/2} - 1)a_{2n-2} + 2(E - [4n+3]a_{11}^{-1/2})a_{2n} + 2(2n+3)(n+1)a_{2n+2} = 0, \tag{III. 38d}$$

etc.

In the case of the odd quantum numbers we wish to terminate the polynomial $P(y)$ to be of order $(2n-1)$. When $n=1$, $P(y)\psi_0(y)$ corresponds to the wavefunction of the first excited state; $n=2$ to that of the third, $n=3$ the fifth, etc. The termination with

$$a_{2n} = a_{2n+2} = 0 \tag{III. 39}$$

is accomplished by setting

$$(4a_{11} - 1) = 2(4n+1)a_{12}a_{11}^{-1/2} = 4\lambda(4n+1)/9a_{11}^{1/2}. \tag{III. 40}$$

Then the set (III. 38) becomes

$$(E - 3a_{11}^{1/2})a_0 + 3\cdot 1 a_2 = 0, \tag{III. 41a}$$

$$(n-1)wa_6 + (E - 7a_{11}^{1/2})a_2 + 5\cdot 2a_4 = 0, \tag{III. 41b}$$

$$(n-2)wa_2 + (E - 11a_{11}^{1/2})a_4 + 7\cdot 3a_6 = 0,$$

.....

$$(1)wa_{2n-4} + [E - (4n-1)a_{11}^{1/2}]a_{2n-2} = 0, \tag{III. 41c}$$

where w is defined by (III. 25).

The condition that this set of equations has a solution is that the determinant of the coefficients of the a_{2j} vanish, i.e.,

$$\begin{vmatrix} (E - 3a_{11}^{1/2}) & 3\cdot 1 & 0 & 0 & 0 \\ (n-1)w & (E - 7a_{11}^{1/2}) & 5\cdot 2 & 0 & 0 \\ 0 & (n-2)w & (E - 11a^{1/3}) & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1\cdot 2 & (E - [4n-1]a_{11}^{1/2}) \end{vmatrix} = 0. \tag{III. 42}$$

The equation for the determination of $a_{11}^{1/2}$ is

$$a_{11}^{3/2} - \tfrac{1}{4}a_{11}^{1/2} - (4n+1)\lambda/9 = 0. \tag{III. 43}$$

Since a_{12} (which appears in w) is $2\lambda/9$, Eqs. (III. 40), (III. 41), (III. 42), and (III. 43) are sufficient to determine a_{11}, the energy level E_{2n-1}, and the coefficients $a_0, a_2, \cdots$ of $P(y)$. To make contact with the beginning of this section, let $n=1$. Then

$$E_1 = 3a_{11}^{1/2}, \tag{III. 44a}$$

$$a_{11}^{3/2} - \tfrac{1}{4}a_{11}^{1/2} - 5\lambda/9 = 0 \tag{III. 44b}$$

which is equivalent to (III. 8).

When $n=2$

$$E^2 - 10a_{11}^{1/2} + 21a_{11} - 12a_{12}a_{11}^{-1/2} = 0. \tag{III. 45}$$

The larger of the two roots of this equation will yield the energy of the third excited state E_3, while the smaller will yield the first, E_1;

$$E = 5a_{11}^{1/2} \pm (4a_{11} + 12a_{12}a_{11}^{-1/2})^{1/2}. \tag{III. 46}$$

In this case $a_{11}^{1/2}$ is a root of

$$a_{11}^{3/2} - \tfrac{1}{4}a_{11}^{1/2} - \lambda = 0. \tag{III. 47}$$

When $\lambda^2 < \frac{1}{432}$ the solution of this cubic is

$$a_{11}^{1/2} = 3^{-1/2}\cos\tfrac{1}{3}[\cos^{-1}(12\lambda\cdot 3^{1/2})]. \tag{III. 48}$$

When $\lambda^2 > \frac{1}{432}$

$$a_{11}^{1/2} = (\tfrac{1}{2}\lambda)^{1/3}\left\{\left[1 + \left(1 - \frac{1}{432\lambda^2}\right)^{1/2}\right]^{1/3} + \left[1 - \left(1 - \frac{1}{432\lambda^2}\right)^{1/2}\right]^{1/3}\right\}. \tag{III. 49}$$

