An upper bound on the number of eigenvalues of an infinite dimensional Jacobi matrix

Jeffrey S. Geronimo
School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 6 April 1981; accepted for publication 4 September 1981)

An upper bound is given on the number of eigenvalues of a class of infinite dimensional Jacobi matrices. The theorem presented is a discrete analog of the celebrated result of V. Bargmann on the number of bound states of the Schrödinger equation

PACS numbers: 02.10.Sp

1. INTRODUCTION

Given sequences $\{a(n)\}$ and $\{b(n)\}$ with $a(n+1)>0$ and $b(n)$ real, $n\geqslant 0$, one can construct a sequence of polynomials $\{p(\lambda,n)\}$ satisfying

$$a(n+1)p(\lambda,n+1)+b(n)p(\lambda,n)+a(n)p(\lambda,n-1)$$
$$=\lambda p(\lambda,n),\quad n\geqslant 0,$$
$$p(\lambda,0)=1,\quad p(\lambda,-1)=0. \tag{1}$$

It is known[1] that these polynomials will be orthogonal with respect to some (not necessarily unique) distribution function $d\rho(\lambda)$. If the following limits exist

$$\lim_{n\to\infty} a(n)=a(\infty)>0 \qquad \lim_{n\to\infty} b(n)=b(\infty), \tag{2}$$

then $d\rho(\lambda)$ will be unique and by a theorem of Blumenthal[2] (see also Nevai[3]) the points

$$a=b(\infty)-2a(\infty) \quad \text{and} \quad b=b(\infty)+2a(\infty) \tag{3}$$

are the smallest and largest limit points of the spectrum $\rho(\lambda)$ [the support of $d\rho(\lambda)$]. It is without loss of generality that one can set $a(\infty)=\frac{1}{2}$ and $b(\infty)=0$, so $a=-1$ and $b=1$.

Since the points in the complement of $[-1,1]$ are in the point spectrum of $\rho(\lambda)$ it is useful to know when there are only a finite number. This question can be reformulated in terms of operator theory in the following manner. Let J be the infinite dimensional Jacobi matrix

$$J=\begin{bmatrix} b(0) & a(1) & 0 & 0 & 0 & \cdots \\ a(1) & b(1) & a(2) & 0 & 0 & \cdots \\ 0 & a(2) & b(2) & a(3) & 0 & \cdots \\ 0 & 0 & a(3) & b(3) & a(4) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix},$$

where the $a(n)$s and $b(n)$s are those given in (1) satisfy (2) with $a(\infty)=\frac{1}{2}, b(\infty)=0$. The infinite dimensional matrix J can be considered a bounded self-adjoint operator acting on l_2 and a question of interest is to find $\psi_m\in l_2$ and λ_m real such that

$$J\psi_m=\lambda_m\psi_m.$$

Our question now becomes—given J, can one find upper bounds on the number of eigenvalues λ_m where $|\lambda_m|>1$? Chihara,[4–6] with his very nice use of chain sequences, Chihara and Nevai,[7] and Geronimo and Case[8] all deal with this question.

The above problem has an important continuum analog: when does the Schrödinger equation with a central force field have only a finite number of bound states? In a spectacular paper written in 1952 Bargmann[9] gave an upper bound on the number of bound states and it is the discrete analog of Bargmann's bound that is the main result of this paper. It has been noted by Reed and Simon[10] that Bargmann's proof is intimately connected with the min–max principle of Weyl, Fisher, and Courant (see Appendix). The proof given here is based on Bargmann's.

2. THE MAIN RESULT

Theorem 1: Let $N(1,\infty), N(-\infty,-1)$ be the number of points in the spectrum of $\rho(\lambda)$ (the number of eigenvalues of J) for $\lambda>1, \lambda<-1$ respectively; then

$$N(1,\infty)\leqslant \sum_{i=1}^{\infty} i\{|1-4a^{+}(i)^2|+|2b^{+}(i-1)|\} \tag{4a}$$

and

$$N(-\infty,-1)\leqslant \sum_{i=1}^{\infty} i\{|1-4a^{-}(i)^2|+|2b^{-}(i-1)|\}, \tag{4b}$$

where

$$1-4a^{+}(i)^2=1-4a^{-}(i)^2=\begin{cases}1-4a(i)^2, & 1-4a^2(1)<0\\ 0, & 1-4a^2(i)\geqslant 0\end{cases}$$

and

$$b^{+}(i)=\begin{cases}b(i), & b(i)>0\\ 0, & b(i)\leqslant 0,\end{cases} \tag{4c}$$
$$b^{-}(i)=\begin{cases}b(i), & b(i)<0\\ 0, & b(i)\geqslant 0\end{cases}$$

Before beginning the proof the equations that are to be used will be derived.

It can be seen that given (1) and (2) with $a(\infty)=\frac{1}{2}$ and $b(\infty)=0$ the polynomials satisfy the following two three-term recursion formulas;

$$p(\lambda,n)=\frac{1}{2a(n)}\left(\frac{\psi(z,n-1)}{z}+[z-2b(n-1)]\,p(\lambda,n-1)\right), \tag{5}$$

 0022-2488/82/060917-05$02.50

$$\psi(z,n) = \frac{1}{2a(n)}\left(\frac{\psi(z,n-1)}{z} + \{[1-4a(n)^2]z - 2b(n-1)\}\,p(\lambda,n-1)\right), \quad (6)$$

$$\psi(z,0) = p(\lambda,0) = 1, \quad (7)$$

with

$$\lambda = \tfrac{1}{2}(z + 1/z).$$

Incrementing (5) by one, letting $z \to 1/z$, and then subtracting the new equation from the original and multiplying by z^{n+1} yields

$$(1-z^2)z^n p\hat{}(\lambda,n) = z^n \psi\hat{}(z,n) - z^{n+2}\psi\hat{}(1/z,n). \quad (8)$$

Here

$$p\hat{}(\lambda,n) = \frac{p(\lambda,n)}{\alpha(n)}, \quad \psi\hat{}(z,n) = \frac{\psi(z,n)}{\alpha(n)}, \quad (9)$$

with

$$\alpha(n) = \prod_{i=1}^{n} \frac{1}{2a(i)}, \quad \alpha(0) = 1. \quad (10)$$

From (6) it follows that

$$z^n\psi\hat{}(z,n) = z^m\psi\hat{}(z,m) + \sum_{i=m}^{n-1}\{[1-4a(i+1)^2]z - 2b(i)\}z^{i+1}p\hat{}(\lambda,i). \quad (11)$$

Using (11) and its analog for $z^{-n}\psi\hat{}(1/z,n)$ in (8) and then using (8) with n equal to m gives

$$z^n p\hat{}(\lambda,n) = z^m p\hat{}(z,m) + z^{m+2}\left(\frac{1-z^{2n-2m}}{1-z^2}\right)\psi\hat{}(1/z,m), + \sum_{i=m}^{n-1}\left[[1-4a(i+1)^2]z^2\left(\frac{1-z^{2n-2i-2}}{1-z^2}\right) - 2b(i)z\left(\frac{1-z^{2n-2i}}{1-z^2}\right)\right]z^i p\hat{}(\lambda,i).$$

Therefore, at $z = 1$,

$$p\hat{}(1,n) = p\hat{}(1,m) + (n-m)\psi\hat{}(1,m) = \sum_{i=m}^{n-1}\{[1-4a(i+1)^2](n-i-1) - 2b(i)(n-i)\}p\hat{}(1,i). \quad (12)$$

Another useful equation is

$$p\hat{}(1,n) - \psi\hat{}(1,n) = 4a^2(n)\,p\hat{}(1,n-1), \quad (13)$$

which can be obtained from (5) and (6). Replacing the original system by the $+$ system, i.e., replacing the coefficients in (12) and (13) by those given in (4), yields

$$p^+(1,n) = p^+(1,m) + (n-m)\psi^+(1,m) - \sum_{i=m}^{n-1}\{|1-4a^+(i+1)^2|(n-i-1) + |2b^+(i)|(n-i)\}\,p^+(1,i) \quad (14)$$

and

$$p^+(1,n) - \psi^+(1,n) = 4a^+(n)^2 p^+(1,n-1). \quad (15)$$

Using the $+$ system the following will now be proved:

Theorem 2: Suppose sign $p^+(1,i)$ = constant for $m < i < n < \infty$ and $p^+(1,n) = 0$ or sign $p^+(1,n)$ $= -$ sign $p^+(1,i)$. Furthermore suppose $p^+(1,m) = 0$ or $-$ sign $p^+(1,m-1) =$ sign $p^+(1,m) =$ sign $p^+(1,i)$ then

$$1 \leqslant \sum_{i=m}^{n-1}(i+1)\{|1-4a^+(i+1)^2| + |2b^+(i)|\}. \quad (16)$$

Proof: The proof breaks up into two cases: case 1, $p^+(1,m) = 0$; and case 2, sign $p^+(1,m) =$ $-$ sign $p^+(1,m-1)$.

Case 1. Stopping (14) at $m+1$ instead of m and dividing by $p^+(1,m+1)$ gives

$$\frac{p^+(1,n)}{p^+(1,m+1)} = 1 + (n-m-1)\frac{\psi^+(1,m+1)}{p^+(1,m+1)} - \sum_{i=m+1}^{n-1}\{|1-4a^+(i+1)^2|(n-i-1) + |2b^+(i)|(n-i)\}\frac{p^+(1,i)}{p^+(1,m+1)}. \quad (17)$$

Since (17) holds for $n > m+1$ and $p^+(1,i)/p^+(1,m+1) > 0$, $m < i < n$, it follows that

$$\frac{p^+(1,i)}{p^+(1,m+1)} \leqslant 1 + (i-m-1)\frac{\psi^+(1,m+1)}{p^+(1,m+1)}, \quad m+1 < i < n. \quad (18)$$

Furthermore from (15) one finds

$$p^+(1,m+1) = \psi^+(1,m+1), \quad (19)$$

and this leads to

$$\frac{p^+(1,i)}{p^+(1,m+1)} \leqslant i-m, \quad m+1 < i < n. \quad (20)$$

Multiplying (17) by -1 then substituting in (20) yields the inequality

$$(n-m) \leqslant \sum_{i=m+1}^{n-1}\{|1-4a^+(i+1)^2|(n-i-1) + |2b^+(i)|(n-i)\}(i-m),$$

where the fact that $p^+(1,n)/p^+(1,m+1) \leqslant 0$ has been used. From the above equation one finds

$$(n-m) \leqslant (n-m)\sum_{i=m+1}^{n-1}\{|1-4a^+(i+1)^2| + |2b^+(i)|\}(i-m), \quad (21)$$

which leads to (16).

Case 2. Dividing (14) by $p^+(1,m)$ gives

$$\frac{p^+(1,n)}{p^+(1,m)} = 1 + (n-m)\frac{\psi^+(1,m)}{p^+(1,m)} - \sum_{i=m}^{n-1}\{|1-4a^+(i+1)^2|(n-i-1) + |2b^+(i)|(n-i)\}\frac{p^+(1,i)}{p^+(1,m)}. \quad (22)$$

As in the previous case the above equation with $n = i > m$ yields the inequality

$$\frac{p^+(1,i)}{p^+(1,m)} \leqslant 1 + (i-m)\frac{\psi^+(1,m)}{p^+(1,m)}, \quad m < i < n. \quad (23)$$

Again from (15) one finds

$$\frac{\psi^+(1,m)}{p^+(1,m)} > 1,$$

which leads to

$$\frac{p^+(1,i)}{p^+(1,m)} \leqslant (i-m+1)\frac{\psi^+(1,m)}{p^+(1,m)}. \tag{24}$$

Multiplying (22) by -1, substituting in (24), and using the fact that $p^+(1,n)/p^+(1,m) \leqslant 0$ gives

$$(n-m)\frac{\psi^+(1,m)}{p^+(1,m)} \leqslant \sum_{i=m}^{n-1} \{|1-4a^+(i+1)^2|(n-i-1) + |(2b^+(i)|(n-i)\}(i+1-m)\frac{\psi^+(1,m)}{p^+(1,m)}, \tag{25}$$

which leads to (16).

Lemma 1: Suppose $\operatorname{sign} p^+(1,i) = \text{const.}$ for $m < i$ and $\lim_{n\to\infty}[p^+(1,n)/n] = 0$. Furthermore suppose $p^+(1,m) = 0$ or $-\operatorname{sign} p^+(1,m-1) = \operatorname{sign} p^+(1,m) = \operatorname{sign} p^+(1,i)$; then

$$1 \leqslant \sum_{i=m}^{\infty} (i+1)\{|1-4a^+(i+1)|^2 + |2b^+(i)|\} \tag{26}$$

Proof: Again the proof involves two cases. Only the case $p^+(1,m) \neq 0$ will be given here because the other case is proved in an analogous manner.

Dividing (14) by $p^+(1,m)(n-m)$ gives

$$\frac{p^+(1,n)}{p^+(1,m)(n-m)} = \frac{1}{n-m} + \frac{\psi^+(1,m)}{p^+(1,m)} - \sum_{i=m}^{n-1}\left\{|1-4a^+(i+1)^2|\left(\frac{n-i-1}{n-m}\right) + |2b^+(i)|\left(\frac{n-i}{n-m}\right)\right\}\frac{p^+(1,i)}{p^+(1,m)}. \tag{27}$$

Multiplying (27) by -1 and then substituting (24) yields

$$\frac{-p^+(1,n)}{p^+(1,m)(n-m)} + \frac{\psi^+(1,m)}{p^+(1,m)} \leqslant \sum_{i=m}^{n-1}\{|1-4a^+(i+1)^2| + |2b^+(i)|\}(i+1)\frac{\psi^+(1,m)}{p^+(1,m)}. \tag{28}$$

Now letting $n \to \infty$ gives (26).

We now give the proof of Theorem 1.

Proof: One begins by replacing the original system by the $+$ system. It is a consequence of the theory of chain sequences[4] that $N^+(1,\infty) \geqslant N(1,\infty)$. One now counts the number of changes in sign of $p^+(1,n)$ as n varies from 0 to ∞, as this number is equal to $N^+(1,\infty)$. Equation (4a) follows from Theorem 2.

To prove the result for $N(-\infty,-1)$ note that the distribution $d\rho^\dagger(\lambda) \equiv -d\rho(-\lambda)$ generates coefficients $a^\dagger(n) = a(n)$ and $b^\dagger(n) = -b(n)$. Replacing this ($\dagger$) system by the $+$ system and then applying the results above complete the proof.

The above theorem has the following immediate corollary:

Corollary 1: If

$$\sum_{i=1}^{\infty}(i+1)\{|1-4a^+(i+1)| + |2b^+(i)|\} < 1,$$

$$\sum_{i=1}^{\infty}(i+1)\{|1-4a^-(i+1)\} + |2b^-(i)|\} < 1,$$

then $N(1,\infty) = 0$, $N(-\infty,-1) = 0$, respectively.

3. FURTHER RESULTS

Consider the systems for which

$$\sum_{n=1}^{\infty} n\{|1-4a(n)^2| + |2b(n-1)|\} < \infty. \tag{29}$$

Then $N(1,\infty)$ and $N(-\infty,-1)$ are finite. Furthermore under these conditions it is known[3,8] that the sequence $\{\frac{1}{2}z^n\psi(z,n)\}$ converges uniformly for $|z| \leqslant 1$ to a function $zf_+(z)$ and that

$$d\rho(\lambda) = \begin{cases} \sigma(\theta)d\lambda, & \lambda = \cos\theta,\ 0 \leqslant \theta \leqslant \pi, \\ \sum_{i=1}^{N}\rho_i\delta(\lambda-\lambda_i), & \lambda \text{ not as above.} \end{cases}$$

Here

$$\sigma(\theta)\,d\lambda = \frac{\sin\theta}{2\pi|f_+(z)|^2}, \quad \lambda = \cos\theta, z = e^{i\theta}, \tag{30}$$

$$\rho_i \alpha \left(\frac{df_+(\lambda_i)}{d\lambda}\right)^{-1}, \quad \lambda = \tfrac{1}{2}(z_i + 1/z_i), \tag{31}$$

where λ_i are the locations of the zeros of $f_+(z)$ for $|z| < 1$ (the branch $z = \lambda - (\lambda^2-1)^{1/2}$ has been chosen). From (31) one finds that under the condition given by (29) $N(1,\infty), N(-\infty,-1)$ equal the number of zeros of $f_+(z)$ for $0 < z < 1, 0 > z > -1$, respectively. Another consequence of (29) is that $f_+(z)$ can have a zero on the unit circle only at $z = 1$ and/or $z = -1$ and these zeros are simple.[8] Let $N[1,\infty), N(-\infty,-1]$ denote the number of zeros of $zf_+(z)$ for $0 < z \leqslant 1$ and $0 > z \geqslant -1$, respectively. The following theorem bounds the number of zeros of $f_+(z)$.

Theorem 4: If (29) holds then

$$N[1,\infty) \leqslant \sum_{i=1}^{\infty} i\{|1-4a^+(i)^2| + |2b^+(i-1)|\}, \tag{32a}$$

and

$$N(-\infty,-1] \leqslant \sum_{i=0}^{\infty} i\{|1-4a^-(i)^2| + |2b^-(i-1)|\}. \tag{32b}$$

Proof: From the min–max principle and the theory of chain sequences $N[1,\infty) \leqslant N^+[1,\infty)$. Furthermore it has been shown by Chihara and Nevai[7] that $f^+(1) = 0$ implies $\lim_{n\to\infty}[p^+(1,n)/n] = 0$. Equation (32a) now follows from Theorem 1 and Lemma 1. (32b) follows by considering the coefficients generated by the distribution $d\rho^\dagger(\lambda) = -d\rho(-\lambda)$.

4. EXAMPLES AND DISCUSSION

Example 1 (Tchebycheff polynomials of the first kind): Let $a(1) = 1/\sqrt{2}$, $a(n) = \frac{1}{2}$, $n > 1$, and $b(n) = 0, n \geqslant 0$. These are the coefficients associated with the polynomials $p(\lambda,0) = 1$, $p(\lambda,n) = \sqrt{2}\cos n\theta, n \geqslant 1, \lambda = \cos\theta$, which are orthonormal with respect to the weight

$$d\rho(\lambda) = (1/\pi)(1-\lambda^2)^{-1/2}\,d\lambda.$$

One finds that

$$f^+(z) = \frac{1}{2\sqrt{2}}\frac{1}{z}(1-z^2).$$

Therefore $N[1,\infty) = N(-\infty,-1] = 1$ and the bound given by Theorem 4 is saturated.

Example 2: $a(n) = \frac{1}{2}, n \geqslant 1, b(0) = \epsilon/2, b(n) = 0,$ $n \geqslant 1$. $f_+(z)$ for this system is

$$f_+(z) = (1/2z)(1 - \epsilon z).$$

With $\epsilon = 1\ (-1)$ this system becomes up to normalization the Jacobi polynomials $p^{(-1/2,1/2)}(\lambda,n),(p^{(1/2,-1/2)}(\lambda,n))$. With $\epsilon = 1$ the bound for $N[1,\infty)$ is saturated and $N(-\infty,-1] = 0$. For $\epsilon > 1$ there is only one mass point for $0 < z < 1$ no matter how large ϵ becomes.

Example 3: $a(n) = \frac{1}{2}$, $n \geqslant 1, b(n) = 0, n \neq m$, $b(m) \neq 0$. Then

$$f_+(z) = \frac{1}{2z}\left[1 - 2b(m)z^{m+1}\frac{z^{m+1} - z^{-(m+1)}}{z - 1/z}\right].$$

For $b(m) = 1/2(m+1), f_+(1) = 0$, and for $b(m) > 1/2(m+1)$, the zero of $f_+(z)$ moves into the interior of the unit circle along the positive real axis.

From the examples given above one can make the following observations:

1. As n increases it takes smaller and smaller deviations of $a^+(n)$ and $b^+(n)$ from their asymptotic values to introduce mass points.

2. The bound is a poor bound if the coefficients $a^+(n)$ and $b^+(n)$ deviate strongly from their asymptotic values.

The above comments hold equally well for $a^-(n)$ and $b^-(n)$.

ACKNOWLEDGMENTS

I would like to thank Professor Michael F. Barnsley for many useful discussions. This work was supported in part by N.S.F. grant MCS-8002731.

Note added in proof: A discrete analog of the Schwinger bound has been obtained.[12]

APPENDIX

As has been noted by Simon, Bargmann's theorem is intimately connected with the min–max principle of Well, Fisher, and Courant. Here we use a special case of the theorem. (For the general case and the proof see Reed and Simon.[10])

Theorem A: Let J be a bounded infinite dimensional Jacobi matrix operating on l_2. Define

$$\lambda_n(J) = \inf_{\phi_1,\phi_2,\ldots,\phi_{n-1}} U_J(\phi_1,\phi_2\ldots\phi_{n-1}),$$

where

$$U_J(\phi_1,\phi_2\ldots\phi_m) = \sup_{\substack{\psi \in D(J);\|\psi\| = 1 \\ (\psi,\phi_i) = 0,\ i = 1,2,\ldots,m.}} (\psi, J\psi)$$

Then

1. there are n eigenvalues above the top of the essential spectrum and $\lambda_n(J)$ is the nth eigenvalue.

or

2. λ_n is the top of the essential spectrum. In that case $\lambda_n = \lambda_{n+1} = \lambda_{n+2}\ldots$ and there are at most $n-1$ eigenvalues above λ_n.

Here $D(J)$ is the domain of J, and the scalar product and norm mentioned above are in the l_2 sense.

The definition of the essential spectrum $\sigma_{\text{ess}}(J)$ can be found in Reed and Simon,[11] p. 236. In most of the cases we are interested in $\sigma_{\text{ess}}(J) = \sigma_{\text{ac}}(J)$ the absolutely continuous part of the spectrum.

Corollary A1: Let

$$\tilde{J} = \begin{bmatrix} -1 & \frac{1}{2} & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & \frac{1}{2} & -1 & \frac{1}{2} & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

and

$$J^p = \begin{bmatrix} b(0) & a(1) - \frac{1}{2} & 0 & 0 & 0 & 0 \\ a(1) - \frac{1}{2} & b(1) & a(2) - \frac{1}{2} & 0 & 0 & 0 \\ 0 & a(2) - \frac{1}{2} & b(2) & a(3) - \frac{1}{2} & 0 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

(note that $\tilde{J}$ is negative and $\sigma_{\text{ess}}(J) = [-2,0]$). Assume $\lim_{n\to\infty} a(n) = \frac{1}{2}$ and $\lim_{n\to\infty} b(n) = 0$; then $\lambda_n(\tilde{J} + \beta J^p)$, $\beta \geqslant 0$ is a monotone nondecreasing continuous function of β. Furthermore if $\lambda_n(\tilde{J} + \beta J^p) > 0$ then it is a strictly increasing function of β.

Proof[10]: Since $\sigma_{\text{ess}}[\tilde{J} + \beta J^p] = [-2,0]$ Theorem A1 implies $\lambda_n(\tilde{J} + \beta J^p) \geqslant 0$ for all n. Therefore

$$\lambda_n = \min_{\phi_1,\phi_2,\ldots,\phi} \max_{\substack{\psi \in D(J) \cap D(J^p) \\ \|\psi\| = 1 \langle\psi,\phi_i\rangle = 0,\ i = 1,2,\ldots,n}} \max[0, \langle\psi, (\tilde{J} + \beta J^p)\psi\rangle].$$

Since $\tilde{J}$ is negative $\max\{0, \langle\psi, (\tilde{J} + BJ^p)\psi\rangle\}$ is either zero or an increasing straight line for fixed ψ. The result now follows by maximizing over the family of lines then minimizing over ϕ_i.

Let $N^\beta(0,\infty)$ be the number of eigenvalues of $\tilde{J} + \beta J^p$ in $(0,\infty)$ then for $\beta > +1, N^\beta(0,\infty) \geqslant N^1[0,\infty)$. This plus the theory of chain sequences imply $N^+[0,\infty) \geqslant N[0,\infty)$.

The techniques developed from the min–max theorem and from perturbation theory are used in obtaining approximations to the eigenvalues and eigenvectors of the Schrödinger operator. These techniques should also prove useful in obtaining approximations to the eigenvalues and eigenvectors of Jacobi matrices. This will be discussed at a later time.

[1] J. Favard, C. R. Acad. Sci. Paris **200**, 2052 (1935).
[2] O. Blumenthal, Inaugeral dissertation, Gottingen, 1898.
[3] P. G. Nevai, Mem. Am. Math. Soc. **18**, 1979.

[4]T. S. Chihara, Trans. Am. Math. Soc. **104**, 1 (1962).
[5]T. S. Chihara, *An Introduction to Orthogonal Polynomials* (Gordon and Breach, New York, 1978).
[6]T. S. Chihara, SIAM J. Math. Anal. **11**, 358 (1980).
[7]T. S. Chihara and P. G. Nevai, to appear in J. Approx. Theory.
[8]J. S. Geronimo and K. M. Case, Trans. Am. Math. Soc. **258**, 467 (1980).
[9]V. Bargmann, Proc. N.A.S. **38**, 961 (1952).
[10]M. Reed and B. Simon, *Methods of Mathematical Physics IV, Analysis of Operators*, 1st ed. (Academic, New York, 1978).
[11]M. Reed and B. Simon, *Methods of Mathematical Physics I, Functional Analysis*, 1st ed. (Academic, New York, 1972).
[12]J. S. Geronimo, in preparation.

Permutational properties of the generalized Clebsch–Gordan coefficients

R. Chatterjee and T. Lulek[a)]
Department of Physics, The University of Calgary, Calgary, AB T2N 1N4, Canada

(Received 22 September 1981; accepted for publication 6 November 1981)

The properties of the permutational symmetry of the $3\Gamma\gamma$ symbols for an arbitrary compact group have been generalized to a multiple Kronecker product of irreducible representations. It has been shown that these properties of the corresponding $N\Gamma\gamma$ symbols under a permutation of their columns are related to a choice of the coupling schemes for the appropriate polyads $\Gamma_1\cdots\Gamma_N$. The simplest permutational symmetry, i.e., associated with a Young diagram for the symmetric group Σ_N, is, in general, incompatible with schemes having definite intermediate representations and, in the case of mixed symmetry, does not preserve the absolute value of a $N\Gamma\gamma$ symbol. The case of $N = 4$ for SU(2) has been studied in detail.

PACS numbers: 02.20. + b

1. INTRODUCTION

Derome and Sharp[1] and Derome[2] generalized the Racah-Wigner angular momentum calculus for the group SU(2) to an arbitrary compact group G and provided a method for analyzing the permutational symmetry of Clebsch–Gordan coefficients. They showed that this symmetry has its origin in an appropriate choice of the system of repetition indices for the resultant representation Γ_3, which appears several times in a Kronecker product $\Gamma_1\times\Gamma_2$. In this paper an attempt has been made to generalize the Derome–Sharp approach to a multiple Kronecker product.

The permutational symmetry of the ordinary Clebsch–Gordan coefficients is usually displayed in the three-column $3\Gamma\gamma$ symbols [$3jm$ Wigner symbols for the case of SU(2)], where there is no distinction between the constituent representations Γ_1 and Γ_2, and the resultant Γ_3. The symmetry of $3\Gamma\gamma$ symbols under a permutation of their columns can be described by a set of unitary matrices which we shall refer to as the Derome–Sharp matrices. The simplest symmetry corresponds to the case when all Derome–Sharp matrices are diagonal, with eigenvalues ± 1. In general, Derome–Sharp matrices depend on both the permutation $\sigma(B\,|A\,)$ relating the $3\Gamma\gamma$ symbols for triads $\Gamma_a\Gamma_{a'}\Gamma_{a''}$ and $\Gamma_b\Gamma_{b'}\Gamma_{b''}$, and on the initial triad $\Gamma_a\Gamma_{a'}\Gamma_{a''}$, with [$aa'a''$] and [$bb'b''$] being any arrangements of the numbers [123].[1–4] A concise generalization of these notions to an arbitrary polyad $\Gamma_1\ldots\Gamma_N$ can be provided using a vocabulary of the theory of permutation representations (e.g., Curtis and Reiner,[5] Wielandt[6]), which is introduced in Sec. 2. In Sec. 3 we define the $N\Gamma\gamma$ symbols and determine their relationship to the generalized Clebsch–Gordan coefficients in an arbitrary coupling scheme. Then in Sec. 4 we provide a detailed discussion of the relations between the choice of a coupling scheme and the corresponding permutational properties of $N\Gamma\gamma$ symbols. In Sec. 5 we demonstrate the general theory for the case $N = 4$ identical representations $D^{(J)}$ of the group SU(2), and in Sec. 6 we discuss the terminology related to $N\Gamma\gamma$ and $n\Gamma$ symbols.

2. DEFINITION OF A PERMUTATION REPRESENTATION ON THE BLOCK OF POLYADS

Let $A = [a(1),\ldots,a(N)]$ be an arrangement (without repetitions) of integers $1,\ldots,N$, related to the initial arrangement $A_0 = [1,\ldots,N]$ through the permutation

$$\sigma_A = \begin{pmatrix} 1 & \ldots & N \\ a(1) & \ldots & a(N) \end{pmatrix} = \begin{pmatrix} \bar{a}(1) & & \bar{a}(N) \\ 1 & \ldots & N \end{pmatrix} = \begin{pmatrix} A_0 \\ A \end{pmatrix} = \begin{pmatrix} \bar{A} \\ A_0 \end{pmatrix} \quad (1)$$

and let

$$\Gamma^A = \Gamma_{a(1)}\cdots\Gamma_{a(N)}, \quad (2)$$

$$\gamma^A = \gamma_{a(1)}\cdots\gamma_{a(N)} \quad (3)$$

be the corresponding arrangement of representations of the initial polyad $\Gamma_1,\ldots,\Gamma_N$ and their basic functions $\gamma_1,\ldots,\gamma_N$. The set of all σ_A forms the symmetric group Σ_N, and the set $\{\Gamma^A;\ \sigma_A\in\Sigma_N\}$ of all polyads obtained from the initial one is a carrier block of a transitive permutation representation Π of this group, defined by

$$\Pi:\ \sigma_D\Gamma^A = \Gamma^B, \quad (4)$$

where the polyad Γ^B has to be formed from Γ^A in a way that the representation occupying the ith place in Γ^A, i.e., $\Gamma_{a(i)}$, is put under the action of σ_D into the $d\,(i)$th place, whereas the ith place is then occupied by $\Gamma_{a(\bar{d}(i))}$, i.e., the representation from the $\bar{d}\,(i)$th place, $i = 1,\ldots,N$. Consequently, we have

$$\sigma_D \equiv \sigma(B\,|A\,) = \sigma_B^{-1}\sigma_A. \quad (5)$$

In particular,

$$\Gamma^A = \sigma_A^{-1}\Gamma^{A_0}. \quad (6)$$

In general, not all the representations Γ_i of the polyad $\Gamma_1\ldots\Gamma_N$ have to be different. Let this polyad has μ_k identical representation $\Gamma^{(k)}$, $k = 1,2,\ldots,p$ so that

$$\Gamma^{(k)}\neq\Gamma^{(k')} \quad \text{for } k\neq k' \text{ and } N = \sum_{k=1}^{p}\mu_k. \quad (7)$$

Let

$$\Gamma^{A_0} = \underbrace{\Gamma^{(1)}\ldots\Gamma^{(1)}}_{\mu_1}\cdots\underbrace{\Gamma^{(p)}\ldots\Gamma^{(p)}}_{\mu_p}. \quad (8)$$

The stability group of the initial polyad Γ^{A_0} is

$$\Sigma(\Gamma^{A_0}) = \Sigma_{\mu_1}\times\ldots\times\Sigma_{\mu_p}, \quad (9)$$

[a)]Present address: Instytut Fizyki UAM, Matejki 48/49, 60-769 Poznań, Poland.

 0022-2488/82/060922-06$02.50

the simple product of symmetric groups acting on location labels of identical representations. Evidently,

$$\mathrm{Ker}\, \Pi = \bigcap_A \Sigma(\Gamma^A) = \sigma_{A_0}, \quad \text{for } p > 1 \tag{10}$$

so that Π is a faithful representation of Σ_N provided the polyad $\Gamma_1 \ldots \Gamma_N$ possesses at least two distinct representations.

Each distinct polyad Γ^F can be put into a one-to-one correspondence with the right cosets of the group Σ_N with respect to the stability subgroup $\Sigma(\Gamma^{A_0})$. Let

$$\Sigma_N = \bigcup_{\sigma_F \in \Xi} \Sigma(\Gamma^{A_0}) \sigma_F \tag{11}$$

be the decomposition of Σ_N into these right cosets, where Ξ is a set of the representatives. Then for any arrangement B we have

$$\sigma_B = \sigma_A \sigma_F, \quad \sigma_A \in \Sigma(\Gamma^{A_0}), \quad \sigma_F \in \Xi \tag{12}$$

so that any permutation of Σ_N can be uniquely expressed by those belonging to $\Sigma(\Gamma^{A_0})$ and Ξ. Evidently, the number of elements of $\Sigma(\Gamma^{A_0})$ is

$$|\Sigma(\Gamma^{A_0})| = \prod_{k=1}^{p} \mu_k! \tag{13}$$

and that of the set Ξ

$$|\Xi| = N! / \prod_{k=1}^{p} \mu_k!, \tag{14}$$

the latter being the number of different polyads generated from Γ^{A_0} by means of Eq. (4).

3. GENERALIZED CLEBSCH–GORDAN COEFFICIENTS AND $N\Gamma\gamma$ SYMBOLS

Let

$$|\Gamma^A \gamma^A\rangle = |\Gamma_{a(1)} \gamma_{a(1)}\rangle \cdots |\Gamma_{a(N)} \gamma_{a(N)}\rangle \tag{15}$$

be the uncoupled basis of the polyad Γ^A. Then the equation

$$|\Gamma^A \Gamma w \gamma\rangle_W = \sum_{\gamma_1 \cdots \gamma_N} \begin{bmatrix} \Gamma^A & \Gamma & w \\ \gamma^A & \gamma & \end{bmatrix}^W |\Gamma^A \gamma^A\rangle \tag{16}$$

defines the generalized Clebsch–Gordan coefficient (the symbol in rectangular brackets), associated with the coupling of N representation of the polyad Γ^A into the resultant representation Γ of G according to a coupling scheme W. Here, w stands for a set of repetition indices, distinguishing subspaces with identical Γ's. For $N = 2$, Eq. (16) is a definition of the ordinary Clebsch–Gordan coefficient.

A coupling scheme depending on consecutive couplings of pairs of representations into definite intermediate representations can be presented by means of a treelike diagram, introduced by Jucys and others,[7,8] with "branches," "boughs," and the "trunk" corresponding to constituent, intermediate, and resultant representations, respectively (cf. an example on Figs. 1(a) and 1(b), presenting the schemes $|\Gamma_1\Gamma_2(\Gamma_{12}w_{12}), \Gamma_3\Gamma_4(\Gamma_{34}\, w_{34}), \Gamma w_{12.34}\gamma\rangle$ and $|\Gamma_1\Gamma_3(\Gamma_{13}w_{13}), \Gamma_2\Gamma_4(\Gamma_{24}w_{24}), \Gamma w_{13,24}\gamma\rangle$, respectively). In the following we shall refer to such coupling schemes to as treelike ones.

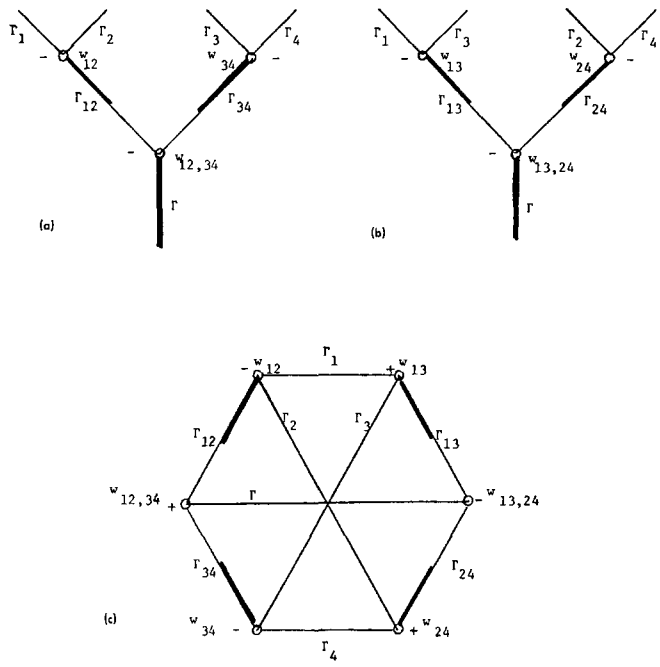


FIG. 1. The graphical representation of some coefficients related to a coupling of a tetrad $\Gamma_1\Gamma_2\Gamma_3\Gamma_4$, according to conventions of Jucys and Bandzaitis.[8] (a)—the generalized Clebsch-Gordan coefficient for the coupling scheme $|\Gamma_1\Gamma_2(\Gamma_{12}w_{12}), \Gamma_3\Gamma_4(\Gamma_{34}w_{34}), \Gamma w_{12.34}\gamma\rangle$, (b)—the corresponding coefficient for $|\Gamma_1\Gamma_3(\Gamma_{13}w_{13}), \Gamma_2\Gamma_4(\Gamma_{24}w_{24}), \Gamma w_{13,24}\gamma\rangle$, (c)—the related Racah recoupling matrix (associated with a 9Γ symbol).

The coupling scheme W does not necessarily have to be treelike, since one can use bases (16) with no definite values of intermediate representations, the repetition indices w being for example the labels of irreducible representations of the group Σ_N and of the corresponding basic functions.

The generalized Clebsch–Gordan coefficient, as defined by Eq. (16), can be substituted by a new symbol through the equation

$$\begin{bmatrix} \Gamma^A & \Gamma & w \\ \gamma^A & \gamma & \end{bmatrix}^W = [-1]^{\Gamma-\gamma}[\Gamma]^{1/2} \begin{pmatrix} \Gamma^A & \Gamma^* \\ \gamma^A & -\gamma \end{pmatrix}_w^{W^*}, \tag{17}$$

where $[-1]^{\Gamma-\gamma}$ is the element of the metric tensor, relating $|\Gamma\gamma\rangle$ to $|\Gamma^* - \gamma\rangle$, the basis of the complex conjugate to Γ, and $[\Gamma]$ is the dimension of Γ. The symbol in parentheses in Eq. (17) will be referred hereafter to as the $(N+1)\Gamma\gamma$ symbol, in a close analogy to the well known $3\Gamma\gamma$ symbol related to $N = 2$. All arguments $\Gamma_i\gamma_i$, $i = 1,\ldots,N+1$, with $\Gamma_{N+1} = \Gamma^*$, $\gamma_{N+1} = -\gamma$, enter into the $(N+1)\Gamma\gamma$ symbol on equal footing, without the distinction between the constituent and resultant representations.

It is sometimes convenient to interpret $N\Gamma\gamma$ symbols in terms of the coupling of N representation $\Gamma_1 \ldots \Gamma_N$ into the resultant unit representation Γ_0 of G, putting

$$\begin{bmatrix} \Gamma^{A_0} & \Gamma_0 & w \\ \gamma^{A_0} & \gamma_0 & \end{bmatrix}^{W'} = \begin{pmatrix} \Gamma^{A_0} & \Gamma_0 \\ \gamma^{A_0} & \gamma_0 \end{pmatrix}_w^{W'^*} = \begin{pmatrix} \Gamma^{A_0} \\ \gamma^{A_0} \end{pmatrix}_w^{W^*}, \tag{18}$$

where the first equality is a consequence of Eq. (17) and an assumption that $[-1]^{\Gamma_0 - \gamma_0} = 1$, whereas the second one defines the scheme W of coupling $\Gamma_1 \ldots \Gamma_{N-1}$ into the resul-

tant Γ_N^* in terms of the scheme W' of coupling $\Gamma_1 \ldots \Gamma_N$ into an arbitrary representation Γ. Such a definition allows for a natural extension of the notion of $N\Gamma\gamma$ symbols for $N = 2$ and 1, namely

$$\begin{pmatrix} \Gamma_1 & \Gamma_2 \\ \gamma_1 & \gamma_2 \end{pmatrix} \equiv \begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_0 \\ \gamma_1 & \gamma_2 & \gamma_0 \end{pmatrix} = \frac{[-1]^{\Gamma_1 - \gamma_1}}{\sqrt{[\Gamma_1]}} \delta_{\Gamma_2, \Gamma_1} * \delta_{\gamma_2, -\gamma_1}, \tag{19}$$

$$\begin{pmatrix} \Gamma \\ \gamma \end{pmatrix} \equiv \begin{pmatrix} \Gamma & \Gamma_0 & \Gamma_0 \\ \gamma & \gamma_0 & \gamma_0 \end{pmatrix} = \delta_{\Gamma\Gamma_0} \delta_{\gamma\gamma_0}. \tag{20}$$

Such quantities were already used by Jucys *et al.* (cf. Sec. 11 of Ref. 7 and Sec. 31 of Ref. 8) in their graphical methods of the angular momentum calculus, and are called $2jm$ and $1jm$ symbols. The $2\Gamma\gamma$ symbols (19) coincide, up to the factor $[\Gamma_1]^{-1/2}$, with the $1jm$ symbols as defined by Butler.[3]

4. THE PERMUTATIONAL SYMMETRY

The basis for the determination of symmetry of $N\Gamma\gamma$ symbols under a permutation of their columns is provided by the invariant summation over the group manifold G, which has the form

$$\frac{1}{|G|} \sum_{g \in G} D^{\Gamma_1}_{\gamma_1 \gamma_1'}(g) \ldots D^{\Gamma_N}_{\gamma_N \gamma_N'}(g) = \sum_{w=1}^{c(\Gamma_1 \ldots \Gamma_N)} \begin{pmatrix} \Gamma^{A_0} \\ \gamma^{A_0} \end{pmatrix}_w^{W*} \begin{pmatrix} \Gamma^{A_0} \\ \gamma'^{A_0} \end{pmatrix}_w^{W}, \tag{21}$$

where $|G|$ is the group volume (for continuous compact groups the summation has to be replaced by the invariant Haar integral), and $c(\Gamma_1 \ldots \Gamma_N)$ is the multiplicity of Γ_0 in the Kronecker product $\Gamma_1 \times \ldots \times \Gamma_N$. The left-hand side of Eq. (21) is manifestly invariant under any permutation of Σ_N, which implies the invariance of the right-hand side, that is the equality

$$\sum_w \begin{pmatrix} \Gamma^A \\ \gamma^A \end{pmatrix}_w^{W*} \times \begin{pmatrix} \Gamma^A \\ \gamma'^A \end{pmatrix}_w^{W} = \sum_x \begin{pmatrix} \Gamma^B \\ \gamma^B \end{pmatrix}_x^{X*} \times \begin{pmatrix} \Gamma^B \\ \gamma'^B \end{pmatrix}_x^{X} \tag{22}$$

for any polyad Γ^A and Γ^B, with arbitrary coupling schemes W and X. Equation (22) can be viewed as a scalar product of two vectors $\Gamma^A\gamma^A$ and $\Gamma^A\gamma'^A$ (or $\Gamma^B\gamma^B$ and $\Gamma^B\gamma'^B$) with components labeled by the repetition indices w (or x) in a $c(\Gamma_1 \ldots \Gamma_N)$-dimensional linear space. The invariance of the scalar product implies a unitary transformation of vector components, i.e.,

$$\begin{pmatrix} \Gamma^B \\ \gamma^B \end{pmatrix}_x^{X} = \sum_x m_{xw}^{XW}(\sigma(B|A), \Gamma^A) \begin{pmatrix} \Gamma^A \\ \gamma^A \end{pmatrix}_w^{W}, \tag{23}$$

where $m_{xw}^{XW}(\sigma(B|A), \Gamma^A)$, with $x, w = 1, \ldots, c(\Gamma_1 \ldots \Gamma_N)$, form a unitary matrix, referred hereafter to as the Derome–Sharp matrix. Equation (23) determines the permutational properties of $N\Gamma\gamma$ symbols (and, consequently, those of generalized Clebsch–Gordan coefficients). It was first given by Derome and Sharp[1] for the case $N = 3$. Evidently, the highest permutational symmetry is achieved when all Derome–Sharp matrices are $c(\Gamma_1 \ldots \Gamma_N)$ dimensional unit matrices.

Using the orthogonality relations for $N\Gamma\gamma$ symbols

$$\sum_{\gamma^A} \begin{pmatrix} \Gamma^A \\ \gamma^A \end{pmatrix}_w^{W*} \begin{pmatrix} \Gamma^A \\ \gamma^A \end{pmatrix}_{w'}^{W} = \delta_{ww'}, \tag{24}$$

we obtain from (23)

$$m_{xw}^{XW}(\sigma(B|A), \Gamma^A) = \sum_{\gamma_1 \ldots \gamma_N} \begin{pmatrix} \Gamma^B \\ \gamma^B \end{pmatrix}_x^{X} \begin{pmatrix} \Gamma^A \\ \gamma^A \end{pmatrix}_w^{W*}; \tag{25}$$

i.e., we explicitly express the elements of Derome–Sharp matrices in terms of $N\Gamma\gamma$ symbols.

A change of the coupling scheme $W \to U$ for the polyad Γ^A can be uniquely determined by a unitary transformation of the basis

$$|\Gamma^A \Gamma_0 u \gamma_0\rangle_{U'} = \sum_w T_{wu}^{WU}(A) |\Gamma^A \Gamma_0 w \gamma_0\rangle_{W'} \tag{26}$$

[cf. Eq. (18)], and the related transformation of $N\Gamma\gamma$ symbols

$$\begin{pmatrix} \Gamma^A \\ \gamma^A \end{pmatrix}_u^{U} = \sum_w T_{wu}^{WU}(A) \begin{pmatrix} \Gamma^A \\ \gamma^A \end{pmatrix}_w^{W}, \tag{27}$$

implying the following transformation of Derome–Sharp matrices

$$m^{YU}(\sigma(B|A), \Gamma^A) = T^{XY}(B) m^{XW}(\sigma(B|A), \Gamma^A) T^{WU}(A)^\dagger, \tag{28}$$

where the dagger denotes the Hermitian conjugate. In general, $T^{XY}(B) \neq T^{WU}(A)$, so a change of the coupling scheme is, in general, associated with nonunitary transformations of Derome–Sharp matrices.

It follows from Eq. (23) that

$$m(\sigma(C|A), \Gamma^A) = m(\sigma(C|B), \Gamma^B) m(\sigma(B|A, \Gamma^A) \tag{29}$$

for any A, B, C (the superscripts of coupling schemes are dropped). Hence not all Derome–Sharp matrices are mutually independent, but any of them can be expressed by matrices

$$M_B = m(\sigma_B^{-1}, \Gamma^{A_0}), \tag{30}$$

with B varying through all N arrangements, since Eq. (29) implies

$$m(\sigma(C|B), \Gamma^B) = M_C M_B^\dagger. \tag{31}$$

We shall refer hereafter to the M_B's (Eq. 30) as the fundamental matrices.

As it was shown by Derome[2] for $N = 3$, the relation between the degree of freedom for a choice of fundamental matrices, and a choice of coupling systems W for different arrangements Γ^A of the initial polyad Γ^{A_0}, depends on the number of identical representations in this polyad. Let Γ^{A_0} be given by Eq. (8). Then for any permutation σ_A of identical representations in Γ^{A_0} we have

$$\sigma_A \Gamma^{A_0} = \Gamma^{A_0}, \quad \sigma_A \in \Sigma(\Gamma^{A_0}) \tag{32}$$

so the corresponding fundamental matrix

$$M_A = \{m_{w'w}^{W_0 W_0}(\sigma_A^{-1}, \Gamma^{A_0}); w, w' (\text{ranging})\} \tag{33}$$

is, according to Eq. (25), uniquely determined by the coupling scheme W_0 for Γ^{A_0}.

The number of independent fundamental matrices is evidently equal to that of different polyads, since for each Γ^F one can choose arbitrarily the corresponding coupling scheme. Consequently, any fundamental matrix M_B, with σ_B given by Eq. (12), can be written in a form

$$M_B = M_F M_A, \quad \sigma_F \in \Xi, \quad \sigma_A \in \Sigma(\Gamma^{A_0}), \tag{34}$$

where M_A is determined by the coupling scheme W_0 for the

initial polyad Γ^{A_0} through Eqs. (29) and (33), and M_F is an arbitrary (unitary) matrix. The choice of M_F, $\sigma_F \in \Xi$ is equivalent to the choice of the coupling scheme W for the polyad Γ^F.

The largest degree of freedom in a choice of fundamental matrices corresponds to the case when all $\Gamma^{(k)}$ are distinct, i.e., when $\mu_k = 1$ for $k = 1,...,N$. One can then choose the coupling scheme W for each of $N!$ polyads separately, so all fundamental matrices M_F, with the exception of

$$M_{A_0} = I, \tag{35}$$

where I is the $c(\Gamma_1...\Gamma_N)$-dimensional unit matrix, are arbitrary. The number of all Derome–Sharp matrices is then $(N!)^2$.