In the large λ regime

$$a_{11}^{1/2} \sim \lambda^{1/3} + (1728\lambda)^{-1/3} + \cdots \sim \lambda^{1/3} + \tfrac{1}{12}\lambda^{-1/3} + \cdots \tag{III. 50}$$

so that

$$\left.\begin{matrix} E_3 \\ E_1 \end{matrix}\right\} \sim \lambda^{1/3}[(5 \pm 2\sqrt{5/3}) + (\tfrac{5}{12} \pm \tfrac{1}{15}\sqrt{5/3})\lambda^{-2/3} + \cdots] \tag{III. 51}$$

$$\sim \begin{cases} 7.582\lambda^{1/3} + 0.5027\lambda^{-2/3} + \cdots \\ 2.418\lambda^{1/3} + 0.3306\lambda^{-2/3} + \cdots \end{cases} \tag{III. 52}$$

which are to be compared with the exact results[1]

$$E_3 \sim 7.336\lambda^{1/3} + 0.6183\lambda^{-2/3} + \cdots, \tag{III. 53a}$$

$$E_1 \sim 2.394\lambda^{1/3} + 0.3578\lambda^{-2/3} + \cdots. \tag{III. 53b}$$

In the limit as $\lambda \to \infty$, the percentage error in our approximate E_3 is 3.3% while that in our approximate E_1 is 1.0%.

The energy levels E_5, E_7, etc. for small odd integers follow in a similar manner from (III. 42) as do the wavefunctions from (III. 41) and (III. 42).

IV. FIRST ORDER LARGE QUANTUM NUMBER STATES

We now consider the case of n large and λ not too small. As was noted in Ref. 7 the "large" λ regime begins for smaller and smaller λ as n increases, the condition being that $\eta \equiv 2\lambda n > 1$, i.e., if $n=20$, $\lambda > \frac{1}{40}$; hence when λ is relatively small, say 0.02.

Since, in Eq. (III. 21), E is always associated with $a_{11}^{1/2}$, we set

$$E \equiv \epsilon a_{11}^{-1/2} \tag{IV. 1}$$

so that we are required to determine ϵ. Our basic equations (III. 21), then, when put in inverted order, have the form

$$[\epsilon - (4n-3)]a_{2n-2} + 4\cdot 1 a_{12}a_{11}^{-1}a_{2n-4} = 0,$$

$$(n-1)(2n-1)a_{11}^{-1/2}a_{2n-2} + [\epsilon - (4n-7)]a_{2n-4} + 4\cdot 2a_{12}a_{11}^{-1}a_{2n-6} = 0,$$

$$(n-2)(2n-3)a_{11}^{-1/2}a_{2n-4} + [\epsilon - (4n-11)]a_{2n-6} + 4\cdot 3a_{12}a_{11}^{-1}a_{2n-8} = 0,$$

..... (IV. 2)

$$(2\cdot 3)a_{11}^{-1/2}a_4+[\epsilon-5]a_2+4(n-1)a_{12}a_{11}^{-1}a_0=0,$$

$$(1\cdot 1)a_{11}^{-1/2}a_2+[\epsilon-1]a_0=0.$$

When n is large, in the first few equations ϵ always appears in the combination $(\epsilon-4n)+O(n)$. Hence we scale it with n and define ζ so that

$$\epsilon=n\zeta. \tag{IV.3}$$

Then it remains to find values of ζ as eigenvalues. Our characteristic equation which must be satisfied if (IV. 2) is to have a solution is

$$\begin{vmatrix} (\zeta-4)+(3/n) & 4\cdot 1a_{12}a_{11}^{-1}/n & 0 & 0 & \cdots \\ (n-1)(2n-1)/(na_{11}^{1/2}) & (\zeta-4)+(7/n) & 4\cdot 2a_{12}a_{11}^{-1}/n & 0 & \cdots \\ 0 & (n-2)(2n-3)/(na_{11}^{1/2}) & (\zeta-4)+(11/n) & 4\cdot 3a_{12}a_{11}^{-1}/n & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{vmatrix}=0. \tag{IV.4}$$

We now show that as $n\to\infty$ and $\lambda\to\infty$, the upper part of the determinant becomes diagonal. While the lower part does not, the scheme is sufficient to find the largest eigenvalue.

Notice that when n is large [cf. (III. 27)]

$$a_{11}^{3/2}\sim 4n\lambda/9 \text{ while generally } a_{12}=2\lambda/9. \tag{IV.5}$$

Hence the upper off-diagonal elements of (IV. 4) are of order

$$O(a_{12}a_{11}^{-1}/n)=O([\lambda^{1/3}n^{-4/3}])/n^{1/3} \tag{IV.6}$$

while the lower off-diagonal elements are

$$O(n^2/na_{11}^{1/2})=O([n^{4/3}\lambda^{-4/3}])/n^{2/3}. \tag{IV.7}$$

If we let $\lambda\to\infty$ and $n\to\infty$ in such a manner that

$$\lambda^{1/3}n^{-4/3}=u=\text{const}, \tag{IV.8}$$

then all off-diagonal elements near the "top" of the determinant (IV. 4) vanish so that part of the determinant becomes diagonal even though the bottom might not.

It should be emphasized that the set of equations (IV. 2) corresponds to the even energy levels $E_0, E_2, E_4, \cdots$, and as was mentioned below, Eq. (III. 21), a given value of n in (IV. 2) yields E_{2n-2} as the largest characteristic value of (III. 26) and (IV. 4).