Another extreme case is that of N identical representations, i.e., $\Gamma_i = \Gamma$ for $i = 1,...,N$. Now there is only one polyad $\Gamma...\Gamma$, so the total number of Derome–Sharp matrices $m(\sigma, \Gamma...\Gamma) \equiv m(\sigma)$ is $N!$. All these matrices are uniquely determined by the choice of the coupling scheme W, and form a $c(\Gamma...\Gamma)$-dimensional representation of the group Σ_N, i.e.,

$$m(\sigma\sigma') = m(\sigma)m(\sigma'). \tag{36}$$

This representation is, in general, reducible, and can be decomposed into irreducible representations $\{\lambda\}$. In general, the representation m of Σ_N contains multidimensional irreducible representations (i.e., $\{\lambda\}$ different from the completely symmetric representation $\{N\}$ or the antisymmetric $\{1^N\}$). According to formula (23), one cannot then choose such a coupling scheme W, for which the absolute value for the $N\Gamma\gamma$ symbols would be unchanged. In the next section we demonstrate it for the case of $4jm$ symbols for the polyad $jjjj$ of SU(2).

5. AN EXAMPLE: THE $4jm$ SYMBOLS FOR IDENTICAL REPRESENTATIONS

Now we consider the case $N = 4$, $G = \mathrm{SU}(2)$, and $p = 1$, that is, the coupling of four representations $D^{(j)}$ into the resultant $D^{(0)}$. Let us choose a treelike coupling scheme W' (Fig. 1) determined by the ket $|jj(J), jj(J), 00\rangle$. Then the $4jm$ symbols [i.e., the $4\Gamma\gamma$ symbols for SU(2)] can be expressed in terms of ordinary $3jm$ Wigner symbols as

$$\begin{pmatrix} j & j & j & j \\ m_1 & m_2 & m_3 & m_4 \end{pmatrix}^W_J = (-1)^{J-M}\sqrt{2J+1}\begin{pmatrix} j & j & J \\ m_1 & m_2 & -M \end{pmatrix}\begin{pmatrix} j & j & J \\ m_3 & m_4 & M \end{pmatrix}, \tag{37}$$

where $M = m_1 + m_2 = -m_3 - m_4$. The elements of the Derome–Sharp matrix are given by

$$\begin{aligned} m_{JJ'}(\sigma_B) \equiv m^{WW}_{JJ'}(\sigma_B, jjjj) = \sum_{m_1,...m_4,M,M'} (-1)^{J-M+J'-M'}\sqrt{(2J+1)(2J'+1)}\begin{pmatrix} j & j & j' \\ m_1 & m_2 & -m' \end{pmatrix} \\ \times\begin{pmatrix} J' & j & j \\ M' & m_3 & m_4 \end{pmatrix}\begin{pmatrix} j & j & j \\ m_{b(1)} & m_{b(2)} & -M \end{pmatrix}\begin{pmatrix} J & j & j \\ M & m_{b(3)} & m_{b(4)} \end{pmatrix} \end{aligned} \tag{38}$$

Let the group Σ_4 decompose as

$$\Sigma_4 = T \cup \sigma_7 T, \quad T = D_2 \cup \sigma_5 D_2 \cup \sigma_6 D_2, \quad D_2 = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}, \tag{39}$$

where

$$\sigma_1 = (1)(2)(3)(4), \quad \sigma_2 = (12)(34), \quad \sigma_3 = (13)(24), \quad \sigma_4 = (14)(23), \quad \sigma_5 = (1)(234), \quad \sigma_6 = (1)(243), \quad \sigma_7 = (1324), \tag{40}$$

and D_2 and T are normal subgroups of Σ_4, denoted by Schoenflies symbols for point groups according to the well-known isomorphism $\Sigma_4 \leftrightarrow 0$. Then we get from (38)

$$m_{JJ'}(\sigma_i) = (-1)^{2J}\delta_{JJ'} \quad \text{for } i = 1,2,3,4, \tag{41}$$

$$m_{JJ'}(\sigma_5) = m_{J'J}(\sigma_6) = (-1)^{J'}\sqrt{(2J+1)(2J'+1)}\begin{Bmatrix} jjJ' \\ jjJ \end{Bmatrix}, \tag{42}$$

$$m_{JJ'}(\sigma_7) = (-1)^{2j-J}\delta_{JJ'}, \tag{43}$$

where the expression in curly brackets in Eq. (42) is the standard $6j$ Wigner symbol.

It follows that the mapping $m : \sigma \to m(\sigma)$ $(\sigma \in \Sigma_4)$ is a homomorphism, with

$$\mathrm{Ker}\, m = D_2, \tag{44}$$

so that the Derome–Sharp matrices $m(\sigma)$ form in this case a faithful representation $m(\pi)$ of the quotient group

$$\Sigma_4/D_2 \cong \Sigma_3; \quad \pi \in \Sigma_3. \tag{45}$$

Let

$$\pi_2 = \pi_3^{-1} \equiv (123) = \mathrm{Im}\,\sigma_5 = \mathrm{Im}\,\sigma_6^{-1}, \quad \pi_4 \equiv (1)(23) = \mathrm{Im}\,\sigma_7 \tag{46}$$

by generators of the group Σ_3. Then the representation m^j of Σ_3 is generated by matrices

$$m^{1/2}(\pi_2) = \begin{array}{c|cc} J \backslash J' & 0 & 1 \\ \hline 0 & -1/2 & -\sqrt{3}/2 \\ 1 & \sqrt{3}/2 & -1/2 \end{array}; \quad m^{1/2}(\pi_4) = \begin{array}{c|cc} J \backslash J' & 0 & 1 \\ \hline 0 & -1 & 0 \\ 1 & 0 & 1 \end{array}, \tag{47}$$

$$m^1(\pi_2) = \begin{array}{c|ccc} J \backslash J' & 0 & 1 & 2 \\ \hline 0 & 1/3 & 1/\sqrt{3} & \sqrt{5/2} \\ 1 & -1/\sqrt{3} & -1/2 & \sqrt{5/2}\sqrt{3} \\ 2 & \sqrt{5/3} & -\sqrt{5/2}\sqrt{3} & 1/6 \end{array};$$

$$m^1(\pi_4) = \begin{array}{c|ccc} J \backslash J' & 0 & 1 & 2 \\ \hline 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 2 & 0 & 0 & 1 \end{array}, \tag{48}$$

and so on. Equations (47) and (48) are the simplest examples of Derome–Sharp matrices for $4jm$ symbols. It is evident that some of these matrices are essentially nondiagonal, for reason of appearance of the $6j$ symbol in Eq. (42), which is frequently different from zero for $J \neq J'$.

The representation m^j of Σ_3 is, in general, reducible. Its decomposition into irreducible representations $\{3\}$, $\{21\}$, $\{1^3\}$ of Σ_3 can be performed by the character theory. Equations (42) and (43) yield the following formulas for characters of Σ_3:

$$\chi^j(1^3) = 2j + 1, \tag{49}$$

$$\chi^j(12) = [1 + (-1)^{2j}]/2 = 1 \text{ or } 0 \text{ for } 2j + 1 = 1 \text{ or } 0 \bmod 2, \tag{50}$$

$$\chi^j(3) = (-i/\sqrt{3})(\omega^{2j+1} - \omega^{-2j-1}) = 1 \text{ or } -1 \text{ or } 0 \text{ for } 2j + 1 = 1 \text{ or } 2 \text{ or } 0 \bmod 3, \tag{51}$$

where $\omega = e^{2\pi i/3}$. We obtain therefore the decomposition

$$m^j = k \text{ reg} + m^\kappa, \tag{52}$$

where

$$2j + 1 = 6k + \kappa;\ 2\kappa = 0,1,2,3,4;\ k = 0,1,2,..., \tag{53}$$

$$m^0 = 0,\ m^{1/2} = \{21\},\ m^1 = \{3\} + \{21\},$$

$$m^{3/2} = \{3\} + \{1^3\} + \{21\},$$

$$m^2 = \{3\} + 2\{21\}, \tag{54}$$

and

$$\text{reg} = \{3\} + 2\{21\} + \{1^3\} \tag{55}$$

is the regular representation of Σ_3.

It follows from (52) that for each $j \neq 0$ the representation m^j contains the two-dimensional representation $\{21\}$ of Σ_3. It is therefore impossible to choose such a coupling scheme W that the absolute value of any $4jm$ symbol remains unchanged.

Properties of Derome–Sharp matrices are related through Eq. (42) to those of $6j$ symbols. That is the unitarity of $m^j(\sigma_5) = m^j(\pi_2)$ implies a particular case of a standard orthogonality relation for $6j$'s, and the group multiplication

$$m^j(\pi_2)m^j(\pi_2) = m^j(\pi_3) \tag{56}$$

leads to a particular case of the Racah backcoupling rule.

The permutational symmetry of the so called "irreducible j tensors" of SU(2) was already investigated by Vanagas and Batarunas[9] (cf. also Shelepin[10]). These quantities are related, in our notation, to $(N + 1)jm$ symbols associated with bases $|j...j\overline{J}\omega M\rangle$, i.e., irreducible bases for the carrier space for the representation $(D^j)^N$ with a fixed value of the resultant angular momentum $\overline{J}$. They gave the general formulas for characters of the corresponding representations $m^{j\overline{J}}$ of the group Σ_N, and decompositions of $m^{j\overline{J}}$ into irreducible representations for $N = 3,4,5$. Our $4jm$ symbols correspond to the case $N = 4, \overline{J} = 0$, or to $N = 3, \overline{J} = j$. Our formulas for characters [Eqs. (49)–(51)] become, after appropriate substitutions, the special cases of formulas of Vanagas and Batarunas.[9] It is worthwhile to note that the homomorphism $\Sigma_4 \to \Sigma_3$, defined by our Eqs. (44)–(45), holds only for $\overline{J} = 0$, and for $\overline{J} \neq 0$ the representation of Σ_4 by Derome–Sharp matrices $m^{j\overline{J}}$ becomes a faithful one.

6. DISCUSSION

We have shown that the permutational symmetry of the $3\Gamma\gamma$ symbol of any compact group G can be generalized to the case of $N\Gamma\gamma$ symbols that are related to a coupling of N irreducible representations $\Gamma_1...\Gamma_N$ into the resultant unit representation Γ_0. We have pointed out that the symmetry of $N\Gamma\gamma$ symbols under a permutation of their columns is immediately related to a choice of the coupling scheme W, and the degree of freedom in a choice of the corresponding Derome–Sharp matrices is strongly dependent on the number of identical representations for the polyad $\Gamma_1...\Gamma_N$.

The case of $N = 3$ has been thoroughly discussed by Derome,[2] who has shown that if the corresponding representation m of the group Σ_3 contains an irreducible representation $\{21\}$, corresponding to mixed permutational symmetry, then it is not possible to choose such a system of repetition indices for which the absolute value of any $3\Gamma\gamma$ symbol would remain invariant under any permutation of its columns. However, there exist numerous simple cases for which the representation $\{21\}$ does not appear, so that one can choose symmetrical systems of repetition indices. Such cases are usually referred to as simple phase groups (Van Zanten and de Vries,[11] Butler[3]), or simple phase triads of representations (Butler and King[12]). As we have shown in the present paper for a relatively simple case of $4jm$ symbols for SU(2), there do not exist analogical simple phase tetrads $jjjj$ for $j \neq 0$, so that the full permutational symmetry of $N\Gamma\gamma$-symbols is, for $N > 3$ identical representations Γ with $[\Gamma] > 1$, an exceptional case. In general, it can be kept only for one-dimensional representations $\{N\}$ and $\{1^N\}$, when the coupling scheme is chosen in such a way that the repetition indices w provide a decomposition of the representation m of Σ_N composed of Derome–Sharp matrices into irreducible representations of Σ_N.

In general, for the polyad $\Gamma...\Gamma$ of N identical representations Γ, the Derome–Sharp matrices form a representation m of Σ_N, with the decomposition

$$m = \sum_{\{\lambda\}} f(m, \{\lambda\})\{\lambda\} \tag{57}$$

into irreducible representations $\{\lambda\}$, with multiplicities $f(m, \{\lambda\})$, and the sum ranging over all partitions $\{\lambda\}$ of N into $[\Gamma]$ or less parts. It is evident that the simple phase $N\Gamma\gamma$ coefficients occur only for the one-dimensional representations $\{\lambda\} = \{N\}$ and $\{1^N\}$, when the assumed system of repetition indices w provides an irreducible basis for the representation m [or, at least, for its symmetric and antisymmetric part, $f(m, \{N\})\{N\}$ and $f(m, \{1^N\})\{1^N\}$, respectively]. Therefore, the notion of simple phase groups and (or) simple phase polyads has to be constrained, for $N > 3$, to simple phase symmetric and antisymmetric part of the polyad $\Gamma...\Gamma$. Any part related to a multidimensional represen-

tation $\{\lambda\}$ of Σ_N is associated with $N\Gamma\gamma$ coefficients, which unavoidably change their absolute values under some permutations of Σ_N.

It is worthwhile to make some comments on the terminology related to several coefficients of Racah algebra, in a light of the permutational symmetry. In this paper we have used the notion of $N\Gamma\gamma$ symbols, which is a natural generalization of that for $3jm$ Wigner symbols, with N being the number of pairs of arguments (Γ_i, γ_i). The presence of γ [or m for SU(2)] points out the dependence of the symbols on a choice of bases of irreducible representations (in this context the frequently used name "$3j$" for the $3jm$ symbol is inconsistent with a more general notion of $3nj$—cf. Butler[3]). The $N\Gamma\gamma$ symbol does not always depend on any intermediate representation, and even if it does (for treelike coupling schemes), it still does not depend on the basis functions of the intermediate representations. The name "$N\Gamma\gamma$" does not therefore involve the coupling scheme, which has to be specified in each particular case.

From the permutational symmetry point of view and according to Butler's[3] suggestions, the Derome–Sharp matrices should be called $N\Gamma$ symbols, since they do not depend on bases of irreducible representations. So for $N = 1$ it would be $\delta_{\Gamma\Gamma_0}$, for $N = 2$—the "2Γ-phase" $[-1]^{2\Gamma}$, related to the symmetry of the metric tensor $[-1]^{\Gamma-\gamma}$, and for $N = 3$—the ordinary Derome–Sharp matrix. The names "$N\Gamma$" and "$N\Gamma\gamma$" involve therefore only the constituent and resultant representations, omitting the coupling schemes, and in particular, any intermediate representations. Unfortunately, such terminology is somehow confusing from the point of view of Racah algebra and graphical methods of angular momentum theory, with the already established notion of $3nj$ symbols, $n = 2, 3, \ldots$. These methods are based on treelike coupling schemes, and the transformation between two such schemes is realized by an appropriate Racah recoupling matrix (or its symmetrical version—a $3n\Gamma$ symbol), the diagrams of the latter being obtained by the connection of free ends of lines of representations with the same label in the corresponding treelike diagrams.[7,8] Figures 1a and 1b present two treelike coupling schemes for the tetrad $\Gamma_1\Gamma_2\Gamma_3\Gamma_4$, and Fig. 1c is the diagram of the corresponding Racah recoupling matrix. Each of the schemes 1(a) and 1(b) is associated with a set of $5\Gamma\gamma$ symbols [the four constituent representations $\Gamma_1\Gamma_2\Gamma_3\Gamma_4$ and the resultant Γ—cf. Eq. (18)], and the corresponding 5Γ symbol describing the transformation between these sets of $5\Gamma\gamma$'s can be expressed in terms of a 9Γ Wigner symbol, the "9" being the total number of lines in Fig. 1(c), including intermediate representations. In general, the Derome–Sharp matrix for a treelike coupling scheme associated with an $N\Gamma\gamma$ symbol is proportional to a Racah recoupling matrix associated with a $3n\Gamma$ symbol, $n \leqslant N-2$, and depending transparently on $3n$ representations: $(n+1)$ constituent, $2(n-1)$ intermediate, and 1 resultant one. The equality $n = N-2$ holds for nontrivial recoupling matrices, i.e., for matrices presented by such diagrams which cannot be divided into two separate parts by a dissection of one, two, or three lines.

[1] J. R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[2] J. R. Derome, J. Math. Phys. 7, 612 (1966).
[3] P. H. Butler, Philos. Trans. Soc. London **A227**, 545 (1975).
[4] B. Lulek and T. Lulek, Acta Phys. Pol. **A54**, 561 (1978).
[5] C. W. Curtis and J. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Interscience, New York, 1962).
[6] H. Wielandt, *Finite Permutation Groups* (Academic, New York, 1968).
[7] A. P. Jucys (Yutsis), J. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum*, (National Publisher for Political and Scientific Literature of Lithuanian SSR, Vilinius, 1960)(In Russian; the English translation by Israel Program for Scientific Translations, Jerusalem, 1962).
[8] A. P. Jucys (Yutsis) and A. A. Bandzaitis, *The Theory of Angular Momentum in Quantum Mechanics* (Mintis, Vilinius, 1965) (in Russian).
[9] V. V. Vanagas and J. W. Batarunas, Liet. Fiz. Rinkinys **1**, 9 (1961) (in Russian).
[10] L. A. Shelepin, Proc. (Trudy) of the P.N. Lebedev Physics Inst., **70**, 1 (in Russian); English translation by Consultants Bureau (Plenum, New York, 1975).
[11] A. J. Van Zanten and E. de Vries, J. Algebra **25**, 475 (1973).
[12] P. H. Butler and R. C. King, Can. J. Math. **26**, 328 (1974).

A new approach to permutation group representation

Jin-Quan Chen
Physics Department, Nanjing University, [a] *Nanjing, People's Republic of China*
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

Mei-Juan Gao
Physics Department, Nanjing University, Nanjing, People's Republic of China

(Received 28 January 1981; accepted for publication 24 April 1981)

The new approach to group representation theory (Ref. 1) is applied to the treatment of the permutation group. It is shown that obtaining (1) the primary characters and the fractional parentage coefficients, (2) the Yamanouchi bases and the Clebsch–Gordan coefficients, and (3) the irreducible matrix elements of the permutation group $S(f)$, are all simplified to a unified procedure — diagonalizing a certain operator in the corresponding representation. The operator to be diagonalized for the above three problems is (1) the 2-cycle class operator C_f of $S(f)$, (2) an appropriate linear combination of the f-1 2-cycle class operators $C_f, C_{f-1},...,C_2$ of the group chain $S(f) \supset S(f-1) \supset ... \supset S(2)$, and (3) an appropriate linear combination of $2f-3$ 2-cycle class operators $C_f, C_{f-1},...,C_2, \mathscr{C}_{f-1},...,\mathscr{C}_2, \mathscr{C}_i$ being the 2-cycle class operator of the subgroup $\mathscr{S}(i)$ of the state permutation group $\mathscr{S}(f)$, respectively. This method, the eigenfunction method, is simpler in concept, yet more powerful in practical calculations.

PACS numbers: 02.20. + b

1. INTRODUCTION

In Ref. 1 we put forward a new approach to group representation theory. The purposes of this new approach are three fold: (1) to give a simple and unified method for calculating the characters (the term character used in this paper always refers to the primary character), the irreducible bases, the Clebsch–Gordan coefficients, the isoscalar factors, etc., (2) to give a unified treatment for finite groups, compact Lie groups and space groups, and (3) by introducing the concept of the complete set of commuting operators (CSCO) into group theory and using the CSCO approach of quantum mechanics, to make the group representation theory more accessible to physicists. In this paper we want to apply this approach to the particular case of the permutation group.

The permutation group has important application in the physics of many-particle systems. The importance lies in the facts that a system of identical particles has permutation symmetry and that there exist many deep and delicate interrelations between the permutation group and the unitary group. The representation theory of the permutation group is well established through the efforts of Young, Frobenious, and Yamanouchi *et al.*[2,3] This theory has many advantages. For example, it gives the branching law for reducing the irreducible representations (hitherto referred to as irrep) of $S(f)$ into those of $S(f-1)$, a simple method for constructing the irreducible matrix elements of the neighboring permutations, and the intuitive and elegent way of labelling the irreps and the irreducible bases by the Young diagrams and the Young tableaux, etc.

However, from the practical point of view, this theory has some serious drawbacks. (1) It is too difficult for physicists to grasp quickly. Soklov[4] pointed out: "Due to the fact that the group theory, especially the theory of representations and characters of the permutation group, is extremely difficult even for the specialists, there occured in the history the tendency against the use of the so called "group pest" in quantum mechanics." (2) When it comes to actual calculation, it is rather tedious to obtain the characters, the Yamanouchi bases, the Clebsch–Gordan coefficients, etc.

The new approach to the permutation group has the advantage of being concise in theory and easily manageable in practice. A process of reducing a representation into the irreducible representations corresponds to a transformation from the nondiagonal representation to the diagonal representation of the CSCO. Thus the calculation of the characters, the irreducible bases, the reduction rules, the Clebsch–Gordan coefficients, the coefficients of fractional parentage (cfp), etc., are all reduced to a single recipe: finding the eigenfunction of a certain kind of CSCO which consists of only one operator.

The new approach to the permutation group is an independent theory in the sense that we can obtain, through straightforward, standard, and easily programmable calculations, all the results without using the traditional Young–Yamanouchi theory. Nevertheless, it has it own drawbacks, namely, we cannot get general conclusions about the dimensionality of irreps, the branching law, the irreducible matrix elements, etc., prior to concrete calculations.

From the above discussion it is seen that the advantages and the disadvantages of the Young–Yamanouchi theory and the new approach of the permutation group are complementary; one's disadvantage is just the other's advantage. Thus, from the practical point of view, we can combine these two to give a powerful method for handling the representation of the permutation group. This is the main purpose of this paper.

For easier accessibility and to establish notation, we present some important theorems given in Ref. 1(a) but omit,

[a] Permanent address.

 0022-2488/82/060928-16$02.50

except for a few special cases, the proofs. The whole content is divided into eleven sections:

1. Introduction.
2. CSCO-I and characters.
3. CSCO-II and the Yamanouchi bases.
4. The Clebsch–Gordan coefficients.
5. The coefficients of fractional parentage.
6. The intrinsic group.
7. CSCO-III and irreducible matrix elements.
8. Reduction of nonregular representations.
9. The state permutation group.
10. The quasistandard basis of the permutation group.
11. Conclusion.

2. CSCO-I AND CHARACTERS

Since the concept of the class operator is crucial to the developing of this new approach, we begin with a brief review of the definition and the properties of the class operators and class space.

A. Group space

Let $\{R_a\} = R_1, R_2,...,R_g$ or $R, S, T,..$ be the elements of a finite group G of order g. The g dimensional linear vector space spanned by the g elements R_a is called the group space V_G. The metric tensor in the group space is defined as

$$\langle R_a | R_b \rangle = \delta_{ab}. \tag{1}$$

B. Class operators and class space

For a finite group G with N classes, we have N class operators. The class operator C_i is equal to the sum of all the group elements belonging to the ith class,

$$C_i = \sum_{l=1}^{g_i} R_l^{(i)}, \tag{2}$$

with g_i the number of elements in the class i.

The N dimensional linear vector space spanned by the N class operators $C_1, C_2,...,C_N$ is called the class space.

From Eqs. (1) and (2), the metric tensor in the class space is seen to be

$$\langle C_i | C_j \rangle = g_i \delta_{ij}. \tag{3}$$

The class operators have three important properties[5]:

(a) They commute with each other,

$$[C_i, C_j] = 0. \tag{4}$$

(b) They commute with any element of the group G,

$$[C_i, R] = 0. \tag{5}$$

(c) They are closed under multiplication,

$$C_i C_j = \sum_k C_{ij}^k C_k. \tag{6}$$

The coefficients C_{ij}^k are called the structure constant of the finite group G. Therefore, the N class operators constitute an algebra — the class algebra.

From Eq. (6) we know that the N "basis vectors" $C_1, C_2,...,C_N$ of the class space carry a representation of the class algebra,

$$C_i C_j = \sum_k \mathscr{D}_{kj}(C_i) C_k. \tag{7a}$$

The matrix representative $\mathscr{D}(C_i)$ of the class operator C_i in the class space is equal to the structure constants,

$$\mathscr{D}_{kj}(C_i) = C_{ij}^k. \tag{7b}$$

C. CSCO-I

Definition 1: A set of l operators $C_1, C_2,...,C_l$ picked out of the N class operators of a group G is called the CSCO-I of G, or simply the CSCO of G, if the set is a complete set of commuting operators in the class space, denoted as

$$C = (C_1, C_2,...,C_l). \tag{8}$$

We use the term complete set of commuting operators (CSCO) in the same sense as used by Dirac,[6] i.e., a set of commuting operators $C = (C_1, C_2...)$ is called a CSCO of a space if, in that space, all the eigenvalues of C are nondegenerate. It must be emphasized that a CSCO is related to a particular space.

In Ref. 1(b) it was shown that the CSCO defined above is a generalization of the Casimir operators of Lie groups.

It can be proved that the set of the N class operators of G is certainly a CSCO in the N-dimensional class space. The next question is can we constuct a CSCO of G out of fewer, say l ($l < N$), class operators? From the practical point of view, we wish the number of operators contained in the CSCO of G as small as possible. The CSCO of G can be found in the following way.

Pick up some class operator C_i, and seek its "eigenvector" Q in the class space. Any "vector" in the class space can be expressed as a linear combination of the basis vectors, therefore

$$Q = \sum_j q_j C_j, \tag{9a}$$

$$C_i Q = \lambda_i Q, \tag{10a}$$

where q_j are coefficients and λ_i are eigenvalues. With the help of Eq. (6), the eigenequation (10a) becomes

$$\sum_j (C_{ij}^k - \lambda_i \delta_{jk}) q_j = 0. \tag{11a}$$

From

$$\det \| C_{ij}^k - \lambda_i \delta_{jk} \| = \prod_\nu (\lambda_i - \lambda_i^{(\nu)})^{m_\nu} = 0, \tag{11b}$$

we can obtain the eigenvalues $\lambda_i^{(\nu)}$ and the corresponding degeneracy m_ν. If all the eigenvalues $\lambda_i^{(\nu)}$ are nondegenerate, i.e., $m_\nu = 1$, for $\nu = 1,2,...,N$, then the single class operator C_i constitute the CSCO of G. It is easy to show that the eigenvectors $Q^{(\nu)}$ of this class operator C_i are also the simultaneous eigenvectors of all the class operators.

If there is any degeneracy in some eigenvalues $\lambda_i^{(\nu)}$, then C_i is not a CSCO of G. We have to add another class operator, for example C_j, and seek the simultaneous eigenvectors of C_i and C_j.

$$\begin{pmatrix} C_i \\ C_j \end{pmatrix} Q = \begin{pmatrix} \lambda_i \\ \lambda_j \end{pmatrix} Q. \tag{10b}$$

If all the sets of the eigenvalues (λ_i, λ_j) are nondegenerate, then (C_i, C_j) will be the CSCO of G. Otherwise we have to add more class operators, until all the degeneracies of the set of eigenvalues are reduced to one. The choice of the CSCO of G is, of course, not unique.

For finite groups with known characters, it is very easy to find their CSCO without any need of the above procedure.[7] We have already found the CSCO for commonly used finite groups [see Tables 1 and 2 of Ref. 1(a)]. The number l of the operators contained in the CSCO of those groups is equal to one, two, or at most three. In other words, l is much less than the class number N. It is precisely this fact that makes the eigenfunction method valuable in practical calculation. The CSCO for the permutation group $S(f)$ is equal to

$$C(f) = C_{(2)}(f) = C_f = \sum_{i<j}^{f}(ij) \quad \text{for } f \leqslant 5 \text{ and } f = 7,$$
$$C(f) = (C_{(2)}(f), C_{(3)}(f)) \quad \text{for } f = 6 \text{ and } 8 \leqslant f \leqslant 14,$$
$$C_{(3)}(f) = \sum_{i<j<k}^{f} [(ijk) + (ikj)], \tag{12}$$

$C_i\,(f)$ being the i-cycle class operator of $S(f)$.

D. Theorem 1

In the class space the eigenoperator $Q^{(\nu)}$ of the CSCO of G is essentially the projection operator on to the irrep (ν) of G.

Proof: Attach superscripts to the symbols Q and q in Eq. (9a),

$$Q^{(\nu)} = \sum_{j=1}^{N} q_j^{(\nu)} C_j. \tag{9b}$$

The eigenvector $Q^{(\nu)}$ of the CSCO satisfies the following set of eigenequations:

$$C_i Q^{(\nu)} = \lambda_i^{(\nu)} Q^{(\nu)}, \quad i = 1,2,...,l. \tag{10c}$$

Equation (10c) can be written more concisely as

$$CQ^{(\nu)} = \lambda^{(\nu)} Q^{(\nu)}, \tag{10d}$$

with

$$\lambda^{(\nu)} = (\lambda_1^{(\nu)}, \lambda_2^{(\nu)}, ..., \lambda_l^{(\nu)}). \tag{13}$$

$q_j^{(\nu)}$ satisfy a set of linear algebraic equations:

$$\sum_{j=1}^{N} (C_{ij}^{k} - \lambda_i^{(\nu)} \delta_{jk}) q_j^{(\nu)} = 0, \quad k = 1,..N, \quad i = 1,2,...l, \tag{14}$$

or written in the form of matrix equations,

$$\mathscr{D}(C_i)\mathbf{q}^{(\nu)} = \lambda_i^{(\nu)} \mathbf{q}^{(\nu)}, \tag{15a}$$

$$\mathbf{q}^{(\nu)} = \begin{pmatrix} q_1^{(\nu)} \\ \cdot \\ \cdot \\ \cdot \\ q_N^{(\nu)} \end{pmatrix}. \tag{15b}$$

From quantum mechanics we know that the simultaneous eigenvectors $q^{(\nu)}$ of a CSCO must satisfy the orthonormality and completeness condition with the metric tensor g_i as weights:

$$\sum_i g_i q_i^{(\nu)*} q_i^{(\mu)} = \delta_{\nu\mu}, \tag{16a}$$

$$\sum_\nu g_i q_i^{(\nu)*} q_j^{(\nu)} = \delta_{ij}. \tag{16b}$$

Using Eqs. (4), (9b), and (10d) we have

$$CQ^{(\nu)}Q^{(\mu)} = \lambda^{(\nu)} Q^{(\nu)} Q^{(\mu)} = \lambda^{(\mu)} Q^{(\nu)} Q^{(\mu)}. \tag{17}$$

Equation (17) shows that $Q^{(\nu)}Q^{(\mu)}$ is an eigenoperator of C with eigenvalue $\lambda^{(\nu)}$ or $\lambda^{(\mu)}$. According to the definition of the CSCO of G, the eigenvalues of C are nondegenerate, therefore

$$Q^{(\nu)}Q^{(\mu)} = \delta_{\nu\mu}\eta_\nu Q^{(\nu)}, \tag{18}$$

where η_ν is a constant depending only on ν. Letting

$$P^{(\nu)} = \eta_\nu^{-1} Q^{(\nu)}, \tag{19}$$

we have

$$P^{(\nu)}P^{(\mu)} = \delta_{\nu\mu} P^{(\nu)}. \tag{20}$$

In other words $P^{(\nu)}$ are idempotents.
From (16b), we get the inverse expansion of Eq. (9b),

$$C_i = \sum_{\nu=1}^{N} g_i q_i^{(\nu)*} Q^{(\nu)}. \tag{21}$$

Multiplying Eq. (21) from right with $Q^{(\mu)}$ and usings Eqs. (10c) and (18),

$$\lambda_i^{(\nu)} = \eta_\nu g_i q_i^{(\nu)*}. \tag{22}$$

Combining Eqs. (19), (21), and (22),

$$C_i = \sum_{\nu=1}^{N} \lambda_i^{(\nu)} P^{(\nu)}. \tag{23}$$

Letting $C_i = e$ (identity) and noting that its eigenvalue $\lambda_e^{(\nu)} \equiv 1$, we finally get

$$e = \sum_{\nu=1}^{N} P^{(\nu)}. \tag{24}$$

This is the decomposition theorem for the identity element of the group G. Thus the operator $P^{(\nu)}$ is the projection operator on to the irrep (ν) of G, and Theorem 1 is proved.

E. Theorem 2

The relations between the eigenvector $q_i^{(\nu)}$, the eigenvalue $\lambda_i^{(\nu)}$, and the character $\chi_i^{(\nu)}$ can be shown to be[7]

$$\chi_i^{(\nu)} = (\sqrt{g}) q_i^{(\nu)*}, \quad \chi_i^{(\nu)} = (h_\nu / g_i)\lambda_i^{(\nu)}. \tag{25}$$

Therefore, the orthonomality and completeness conditions, Eq. (16), are just the first and second orthogonality theorem of the characters

$$\sum_i (g_i/g)\chi_i^{(\nu)*}\chi_i^{(\mu)} = \delta_{\nu\mu}, \tag{26a}$$

$$\sum_\nu (g_i/g)\chi_i^{(\nu)*}\chi_j^{(\nu)} = \delta_{ij}. \tag{26b}$$

From Eqs. (15) and (25) one has

Theorem 2: The eigenvectors of the CSCO of G in the class space are proportional to the complex conjugate of the

characters:

$$\mathscr{D}(C_i)\chi^{(\nu)*} = \lambda_i^{(\nu)}\chi^{(\nu)*}, \quad i = 1,2,...,l, \tag{27}$$

$$\chi^{(\nu)} = \begin{pmatrix} \chi_1^{(\nu)} \\ \cdot \\ \cdot \\ \cdot \\ \chi_N^{(\nu)} \end{pmatrix}.$$

Therefore a simultaneous diagonalization of the l matrices $\mathscr{D}(C_1)...\mathscr{D}(C_l)$ with the normalization condition (26a) gives all the characters.

F. The labeling of inequivalent irreps

The CSCO of G must have N different sets of eigenvalues $\lambda^{(\nu)}$, $\nu = 1,2,...,N$, since it is a CSCO in the class space. Corresponding to each $\lambda^{(\nu)}$, there is only a unique projection operator $P^{(\nu)}$. Therefor, one can use the eigenvalue ν to label the nonequivalent irreps.

For the permutation group $S(f)$, the relation between the new label $\lambda^{(\nu)} = (\lambda_{(2)}^{(\nu)}, \lambda_{(3)}^{(\nu)})$ and the old label, the partition $[\nu] = [\nu_1, \nu_2,...\nu_f]$, is[8]

$$\lambda_{(2)}^{(\nu)} = \frac{f}{2} + \tfrac{1}{2}\sum_l \nu_l(\nu_l - 2l),$$

$$\lambda_{(3)}^{(\nu)} = \frac{1}{3}\left(2f - \frac{3}{2}f^2 + \sum_l \nu_l\left[\nu_l^2 - (3l - \frac{3}{2})\nu_l + 3l(l-1)\right]\right). \tag{28}$$

The relations between the eigenvalues of the conjugated irreps (interchange rows and columns in Young diagrams) are

$$\lambda_{(2)}^{[\tilde{\nu}]} = -\lambda_{(2)}^{[\nu]}, \quad \lambda_{(3)}^{[\tilde{\nu}]} = \lambda_{(3)}^{[\nu]}. \tag{29}$$

For self-conjugate representations $[\nu] = [\tilde{\nu}]$, one has

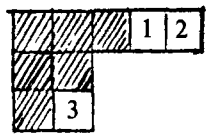
$$\lambda_{(2)}^{[\nu]} = 0. \tag{30}$$

The eigenvalue $\lambda_{(2)}^{[\nu]}$ has a simple meaning[9]: it equals the difference between the numbers of the symmetric bonds and antisymmetric bonds in any irreducible basis vector of $[\nu]$,

$$\lambda_{(2)}^{[\nu]} = n_s - n_a. \tag{31}$$

$\lambda_{(2)}^{[\nu]}$ can be calculated directly in the following way[9]: One ascribes symmetric bonds to all the boxes in the same row of a Young diagram $Y^{[\nu]}$ and antisymmetric bonds to boxes in the same column, while boxes that are neither in the same row nor the same column are counted as half-symmetric and half-antisymmetric, and therefore do not contribute to $\lambda_{(2)}^{[\nu]}$; the self-conjugate diagram in $Y^{[\nu]}$ also does not contribute to $\lambda_2^{[\nu]}$.

Example:



$$\lambda_{(2)}^{[522]} = 0 + 3 + 4 + 1 - 2 = 6.$$

The hatched diagram contributes zero, adding successively box 1 and box 2 gives three and four symmetric bonds, while adding box 3 gives 1 symmetric and 2 antisymmetric bonds.

In Table I, we list the new and old labels for those irreps, whose row length is greater or equal to the column length, of the permutation group $S(2) - S(8)$. $\lambda_{(3)}^{[\nu]}$ are listed only when $\lambda_{(2)}^{[\nu]}$ are degenerate.

G. Another form of the CSCO

From relation (28), we can easily find a suitable linear combination of the 2- and 3-cycle class operator

$$C(f) = a_f C_{(2)}(f) + b_f C_{(3)}(f), \tag{32}$$

where a_f and b_f are coefficients, such that the single operator $C(f)$ has N different eigenvalues

$$\lambda^{(\nu)} = a_f\lambda_{(2)}^{(\nu)} + b_f\lambda_{(3)}^{(\nu)}, \quad \nu = 1,2,...N. \tag{33}$$

Therefore, the single operator $C(f)$ of Eq. (32) constitute the CSCO of $S(f)$ with $f \leqslant 14$. For example, $S(6)$ has $N = 11$ classes. Using relation (28), we obtained the eigenvalues $\lambda_{(2)}^{(\nu)}$ and $\lambda_{(3)}^{(\nu)}$ listed in Table II. The single operator

$$C(6) = C_{(2)}(6) + 3C_{(2)}(6) \tag{34}$$

has 11 different eigenvalues, as can be seen from Table II. Thus $C(6)$ is the CSCO of $S(6)$. Obviously, there are infinite ways of contructing such a linear combination.

TABLE I. The new and old labeling of IRREPS of the permutation group $S(f)$ with $f \leqslant 8$.

	S_2	S_3		S_4			S_5				
$[\nu]$	[2]	[3]	[21]	[4]	[31]	[22]	[5]	[41]	[32]	[31²]	
$\lambda_{(2)}$	1	3	0	6	2	0	10	5	2	0	

	S_6						S_7							
$[\nu]$	[6]	[51]	[42]	[41²]	[33]	[321]	[7]	[61]	[52]	[51²]	[43]	[421]	[3²1]	[41³]
$\lambda_{(2)}$	15	9	5	3	3	0	21	14	9	7	6	3	1	0
$\lambda_{(3)}$				4	−8									

	S_8												
$[\nu]$	[8]	[71]	[62]	[61²]	[53]	[521]	[51³]	[4²]	[431]	[42²]	[4211]	[3²2]	
$\lambda_{(2)}$	28	20	14	12	10	7	4	8	4	2	0	0	
$\lambda_{(3)}$							16		−8		0	−16	

TABLE II. The eigenvalues of two kinds of CSCO-I, $(C_{(2)}, C_{(3)})$ and $3C_{(3)} + C_{(2)}$, of $S(6)$.

λ \ $[\nu]$	[6]	[51]	[42]	[411]	[33]	[321]	[222]	[31^3]	[2^21^2]	[21^4]	[1^6]
$\lambda_{(2)}$	15	9	5	3	3	0	−3	−3	−5	−9	−15
$\lambda_{(3)}$	40	16	0	4	−8	−5	−8	4	0	16	40
$3\lambda_{(3)} + \lambda_{(2)}$	135	57	5	15	−21	−15	−27	9	−5	39	105

From now on, the CSCO of $S(f)$ can be understood as either a set of class operators of Eq. (8) or a single operator of Eq. (32). Likewise, the eigenvalue $\lambda^{(\nu)}$ can be understood as either a set of eigenvalues or just a single eigenvalue of Eq. (33).

H. Calculation of characters

As mentioned before, to obtain the charaters one has to diagonalize simultaneously the matrix representatives of the class operators $(C_{(2)}, C_{(3)},...)$, which constitute the CSCO of $S(f)$. However, if we use the known result of the dimensionality of the irreps from the traditional theory of the permutation group, we can still obtain the characters by diagonalizing only the matrix $\mathscr{D}(C_{(2)})$ of the two-cycle class operator. It can be done in the following way.

Supposing the eigenvalue $\lambda^{(\nu)}$ has twofold degeneracy, in other words there are two partitions, say $[\nu]$ and $[\nu']$, corresponding to the same eigenvalue $\lambda^{(\nu)}$. For example, $[\nu] = [411]$ and $[33]$ correspond to the same eigenvalue $\lambda^{(\nu)}_{(2)} = 3$. For this eigenvalue $\lambda^{(\nu)}$, one gets two independent solutions from Eq. (14), which can be orthonormalized according to (16a). Let these two orthonormalized eigenvector be $\mathbf{q}^{(\nu)}$ and $\mathbf{q}^{(\nu)'}$. According to Eq. (25), the character $\chi^{[\nu]}$ can be expressed as

$$\chi^{[\nu]} = (\sqrt{g})(a\mathbf{q}^{(\nu)} + b\mathbf{q}^{(\nu)'}), \tag{35}$$

where a and b are coefficients and satisfy the condition

$$a^2 + b^2 = 1. \tag{36}$$

From Eq. (35) one has

$$h_\nu = \chi^{[\nu]}_e = (\sqrt{g})(aq^{(\nu)}_e + bq^{(\nu)'}_e). \tag{37}$$

From the known dimension h_ν, and Eqs. (36) and (37), we can calculate the coefficients a and b. Due to the orthogonality property, the other character vector must be

$$\chi^{[\nu']} = \sigma(\sqrt{g})(b\mathbf{q}^{(\nu)} - a\mathbf{q}^{(\nu)'}), \tag{38a}$$

where

$$\sigma = \text{sign}(bq^{(\nu)}_e - aq^{(\nu)'}_e). \tag{38b}$$

It is easy to generalize this procedure to the case when the degeneracy of $\lambda_{(2)}$ is greater than 2. By this method we calculate all the characters of the group S_{11}. The essential difference between the eigenfunction method and the Jones' method[10] or Boerner's[5] method for calculation of characters is that we introduced the CSCO of G and thus it is much simpler. In fact their methods are of prohibitory difficulty for use in calculating characters of high-order finite groups.

I. Example. Group $S(3)$

$S(3)$ has three classes, $C_1 = e$, $C_2 = (12) + (23) + (13)$, $C_3 = (123) + (132)$.
From the multiplication table

$$C_2C_2 = 3(C_1 + C_3), \quad C_2C_3 = 2C_2 \tag{39}$$

and Eq. (7a), we obtain the matrix representative of C_2:

$$\mathscr{D}(C_2) = \begin{pmatrix} 0 & 3 & 0 \\ 1 & 0 & 2 \\ 0 & 3 & 0 \end{pmatrix}. \tag{40}$$

Diagonalizing $\mathscr{D}(C_2)$, one gets three different eigenvalues: $\lambda_{(2)} = 3, 0, -3$. Therefore C_2 is a CSCO of $S(3)$. The three eigenvectors of $\mathscr{D}(C_2)$, normalized according to Eq. (26a), give the three character vectors

$$\begin{array}{lccc} & 3 & 0 & -3 \\ \text{partition} & [3] & [21] & [1^3] \\ \text{eigenvectors} & \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} & \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} & \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \end{array}. \tag{41}$$

If we at first pick out the operatore C_3 and diagonalize the matrix

$$\mathscr{D}(C_3) = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \tag{42}$$

we would find the eigenvalue $\lambda_{(3)} = 2, -1, 2$. There is a twofold degeneracy in $\lambda_{(3)} = 2$, therefore C_3 is not a CSCO of $S(3)$.

3. CSCO-II AND THE YAMANOUCHI BASES

A. Theorems 3 and 4

Theorem 3: A necessary and sufficient condition for $\psi^{(\nu)}$ to belong to the irrep (ν) of a group G is that $\psi^{(\nu)}$ is an eigenfunction of the CSCO-I of G:

$$C\psi^{(\nu)} = \nu\psi^{(\nu)}. \tag{43}$$

Hereafter, we use ν to denote both the eigenvalue and the label of the irrep. (Sometimes we also use the partition $[\nu]$ as the irrep label). A generalization of Theorem 3 is

Theorem 4: A necessary and sufficient condition for $\psi^{\nu}_{\mu_1\mu_2...}$ to belong to the irreps $\nu, \mu_1, \mu_2,...$ of the group chain $G \supset G_1 \supset G_2 \cdots$ is that $\psi^{(\nu)}_{\mu_1\mu_2...}$ satisfies the following

eigenequations:

$$\begin{pmatrix} C \\ C(1) \\ C(2) \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \psi^{(\nu)}_{\mu_1\mu_2\cdots} = \begin{pmatrix} \nu \\ \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \psi^{(\nu)}_{\mu_1\mu_2\cdots}, \tag{44}$$

where $C(i)$ is the CSCO of G_i, μ_i is its eigenvalue. Equation (44) can be written more compactly as

$$\begin{pmatrix} C \\ C(s) \end{pmatrix} \psi^{(\nu)}_m = \begin{pmatrix} \nu \\ m \end{pmatrix} \psi^{(\nu)}_m, \quad C(s) = (C(1), C(2), \cdots), \tag{45}$$

$m = (\mu_1, \mu_2, \cdots)$.

$C(s)$ will be called the CSCO of the subgroup chain $G(s) = G_1 \supset G_2 \supset \cdots$. $\psi^{(\nu)}_m$ will be called the irreducible basis in the $G \supset G(s)$ classification.

B. Definition 2

If $(C, C(s))$ is a CSCO in each irreducible space of G, then $(C, C(s))$ will be called CSCO-II of G. (It should be remembered that CSCO-I is a CSCO in the class space.)

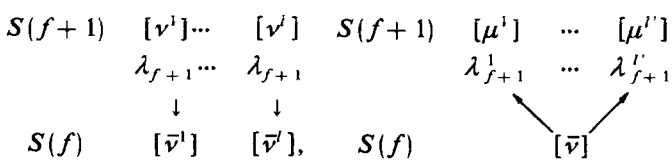
C. CSCO-II of $S(f)$

Since the group chain $S(f) \supset S(f-1) \supset \cdots \supset S(2)$ is a canonical subgroup chain (i.e., the irreps of the group chain provide a complete solution to the labeling problem), the set of operators $M = (C(f), C(f-1), \cdots, C(2))$ is a CSCO-II of the group $S(f)$, $C(i)$, being the CSCO of $S(i)$. From Theorem 4, we know that the simultaneous eigenfunctions of M are the Yamanouchi bases.

Now we want to prove that the set of the operators $(C(f) \cdots C(2))$ is overcomplete, i.e., the number of the operators contained in it is more than enough to make it a CSCO in each irreducible space of $S(f)$. In fact we have

Theorem 5: The $(f-1)$ 2-cycle class operators $(C_f, C_{f-1}, \cdots, C_2)$ of the group chain $S(f) \supset S(f-1) \supset \cdots \supset S(2)$ constitute a CSCO-II of the group $S(f)$.

Proof: To prove the theorem is equivalent to prove that a Yamanouchi basis of $S(f)$ can be labeled uniquely by the eigenvalues $\lambda_f \cdots \lambda_2$ [Notice that we changed from the notation $\lambda_{(2)}(f)$ to λ_f for the eigenvalue of 2-cycle class operator C_f]. From Table I we know that for $f \leqslant 5$, a single operator C_f constitutes the CSCO of $S(f)$, in other words, $C(f) = C_f$; so the theorem is trivial for $f \leqslant 5$. Now supposing it holds for $S(f)$, we want to prove it also holds for $S(f+1)$. According to the relationship (28) between the Young diagram and the eigenvalue λ_f, and the branching law, this in turn amounts to proving that if there are l Young diagrams $[\nu^1] \cdots [\nu^l]$ of $S(f+1)$ corresponding to the same eigenvalue λ_{f+1} and supposing $[\bar{\nu}^i]$ be the Young diagram resulting from removing one box from the Young diagram $[\nu^i]$ (see schematic diagram below), then $[\bar{\nu}^i] \neq [\bar{\nu}^j]$ for $i \neq j$.



It again amounts to proving that all the eigenvalues λ^i_{f+1} corresponding to the Young diagrams $[\mu^i]$ which result from adding one box to the same Young diagram $[\bar{\nu}]$ of $S(f)$ are different. Let $[\mu^i] = [f_1, f_2, \cdots f_{i-1}, f_i + 1, f_{i+1} \cdots]$ be the Young diagram resulting from adding one box in the ith row of the Young diagram $[\bar{\nu}] = [f_1, \cdots, f_{i-1} f_i f_{i+1} \cdots]$. From (28) one has

$$\lambda^i_{f+1} = \frac{f+1}{2} + \tfrac{1}{2}\sum_l (f_l + \delta_{li})(f_l + \delta_{li} - 2l). \tag{46}$$

If $\lambda^i_{f+1} = \lambda^j_{f+1}$ one gets

$$f_i + j = f_j + i, \tag{47}$$

from (46). Supposing $i < j$, it must have $f_i > f_j$ due to the rule for the Young diagrams, and thus (47) holds only when $i = j$. The theorem is proved.