The largest characteristic value of (IV. 4) is $\zeta=4$. Hence, from (IV. 3) and (IV. 1), as $n\to\infty$

$$E_{2n}\sim 4na_{11}^{1/2}\sim 4n(4n\lambda/9)^{1/3}$$
$$\sim 1.21(2n)^{4/3}\lambda^{1/3}. \tag{IV.9}$$

The dependence of E_{2n} on $(2n)$ and λ is precisely that obtained in a WKB type calculation.[2,8] The coefficient 1.21 is 13% below the exact value 1.38. Similar results can be found for the odd quantum numbers.

We now proceed to show how a second order approximation can be developed for excited states.

V. EXCITED STATES, SECOND ORDER

The procedure for finding second and higher order approximations to wavefunctions and energy levels of excited states follows in a similar manner. We first derive the second approximation to the first excited state quantities by choosing

$$\psi_1^{(2)}(y)=y\exp[-(a_{21}y^8+a_{22}y^{10}+a_{23}y^{12})^{1/4}] \tag{V.1}$$

with

$$a_{23}=(2\lambda/9)^2. \tag{V.2}$$

In the small y regime

$$\psi_1^{(2)}(y)\sim y\exp[-(c_2y^2+c_4y^4+c_6y^6+\cdots)] \tag{V.3}$$

with

$$c_2=a_{21}^{1/4},\quad c_4=\tfrac{1}{4}a_{22}a_{21}^{-3/4}, \tag{V.4a}$$

$$c_6=\tfrac{1}{4}(a_{23}a_{21}^{-3/4}-\tfrac{3}{8}a_{22}^2a_{21}^{-7/4}), \tag{V.4b}$$

so that

$$c_6=(\lambda^2/81c_2^3)-\tfrac{1}{2}(3c_4^2/c_2). \tag{V.4c}$$

Since

$$\frac{d^2\psi_1^{(2)}}{dy^2}=[-6c_2+4(c_2^2-5c_4)y^2+(16c_2c_4-42c_6)y^4+\cdots]\psi_1^{(2)}, \tag{V.4d}$$

the Schrödinger equation

$$\frac{d^2\psi_1}{dy^2}=2(-E_1+\tfrac{1}{2}y^2+\lambda y^4)\psi_1 \tag{V.5}$$

implies that

$$E_1=3c_2,\quad 4(c_2^2-5c_4)=1, \tag{V.6}$$

$$\lambda=8c_2c_4-21c_6. \tag{V.7}$$

By combining (V. 7) with (V. 4c) a relation between c_2 and c_4 follows,

$$c_6=\tfrac{8}{21}c_2c_4-\frac{\lambda}{21}=\frac{\lambda^2}{81c_2^5}-\frac{3}{2}\frac{c_4^2}{c_2}. \tag{V.8}$$

Since from (V. 6),

$$c_2 = \tfrac{1}{3}E_1 \quad \text{and} \quad c_4 = \frac{1}{5}\left(\frac{E_1^2}{9} - \frac{1}{4}\right), \tag{V. 9}$$

we obtain the characteristic equation for E_1,

$$E_1^6 - \tfrac{927}{286}E_1^4 - \tfrac{1350}{143}\lambda E_1^3 + \tfrac{5103}{2288}E_1^2 - \tfrac{9450}{143}\lambda^2 = 0, \tag{V. 10}$$

which has the equivalent form

$$E_1 = \frac{3}{2} + \frac{80(135)\lambda(7\lambda + E_1^3)}{(2E_1+3)(572E_1^2 - 567)E_1^2}. \tag{V. 11}$$

Clearly as $\lambda \to 0$, $E_1 \to \frac{3}{2}$ as it should. Once E_1 is known for a given value of λ, c_2 and c_4 follow (V. 9), and c_6 from (V. 8). From these c's, a_{21} and a_{22} are to be obtained from (V. 4a).

The large λ asymptotic expression for E_1 is obtained by substituting

$$E_1 \sim \epsilon_1\lambda^{1/3} + \epsilon_2\lambda^{-1/3} + \cdots \tag{V. 12}$$

into (V. 10) and setting the coefficient of λ^2 and $\lambda^{4/3}$ equal to zero. One finds

$$143\epsilon_1^6 - 1350\epsilon_1^3 - 9450 = 0, \tag{V. 13}$$

$$\epsilon_2 = 309\epsilon_1^2/[4(143\epsilon_1^3 - 675)] \tag{V. 14}$$

so that

$$E_1 \sim 2.417\lambda^{1/3} + 0.336\lambda^{-1/3} + \cdots \tag{V. 15}$$

which is considerably closer to the exact asymptotic expression (III. 13) than is the first approximation (III. 12b). As $\lambda \to \infty$ it deviates by less than 1% from the exact result.