Therefore one can label a Yamanouchi basis of $S(f)$ either by the $f-1$ Yamanouchi symbols $(r_f, r_{f-1}, \cdots, r_2, r_1 \equiv 1)$ or by the $f-1$ eigenvalues $(\lambda_f \cdots \lambda_2)$. They have one to one correspondence. This correspondence can be seen more clearly from the branching diagram. Figure 1 gives the branching rule for $S(f)$ with $f \leqslant 6$. The numbers below each partitions are the eigenvalues λ_f. Starting from each partition of $S(f)$, each route along the arrows corresponds to a Yamanouchi basis vector of $S(f)$. We can use either a string of partitions or a string of eigenvalues through which the route passes to label a Yamanouchi basis vector. For example,

1	2	4	5
3			
6			

[411]→[41]→[31]→[21]→[2]
(3, 5, 2, 0, 1)

1	2	4	
3	5	6	

[33]→[32]→[31]→[21]→[2]
(3, 2, 2, 0, 1)

From the above example one can see that although both the irreps [411] and [33] correspond to the same eigenvalue 3 of C_6, their reduction routes are different. The possible degeneracy in the eigenvalue λ_6 does not prevent us from labeling the Yamanouchi basis uniquely by the set of eigenvalues $(\lambda_6, \lambda_5 \cdots \lambda_2)$.

According to the branching diagram, like Fig. 1, by taking a suitable linear combination of the $f-1$ 2-cycle

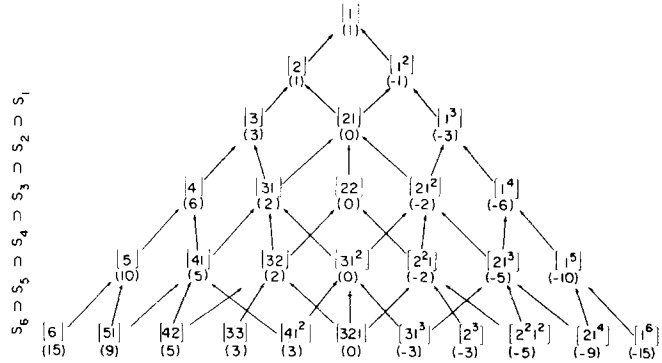


FIG. 1. The branching diagram of $S(f)$, $f \leqslant 6$. The numbers below each partition are the eigenvalues λ_f.

TABLE III. The one-to-one correspondence between the Young tableaux, and the eigenvalues of two kinds of CSCO-II of $S(5)$, i.e., $(\lambda_5, \lambda_4, \lambda_3, \lambda_2)$ and $\Sigma_{n=2}^{5} n\lambda_n$.

		1234	1235	1245	1345	123	124	134
Y tableau	12345	5	4	3	2	45	35	25
$(\lambda_5\lambda_4\lambda_3\lambda_2)$	(10,6,3,1)	(5,6,3,1)	(5,2,3,1)	(5,2,0,1)	(5,2,0, − 1)	(2,2,3,1)	(2,2,0,1)	(2,2,0, − 1)
λ	85	60	44	35	31	29	20	16
	125	135	123	124	134	125	135	145
Y tableau	34	24	4	3	2	3	2	2
			5	5	5	4	4	3
$(\lambda_5\lambda_4\lambda_3\lambda_2)$	(2,0,0,1)	(2,0,0, − 1)	(0,2,3,1)	(0,2,0,1)	(0,2,0, − 1)	(0, − 2,0,1)	(0, − 2,0, − 1)	(0, − 2, − 3, − 1)
λ	12	8	19	10	6	− 6	− 10	− 19

class operators, one can easily construct a single operator,

$$A = \sum_{n=2}^{f} k_n C_n, \tag{48}$$

such that A is a CSCO-II of $S(f)$. To this end, we only need to choose the coefficients k_n properly so as to make the eigenvalues λ of the operator A all different for each Yamanouchi basis vector of $S(f)$,

$$\lambda = \sum_{n=2}^{f} k_n \lambda_n. \tag{49}$$

For example,

$$A = \sum_{n=2}^{f} nC_n \tag{50}$$

is a CSCO-II of $S(f)$ for $2 \leqslant f \leqslant 5$. The eigenvalues λ for $S(5)$ are listed in Table III, along with $(\lambda_f \cdots \lambda_2)$ and the Young tableaux. There is a one to one correspondence between them.

D. Eigenfunction method for finding the Yamanouchi basis

The eigenfunction method for finding the Yamanouchi basis can be described as follows: Suppose there are N orthonormalized (it is trivial to extend to the nonorthonormalized cases by introducing the so-called dual basis.) f-particle wave functions

$$\varphi_a = \varphi_a(x_1 \cdots x_f), \quad a = 1,2,\ldots,\mathcal{N}, \tag{51}$$

carrying a reducible representation of $S(f)$. We take a linear combination of φ_a to form a Yamanouchi basis,

$$\psi_\lambda = \sum_{a=1}^{\mathcal{N}} U_{\lambda a} \varphi_a. \tag{52}$$

According to Theorems 4 and 5, the Yamanouchi bases are eigenfunctions of the CSCO-II of the permutation group, i.e., ψ_λ satisfy the following simultaneous eigenequations:

$$C_n \psi_\lambda = \lambda_n \psi_\lambda, \quad n = f, f-1,\ldots,2. \tag{53}$$

or

$$\begin{pmatrix} C \\ C(s) \end{pmatrix} \psi_m^{(\nu)} = \begin{pmatrix} \nu \\ m \end{pmatrix} \psi_m^{(\nu)},$$

$$(C; C(s)) = (C_f; C_{f-1}, \cdots, C_2),$$

$$C_n = \sum_{i<j}^{n} (ij), \tag{54}$$

$$\lambda = (\lambda_f; \lambda_{f-1}, \cdots \lambda_2) = (\nu; m).$$

The expansion coefficients $U_{\lambda a}$ can be found by solving the following eigenequations:

$$\sum_b [\langle \varphi_a | C_n | \varphi_b \rangle - \lambda_n \delta_{ab}] U_{\lambda b} = 0, \quad n = f, \cdots, 2. \tag{55}$$

The matrix of C_n is a symmetric matrix.

If we choose a single operator A of Eq. (48) as the CSCO-II of $S(f)$, then the set of eigenequations (53) or (55) is replaced by a single eigenequation

$$A\psi_\lambda = \lambda \psi_\lambda,$$

$$\sum_b [\langle \varphi_a | A | \varphi_b \rangle - \lambda \delta_{ab}] U_{\lambda b} = 0. \tag{56}$$

The problem of simultaneously diagonalizing $f-1$ matrices is thus reduced to that of diagonalizing a single matrix. For computer calculation, it greatly simplifes the problem. However, for hand calculation, it is preferable to use Eq. (53). Because we can start with $n = 2$, and successively use the eigenequation of C_{n-1} to eliminate the unknown variables in the eigenequation of C_n.

In many cases we can know beforehand which irreps the given reducible representation will decompose into. In such cases, the task of finding the eigenvalues from Eqs. (55) or (56) can be skipped. According to the known decomposition rule, and the correspondence between the eigenvalue and the Young tableau, like Table III, we can write down the eigenvalue and substitute it into Eqs. (55) or (56), then solve the resulting algebraic equations.

a. Multiplicity free case

If the eigenvalue λ is nondegenerate, it means that the corresponding irrep $[\nu]$ occurs only once in the given reducible representation. From Eq. (56), we can obtain the Yamanouchi basis except for a possible difference in the relative phase. The Young–Yamanouchi phase convention stipulates that the off diagonal matrix elements of the neighboring permutation must be positive. Using this condition the relative phases of the eigenfunctions $\psi_m^{(\nu)}$ belonging to the same irrep $[\nu]$ can be determined. Other ways for obtaining the eigenfunctions with the Yamanouchi phase convention are given in the following Sub Sec. (*b*) and Sec. 7.

b. General case

If an eigenvalue λ has τ_ν-fold degeneracy, it implies that the corresponding irrep ν occurs τ_ν times in the given

reducible representation. For the eigenvalue $\lambda = (\nu, m)$, we can get τ_ν linearly independent solutions

$$\psi_\lambda^\tau = \psi_m^{(\nu)\tau} = \sum_{a=1}^{f} U_{\lambda a}^{(\tau)} \varphi_a, \quad \tau = 1,2,\ldots,\tau_\nu. \tag{57}$$

Apart from the requirement that ψ_λ^τ be orthogonal on the multiplicity label τ,

$$\langle \psi_{\lambda'}^{\tau'} | \psi_\lambda^\tau \rangle = \delta_{\tau\tau'}\delta_{\lambda\lambda'}, \tag{58}$$

we are free to choose the τ_ν solutions. However, it must emphasized that these randomly chosen eigensolutions with the same (ν) and (τ) but different (m) will in general not span an irrep, since the label τ is chosen arbitrarily for each (ν,m). This problem can be solved in the following way.

We know that once a certain component $\psi_m^{(\nu)}$ of an irrep (ν) is known, all the other components of the irrep (ν) can be deduced by the action of an appropriate element, denoted as $F_{m'm}^{(\nu)}(R)$, of the group algebra,

$$\psi_{m'}^{(\nu)} = F_{m'm}^{(\nu)}(R)\psi_m^{(\nu)}. \tag{59a}$$

The explicit form of $F_{m'm}^{(\nu)}(R)$ is easy to find out for each group. The Young–Yamanouchi matrix elements of the neighboring permutation has the following properties:

$$D_{m'm}^{[\nu]}(i-1,i) = \begin{cases} 1/\sigma, & m' = m \\ (\sigma^2-1)^{1/2}/|\sigma|, & \text{for } Y_{m'}^{[\nu]} = (i-1,i)Y_m^{[\nu]} \\ 0, & \text{otherwise} \end{cases} \tag{60}$$

where σ is the axial distance[2] from the ordinal i to $i-1$ in the Young tableau $Y_m^{[\nu]}$. Therefore, for the permutation group, Eq. (59a) takes the form

$$\psi_{m'}^{[\nu]\tau} = \{[p_{i\,i-1} - D_{mm}^{[\nu]}(i-1,i)]/D_{m'm}^{[\nu]}(i-1,i)\}\psi_m^{[\nu]\tau}, \tag{59b}$$

where the Young diagram $Y_{m'}^{[\nu]}$ is obtained from $Y_m^{[\nu]}$ by interchanging the number i with $i-1$. Thus the finding of the Yamanouchi bases can proceed in the following way.

For a certain component, say the one with the maximum Yamanouchi symbol[2] [which we call the first $(m=1)$ components—the label m can be regarded as the set of quantum numbers $(\lambda_{f-1}\ldots\lambda_2)$ of the Yamanouchi symbol, or the index specifying the order of the Yamanouchi bases with decreasing page-order convention for the Yamanouchi symbol] of each irrep (ν), find the τ_ν orthogonal eigensolutions $\psi_m^{(\nu)}$ from Eq. (56). Choosing appropriate neighboring permutations and using Eqs. (60) and (59b), we can get all the other components $\psi_m^{(\nu)}$, successively. For example, for the irrep [32] of S_5, from the Yamanouchi basis $\left|\begin{smallmatrix}123\\45\end{smallmatrix}\right\rangle$ we can obtain all the other components by applying the neighboring permutation in the following way:

$$\left|\begin{smallmatrix}123\\(45)\end{smallmatrix}\right\rangle \xrightarrow{34} \left|\begin{smallmatrix}124\\35\end{smallmatrix}\right\rangle \xrightarrow{(23)} \left|\begin{smallmatrix}134\\25\end{smallmatrix}\right\rangle \xrightarrow{(45)} \left|\begin{smallmatrix}135\\24\end{smallmatrix}\right\rangle$$

$$\downarrow (45)$$

$$\left|\begin{smallmatrix}125\\34\end{smallmatrix}\right\rangle$$

The advantages of this procedure are twofold. (1) For each irrep (ν) one only needs to solve Eq. (56) once instead of h_ν times. (2) Using the Yamanouchi's matrix elements, Eq. (60), ensures that our solutions are consistent with the Young–Yamanouchi phase convention.

Equation (59b) is, of course, also applicable to the multiplicity free case.

For another way of handling the multiplicity problem see Sec. 7.

4. THE CLEBSCH–GORDAN COEFFICIENTS

Supposing $\psi_{m_1}^{(\nu_1)}(x)$ and $\psi_{m_2}^{(\nu_2)}(\xi)$ are the two wave functions of the same f particles in the x- and ξ-space, and they are the Yamanouchi bases of the permutation group $S^x(f)$ and $S^\xi(f)$, respectively, where $S^x(f)$ ($S^\xi(f)$) is the permutation group in the x-(ξ-) space.

The product wave functions $\varphi_{m_1}^{(\nu_1)}(x)\psi_{m_2}^{(\nu_2)}(\xi)$ carry a reducible representation, the uncoupled representation, of the permutation group $S^q(f)$ with $q = (x,\xi)$. By means of the Clebsch–Gordan coefficients (CGC), $C_{\nu_1 m_1,\nu_2 m_2}^{(\nu)\tau,m}$, of $S(f)$, they can be linearly combined into Yamanouchi bases of $S^q(f)$:

$$\begin{aligned}\Psi_m^{(\nu)\tau}(q) &= \sum_{m_1 m_2} C_{\nu_1 m_1,\nu_2 m_2}^{(\nu)\tau,m} \varphi_{m_1}^{\nu_1}(x)\psi_{m_2}^{\nu_2}(\xi) \\ &= \sum_{m_1 m_2} C_{\nu_1 m_1,\nu_2 m_2}^{(\nu)\tau,m} |m_1 m_2\rangle,\end{aligned} \tag{61}$$

where τ is a multiplicity label. The CGC satisfy the unitarity condition:

$$\sum_{m_1 m_2} C_{\nu_1 m_1,\nu_2 m_2}^{(\nu)\tau,m} C_{\nu_1 m_1,\nu_2 m_2}^{(\nu')\tau',m'} = \delta_{\nu\nu'}\delta_{mm'}\delta_{\tau\tau'}, \tag{62a}$$

$$\sum_{\nu m\tau} C_{\nu_1 m_1,\nu_2 m_2}^{(\nu)\tau,m} C_{\nu_1 m_1',\nu_2 m_2'}^{(\nu)\tau,m} = \delta_{m_1 m_1'}\delta_{m_2 m_2'}. \tag{62b}$$

From Eq. (54), we obtain the eigenequations satisfied by the CGC.

$$\sum_{m_1 m_2}\left[\left\langle m_1' m_2' \left| \begin{matrix} C \\ C(s) \end{matrix} \right| m_1 m_2 \right\rangle - \binom{\nu}{m}\delta_{m_1 m_1'}\delta_{m_2 m_2'} C_{\nu_1 m_1,\nu_2 m_2}^{(\nu)\tau,m}\right] = 0. \tag{63}$$

Using

$$C_n = C_{n-1} + \sum_{i=1}^{n-1}(in), \tag{64}$$

the matrix elements of C_n can be calculated by the following recursive formula:

$$\begin{aligned}&\langle m_1' m_2' | C_n | m_1 m_2 \rangle \\ &\quad = \langle m_1' m_2' | C_{n-1} | m_1 m_2 \rangle + \sum_{i=1}^{n-1} D_{m_1' m_1}^{(\nu_1)}(in) D_{m_2' m_2}^{(\nu_2)}(in).\end{aligned} \tag{65}$$

It is readily seen that the matrix $\langle m_1' m_2' | C_n | m_1 m_2 \rangle$ is a real symmetric matrix.

The following two important facts should be mentioned.

(1) The operator $C_2 = (12)$ of $S^q(f)$ is diagonalized in the uncoupled representation.

Proof: Since $\varphi_{m_1}^{\nu_1}$ and $\psi_{m_2}^{\nu_2}$ are Yamanouchi bases, we have

$$\begin{aligned}&{}^x C_2 \varphi_{m_1}^{\nu_1} = \lambda_2^{(1)}\varphi_{m_1}^{\nu_1}, \quad {}^\xi C_2 \psi_{m_2}^{\nu_2} = \lambda_2^{(2)}\psi_{m_2}^{\nu_2}, \\ &\lambda_2^{(i)} = \pm 1, \quad i = 1,2.\end{aligned} \tag{66}$$

From Eq. (66), it follows immediately that

$$\begin{aligned}&C_2 |m_1 m_2\rangle = \lambda_2 |m_1 m_2\rangle, \\ &\lambda_2 = \lambda_2^{(1)}\lambda_2^{(2)}.\end{aligned} \tag{67}$$

TABLE IV. The eigenvalues β of the quasi-CSCO-II, $\Sigma_f^{n=3} nC_n$, of $S(f)$ for the first Yamanouchi basis vector of each irrep, $f \leqslant 6$.

$[\nu]$	[3]	[21]	$[1^3]$	[4]	[31]	[22]	[211]	$[1^4]$	[5]	[41]	[32]	[311]	[221]	$[21^3]$	$[1^5]$
β	9	0	-9	33	17	0	-8	-33	83	58	27	17	-10	-33	-83
$[\nu]$	[6]	[51]	[42]	[411]	[33]	[321]	$[31^3]$	$[2^3]$	$[2^21^2]$	$[21^4]$	$[1^6]$				
β	173	137	88	76	45	27	-1	-28	-40	-87	-173				

For later convenience, λ_2 will be referred to as the (permutation) parity.

Therefore, to obtain CGC we only need to diagonalize the f-2 class operator $C_f \ldots C_3$

(2) If we divided the product bases $|m_1 m_2\rangle$ according to the $\lambda_2 = \pm 1$ into the two groups $|m_1 m_2\rangle_{(+)}$ and $|m_1 m_2\rangle_{(-)}$, then

$$ {}_{(+)}\langle m_1' m_2' | C_n | m_1 m_2 \rangle_{(-)} \equiv 0. \tag{68}$$

This can be easily proved by using the fact that $[C_n, C_2] = 0$.

The property Eq. (68) greatly simplifies both the calculation and the tabulation of the CGC. Each CGC table can be decomposed into two subtables, one corresponding to $\lambda_2 = 1$, and the other corresponding to $\lambda_2 = -1$. (See the Tables of the CGC of permutation groups in Ref. 11)

From the discussion in Sec. 3, it is known that we only need to calculate the first component $C^{(\nu)\tau,\, m=1}_{\nu_1 m_1, \nu_2 m_2}$ of CGC from Eq. (63). Since the first components of Yamanouchi bases of all the irreps, except the totally antysymmetric irrep, have $\lambda_2 = 1$, and the CGC for the totally antisymmetric irrep are very simple, we can always restrict ourselves to diagonalize the $f-2$ operators $C_f, C_{f-1}, \ldots, C_3$ in the positive parity space. To facilitate computer calculation, in analogy with Eq. (48) we can introduce a single operator

$$B = \sum_{n=3}^{f} \kappa_n C_n, \tag{69}$$

and require that its eigenvalue corresponding to the first component of each irrep be unique, i.e., no eigenvalues corresponding to any other Yamanouchi bases of $S(f)$ are identical to it.

The difference between the operator A [Eq. (48)] and B should be noted. A is a CSCO-II of $S(f)$, while B is not. The eigenvalues of B corresponding to the nonfirst components of Yamanouchi bases may be identical. For simplicity, we call B the quasi-CSCO-II of $S(f)$. For example,

$$B = \sum_{n=3}^{f} nC_n \tag{70}$$

is a quasi CSCO-II for S_f with $f \leqslant 6$. The eigenvalues of B for the first Yamanouchi basis vector of each irrep are listed in Table IV.

In summary, the steps for calculating CGC of the permutation group are

(1) Using the following eigenequation to calculate the first component CGC:

$$\sum_{m_1 m_2} \left[{}_{(+)}\langle m_1' m_2' | B | m_1 m_2 \rangle_{(+)} - \beta \delta_{m_1 m_1'} \delta_{m_2 m_2'} \right] C^{(\nu)\tau,\, m=1}_{\nu_1 m_1, \nu_2 m_2} = 0. \tag{71}$$

If β is a τ_ν-fold root of the secular equation, it means that the irrep $[\nu]$ occur τ_ν times in the product representation $[\nu_1] \times [\nu_2]$. There are τ_ν independent eigenvectors $\{C^{(\nu)\tau, m=1}_{\nu_1 m_1, \nu_2 m_2}\}, \tau = 1, 2, \ldots, \tau_\nu$ (with m_1, m_2 as the row label). They are to be made orthogonal with respect to the label τ [see Eq. (62a)].

From the known results of the Clebsch–Gordan series of the permutation group given by Itzykson and Nauenberg,[12] the task of finding the eigenvalues can again be skipped. We just substitute the known eigenvalue β into Eq. (71), and solve a set of linear homogeneous algebraic equations.

(2) Having obtained the first component solution, we use Eq. (59b) to find other components. Substitute $\Psi^{(\nu)\tau}_m$ and $\Psi^{(\nu)\tau}_{m'}$ for $\psi^{(\nu)\tau}_m$ and $\psi^{(\nu)\tau}_{m'}$ in Eq. (59b), respectively, and multiplying from left with ${}_{(+)}\langle m_1 m_2|$ we have

$$\begin{aligned} C^{(\nu)\tau, m'}_{\nu_1 m_1', \nu_2 m_2'} = {} & [D^{[\nu]}_{m'm}(i-1, i)]^{-1} \\ & \times \sum_{m_1 m_2} \left[D^{\nu_1}_{m_1' m_1}(i-1, i) D^{\nu_2}_{m_2' m_2}(i-1, i) \right. \\ & \left. - D^{\nu}_{mm}(i-1, i) \delta_{m_1' m_1} \delta_{m_2' m_2} \right] C^{(\nu)\tau, m}_{\nu_1 m_1, \nu_2 m_2}. \end{aligned} \tag{72}$$

Using this method we have calculated the CGC of the permutation group S_f with $f \leqslant 6$. The numerical tables have been published.[12] All coefficients are expressed in the form of square roots of rationals. An alternative, and incidently, better way of calculating the CG coefficients of the permutation group is given in Ref. 19.

5. THE COEFFICIENTS OF FRACTIONAL PARENTAGE

Let $\psi\begin{pmatrix}[\nu]\\ m, \alpha \Lambda M\end{pmatrix}$ be a Yamanouchi basis $[\nu]m$ of $S(f)$ and an irreducible basis of the group chain SU $(n) \supset G \supset G(s)$, where $G(s)$, is a subgroup chain of G, $[\nu]$, Λ, and M are the irrep labels for SU(n), G, and $G(s)$, respectively, α is a multiplicity label taking care of possible multiplicities in the reduction from $[\nu]$ to Λ. The wave function of an f-particle system can be expressed in terms of that for $f-1$ particles and the f particle in the same group theoretical classification:

$$\psi\begin{pmatrix}[\nu]\\ m, \alpha \Lambda M\end{pmatrix} = \sum_{\alpha_1 \Lambda_1} C^{[\nu], \alpha \Lambda}_{[\nu_1] \alpha_1 \Lambda_1, [1] \Lambda_0} \left[\psi\begin{pmatrix}[\nu_1]\\ m_1, \alpha_1 \Lambda_1\end{pmatrix} \psi^{[1]}_{\Lambda_0}(f) \right]^{\Lambda}_{M}, \tag{73}$$

where $C^{[\nu], \alpha \Lambda}_{[\nu_1] \alpha_1 \Lambda_1, [1] \Lambda_0}$ is the one particle CFP [or SU(n) $\supset G$ isoscalar factor) which is independent of m_1. $([\nu]m)$ and $([\nu_1]m_1)$ are labels for the Yamanouchi bases of $S(f)$ and $S(f-1)$, respectively, the square bracket indicated that the product bases are to be combined into the irrep ΛM of the group chain $G \supset G(s)$ by the CGC of the group G, Λ_0 is the irrep label of G for the defining representation[1] of SU(n).

Since the right-hand side of Eq. (73) is already Yamanouchi bases $[\nu_1]m_1$ of $S(f-1)$, the requirement that the left-hand side of Eq. (73) be a Yamanouchi basis $[\nu]m$ amounts to the requirement that it be an eigenfunction of the 2-cycle

class operator C_f of $S(f)$. Using Eq. (64) one has

$$\sum_{i=1}^{f-1}(if)\psi\begin{pmatrix}[\nu]\\ m,\alpha\Lambda M\end{pmatrix}=(\lambda_f^{[\nu]}-\lambda_{f-1}^{[\nu_1]})\psi^{[\nu]}\begin{pmatrix}[\nu]\\ m,\alpha\Lambda M\end{pmatrix}, \tag{74}$$

namely, a diagonalization of the operator $\Sigma_{i=1}^{f-1}(if)$ in the bases with fixed $[\nu_1]m_1$,

$$|\alpha_1\Lambda_1\rangle_{\nu_1 m_1}^{\Lambda}\equiv\left[\psi\begin{pmatrix}[\nu_1]\\ m_1,\alpha_1\Lambda_1\end{pmatrix}\psi_{A_0}^{[1]}(f)\right]_M^{\Lambda}, \tag{75}$$

gives the one-particle CFP:

$$\sum_{\alpha_2\Lambda_2}\left[\left\langle\alpha_1\Lambda_1\left|\sum_{i=1}^{f-1}(if)\right|\alpha_2\Lambda_2\right\rangle_{\nu_1}^{\Lambda}-\delta_{\alpha_1\alpha_2}\delta_{\Lambda_1\Lambda_2}(\lambda_f^{[\nu]}-\lambda_{f-1}^{[\nu_1]})\right]C_{[\nu_1]\alpha_2\Lambda_2,[1]A_0}^{[\nu],\alpha\Lambda}=0. \tag{76}$$

If $\lambda_f^{[\nu]}$ is a τ_ν-fold root of the secular equation, there are τ_ν independent eigenvectors $\{C_{[\nu_1]\alpha_1\Lambda_1,[1]A_0}^{[\nu],\alpha\Lambda}\},\alpha=1,2,\ldots,\tau_\nu$. They can always be chosen to be orthogonal with respect to the index α.