The general improvement over the first approximation is evident from the table below for intermediate values of λ [the first approximation and exact values being extracted from the table below Eq. (III. 13)].

λ	E_1 (1st approx.)	E_1 (2nd approx.)	E_1 (exact)[1,2]
1	2.769	2.747	2.738
2	3.348	3.309	3.293
50	9.168	8.995	8.915

The basic form which we choose for the second order approximation to excited states is

$$\psi(y) = P(y)\exp[-(a_{21}y^8 + a_{22}y^{10} + a_{23}y^{12})^{1/4}] \equiv P(y)\psi_0^{(2)}(y), \tag{V. 16}$$

where the correct approach to zero as $|y| \to \infty$ is guaranteed by defining a_{23} by (V. 2).

When $\psi''(y)$ is calculated [using the form of $\psi_0^{(2)}$ (for small y) given in (V. 3)], retaining terms to order y^4, and set into

$$\frac{d^2\psi}{dy^2} = 2(-E + \tfrac{1}{2}y^2 + \lambda y^4)\psi, \tag{V. 17}$$

it is found that the function $P(y)$ satisfies the differential equation

$$\frac{d^2P}{dy^2} - 4y\frac{dP}{dy}(c_2 + 2c_4y^2 + 3c_6y^4) - P[(2c_2 - E) + y^2(1 - 4c_2^2 + 12c_4) + 2y^4(\lambda + 15c_6 - 8c_2c_4)] = 0. \tag{V. 18}$$

This is the analog of (III. 17) except that now terms to order y^4 are retained. If we were to proceed to a third order calculation we would use (II. 3) and retain terms to order y^6.

We seek polynomial solutions of (V. 8). To this end we construct the series expansion

$$P(y) = \sum_{j=0}^{\infty} a_{2j}y^{2j+\rho} \tag{V. 19}$$

and choose energy levels in such a way that the series terminates into polynomials. The parameter ρ [in (V. 9)] is set equal to zero in the consideration of the even order states $E_0, E_2, E_4, \cdots$ and 1 for odd order states $E_1, E_3, \cdots$.

By substituting (when $\rho = 0$) (V. 19) into (V. 18) and setting the coefficients of various powers of y equal to zero we find that

$$2a_2 + 2a_0(E - c_2) = 0, \tag{V. 20a}$$

$$12a_4 + 2a_2(E - 5c_2) - a_0(1 + 12c_4 - 4c_2^2) = 0, \tag{V. 20b}$$

$$30a_6 + 2a_4(E - 9c_2) - a_2(1 + 28c_4 - 4c_2^2) - 2a_0(15c_6 - 8c_2c_4 + \lambda) = 0, \tag{V. 20c}$$

etc.,

with the general recurrence formula connecting the a_{2j} being

$$(2j+2+\rho)(2j+1+\rho)a_{2j+2} + 2a_{2j}[E - c_2(1 + 4j + 2\rho)] - a_{2j-2}[1 - 4c_2^2 + (16j - 4 + 8\rho)c_4] - a_{2j-4}[-16c_2c_4 + 2\lambda + 6(4j - 3 + 2\rho)c_6] = 0. \tag{V. 21}$$

Let us find the energy and wavefunction associated with the second excited state of our oscillator. From Eq. (V. 20), we wish to terminate the series (V. 19) with the term $j = 1$ (and $\rho = 0$). To this end we set

$$a_4 = a_6 = 0. \tag{V. 22}$$

Then the hierarchy (V. 7) is restricted to the four equations

$$2a_2 + 2a_0(E - c_2) = 0, \tag{V. 23a}$$

$$2a_2(E - 5c_2) - a_0(1 + 12c_4 - 4c_2^2) = 0, \tag{V. 23b}$$

$$a_2(1 + 28c_4 - 4c_2^2) + 2a_0(15c_6 - 8c_2c_4 + \lambda) = 0, \tag{V. 23c}$$

$$a_2(-16c_2c_4 + 2\lambda + 54c_6) = 0. \tag{V. 23d}$$

The four unknowns in our problem are E_2, a_2/a_0, a_{22}, and a_{21} [see Eqs. (V. 1)– (V. 4)]. While the three c_j's, c_2, c_4, and c_6 appear in the above four equations and (V. 4), only two are independent since a_{23} in (V. 4) given by (V. 2).