The matrix element in Eq. (76) can be calculated by the following formula:

$$\left\langle\alpha_1\Lambda_1\left|\sum_{i=1}^{f-1}(if)\right|\alpha_2\Lambda_2\right\rangle_{\nu_1}^{\Lambda}=\frac{f-1}{h_{\nu_1}}\sum_{m_1}\langle\alpha_1\Lambda_1|(f,f-1)|\alpha_2\Lambda_2\rangle_{\nu_1 m_1}^{\Lambda}, \tag{77}$$

h_{ν_1} being the dimension of the irrep $[\nu_1]$ of $S(f-1)$. In deriving Eq. (77) we have used the fact that the left-hand side of Eq. (77) is independent of row label m_1. The matrix element of the permutation $(f,f-1)$ can be calculated by the same method as used by Jahn,[13]

$$\langle\alpha_1\Lambda_1|(f,f-1)|\alpha_2\Lambda_2\rangle_{\nu_1 m_1}^{\Lambda}=\sum_{\alpha'\Lambda'}C_{[\nu']\alpha'\Lambda',[1]A_0}^{[\nu_1],\alpha_1\Lambda_1}C_{[\nu']\alpha'\Lambda',[1]A_0}^{[\nu_1],\alpha_2\Lambda_2}\times\langle\Lambda'\Lambda_1\Lambda|(f,f-1)|\Lambda'\Lambda_2\Lambda\rangle, \tag{78}$$

where $C_{[\nu']\alpha'\Lambda',[1]A_0}^{[\nu_1],\alpha_1\Lambda_1}$, etc., are the CFP for the $(f-1)$-particle system and $[\nu']$ is the irrep of $S(f-2)$. $[\nu']$ is determined by the Young diagram left after deleting the box $f-1$ in the Young tableau $Y_{m_1}^{[\nu_1]}$. Therefore, the left-hand side of Eq. (78) depends on m_1, though the left-hand side of Eq. (77) is independent of m_1. The factors $\langle\Lambda'\Lambda_1\Lambda|(f,f-1)|\Lambda'\Lambda_2\Lambda\rangle$ are essentially the Racah coefficients of the group G. For example, for the group $G=SO_3$ it reads

$$\langle L'L_1L|(f,f-1)|L'L_2L\rangle=(-1)^{L'+L+L_1+L_2}U(lL'Ll;L_1L_2), \tag{79}$$

and for $G=SU_3$ it reads

$$\langle(\lambda'\mu')(\lambda_1\mu_1)(\lambda\mu)|(f,f-1)|(\lambda'\mu')(\lambda_2\mu_2)(\lambda\mu)\rangle=(-1)^{\lambda'-\mu'+\lambda-\mu+\lambda_1-\mu_1+\lambda_2-\mu_2}\times U((10)(\lambda'\mu')(\lambda\mu)(10);(\lambda_1\mu_1)(\lambda_2\mu_2)). \tag{80}$$

This technique for calculating the CFP is simpler than the one used by Jahn.[13] So and Strottman[14] used a similar method to calculate the SU (6) $\supset$ SU (3)$\times$SU (2) isoscalar factor (ISF). [Their Eq. (13) is not correct. It should read as

$$\frac{N-1}{h_{f^{(1)}}}\sum_{\mu^{(1)}}\left\langle\begin{matrix}[f^{(1)}]\\ (\lambda_1\mu_1)S_1\omega_1,r^{(1)},\end{matrix}\begin{matrix}[1]\\ (10)1/2;\end{matrix}(\lambda\mu)S\omega\left|p_{n-1,n}\right|\begin{matrix}[f^{(1)}]\\ (\bar\lambda_1\bar\mu_1)(\bar S_1\bar\omega_1,r^{(1)},\end{matrix}\begin{matrix}[1]\\ (10)1/2;\end{matrix}(\lambda\mu)S\omega\right\rangle.$$

Only if the individual integral in the above equation were independent of $r^{(1)}$ (in our notation, m_1), the above equation would have reduced to their equation (13). However, the integral does depends on $r^{(1)}$, as explained before.]

6. THE INTRINSIC GROUP

As mentioned before, if the set of eigenvalues (ν,m) of the CSCO-II of G in Eq. (45) is degenerate with degeneracy τ_ν, for this (ν,m) there will be τ_ν independent solutions $\psi_m^{(\nu)1}\cdots\psi_m^{(\nu)\tau_\nu}$. In the above discussion, these τ_ν solutions are chosen arbitrarily. In order to determine them uniquely, one has to seek new operators which commute with the CSCO-II. The eigenvalues of such operators can be used to distinguish these τ_ν sets of irreducible bases which have exactly the same transformation properties under the group G. Obviously it is impossible to find such operators from the group G but, as will be seen below, one can find them from the so called "intrinsic group" $\bar G$.

For any group element R, we can define a corresponding operator $\bar R$ in the group space V_G by the following equation:

$$\bar RS=SR\quad\text{for any } S\in V_G. \tag{81}$$

In other words, the action of the operator $\bar R$ on any "vector" S in V_G is changing it into another vector SR. It must be emphasized that Eq. (81) is a defining equation for the operator $\bar R$, rather than an identity relation of operators. Therefore, it is not permissible to multiply Eq. (81) from the right by another vector T of V_G, i.e.,

$$\bar RST\neq SRT. \tag{82}$$

Instead, ST must be considered as a new vector in V_G, then using definition Eq. (81), one gets

$$\bar RST=\bar R(ST)=STR. \tag{83}$$

We are going to prove that the totality of the operators $\bar R$ constitute a group $\bar G$—the intrinsic group. Suppose for group G we have the multiplication relation

$$R_1R_2=R_3. \tag{84}$$

From Eqs. (81) and (84) one has

$$\bar R_2\bar R_1S=\bar R_2SR_1=SR_1R_2=SR_3=\bar R_3S. \tag{85}$$

Since S is an arbitrary vector in V_G, therefore,

$$\bar R_2\bar R_1=\bar R_3. \tag{86}$$

It means that there is a one to one correspondence between the operator $\bar R$ and R. Equations (84) and (86) show that the totality of $\bar R$ does constitute a group which is anti-isomorphic to the original group G. The group $\bar G$ is called the intrinsic group of G, or simply the intrinsic group, if no confusion will arise. The reason for naming it the intrinsic group was explained in Ref. 1(b).

Equation (81) gives a representation of the intrinsic group $\bar{G}$ in the group space,

$$\bar{R}_b R_c = \sum_a D_{ac}(\bar{R}_b) R_a = R_c R_b. \tag{87}$$

Boerner[5] called $D(\bar{R}_b)$ the inverted representation of G. We prefer to call it the representation of the intrinsic group $\bar{G}$ in the group space V_G. Our main interest is the operator group $\bar{G}$ and its representations in various spaces, e.g., the group space, configuration space, the space of functions on group manifold, etc.

A few more points worth mentioning are

(1) $\bar{R}$ is not the conjugate element of the group element R. From Eq. (81), we obtain the relation between $\bar{R}$ and R:

$$\bar{R} = SRS^{-1} \quad \text{when } \bar{R} \text{ acts on } S. \tag{88}$$

It must be emphasized that the equality holds only when $\bar{R}$ acts on S. When $\bar{R}$ acts on another element T, $\bar{R}$ would be equal to TRT^{-1}. Therefore, although at first sight $\bar{R}$ in Eq. (88) looks like a conjugate element of R, in fact the meaning is absolutely different.

(2) It is important to distinguish between the subgroup $\bar{G}_1$ of the intrinsic group $\bar{G}$ of G, and the intrinsic group $\bar{G}'_1$, of the subgroup G_1 of G. $\bar{G}_1$ is defined in the group space of G, and commutes with the whole group G (see below), while $\bar{G}'_1$ is defined in the group space of G_1 and commutes only with the subgroup G_1.

The intrinsic group $\bar{G}$ has many important properties.

(1) The groups $\bar{G}$ and G are anti-isomorphic [See Eqs. (84) and (86)].

(2) The elements of $\bar{G}$ commute with those of G.

Proof: From Eq. (81), one has

$$S\bar{R}T = STR. \tag{89}$$

Comparing Eq. (89) with (83), and noting that T is an arbitrary element in V_G, one has

$$\bar{R}S = S\bar{R} \tag{90a}$$

or

$$[\bar{R}, S] = 0. \tag{90b}$$

Therefore the property is proved.

Note the significant difference between Eqs. (81) and (90a). In Eq. (81) S is a basis, while in Eq. (90a), S is an operator. The rule for determining whether a group element S is to be regarded as a basis or as an operator is very simple: If S is the last one behind an intrinsic group element, then S should be regarded as a basis; if there are other group elements of G behind S, then S should be regarded as an operator.

(3) Since $\bar{G} = \{\bar{R}\}$ is anti-isomorphic with $G = \{R\}$, the group $\bar{G}$ is isomorphic with the group $\{R^{-1}\}$. Furthermore, the group $\{R^{-1}\}$ is essentially the group G; the only difference is in the name of each group element. Therefore, the intrinsic group $\bar{G}$ is essentially isomorphic to the group G. Thus all the conclusions we obtained in the above sections also hold for the intrinsic group $\bar{G}$. For example, if G has a group chain $G \supset G_1 \supset G_2 \supset \ldots$, $\bar{G}$ has the corresponding group chain $\bar{G} \supset \bar{G}_1 \supset \bar{G}_2 \supset \ldots$. if $G \supset G(s)$ is a canonical subgroup chain, $\bar{G} \supset \bar{G}(s)$ must be also a canonical subgroup chain. If $C = \Sigma_i k_i C_i$ is the CSCO-I of G, $\bar{C} = \Sigma_i k_i \bar{C}_i$ must be the CSCO-I of $\bar{G}$. If $(C, C(s))$ is the CSCO-II of G, then $(\bar{C}, \bar{C}(s))$ is the CSCO-II of $\bar{G}$, etc.

(4) The CSCO of G and $\bar{G}$ are equal.

$$C = \bar{C}. \tag{91}$$

Proof: To prove Eq. (91), we only need to prove that the class operator of G and $\bar{G}$ are equal. From Eq. (81), one has

$$\bar{C}_i R = RC_i. \tag{92a}$$

From the property of the class operator, Eq. (5), one has

$$C_i R = RC_i. \tag{92b}$$

Comparing Eq. (92a) with (92b), and noting that R is an arbitrary vector in V_G, we arrive at

$$\bar{C}_i = C_i. \tag{93}$$

According to Theorem 3 and Eq. (93), we know that if $\psi^{(\nu)}$ belongs to an irrep (ν) of G, it must also belong to the irrep (ν) of $\bar{G}$.

(5) If G_1 is a subgroup of G, and correspondingly $\bar{G}_1$ is the subgroup of $\bar{G}$, then G_1 and $\bar{G}_1$ have the same structure constant.

Let $C_\alpha, C_\beta, \ldots$ and $\bar{C}_\alpha, \bar{C}_\beta, \ldots$ be the class operators of G_1 and $\bar{G}_1$, respectively. The above property means that corresponding to

$$C_\alpha C_\beta = \sum_\gamma C^\gamma_{\alpha\beta} C_\gamma, \tag{94}$$

we have

$$\bar{C}_\alpha \bar{C}_\beta = \sum_\gamma C^\gamma_{\alpha\beta} \bar{C}_\gamma. \tag{95}$$

Equation (95) can be easily proved by multiplying it from the left by an arbitrary vector R in V_G.

It must be pointed out that although the class operators of G and $\bar{G}$ are equal, the class operators of the subgroup G_1 and $\bar{G}_1$ are not equal. This is readily seen from the fact that $\bar{C}_\alpha$ commutes with the whole group G, while C_α commutes only with the subgroup G_1. From Eq. (11b) we know that the possible eigenvalues (or the eigenvalue spectrum) of a class operator are determined by the structure constant of the group. G_1 and $\bar{G}_1$ have the same structure constant, therefore $C(1)$ and $\bar{C}(1)$ have the same eigenvalue spectrum. Similarly, the set of operators $C(s) = (C(1), C(2), \ldots)$ and $\bar{C}(s) = (\bar{C}(1)\ \bar{C}(2), \ldots)$ have the same eigenvalue spectrum. $\bar{C}(s)$ commutes with CSCO-II of G, thus $\bar{C}(s)$ is the set of new operators which we are looking for.

7. CSCO-III AND THE IRREDUCIBLE MATRIX ELEMENTS

Definition: If the set of operators $K = (C, C(s), \bar{C}(s))$ is a CSCO in the group space, then K is called the CSCO-III of the group G.

It is easy to prove that if $(C, C(s))$ is the CSCO-II of G, then $(C, C(s), \bar{C}(s))$ must be a CSCO-III of G.

Theorem 6: In the group space, the orthonormalized eigenoperator of the CSCO-III is the normalized generalized projection operator $\mathring{P}^{(\nu)k}_m$.

$$\begin{pmatrix} C \\ C(s) \\ \bar{C}(s) \end{pmatrix} \mathring{P}^{(\nu)k}_m = \begin{pmatrix} \nu \\ m \\ k \end{pmatrix} \mathring{P}^{(\nu)k}_m, \tag{96a}$$

$$\mathring{P}_m^{(\nu)k} = \sqrt{\frac{h_\nu}{g}} \sum_{a=1}^{g} D_{mk}^{(\nu)*}(R_a) R_a, \quad \begin{matrix} \nu = 1,2,...,N, \\ m,k = m_1 ... m_{h_\nu} \end{matrix} \tag{96b}$$

$$\langle \mathring{P}_{m'}^{(\nu')k'} | \mathring{P}_m^{(\nu)k} \rangle = \delta_{\nu\nu'} \delta_{mm'} \delta_{kk'}, \tag{96c}$$

where $D_{mk}^{(\nu)}(R)$ are the irreducible matix elements in the $G \supset G(s)$ classification.

C has N different sets of eigenvalues. In the eigenspace of C, which is a subspace of the group space, with the eigenvalue ν, $C(s)$ and $\bar{C}(s)$ both have h_ν different eigenvalues. The total number of eigenoperators $\mathring{P}_m^{(\nu)k}$ is

$$g = \sum_{\nu=1}^{N} h_\nu^2. \tag{97}$$

From Theorem 4 and Eq. (96a), we know that $\mathring{P}_m^{(\nu)k}$ is $G \supset G(s)$ irreducible basis $(\nu)m$ and $\bar{G} \supset \bar{G}(s)$ irreducible basis $(\nu)k$. It can be proved that $\mathring{P}_m^{(\nu)k}$ has the following transformation properties[7]:

$$R\mathring{P}_m^{(\nu)k} = \sum_{m'} D_{m'm}^{(\nu)}(R) \mathring{P}_{m'}^{(\nu)k}, \tag{98a}$$

$$\bar{R}\mathring{P}_m^{(\nu)k} = \sum_{k'} D_{k'k}^{(\nu)}(\bar{R}) \mathring{P}_m^{(\nu)k'}, \tag{98b}$$

and

$$D_{k'k}^{(\nu)}(\bar{R}) = D_{kk'}^{(\nu)}(R). \tag{99}$$

It is seen that R only changes the "external quantum number" m and the intrinsic group element $\bar{R}$ only changes the "intrinsic quantum number" k. The h_ν irreducible bases $\psi_m^{(\nu)1} ... \psi_m^{(\nu)h_\nu}$ have exactly the same transformation property under the group G. Hence the intrinsic quantum number k provides the additional quantum number to distinguish the h_ν equivalent irreps which occur in the reduction of the regular representation of G.

Theorem 7: In the group space, the eigenvectors of the CSCO-III are proportional to the complex conjugate of the irreducible matrix elements.

$$\sum_b \left[\left\langle R_a \left| \begin{matrix} C \\ C(s) \\ \bar{C}(s) \end{matrix} \right| R_b \right\rangle - \begin{pmatrix} \nu \\ m \\ k \end{pmatrix} \delta_{ab} \right] D_{mk}^{(\nu)*}(R_b) = 0. \tag{100}$$

Therefore, from Eq. (100) and the normalization condition,[2]

$$\frac{h_\nu}{g} \sum_a |D_{mk}^{(\nu)}(R_a)|^2 = 1, \tag{101}$$

we can calculate all the irreducible matrix elements in the $G \supset G(s)$ classification.

Equation (100) can be generalized to calculate the matrix elements of the group operators between two irreducible bases which are classified according to different subgroup chains.

Suppose $\psi_m^{(\nu)}$ is a $G \supset G(s)$ irreducible basis, $\varphi_k^{(\nu)}$ is a $G \supset G(s)'$ irreducible basis, and $C(s)'$ is the CSCO of the subgroup chain $G(s)'$; then the matrix elements'

$$\mathscr{D}_{mk}^{(\nu)}(R_a) = \langle \psi_m^{(\nu)} | R_a | \varphi_k^{(\nu)} \rangle \tag{102}$$

satisfy the following eigenequations:

$$\begin{pmatrix} C \\ C(s) \\ \bar{C}(s)' \end{pmatrix} \mathscr{D}_{mk}^{(\nu)*}(R_a) = \begin{pmatrix} \nu \\ m \\ k \end{pmatrix} \mathscr{D}_{mk}^{(\nu)*}(R_a). \tag{103}$$

From Eqs. (101) and (103), we can calculate $\mathscr{D}_{mk}^{(\nu)}(R_a)$.

Since $(C_f ... C_{f-1})$ is a CSCO-II of $S(f)$, the $2f-3$ operators

$$(C_f, C_{f-1}, ..., C_2, \bar{C}_{f-1} ... \bar{C}_2) \tag{104}$$

constituted a CSCO-III of $S(f)$, where

$$\bar{C}_n = \sum_{i<j}^{n} (\overline{ij}), \tag{105}$$

$(\overline{ij})$ being a permutation operator of the intrinsic group $\bar{S}(f)$. We can also choose a single operator

$$K = \sum_{k=2}^{f} a_n C_n + \sum_{n=2}^{f-1} b_n \bar{C}_n \tag{106}$$

as a CSCO-III of $S(f)$. The coefficients a_n and b_n are chosen so that the operator K has g different eigenvalues,

$$\mu = \sum_{n=2}^{f} a_n \lambda_n + \sum_{n=2}^{f-1} b_n \bar{\lambda}_n. \tag{107}$$

With the help of the branching diagram, like Fig. 1, and the fact that C_n and $\bar{C}_n$ have the same eigenvalue spectrum, it is easy to find the coefficients a_n and b_n for a given f (there are infinite ways to do this).

With K as CSCO-III, Eq. (100) now becomes

$$\sum_b [\langle R_a | K | R_b \rangle - \mu \delta_{ab}] D_\mu^*(R_b) = 0. \tag{108}$$

Again, the eigenvalues μ can be known beforehand, thus one merely needs to solve a set of g linear homogeneous algebraic equations to obtain the irreducible matrix elements. The only remaining freedom is the choice of phase. It can be shown[7] that to obtain the Young–Yamanouchi irreducible matrix elements one only need use the following rule.

(a) From the quantum number μ, trace back to the quantum numbers (ν, m, k). Find the corresponding Young tableaux $Y_m^{[\nu]}$ and $Y_k^{[\nu]}$. There is one and only one group element, say R_0, which transfers the Young tableau $Y_k^{[\nu]}$ to $Y_m^{[\nu]}$, i.e.,

$$Y_m^{[\nu]} = R_0 Y_k^{[\nu]}. \tag{109}$$

(b) The absolute phase of the eigenvector $\{D_{mk}^{(\nu)}(R_a)\} = (D_\mu(R_1), D_\mu(R_2) \cdots)$ is dictated by the requirement that its component $D_\mu(R_0)$ must be positive.

Example: Group $S(3)$

From Fig. 1, it is seen that (ν, m) has four sets of values (3, 1), (0, 1), (0, −1), and (−3, −1). Since the possible values of k are the same as those of m, the CSCO-III has six sets of eigenvalues $(\nu, m, k) = (3, 1, 1), (0, 1, 1), (0, -1, 1), (0, 1, -1), (0, -1, -1), (-3, -1, -1)$. We can choose

$$K = 3C_3 + 2C_2 + \bar{C}_2 \tag{110}$$

as a CSCO-III. K has six different eigenvalues,

$$\mu = 3\nu + 2m + k = 12, 3, 1, -1, -3, -12. \tag{111}$$

It is easy to construct the representative matrix of K in the six dimensional group space of $S(3)$ with basis vectors ordered as e, (12), (13), (23), (123), and (132):

$$K = \begin{pmatrix} 0 & 6 & 3 & 3 & 0 & 0 \\ 6 & 0 & 0 & 0 & 3 & 3 \\ 3 & 0 & 0 & 0 & 4 & 5 \\ 3 & 0 & 0 & 0 & 5 & 4 \\ 0 & 3 & 4 & 5 & 0 & 0 \\ 0 & 3 & 5 & 4 & 0 & 0 \end{pmatrix}. \tag{112}$$

Using the known eigenvalues Eq. (111) and the normalization Eq. (101), we can calculate the six eigenvectors of K:

$$\mathbf{D}^{(3)}_{11}=\begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix},\quad \mathbf{D}^{(0)}_{11}=\begin{pmatrix}1\\1\\-(\sqrt{3})/2\\-(\sqrt{3})/2\\-(\sqrt{3})/2\\-(\sqrt{3})/2\end{pmatrix},$$

$$\mathbf{D}^{(0)}_{-11}=\frac{\sqrt{3}}{2}\begin{pmatrix}0\\0\\-1\\1\\-1\\1\end{pmatrix},\tag{113}$$

$$\mathbf{D}^{(0)}_{1-1}=\frac{\sqrt{3}}{2}\begin{pmatrix}0\\0\\-1\\1\\1\\-1\end{pmatrix},\quad \mathbf{D}^{(0)}_{-1-1}=\begin{pmatrix}1\\-1\\(\sqrt{3})/2\\(\sqrt{3})/2\\-(\sqrt{3})/2\\-(\sqrt{3})/2\end{pmatrix},$$

$$\mathbf{D}^{(-3)}_{-1-1}=\begin{pmatrix}1\\-1\\-1\\-1\\1\\1\end{pmatrix}.$$

The phases are determined by the rule given above. For example, $(\nu, m)=(0,\ -1)$ corresponds to $\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}$. $(\nu,k)=(0,1)$ corresponds to $\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}$ According to Eq. (109), $R_0=(23)$, therefore the absolute phase of the eigenvector $D^{(0)}_{-11}$ is dictated by the requirement that the component $D^{(0)}_{-11}(23)$ must be positive.

8. REDUCTION OF NONREGULAR REPRESENTATIONS

A. Intrinsic state regular representation case

Equation (81) defines the action of the intrinsic group elements in the group space. We now address ourselves to the question of defining the action of the intrinsic group elements in the configuration space. The action of the group element R of G on the configuration wave function $\Phi(X)$ is usually defined as

$$R\Phi(X)=\Phi(R^{-1}X).\tag{114}$$

To define the action of the intrinsic group element on $\Phi(X)$ is more complicated. Let us first discuss the case of the regular representation. Suppose there are g functions $\Phi_1(X)\cdots\Phi_g(X)$ which carry a regular representation of G. Among them we pick out any one, say $\Phi_{i_0}\equiv\Phi_0$, and define the action of the intrinsic group element $\bar{R}$ on Φ_0 to be equal to that of the group element R,

$$\bar{R}\Phi_0(X)=R\Phi_0(X)\quad \text{for } R\in G.\tag{115}$$

We call the state $\Phi_0(X)$ which satisfies Eq. (115) the intrinsic state of the group G.