Since a_2 is not to vanish we have from (V. 23d) the following relation for c_6,

$$27c_6 = 8c_2c_4 - \lambda, \tag{V. 24}$$

which when combined with (V. 4c) yields an expression for c_4 in terms of λ and c_2,

$$\frac{c_4}{c_2^2} = -\frac{8}{81} + \left[\left(\frac{8}{81}\right)^2 + \frac{2}{81}\left(\frac{1}{3}\frac{\lambda^2}{c_2^6} + \frac{\lambda}{c_2^3}\right)\right]^{1/2} \tag{V. 25}$$

which vanishes as $\lambda \to 0$. We eliminate the ratio a_2/a_0 between (V. 23a) and (V. 23b), and between (V. 23c) and (V. 23b). The resulting equations are

$$2(E-5c_2)(E-c_2)+(1-4c_2^2)=-12c_4, \quad \text{(V. 26a)}$$

$$9(E-c_2)(1+28c_4-4c_2^2)+8(8c_2c_4-\lambda)=0. \quad \text{(V. 26b)}$$

From each of these equations one could solve for $(c_4/c_2^2)+\frac{8}{81}$ and eliminate c_4 with the aid of (V. 25) obtaining two nonlinear equations for c_2, one being of the eight degree and the other of the sixth, with the coefficients of powers of c_2 being polynomials in E. The characteristic determinant for E which in this case is of 14×14 is obtained by Sylvesters' dialytic method of elimination.[9] Since the algebra is long and tedious we sketch a numerical successive approximation alternative.

We prefer to expand each variable in the set (V. 25)–(V. 26) around a good estimate of its value as perhaps would be obtained from the first order calculations (III. 29a). We thus write

$$E_2=\epsilon_0(1+\epsilon), \quad c_2=c_0(1+\gamma), \quad \text{(V. 27a)}$$

$$c_4=c_4^{(0)}(1+\alpha\gamma), \quad \text{(V. 27b)}$$

where ϵ_0, c_0, and $c_4^{(0)}$ are postulated to be good estimates of E_2, c_2, and c_4 so that ϵ and γ can be treated as numbers so small that ϵ^2, γ^2, and $\epsilon\gamma$ can at first be neglected. The values of ϵ_0 and c_0 might be chosen from (III. 29a) for given values of λ. The linearized basic equations are obtained by substituting (V. 27) into (V. 26) retaining only first order terms in ϵ and γ. Then

$$(\epsilon_0^2-6\epsilon_0c_0+3c_0^2+6c_4^{(0)}+\tfrac{1}{2})+2\epsilon(\epsilon_0^2-3c_0\epsilon_0)+6\gamma(c_0^2-c_0\epsilon_0+\alpha c_4^{(0)})=0, \quad \text{(V. 28a)}$$

$$[9(\epsilon_0-c_0)(1-4c_0^2)+4(63\epsilon_0-47c_0)c_4^{(0)}-8\lambda]+9\epsilon\epsilon_0(1-4c_0^2+28c_4^{(0)})-\gamma[9c_0(1-4c_0^2)+72c_0^2(\epsilon_0-c_0)+188c_0c_4^{(0)}-4\alpha c_4^{(0)}(63\epsilon_0-47c_0)]=0. \quad \text{(V. 28b)}$$

The values of $c_4^{(0)}$ and α in (V. 27b) are obtained by substituting $c_2=c_0(1+\gamma)$ into (V. 25) and obtaining the constant term and that linear term in γ.

We sketch the procedure for using the above equations by considering the case $\lambda=2$. Then from Eqs. (III. 30b) and (III. 31) useful estimates for ϵ_0 and c_0 are $\epsilon_0=6.4$ and $c_0=1.2$. Then

$$c_4^{(0)}=0.1777 \quad \text{and} \quad \alpha=-0.7701.$$

Upon substituting these parameters into (V. 28) we find the pair of equations for ϵ and γ,

$$0.02003+0.93672\epsilon=\gamma, \quad \text{(V. 29a)}$$

$$0.01078+0.01730\epsilon=\gamma, \quad \text{(V. 29b)}$$

which yield

$$\epsilon=-0.01006 \quad \text{and} \quad \gamma=0.01061, \quad \text{(V. 30)}$$

$$E_2=6.34 \quad \text{and} \quad c_2=1.21.$$

When ϵ_0 is chosen to be 6.34 and $c_0=1.21$, and the cycle is repeated, the calculated energy level E_2 becomes

$$E_2=6.335 \quad \text{(V. 31)}$$

which deviates by only 0.5% from the exact result 6.304 tabulated below Eq. (III. 34b) as compared with the first order approximation 6.430 which was high by 2%.

We close this section by making a second order calculation of E_2 in the regime of large λ. We start with the first order estimates and

$$E_2\sim4.8\lambda^{1/3} \quad \text{and} \quad c_2\sim a_{11}^{1/2}\sim(7\lambda/9)^{1/3}\approx0.92\lambda^{1/3} \quad \text{(V. 32)}$$

from which we deduce from (V. 25), choosing $\epsilon_0\sim4.8\lambda^{1/3}$ and $c_0\sim0.92\lambda^{1/3}$,

$$c_4^{(0)}\sim0.11497\lambda^{2/3} \quad \text{and} \quad \alpha\sim-0.7704. \quad \text{(V. 33)}$$

Then Eqs. (V. 28a) and (V. 28b) imply that

$$0.89225\epsilon-\gamma=0.0103411, \quad \text{(V. 34a)}$$

$$0.02246\epsilon+\gamma=-0.021998, \quad \text{(V. 34b)}$$

so that

$$\epsilon=-0.01274 \quad \text{and} \quad \gamma=-0.0217. \quad \text{(V. 35)}$$

Therefore, the new approximation to E_2 is

$$E_2=4.8\lambda^{1/3}[1-0.01274]\approx4.739\lambda^{1/3} \quad \text{(V. 36)}$$

which deviates by only 0.9% from the exact value 4.697.

VI. REMARKS ON THE POTENTIAL $\frac{1}{2}y^2+\alpha y^4+\lambda y^6=V(y)$

A potential whose highest power is y^6 (with $\lambda>0$) can be treated in the manner employed in previous sections of our discussion of potentials with a quartic anharmonicity. While we limit ourselves here to some remarks about the ground state of our new oscillator, it will be evident that the analysis can be extended to cover excited states as well.

We consider a ground state wavefunction of the form

$$\psi_0(y)=\exp[-(a_{11}y^4+a_{12}y^6+a_{13}y^8)^{1/2}] \quad \text{(VI. 1)}$$

which has the proper behavior for large $|y|$ provided that we choose

$$a_{13}=\lambda/8 \quad \text{(VI. 1a)}$$

so that (V. 1) is consistent with Eqs. (I. 5) and (I. 6) when $\alpha=3$.

Following the practice of previous sections, for small y, we set

$$\psi_0(y)\sim\exp(-c_2y^2-c_4y^4-c_6y^6\cdots) \quad \text{(VI. 2)}$$

where

$$c_2=a_{11}^{1/2}, \quad c_4=\tfrac{1}{2}a_{12}a_{11}^{-1/2} \quad \text{(VI. 3a)}$$

$$c_6=(\lambda/16a_{11}^{1/2})-(a_{12}^2/8a_{11}^{3/2}), \quad \text{(VI. 3b)}$$

so that

$$\frac{1}{2}\frac{d^2\psi_0}{dy^2}=[-c_2+2(c_2^2-3c_4)y^2-(15c_6-8c_2c_4)y^4+\cdots]\psi_0. \quad \text{(VI. 4a)}$$

When this is compared with the Schrödinger equation

$$\frac{1}{2}\frac{d^2\psi_0}{dy^2}=(-E_0+\tfrac{1}{2}y^2+\alpha y^4+\lambda y^6)\psi_0, \quad \text{(VI. 4b)}$$

it is apparent that to obtain agreement to terms of order y^4 we set

$$E_0 = c_2, \quad c_2^2 - 3c_4 = \tfrac{1}{4}, \tag{VI. 5}$$

$$\alpha = 8c_2c_4 - 15c_6. \tag{VI. 6}$$

The λy^6 term which is important for large y is accounted for in our choice of a_{13}.

The coefficient c_6 is eliminated by combining (V. 6) and (V. 3b) to yield a relationship between c_4 and c_2,

$$\frac{\lambda}{16c_2} - \frac{c_4^2}{2c_2} = \frac{8c_2c_4}{15} - \frac{\alpha}{15}, \tag{VI. 7}$$

since from (V. 5)

$$c_4 = \tfrac{1}{3}(E_0^2 - \tfrac{1}{4}). \tag{VI. 8}$$

This yields the following characteristic equation for the ground state energy,

$$(E_0^2 - \tfrac{1}{4})(\tfrac{7}{5}E_0^2 - \tfrac{1}{12}) = \frac{3\lambda}{8} + \frac{2\alpha E_0}{5}. \tag{VI. 9}$$

This gives the energy level $E_0 = \frac{1}{2}$ in the limit $\lambda \to 0$, $\alpha \to 0$ as it should. The expected worst case is that of $\alpha = 0$ and λ very large. Then E_0 is very large and

$$\tfrac{7}{5}E_0^4 \sim \tfrac{3}{8}\lambda \tag{VI. 10}$$

so that

$$E_0 \sim (\tfrac{15}{56})^{1/4}\lambda^{1/4} = 0.719\lambda^{1/4} \tag{VI. 11}$$

which is to be compared with the exact result[7]

$$E_0 \sim 0.681\lambda^{1/4}. \tag{VI. 12}$$

As in the case of the quartic oscillator this can be improved by adding another parameter to the approximating wavefunction, putting it into the form

$$\psi_0(y) = \exp[-(a_{21}y^6 + a_{22}y^8 + a_{23}y^{10} + a_{24}y^{12})^{1/3}]. \tag{VI. 13}$$