We now want to show that Eq. (115) suffices to define the action of the intrinsic group elements on any other function $\Phi_a(X)$. $\Phi_a(X)$ can be obtained from the chosen intrinsic state $\Phi_0(X)$ through the action of a certain element R_a of G.

$$\Phi_a(X)=R_a\Phi_0(X).\tag{116}$$

From Eqs. (115) and (81), we get the operation of the intrinsic group element $\bar{R}$ on any function $\Phi_a(X)$:

$$\bar{R}\Phi_a(X)=\bar{R}R_a\Phi_0(X)=R_aR\Phi_0(X).\tag{117}$$

The physical meaning of Eq. (117) can be seen more clearly by rewriting it in the following form:

$$\bar{R}\Phi_a(X)=R_aRR_a^{-1}\Phi_a(X).\tag{118}$$

It is seen that the action of the intrinsic group element $\bar{R}$ on $\Phi_a(X)$ can be carried out in three steps. (a) Through the operation R_a^{-1}, bring the state $\Phi_a(X)$ to the intrinsic state Φ_0. (b) Perform the operation R on the intrinsic state Φ_0. (c) through the operation R_a, bring the system back to the original "orientation."

The remaining question is which state should be chosen as the intrinsic state. In principle one can choose any state from $\Phi_1\cdots\Phi_g$ as the intrinsic state, therefore one can choose the one that is most suitable for a given physical problem.

Having defined the action of the intrinsic group element in the configuration space, we can carry out the reduction of the regular representation in the configuration space just as we have done in the group space. After the substitutions

$$R_a\to\Phi_a(X),\quad \mathring{P}^{(\nu)k}_m\to\psi^{(\nu)k}_m(X),\tag{119}$$

all the formulas in Sec. 7 are valid in the configuration space. The relation between the orthonormalized irreducible wave function $\psi^{(\nu)k}_m(X)$ and the generalized projection operator $\mathring{P}^{(\nu)k}_m$ is

$$\psi^{(\nu)k}_m(X)=\mathring{P}^{(\nu)k}_m\Phi_0(X),\tag{120}$$

where $\Phi_0(X)$ is the chosen intrinsic state.

B. Intrinsic state nonregular representation case

In reducing a regular representation, the h_ν sets of the intrinsic quantum number k provide just enough labels to distinguish the h_ν equivalent irreps (ν). While in reducing a nonregular representation, an irrep (ν) may occur only $\tau_\nu<h_\nu$ times. In such a case there are too many intrinsic quantum numbers. The question is, can we still use the intrinsic quantum number to distinguish the τ_ν equivalent irreps? The answer is that, if we can find an intrinsic state Φ_0 such that any other state φ_a in this reducible representation can be generated from Φ_0 through the action of appropriate group elements (i.e., φ_a can be written as $\varphi_a=R_a\Phi_0$), then after slight modification we can still use the intrinsic quantum numbers. Otherwise the intrinsic quantum number is meaningless.

We now proceed to justify this assertion. Suppose $\{\varphi_a\}$, $a=1,2,\ldots\mathcal{N}$, carry a nonregular representation. Pick up any one, say φ_1, as the intrinsic state Φ_0. We assume that the action of a group element on Φ_0 either leaves Φ_0 unchanged or transforms Φ_0 to another function in the set $\{\varphi_a\}$. The

totality of the operators T_α which leaves Φ_0 unchanged forms a group H which is a subgroup of G.

$$T_\alpha \Phi_0 = \Phi_0, \quad \alpha = 1,2,...,m, \quad T_\alpha \in H. \tag{121}$$

The order m of the subgroup H is equal to

$$m = g/\mathcal{N}. \tag{122}$$

H will be called the symmetry group of the intrinsic state Φ_0. If an operator R_a transforms Φ_0 into φ_a, there must be m operators that transform Φ_0 into the same φ_a:

$$\varphi_a(X) = R_a \Phi_0(X) = R_a T_\alpha \Phi_0(X), \quad \alpha = 1,2,...,m. \tag{123}$$

Therefore applying g group elements to Φ_0, we can get only $\mathcal{N}$ functions $\{\varphi_a\}$.

Now let us look at the action of the intrinsic group elements on the basis vector φ_a of the reducible representation. According to Eqs. (123) and (115),

$$\bar{R}\varphi_a(X) = \bar{R}R_a \Phi_0(X) = R_a R \Phi_0(X). \tag{124}$$

Using Eq. (121), on the other hand, one has

$$\bar{R}\varphi_a(X) = \bar{R}R_a T_\alpha \Phi_0(X) = R_a T_\alpha R \Phi_0(X). \tag{125}$$

Since $R_a R\Phi_0(X) \neq R_a T_\alpha R \Phi_0(X)$, unless $R \in H$ (in such a case $\bar{R}$ is equivalent to the identity), Eqs. (124) and (125) tell us that the intrinsic group elements $\bar{R}$, generally speaking, have no definite meaning in the nonregular configuration space, except those belonging to the symmetric group $\bar{H}$. Now the question is raised: Can we find some operators out of $\bar{G}$ which have a definite meaning? The answer is that if the CSCO $C(i)$ of a subgroup G_i of G commutes with the symmetric group H, i.e., $[C(i), T_\alpha] = 0$, $T_\alpha = 1,2,...,m$, then the corresponding intrinsic operator $\bar{C}(i)$ has a definite meaning.

Proof:

$$\bar{C}(i)\varphi_a(X) = \bar{C}(i)R_a \Phi_0(X) = R_a C(i)\Phi_0(X), \tag{126a}$$

$$\bar{C}(i)\varphi_a(X) = \bar{C}(i)R_a T_\alpha \Phi_0(X) = R_a T_\alpha C(i)\Phi_0(X) = R_a C(i) T_\alpha \Phi_0(X) = R_a C(i)\Phi_0(X). \tag{126b}$$

In deriving the second equation, the condition $[C(i), T_\alpha] = 0$ has been used. Equation (126) shows that the action of $\bar{C}(i)$ on $\varphi_a(X)$ always leads to $R_a\, C(i)\Phi_0(X)$, therefore $\bar{C}(i)$ has a definite meaning.

For example, for an intrinsic state

$$\Phi_0(X) = \psi_\alpha(x_1)\psi_\beta(x_2)\psi_\beta(x_3)\psi_\gamma(x_4) \equiv |\alpha\beta\beta\gamma\rangle \tag{127}$$

of the permutation group $S(4)$, the symmetry group of the intrinsic state is $H = \{e, (23)\}$. Since the CSCO of $S(3)$ commutes with H, $\bar{C}(3)$ has a definite meaning, while the CSCO of $\bar{S}(2)$, i.e., the permutation (12), has no definite meaning, as can be seen clearly in the following equations:

$$(\overline{12})\Phi_0 = (\overline{12})|\alpha\beta\beta\gamma\rangle = |\beta\alpha\beta\gamma\rangle,$$

$$(\overline{12})\Phi_0 = (\overline{12})(23)\Phi_0 = (23)(12)\Phi_0 = |\beta\beta\alpha\gamma\rangle. \tag{128}$$

In the case when a single group element of the intrinsic group has no definite meaning, the ordinary definition of the irreducible representation—the minimum invariant subspace under the action of the group—loses its meaning. Due to the fact that the CSCO of $\bar{G}$ along with those of certain subgroups $\bar{G}_i$ of $\bar{G}$ do still have a definite meaning, we might as well call the simultaneous eigenfunctions of the set of operators $(\bar{C}, \bar{C}(i))$ the irreducible bases in the $\bar{G} \supset \bar{G}_i$ classification.

Summary: For a nonregular reducible space, not every $\bar{C}(i)$ in the set of operators $\bar{C}(s) = (\bar{C}(1), \bar{C}(2),...)$ has a definite meaning, but only the CSCO of some particular intrinsic subgroups determined by the symmetry property of the intrinsic state have a definite meaning. The remaining operator set after discarding the meaningless operators in $\bar{C}(s)$ will be denoted by $\bar{C}(s')$. The eigenequations satisfied by the irredubible bases, instead of Eq. (96a), should be modified to

$$\begin{pmatrix} C \\ C(s) \\ \bar{C}(s') \end{pmatrix} \psi_m^{(\nu)\kappa} = \begin{pmatrix} \nu \\ m \\ \kappa \end{pmatrix} \psi_m^{(\nu)\kappa}, \tag{129a}$$

$$\psi_m^{(\nu)\kappa} = \sum_{a=1}^{f} U_{m,a}^{(\nu)\kappa} \varphi_a, \tag{129b}$$

$$\langle \psi_m^{(\nu)\kappa} | \psi_{m'}^{(\nu')\kappa'} \rangle = \delta_{\nu\nu'}\delta_{mm'}\delta_{\kappa\kappa'}, \tag{129c}$$

where $\kappa = \kappa_1, \kappa_2....\kappa_{\tau_\nu}$ again serves as the additional quantum number to distinguish the $1 < \tau_\nu < h_\nu$ equivalent irreps.

9. THE STATE PERMUTATION GROUP

We adopt the definition of the state permutation group $\mathcal{S}_f$ given by Bohr and Mottelson.[9] For a configuration $(\alpha_1)^{f_1}(\alpha_2)^{f_2}...(\alpha_n)^{f_n}$, with $f_1 + f_2 + ...f_n = f$, the n single particle states $\alpha_1,\alpha_2,...,\alpha_n$ are assigned to the state indices $i_1 i_2...i_f$ in the following way:

$$i_1 = ... = i_{f_1} = \alpha_1,$$

$$i_{f_1+1} = ... = i_{f_1+f_2} = \alpha_2,...i_{f-f_n+1} = ... = i_f = \alpha_n.$$

The permutation operator $\mathcal{P}$ of the group $\mathcal{S}(f)$ is defined as [It differs from the definition given by Hamermesh (Ref. 2, Sec. 10-2). What Hamermesh defined as the so-called permutation operator on state indices is in fact the inverse operator of the permutation operator on coordinate indices.]

$$\mathcal{P}|i_1 i_2...i_f\rangle = |i_{p(1)} i_{p(2)}...i_{p(f)}\rangle, \tag{130}$$

where $|i_1 i_2...i_f\rangle$ is an f-particle product state [similar to Eq. (127)]. In other words, the permutation operator $\mathcal{P}$ permutes the state indices, in contrast to the permutation operator p of $S(f)$, which permutes the coordinate indices.

It is straightforward to prove that the state permutation group has the properties:

(a) $\mathcal{S}(f)$ and $S(f)$ are commutative,

(b) $\mathcal{S}(f)$ and $S(f)$ are isomorphic.

The f-particle product state,

$$|\omega_0\rangle = |i_1 i_2...i_f\rangle = \left| \overbrace{\alpha_1...\alpha_1}^{f_1} \cdots \overbrace{\alpha_n...\alpha_n}^{f_n} \right\rangle, \tag{131}$$

will be called the "normal order state." It is convenient to choose the normal order state as the intrinsic state of the permutation group $S(f)$, since under such a choice, the intrinsic permutation operator is equal to the inverse operator of the state permutaion.

Proof: Let φ_a be a state resulting from applying an arbitrary permutation p_a of $S(f)$ to $|\omega_0\rangle$,

$$\varphi_a = p_a|\omega_0\rangle. \tag{132}$$

TABLE V. The SU(3) wave functions of the elementary particles in the flavor space.

ν,m,k		$\|uds\rangle$	$\|dus\rangle$	$\|sdu\rangle$	$\|usd\rangle$	$\|sud\rangle$	$\|dsu\rangle$
3,1,1	$\|1\ 2\ 3,\Sigma^{*\circ}\rangle$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$
0,1,1	$\left\|\begin{matrix}1&2\\3\end{matrix},\Sigma^{\circ}\right\rangle$	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	$-\frac{1}{\sqrt{12}}$	$-\frac{1}{\sqrt{12}}$	$-\frac{1}{\sqrt{12}}$	$-\frac{1}{\sqrt{12}}$
0, −1,1	$\left\|\begin{matrix}1&3\\2\end{matrix},\Sigma^{\circ}\right\rangle$	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
0,1, −1	$\left\|\begin{matrix}1&2\\3\end{matrix},\Lambda^{\circ}\right\rangle$	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
0, −1, −1	$\left\|\begin{matrix}1&3\\2\end{matrix},\Lambda^{\circ}\right\rangle$	$\frac{1}{\sqrt{3}}$	$-\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{12}}$	$\frac{1}{\sqrt{12}}$	$-\frac{1}{\sqrt{12}}$	$-\frac{1}{\sqrt{12}}$
−3, −1, −1	$\left\|\begin{matrix}1\\2,[0]\\3\end{matrix}\right\rangle$	$\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$	$-\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$	$\frac{1}{\sqrt{6}}$

Applying an intrinsic permutation operator $\bar{p}$ to φ_a, one has

$$\bar{p}\varphi_a = \bar{p}p_a|\omega_0\rangle = p_a p|\omega_0\rangle$$
$$= p_a \not{r}^{-1}|\omega_0\rangle = \not{r}^{-1}p_a|\omega_0\rangle = \not{r}^{-1}\varphi_a, \tag{133}$$

where we used Eq. (115) and the relation $p|\omega_0\rangle = \not{r}^{-1}|\omega_0\rangle$. Since φ_a is arbitrary, from Eq. (133) it follows that

$$\bar{p} = \not{r}^{-1}. \tag{134}$$

Therefore, the state permutation group $\mathscr{S}(f)$ is a realization of the intrinsic permutation group $\bar{S}(f)$. In the f-particle product function space, it is more convenient to use the state permutation group than to use the intrinsic permutation group, since the operator $\not{r}$ is easier to manipulate than the operator $\bar{p}$.

From Eq. (134), we know that the class operators of $\mathscr{S}(n)$ and those of $\bar{S}(n)$ are equal, therefore

$$\mathscr{C}(s) = \bar{C}(s), \tag{135a}$$

$$\mathscr{C}(s) = (\mathscr{C}_{f-1},\cdots,\mathscr{C}_2), \tag{135b}$$

where $\mathscr{C}_i$ is the 2-cycle class operator of $\mathscr{S}(i)$.

All the formulas in Sec. 7 are valid after the following substitution:

$$R_a \to \varphi_a, \quad \mathring{p}^{(\nu)k}_m \to \psi^{(\nu)k}_m, \tag{136}$$
$$\bar{C}(s) \to \mathscr{C}(s),$$

and Eq. (98b) is changed into

$$\not{r}\psi^{(\nu)k}_m = \sum_{k'} D^{(\nu)}_{k'k}(p)\psi^{(\nu)k'}_m. \tag{137}$$

Example: Consider a 3-quark system with three flavor states u, d, s. The order of the states is taken as $i_1 = u, i_2 = d$, and $i_3 = s$. This corresponds to the so-called I-spin representation. By diagonalizing the operator $K = 3C_3 + 2C_2 + \mathscr{C}_2$ in the regular representation spanned by $\{p|uds\rangle\}$, we obtain the six irreducible bases listed in Table V. The values of the coefficients in Table V are of course identical to the eigenvectors **D** in Eq. (113), except for the factor $(h_\nu/g)^{1/2}$ as can be seen from Eqs. (120) and (96b).

10. THE QUASISTANDARD BASIS OF THE PERMUTATION GROUP

In Sec. 8 we showed that for nonregular representations, the single intrinsic group element lost meaning, but under the condition imposed by Eq. (123), the CSCO of certain intrinsic subgroup still have a definite meaning. Since the nonregular representation spanned by the f-particle product wave functions $\{p|\omega_0\rangle\}$ in the configuration $(\alpha_1)^{f_1}(\alpha_2)^{f_2}\cdots(\alpha_n)^{f_n}$ satisfy this condition, furthermore the intrinsic permutation group and the state permutation group are anti-isomorphic, thus a single element $\not{r}$ of the state permutation group also has no definite meaning, but the CSCO of the subgroups contained in the group chain

$$\mathscr{S}(f) \supset \mathscr{S}(f-f_n) \supset \cdots \supset \mathscr{S}(f_1)$$

have a definite meaning. The set of these meaningful CSCO will be denoted as

$$(\mathscr{C}, \mathscr{C}(s')) = (\mathscr{C}(f), \mathscr{C}(f-f_n),\cdots,\mathscr{C}(2)), \tag{138}$$

$\mathscr{C}(i)$ being the CSCO-I of $\mathscr{S}(i)$. In other words, the result of the action of these operators is independent of which α_1 is assigned as $i_1,\cdots$ which α_1 is assigned as i_{f_1}, and which α_2 is assigned as $i_{f_1+1}\ldots$, which α_2 is assigned as $i_{f_1+f_2},\ldots$etc. (there are f_i particles in the single particle state α_i). For example, for the $\alpha(\beta)^2\gamma$ configuration ($\alpha = \alpha_1, \beta = \alpha_2, \gamma = \alpha_3$),

$$\mathscr{C}(3)|\beta\alpha\beta\gamma\rangle = \mathscr{C}(3)|i_2 i_1 i_3 i_4\rangle = \mathscr{C}(3)|i_3 i_1 i_2 i_4\rangle$$
$$= |\beta\beta\alpha\gamma\rangle + |\beta\alpha\beta\gamma\rangle + |\alpha\beta\beta\gamma\rangle. \tag{139}$$

Equation (139) shows that the action of $\mathscr{C}(3)$ is independent of which β is assigned as i_2 or i_3.

Definition: The simultaneous eigenfunction $\psi^{(\nu)}_\kappa$ of the set of operators $(\mathscr{C}(f), \mathscr{C}(s'))$ is called the quasistandard basis of the permutation group.

$$\begin{pmatrix}\mathscr{C}(f)\\ \mathscr{C}(s')\end{pmatrix}\psi^{(\nu)}_\kappa = \begin{pmatrix}\nu\\ \kappa\end{pmatrix}\psi^{(\nu)}_\kappa,$$

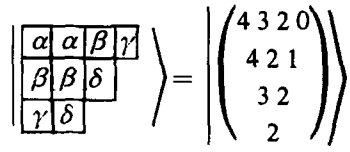
$$\nu = \lambda(f), \quad \kappa = (\lambda(f-f_n)\cdots\lambda(f_1+f_2), \lambda(f_1)). \tag{140}$$

$\psi^{(\nu)}_\kappa$ belongs to the irrep $\lambda(f), \lambda(f-f_n),\ldots,\lambda(f_1)$ of the group chain $\mathscr{S}(f)\supset\mathscr{S}(f-f_n)\supset\ldots\supset\mathscr{S}(f_1)$, in contrast to the standard (i.e., Yamanouchi) basis $\psi^{(\nu)}_m$ which belongs to the irrep $\lambda(f), \lambda(f-1),\ldots,\lambda(2)$ of the canonical group chain $\mathscr{S}(f)\supset\mathscr{S}(f-1)\supset\ldots\supset\mathscr{S}(2)$.

It was proved in Ref. 15 that the quasistandard bases of the permutation group are just the Gel'fand bases of the uni-

tary group. For example,

Weyl tableau[16] Gel'fand symbol[17]



belongs to

$\mathscr{S}(9)$	$\supset$	$\mathscr{S}(7)$	$\supset$	$\mathscr{S}(5)$	$\supset$	$\mathscr{S}(2)$
SU(4)	$\supset$	SU(3)	$\supset$	SU(2)	$\supset$	U(1)
[432]		[421]		[32]		[2].

Therefore, the eigenfunction method can be used to calculate the Gel'fand bases of the unitary group.

When all the single particle states in a Weyl tableau are different, the Gel'fand basis becomes the special Gel'fand basis, and the quasistandard basis becomes the Yamanouchi basis. It means that the Yamanouchi basis of the permutation group is just the special Gel'fand basis of the unitary group, a fact proved by Moshinsky[18] in 1966.

11. CONCLUSIONS

Using the new approach to the permutation group representation, the calculation of characters CFP, CGC, the Yamanouchi basis, and irreducible matrix elements of the permutation group, and the Gel'fand bases of the unitary group, etc., are all reduced to the problem of solving the eigenequation of a certain single operator. With the help of the traditional theory of the permutation group, we can know the eigenvalues of these operators prior to the concrete calculation. Therefore, all one has to do is to solve a set of linerly homogeneous algebraic equations, which is very easy to calculate.

The eigenfunction methods for calculating the $SU(mn)\supset SU(m)\times SU(n)$ isoscalar factors, $SU(m+n)\supset SU(m)\otimes SU(n)$ isoscalar factors, and the transformation coefficients from the Yamanouchi basis of $S(f)$ to the $S(f)\supset S(f_1)\times S(f_2)$ basis with $f=f_1+f_2$ are discussed separately in Refs. 19–21.

ACKNOWLEDGMENT

One of the authors (Chen) would like to express his gratitude to Drs. Da Hsuan Feng, and Michel Vallieres for many helpful discussions and for reading the manuscript, and to Professor B. F. Bayman and Professor M. Hamermesh for their interest in this work and many valuable suggestions. Financial support from Drexel University is acknowledged.

[1] [(a)]Jin-Quan Chen, Fan Wang, and Mei-Juan Gao, Acta Phys. Sinica (in Chinese) **26**, 307 (1977). [(b)]Jin-Quan Chen, Fan Wang, and Mei-Juan Gao, "Intrinsic Lie Groups and Nuclear Collective Rotation About Intrinsic Axes" (to be published).

[2] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison–Wesley, Reading, MA, 1962).

[3] D. E. Rutherfold, *Substitutional Analysis* (Hafner, New York, 1968).

[4] A. V. Soklov and V. P. Shirokovsky, Usp. Fiz. Nauk. **60**, 617 (1956).

[5] H. Boerner, *Group Representation* (North-Holland, Amsterdam, 1963).

[6] P. A. M. Dirac, *The Principle of Quantum Mechanics* (Oxford, U.P., Oxford 1958).

[7] Jin-Quan Chen, "A New Approach to Group Representation Theory," Lecture Notes (to be published).

[8] A. Partensky, J. Math. Phys. **13**, 1503 (1972).

[9] A. Bohr and B. R. Mottelson, *Nuclear Structure*, Vol. 1 (Benjamin, New York, 1969).

[10] H. Jones, *The Theory of Brillouin Zones* (North-Holland, Amsterdam, 1960), p. 76.

[11] Jin-Quan Chen and Mei-Juan Gao, *Reduction Coefficients of Permutation Groups and Their Applications*, (Beijing Science Pub. Co., 1981).

[12] C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).

[13] H. A. Jahn, Proc. R. Soc. London Ser. A **209**, 502 (1951).

[14] S. I. So and D. Strottman, J. Math. Phys. **20**, 153 (1979).

[15] Jin-Quan Chen, Fan Wang, and Mei-Juan Gao, Acta Phys. Sinica **26**, 427 (1977).

[16] J. Paldus, *Theoretical Chemistry*, Vol. 2, Advances and Perspective (Academic, New York, 1976), p. 167.

[17] I. M. Gel'fand and M. L. Zetlin, Dok. Akad. Nauk. SSSR **71**, 825 (1950).

[18] M. Moshinsky, J. Math. Phys. **7**, 691 (1966).

[19] Jin-Quan Chen, J. Math. Phys. **22**, 1 (1981).

[20] Jin-Quan Chen (submitted to J. Math. Phys.).

[21] Jin-Quan Chen *et al.* (in preparation).

Structure and representations of the symmetry group of the four-dimensional cube

M. Baake, B. Gemünden, and R. Oedingen
Physikalisches Institut, Universität Bonn, D-5300 Bonn 1, West Germany

(Received 6 August 1981; accepted for publication 18 December 1981)

In this paper we give, explicitly, the description of the structure, the characters, and the complete system of irreducible representations of the hyperoctahedral group in four dimensions, which we call W_4 (from the German "Würfel"). In a second step, we do the same for the subgroup SW_4, which is formed by all pure rotations contained in W_4.

PACS numbers: 02.20.Qs, 02.20.Rt

INTRODUCTION

In the last years a lot of work has been done on four-dimensional lattices in order to get discrete approximations of the continuous case. It is an important problem to find all symmetries of the regarded lattices, the complete system of all irreducible representations of the appropriate symmetry groups, and their classification. For the connection between abstract groups and the symmetries of (discrete) point sets, especially lattices, see Chap. 1.4 of Miller.[1]

In our case we will consider the hypercubical lattice, generated by

$$\mathbf{e}_1 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix}0\\1\\0\\0\end{pmatrix},$$

$$\mathbf{e}_3 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad \mathbf{e}_4 = \begin{pmatrix}1\\0\\0\\1\end{pmatrix}.$$

Every point of the lattice can be expressed as an integral linear combination of these four vectors. We shall restrict ourselves to the symmetry transformations of the lattice which leave the origin $\mathbf{0} = (0, 0, 0, 0)'$ fixed. It should be noted that unlike the three-dimensional case the hypercubical lattice does not have the largest symmetry of all possible lattices in four dimensions. The interested reader is referred to Chen and Birman.[2] In spite of this fact we confine ourselves to the symmetry group of the hypercubical lattice, because the symmetry group of the "larger" lattice is just an extension of the group mentioned here. The "larger" lattice is generated by

$$\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \text{ and } \mathbf{f} = (1/2, 1/2, 1/2, 1/2)'.$$

It turns out that the symmetry group of the four-dimensional cube W_4, which is obviously a subgroup of O(4), is the "largest" crystallographic point group of the hypercubical lattice at the point **0**. Therefore, W_4 is called the hypercubical holohedry at the point **0** (cf. Chap. 2.7 of Miller[1]). Clearly, W_4 has been well known for many years and has frequently been discussed by several authors. Some of them mention W_4 while classifying crystallographic point groups in four dimensions (Goursat,[3] Hurley,[4] and Chen and Birman[2]).