When certain relations exist between the quartic and sextic force constants α and λ, an exact ground state wavefunction can be found of the form

$$\psi_0(\psi) = \exp[-(ay^2 + by^4)]. \tag{IV. 14}$$

In this case

$$\frac{1}{2}\frac{d^2\psi}{dy^2} = [-a + (2a^2 - 6b)y^2 + 8aby^4 + 8b^2y^6]\psi \tag{VI. 15}$$

which, when compared with (VI. 4b), yield

$$E = a, \tag{VI. 16a}$$

$$\tfrac{1}{4} = a^2 - 3b, \tag{VI. 16b}$$

$$\alpha = 8ab, \quad \lambda = 8b^2, \tag{VI. 16c}$$

since

$$b = \left(\frac{\lambda}{8}\right)^{1/2} \quad \text{and} \quad a = \frac{\alpha}{8b} = \frac{\alpha}{8}\left(\frac{8}{\lambda}\right)^{1/2} = \left(\frac{\alpha^2}{8\lambda}\right)^{1/2}. \tag{VI. 17}$$

(VI. 14) is an exact wavefunction when α and λ are related by

$$\frac{1}{4} = \frac{\alpha^2}{8\lambda} - 3\left(\frac{\lambda}{8}\right)^{1/2} \tag{VI. 18}$$

or

$$\alpha^2 = 8\lambda[\tfrac{1}{4} + 3(\lambda/8)^{1/2}], \tag{VI. 19}$$

when

$$\alpha = \{2\lambda + 3\lambda(8\lambda)^{1/2}\}^{1/2}, \tag{VI. 20}$$

$$E = \tfrac{1}{2}[1 + 3(2\lambda)^{1/2}]^{1/2} \to \begin{cases} \frac{1}{2} & \lambda \to 0, \\ (\frac{9}{8})^{1/4}\lambda^{1/4} & \text{as } \lambda \to \infty. \end{cases} \tag{VI. 21}$$

The exact wavefunction (VI. 14) with α and λ related by (VI. 18) has been found independently by Gillespie.[10]

[1] F. T. Hioe and E. W. Montroll, J. Math. Phys. **16**, 1945 (1975).
[2] S. Biswas, K. Datta, R. Saxena, P. Srivastava, and V. Varma, J. Math. Phys. **14**, 1190 (1973).
[3] C. Bender, G. Guralnik, R. Keener, and K. Olaussen, Phys. Rev. **14**, 2590 (1976).
[4] J. Loeffel, A. Martin, B. Simon, and A. Wightman, Phys. Lett. B 30, 656 (1969).
[5] B. Simon, Ann. Phys. **58**, 76 (1970).
[6] C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969); Phys. Rev. D 7, 1620 (1973).
[7] F. T. Hioe, Don MacMillen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976).
[8] E. C. Titchmarsh, *Eigenfunction Expansions, Part I* (Oxford U. P., Oxford, 1961), 2nd ed.
[9] Cf. M. Bôcher, *Introduction to Higher Algebra* (MacMillan, New York, 1931), p. 198.
[10] G. H. Gillespie, private communication.

The Gaussian version of Ampère's law

E. J. Post

8606 Pershing Drive, Playa del Rey, California 90291
(Received 5 July 1977)

Ampère's law of magnetostatics and Gauss' law of electrostatics are one- and two-dimensional period integrals in the sense of de Rham,[1] by virtue of their invariance under deformation of their cyclic integration domains. The spacetime 2-form $\tilde{G}$ (tilde denoting impair properties) defined by the fields **D** and **H**, permits the two spatial laws to be simultaneously expressed by the two-dimensional period integral in spacetime,

$$\oint_{C_2}\tilde{G} = \text{algebraic sum of charges with world lines linked by } C_2. \tag{1}$$

If the cycle C_2 is a closed spatial surface, Eq. (1) expresses the traditional Gauss law. If C_2 is a closed surface involving a direction of time, Eq. (1) can reduce to the historic version of Ampère's law, provided the Maxwell displacement vanishes. In this Gaussian version of Ampère's law a constant circuital current relates to a cyclic time domain C_t, the magnitude of which is given by a complete circulation of the charge in the circuit. The cyclic domain C_2 so becomes the topological product of C_t with a meridional Ampère loop C_m; i.e., $C_2 = C_t \times C_m$. The reader will note that C_m only links with the current loop in three dimensions; in four dimensions a two-dimensional cycle C_2 is needed to link with the current loop.

A clean application of the Gaussian version of Ampère's law is provided by a current circulating in a superconducting ring; say a superconductor of the first kind so as to have a well-defined penetration domain.

The collective quantum state of the BCS charge carriers makes it meaningful to postulate a unique circulation time t for the collective of BCS pairs, which in turn gives a well defined C_t so that $C_2 = C_t \times C_m$. If s is the number of BCS pairs with a charge $2e$, Eq. (1) yields

$$\oint_{C_2}\tilde{G} = s(2e). \tag{2}$$

Hence the Gaussian version of Ampère's law counts the number of BCS pairs partaking in the supercurrent.

The result given by Eq. (2) can be used to calculate the periods of the 3-form of "spin" given by Kiehn[2]:

$$\oint_{C_3}\tilde{S} = \oint_{C_3} A \wedge \tilde{G}. \tag{3}$$

The three-dimensional cyclic domain C_3 is now taken to be the topological product of C_2 in Eq. (2) and the equatorial integration loop C_e of the 1-form A, thus permitting the decomposition rule

$$\oint_{C_3} A \wedge \tilde{G} = \oint_{C_e} A \cdot \oint_{C_2} \tilde{G} \quad \text{with } C_3 = C_e \times C_2. \tag{4}$$

Experiment has confirmed the period integral $\oint A$ to be a multiple, say n, of the flux quantum $(h/2e)$, hence substitution in Eq. (4) and using Eq. (2) yields, as periods of Eq. (3),

$$\oint_{C_3}\tilde{S} = (sn)h, \tag{5}$$

the orbital angular momentum associated with the charge carriers.

Since the gyromagnetic ratio of a superconductor is known to be of the orbital type, result (5) then implies a magnetic moment,

$$|\mu| = (sn) \text{ Bohr magnetons}, \tag{6}$$

with s the number of BCS pairs and n the flux quantum state of the superconducting ring.

A careful inquiry among solid state experimentalists has not led to an argument against or in favor of the, in principle, observable result (6).

The reader will note that the quantum state of $\tilde{S}$ is given by the product of the quantum states of A and $\tilde{G}$, hence one has the corollary that prime number states of $\tilde{S}$ imply that either A or $\tilde{G}$ is in a ground state with quantum number 1.

Since period laws are not restricted by metric considerations of size, it is mathematically permissible to make the step of extrapolating their validity into the subatomic domains. Following Jehle[3] one may associate a flux unit (h/e) with particles having a magnetic moment. The spin features of a particle demand a 4π cycle of C_e, which in turn demands C_e to be a self-knotted, double loop (trefoil). The self-knotting is essential to prevent a topological equivalence of the 4π cycle to a 2π cycle through a continuous unfolding of the double loop to a single loop. The actual situation is, however, more involved, because a 1-cycle can be unknotted in four dimensions.

This very preliminary topological model of a Dirac particle permits an asymptotic interpretation of its stability in terms of a "balance" between attractive magnetic forces and repelling Coulomb forces between the two branches of the double, self-knotted loop; of course, falling short of telling what is being attracted and repelled. Such speculative thought has some potential for identifying the nature of mass as a carrier of charge, yet its further pursuit is outside the typical realm of mathematical physics and should be made contingent on a test of the physical relevance of Eqs. (5) and (6) in a macroscopic situation.

[1] G. de Rham, *Variétés Differentiables* (Herman, Paris, 1955).
[2] R. M. Kiehn, J. Math. Phys. **18**, 614 (1977). [See also comments by E. J. Post, J. Math. Phys. **18**, 2084 (1977).]
[3] H. Jehle, Phys. Rev. D **3**, 306 (1971); **6**, 441 (1972); **11**, 2147 (1975).

 0022-2488/78/1901-0347$1.00

ERRATA

Erratum: The n-bubble series in the theory of the classical one-component plasma [J. Math. Phys. 18, 292 (1977)]

Y. Furutani[a] and C. Deutsch

Laboratoire de Physique des Plasmas, Université de Paris-Sud, Centre d'Orsay, 91405 Orsay, France
(Received 30 August 1977)

Eq. (1): Sum over n should read: $\sum_{n=2}^{\infty}$

Eq. (4): Last line, criterion for α should read: $\alpha > 0$.

Eq. (10): In third term of rhs, $\theta_n \cos((n-1)\theta_n)$ should read: $\theta_n \cos((n-2)\theta_n)$.

Eq. (19): lhs. should read: $k^2 G(k)/4\pi\Lambda$.

Below Eq. (30): $\sum_{l=0}^{2(2\nu+1)}$ should read $\sum_{l=0}^{2\nu+1}$.

Page 297: On the top of the second column, $f''(\infty)$ should read $f'(\infty)$.

Page 299: Eq. (B4) should read:

$$\mathcal{A}_l(\Lambda) \overset{\Lambda\to\infty}{\sim} \frac{1}{\Lambda}\left(1 - \frac{1}{(2l+1)!(2l+3)}\left[I_{2l+1}(\Lambda) + I_{2l+2}(\Lambda)\right]\right)$$

[a] Present address: Department of Electronics, Okayama University, Okayama 700, Japan.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*

 0022-2488/78/1901-0348$1.00