Others found W_4 by working on wreath products and generalized symmetric groups (Young,[5] Robinson,[6] Frame,[7] Kerber,[8,9] Maier,[10,11] and Osima[12,13]). However, none of them gives an explicit description of group characters and irreducible representations; only Chen and Birman give the character table of W_4 obtained by a computational method. The subgroup SW_4 is a crystallographic point group, too, and therefore mentioned by Hurley,[4] but there is no comment on its structure. As SW_4 cannot be regarded as a generalized symmetric group, it does not appear in papers about this subject. Our main results are listed in several tables at the end of this paper. The relation between the irreducible representations of W_4, resp. SW_4, and those of the full orthogonal group O(4), resp. the special orthogonal group SO(4), will be given in another paper together with an elementary classification of the representations of O(4) and SO(4) (Baake, Gemünden, and Oedingen[14]).

Obviously, quite a lot of our results are already known. Especially, some of them can be found in the paper of Chen and Birman,[2] too. But there is a remarkable difference between the methods used. First of all, our method can easily be generalized for the analogous symmetry observations in n dimensions and, for example, the description of the symmetry group of the three-dimensional cube, $O_h \cong W_3$, becomes very elementary.

Then, by the connections of W_4 to the symmetric group S_8, which are pointed out in Sec. II and can also be generalized for arbitrary dimensions, it is very simple to apply this group to special symmetry problems, using a computer.

Last but not least, by our considerations we straightforwardly obtain minimal systems of generators for W_4 and SW_4, i.e., we get three generators for W_4 (Chen and Birman[2] use nine of them) and two generators for SW_4.

I. PRELIMINARIES

In our notation we follow Miller.[1] Especially, we denote the symmetric group of degree n by S_n, and the cyclic group of order two by Z_2 together with the symbol "$+_2$" for addition modulo 2. Furthermore, we use $Z_2^4 = Z_2 \otimes Z_2 \otimes Z_2 \otimes Z_2$ in the form

$$Z_2^4 = \left\{ \mathbf{a} = \begin{pmatrix}a_1\\a_2\\a_3\\a_4\end{pmatrix} \middle| a_1, a_2, a_3, a_4 \in Z_2 \right\}$$

 0022-2488/82/060944-10$02.50

e_1	$-e_1$
e_2	$-e_2$
e_3	$-e_3$
e_4	$-e_4$

FIG. 1.

1	5
2	6
3	7
4	8

FIG. 2.

together with

$$\mathbf{a} +_2 \mathbf{b} := \begin{pmatrix} a_1 & +_2 & b_1 \\ a_2 & +_2 & b_2 \\ a_3 & +_2 & b_3 \\ a_4 & +_2 & b_4 \end{pmatrix}.$$

Finally, if the symbol " $=$ " is used for representations, this only means the equivalence of the representations and not actual identity.

II. STRUCTURE OF W_4

Let us consider the following discrete subset L_4 of the four-dimensional Euclidean space $\mathbb{R}^4$:

$$L_4 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}; \right\}$$

$$= \{\mathbf{e}_1; -\mathbf{e}_1; \mathbf{e}_2, -\mathbf{e}_2, \mathbf{e}_3, -\mathbf{e}_3, \mathbf{e}_4, -\mathbf{e}_4\}.$$

We call two elements $\mathbf{x}$, $\mathbf{y}$ opposite, if

$$\mathbf{y} = -\mathbf{x}.$$

In this paper we consider the group W_4 of permutations of the elements of L_4, which leave opposite elements opposite. This group turns out to be the symmetry group of the four-dimensional cube. This can be seen if one regards the elements of L_4 as the centers of the eight faces of the four-dimensional cube. This group is identical with that one formed by permutations of the diagram of Fig. 1, which obey the two following rules: (1) Two elements of the same row may be interchanged; (2) only complete rows may be permuted.

From these considerations we conclude that W_4 appears to be the wreath product $Z_2 \sim S_4$. Therefore, we consider

$$W_4 = \{(\mathbf{a}, \pi) \,|\, \mathbf{a} \in Z_2^4,\ \pi \in S_4\}$$

together with the multiplication rule

$$(\mathbf{a}, \pi)\cdot(\mathbf{b}, \sigma) := (\mathbf{a}_\sigma +_2 \mathbf{b}, \pi \circ \sigma)$$

with

$$(\mathbf{a}_\sigma)_k := a_{\sigma(k)} \quad (k \text{ th component of } \mathbf{a}_\sigma).$$

This multiplication rule produces the wreath product structure (for details see Kerber[8,9] and Marcu, Regev, and Rittenberg[15]). It is easy to check that W_4 is a finite group of order 384. Direct calculation shows that

$$C := \{(\mathbf{0}, \mathrm{id}), (\mathbf{1}, \mathrm{id})\};\ \mathbf{0} := \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix};\ \mathbf{1} := \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

is the center of W_4.

Since L_4 consists of eight elements, it is obvious that W_4 is

TABLE I. Classes of conjugate elements of W_4.

S_8 - cycles	1^8	$1^6 2$	$1^4 4$	$1^4 2^2$	$1^2 6$	$1^2 2 4$	$1^2 2^3$	$1^2 3^2$	8	4^2		$2^2 4$		26	$2 3^2$	2^4		
S_4 - cycles	1^4	1^4	$1^2 2$	$1^2 2$	1^4	13	$1^2 2$	$1^2 2$	1^4	13	4	4	2^2	2^2	$1^2 2$	13	13	2^2	$1^2 2$	1^4
order	1	4	12	12	6	32	24	24	4	32	48	48	12	24	12	32	32	12	12	1

TABLE II. Number of elements of a given order for the group W_4.

order	1	2	3	4	5	6	7	8
number	1	75	32	132	0	96	0	48

isomorphic with a subgroup of S_8. If we label the squares of Fig. 1 as shown in Fig. 2, and admit the same manipulations as explained for Fig. 1, we get the following subgroup of S_8:

$$P_4 = \left\{ \Pi \in S_8 \,\middle|\, \bigwedge_{1 \leqslant k \leqslant 4} |\Pi(k+4) - \Pi(k)| = 4 \right\}.$$

P_4 is a subgroup of S_8 of order 384. By means of the mapping

$$\Phi: W_4 \to P_4,$$

$$\Phi(\mathbf{a}, \pi)(k) := \begin{cases} \pi(k) + 4a_k & \text{if} \quad 1 \leqslant k \leqslant 4, \\ \pi(k-4) + 4(1 - a_{k-4}) & \text{if} \quad 5 \leqslant k \leqslant 8. \end{cases}$$

one can prove W_4 and P_4 to be isomorphic. P_4 can be thought of as a representation of W_4 by permutations. We now have all information to obtain the classes of conjugate elements. W_4 contains 20 classes which are uniquely determined in the following way:

Let $(\mathbf{a}, \pi), (\mathbf{b}, \sigma) \in W_4$. Then $(\mathbf{a}, \pi) \sim_{W_4} (\mathbf{b}, \sigma)$ if and only if π and σ have the same cycle structure in S_4 and $\Phi(\mathbf{a}, \pi)$ and $\Phi(\mathbf{b}, \sigma)$ have the same cycle structure in S_8. Calculating, moreover, the order of the classes of W_4, one obtains Table I. For a permutation $\pi \in S_8$ with cycle structure

$$(m_1^{\mu_1}, \ldots, m_l^{\mu_l}), \quad \bigwedge_{1 \leqslant k \leqslant l} \mu_k \neq 0, \text{ you have}$$

$$\operatorname{ord}\pi = \operatorname{lcm}\{m_1, \ldots, m_l\},$$

i.e.,

$$\bigwedge_{1 \leqslant n < \operatorname{ord}\pi} \pi^n \neq \mathrm{id}, \quad \pi^{\operatorname{ord}\pi} = \mathrm{id}.$$

Since conjugate elements have the same order, we obtain Table II, containing the numbers of elements of W_4 of a given order. Table II contains the special result that there is no element of an order higher than 8. With the aid of Table II one can easily compute the number of unequal cyclic subgroups of W_4 (see Table III).

Generalizing the geometrical concept of "axes" in three dimensions, we now define axes as maximal cyclic subgroups, i.e., all those cyclic subgroups that are not contained in any larger one. After the calculation of their numbers we obtain Table IV.

III. REPRESENTATIONS OF W_4

The wreath product structure of W_4 implies the following definition of a four-dimensional representation:

$$T: W_4 \to \operatorname{aut}\mathbb{R}^4$$

where

$$T(\mathbf{a}, \pi)\mathbf{e}_i := (-1)^{a_i}\mathbf{e}_{\pi(i)}, \quad 1 \leqslant i \leqslant 4.$$

TABLE III. Number of cyclic subgroups of a given order for the group W_4.

order	1	2	3	4	5	6	7	8
number	1	75	16	66	0	48	0	12

TABLE IV. Number of axes of a given order for the group W_4.

order	1	2	3	4	5	6	7	8
number	0	48	0	60	0	48	0	12

The representation property is evident:

$$\begin{aligned} T(\mathbf{a}, \pi) \cdot T(\mathbf{b}, \sigma)\mathbf{e}_i &= T(\mathbf{a}, \pi)(-1)^{b_i}\mathbf{e}_{\sigma(i)} \\ &= (-1)^{b_i + 2a_{\sigma(i)}}\mathbf{e}_{\pi(\sigma(i))} \\ &= T(\mathbf{a}_\sigma +_2 \mathbf{b}, \pi\sigma)\mathbf{e}_i \\ &= T[(\mathbf{a}, \pi) \cdot (\mathbf{b}, \sigma)]\mathbf{e}_i; \end{aligned}$$

furthermore,

$$T(\mathbf{0}, \mathrm{id})\mathbf{e}_i = (-1)^0\mathbf{e}_i = \mathbf{e}_i \Rightarrow T(\mathbf{0}, \mathrm{id}) = \mathrm{Id}_{\mathbb{R}^4}.$$

This representation is faithful, orthogonal, and irreducible. It is obvious that W_4 is isomorphic with a finite subgroup of O(4). Especially, the corresponding matrices of T are elements of O(4). From $T(\mathbf{a}, \pi)\mathbf{e}_i = \sum_{j=1}^4 T_{ij}(\mathbf{a}, \pi)\mathbf{e}_j$, we obtain

$$T_{ji}(\mathbf{a}, \pi) = (-1)^{a_i}\delta_{j, \pi(i)}.$$

Since the matrix elements are all integers, it is clear that W_4 is isomorphic with a subgroup of

$$O(4, \mathbb{Z}) := \{A \in O(4) \,|\, \bigwedge_{1 \leqslant i, j \leqslant 4} A_{ij} \in \mathbb{Z}\}.$$

As $\operatorname{ord}(O(4, \mathbb{Z})) = 384$, we obtain $W_4 \cong O(4, \mathbb{Z})$.

It is possible to extend the irreducible representations of S_4 onto W_4. The multiplication rule in W_4 yields that the mapping $S: W_4 \to S_4$, $S(\mathbf{a}, \pi) := \pi$ defines a group homomorphism. Therefore, if V is a representation of S_4, $V \circ S$ is a representation of W_4. This definition yields an algorithm to transfer the S_4-characters to the character table of W_4. The classes of W_4 with the same cycle structure relative to S_4 get the appropriate S_4-character. The irreducible characters of S_4 are given in Table V.

The following one-dimensional representations are obtained:

the trivial representation	("identity"),
$(\mathbf{a}, \pi) \to \operatorname{sgn}\pi$	("signum S_4"),
$(\mathbf{a}, \pi) \to \det[T(\mathbf{a}, \pi)]$	("determinant").

We state the following properties:

$$\operatorname{sgn}[\Phi(\mathbf{a}, \pi)] = (-1)^{\sum_{i=1}^4 a_i},$$

$$\begin{aligned} \det[T(\mathbf{a}, \pi)] &= (-1)^{\sum_{i=1}^4 a_i} \cdot \operatorname{sgn}\pi \\ &= \operatorname{sgn}[\Phi(\mathbf{a}, \pi)] \cdot \operatorname{sgn}\pi. \end{aligned}$$

At this point we know eight of 20 irreducible representa-

TABLE V. Characters of S_4.

cycles	1^4	$1^2 2$	2^2	13	4
order	1	6	3	8	6
	1	1	1	1	1
	1	-1	1	1	-1
	2	0	2	-1	0
	3	1	-1	0	-1
	3	-1	-1	0	1

TABLE VI. Characters of W_4.

S-8-cycles	1^8	$1^6 2$	$1^4 4$	$1^4 2^2$	1^6	$1^2 2^4$	$1^2 2^3$	$1^2 3^2$	8	4^2		$2^2 4$		26	2^3	2^4		
S-4-cycles	1^4	1^4	$1^2 2$	$1^2 2$	1^4	13	$1^2 2$	$1^2 2$	1^4	13	4	4	2^2	2^2	$1^2 2$	13	13	2^2	$1^2 2$	1^4
order	1	4	12	12	6	32	24	24	4	32	48	48	12	24	12	32	32	12	12	1
$\chi_1^{(1)}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\chi_2^{(1)}$	1	-1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	-1	1	-1	1	1	1
$\chi_3^{(1)}$	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	1	1	1	-1	1
$\chi_4^{(1)}$	1	-1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	1	1	-1	1	-1	1
$\chi_1^{(2)}$	2	2	0	0	2	-1	0	0	2	-1	0	0	2	2	0	-1	-1	2	0	2
$\chi_2^{(2)}$	2	-2	0	0	2	1	0	0	-2	-1	0	0	2	-2	0	-1	1	2	0	2
$\chi_1^{(3)}$	3	3	1	1	3	0	1	1	3	0	-1	-1	-1	-1	1	0	0	-1	1	3
$\chi_2^{(3)}$	3	-3	-1	1	3	0	1	-1	-3	0	1	-1	-1	1	-1	0	0	-1	1	3
$\chi_3^{(3)}$	3	3	-1	-1	3	0	-1	-1	3	0	1	1	-1	-1	-1	0	0	-1	-1	3
$\chi_4^{(3)}$	3	-3	1	-1	3	0	-1	1	-3	0	-1	1	-1	1	1	0	0	-1	-1	3
$\chi_1^{(4)}$	4	2	2	2	0	1	0	0	-2	1	0	0	0	0	-2	-1	-1	0	-2	-4
$\chi_2^{(4)}$	4	-2	-2	2	0	-1	0	0	2	1	0	0	0	0	2	-1	1	0	-2	-4
$\chi_3^{(4)}$	4	2	-2	-2	0	1	0	0	-2	1	0	0	0	0	2	-1	-1	0	2	-4
$\chi_4^{(4)}$	4	-2	2	-2	0	-1	0	0	2	1	0	0	0	0	-2	-1	1	0	2	-4
$\chi_1^{(6)}$	6	0	2	0	-2	0	0	-2	0	0	0	0	2	0	2	0	0	-2	0	6
$\chi_2^{(6)}$	6	0	-2	0	-2	0	0	2	0	0	0	0	2	0	-2	0	0	-2	0	6
$\chi_3^{(6)}$	6	0	0	2	-2	0	-2	0	0	0	0	0	-2	0	0	0	0	2	2	6
$\chi_4^{(6)}$	6	0	0	-2	-2	0	2	0	0	0	0	0	-2	0	0	0	0	2	-2	6
$\chi_1^{(8)}$	8	4	0	0	0	-1	0	0	-4	-1	0	0	0	0	0	1	1	0	0	-8
$\chi_2^{(8)}$	8	-4	0	0	0	1	0	0	4	-1	0	0	0	0	0	1	-1	0	0	-8

tions. We calculate their characters and, furthermore, the characters of the Kronecker products of all possible pairs of representations.

Besides reducible representations one obtains eight characters of irreducible representations which are inequivalent to all those that are already known. Their dimensions are

2 (single); 3 (two times); 4 (three times); 8 (two times).

Four representations are still missing; they can be obtained by reducing the twofold Kronecker product of the four-dimensional canonical representation T. By this method we get two inequivalent six-dimensional representations, one three-dimensional and one one-dimensional representation, the last two representations being already known. The two representations still missing are obtained by multiplication of the six-dimensional representations with the one-dimensional ones. Now we have all data for the character table (Table VI).

Obviously, this character table is, except for arrangement, identical with Table III of Chen and Birman,[2] but, for a complete presentation, we give this table once more in our notation. Last but not least, this table is the straightforward consequence of the method described above.

In the following we list and explain the symbols of the character table, and state some properties of the characters.

The characters are denoted by $\chi_k^{(n)}$ where the upper index, n, indicates the degree of the appropriate representation and the lower one, k, enumerates the characters of different representations of the same degree.

$\chi_1^{(1)}$: "identity,"

$\chi_2^{(1)}$: "signum S_8,"

$\chi_3^{(1)}$: "signum S_4,"

$\chi_4^{(1)}$: "determinant,"

$\chi_4^{(1)} = \chi_2^{(1)} \cdot \chi_3^{(1)}$.

Note: The characters of the one-dimensional representations are the representations themselves:

$\chi_1^{(2)}, \chi_1^{(3)}$, and $\chi_3^{(3)}$ are deduced from the character table of S_4;

$\chi_3^{(3)} = \chi_1^{(3)} \cdot \chi_3^{(1)}$;

$\chi_1^{(4)} = \mathrm{tr} T$;

$\chi_1^{(8)} = \chi_1^{(4)} \cdot \chi_1^{(2)}$.

All other characters of degree 2, 3, 4, and 8 can be received by the multiplication rule

$\chi_k^{(n)} = \chi_1^{(n)} \cdot \chi_k^{(1)}$:

$\chi_1^{(6)}$ is deduced from the Kronecker product $T \otimes T$ (skew-symmetric part);

$\chi_2^{(6)} = \chi_1^{(6)} \cdot \chi_2^{(1)}$;

$\chi_3^{(6)}$ is also deduced from $T \otimes T$ (symmetric part);

$\chi_4^{(6)} = \chi_3^{(6)} \cdot \chi_4^{(1)}$.

IV. KRONECKER PRODUCTS

We label the representation belonging to the character $\chi_k^{(l)}$ (which is uniquely determined, except for equivalence) as $\tau_k^{(l)}$. In this notation the four-dimensional representation T is labeled as $\tau_1^{(4)}$. With the aid of the character table we can decide all four-dimensional and eight-dimensional representations—and only these—to be faithful.

TABLE VII. Multiplicity of irreducible representations in n-fold Kronecker products of T with itself, $1 \leqslant n \leqslant 10$, for the group W_4.

	$\tau_1^{(1)}$	$\tau_2^{(1)}$	$\tau_3^{(1)}$	$\tau_4^{(1)}$	$\tau_1^{(2)}$	$\tau_2^{(2)}$	$\tau_1^{(3)}$	$\tau_2^{(3)}$	$\tau_3^{(3)}$	$\tau_4^{(3)}$	$\tau_1^{(4)}$	$\tau_2^{(4)}$	$\tau_3^{(4)}$	$\tau_4^{(4)}$	$\tau_1^{(6)}$	$\tau_2^{(6)}$	$\tau_3^{(6)}$	$\tau_4^{(6)}$	$\tau_1^{(8)}$	$\tau_2^{(8)}$
$\tau_1^{(4)} = T$	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
$\otimes^2 T$	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0
$\otimes^3 T$	0	0	0	0	0	0	0	0	0	0	4	1	0	1	0	0	0	0	3	2
$\otimes^4 T$	4	1	0	1	3	2	7	3	3	3	0	0	0	0	10	6	10	6	0	0
$\otimes^5 T$	0	0	0	0	0	0	0	0	0	0	31	20	15	20	0	0	0	0	45	40
$\otimes^6 T$	31	20	15	20	45	40	76	60	60	60	0	0	0	0	136	120	136	120	0	0
$\otimes^7 T$	0	0	0	0	0	0	0	0	0	0	379	336	315	336	0	0	0	0	693	672
$\otimes^8 T$	379	336	315	336	693	672	1072	1008	1008	1008	0	0	0	0	2080	2016	2080	2016	0	0
$\otimes^9 T$	0	0	0	0	0	0	0	0	0	0	5611	5440	5355	5440	0	0	0	0	10965	10880
$\otimes^{10} T$	5611	5440	5355	5440	10965	10880	16576	16320	16320	16320	0	0	0	0	32896	32640	32896	32640	0	0

By a theorem of Burnside and Brauer (see Isaacs[16]) it is possible to obtain all irreducible representations by decomposing multiple Kronecker products of an arbitrary faithful representation with itself. We have done this with the canonical representation $T = \tau_1^{(4)}$ by means of the characters. The result is listed in Table VII.

Additionally, we give decomposition rules for the reduction of twofold Kronecker products of arbitrary irreducible representations:

$$\tau_1^{(1)} \otimes \tau_k^{(n)} = \tau_k^{(n)} \quad \text{for all possible } n,k;$$

$$\tau_k^{(1)} \otimes \tau_k^{(1)} = \tau_1^{(1)}, \quad 1 \leqslant k \leqslant 4;$$

$$\tau_2^{(1)} \otimes \tau_3^{(1)} = \tau_3^{(1)} \otimes \tau_2^{(1)} = \tau_4^{(1)};$$

$$\tau_2^{(1)} \otimes \tau_4^{(1)} = \tau_4^{(1)} \otimes \tau_2^{(1)} = \tau_3^{(1)};$$

$$\tau_3^{(1)} \otimes \tau_4^{(1)} = \tau_4^{(1)} \otimes \tau_3^{(1)} = \tau_2^{(1)}.$$

We know that

$$\tau_k^{(n)} = \tau_1^{(n)} \otimes \tau_k^{(1)} \quad \text{for } n \neq 6,$$

TABLE VIII. Decomposition of mixed Kronecker products for the group W_4.

	$\tau_1^{(1)}$	$\tau_2^{(1)}$	$\tau_3^{(1)}$	$\tau_4^{(1)}$	$\tau_1^{(2)}$	$\tau_2^{(2)}$	$\tau_1^{(3)}$	$\tau_2^{(3)}$	$\tau_3^{(3)}$	$\tau_4^{(3)}$	$\tau_1^{(4)}$	$\tau_2^{(4)}$	$\tau_3^{(4)}$	$\tau_4^{(4)}$	$\tau_1^{(6)}$	$\tau_2^{(6)}$	$\tau_3^{(6)}$	$\tau_4^{(6)}$	$\tau_1^{(8)}$	$\tau_2^{(8)}$
$\tau_1^{(2)} \otimes \tau_1^{(2)}$	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\tau_1^{(2)} \otimes \tau_1^{(3)}$	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
$\tau_1^{(2)} \otimes \tau_1^{(4)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
$\tau_1^{(2)} \otimes \tau_1^{(6)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
$\tau_1^{(2)} \otimes \tau_3^{(6)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
$\tau_1^{(2)} \otimes \tau_1^{(8)}$	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0
$\tau_1^{(3)} \otimes \tau_1^{(3)}$	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
$\tau_1^{(3)} \otimes \tau_1^{(4)}$	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
$\tau_1^{(3)} \otimes \tau_1^{(6)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0
$\tau_1^{(3)} \otimes \tau_3^{(6)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
$\tau^{(3)} \otimes \tau^{(8)}$	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2	0
$\tau_1^{(4)} \otimes \tau_1^{(4)}$	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0
$\tau_1^{(4)} \otimes \tau_1^{(6)}$	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1
$\tau_1^{(4)} \otimes \tau_3^{(6)}$	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1
$\tau_1^{(4)} \otimes \tau_1^{(8)}$	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1	1	1	1	0	0
$\tau_1^{(6)} \otimes \tau_1^{(6)}$	1	0	0	1	1	1	1	0	0	1	0	0	0	0	1	1	1	1	0	0
$\tau_1^{(6)} \otimes \tau_3^{(6)}$	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
$\tau_1^{(6)} \otimes \tau_1^{(8)}$	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
$\tau_3^{(6)} \otimes \tau_3^{(6)}$	1	1	0	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0
$\tau_3^{(6)} \otimes \tau_1^{(8)}$	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	2	2
$\tau_1^{(8)} \otimes \tau_1^{(8)}$	1	0	1	0	1	0	2	0	2	0	0	0	0	0	2	2	2	2	0	0

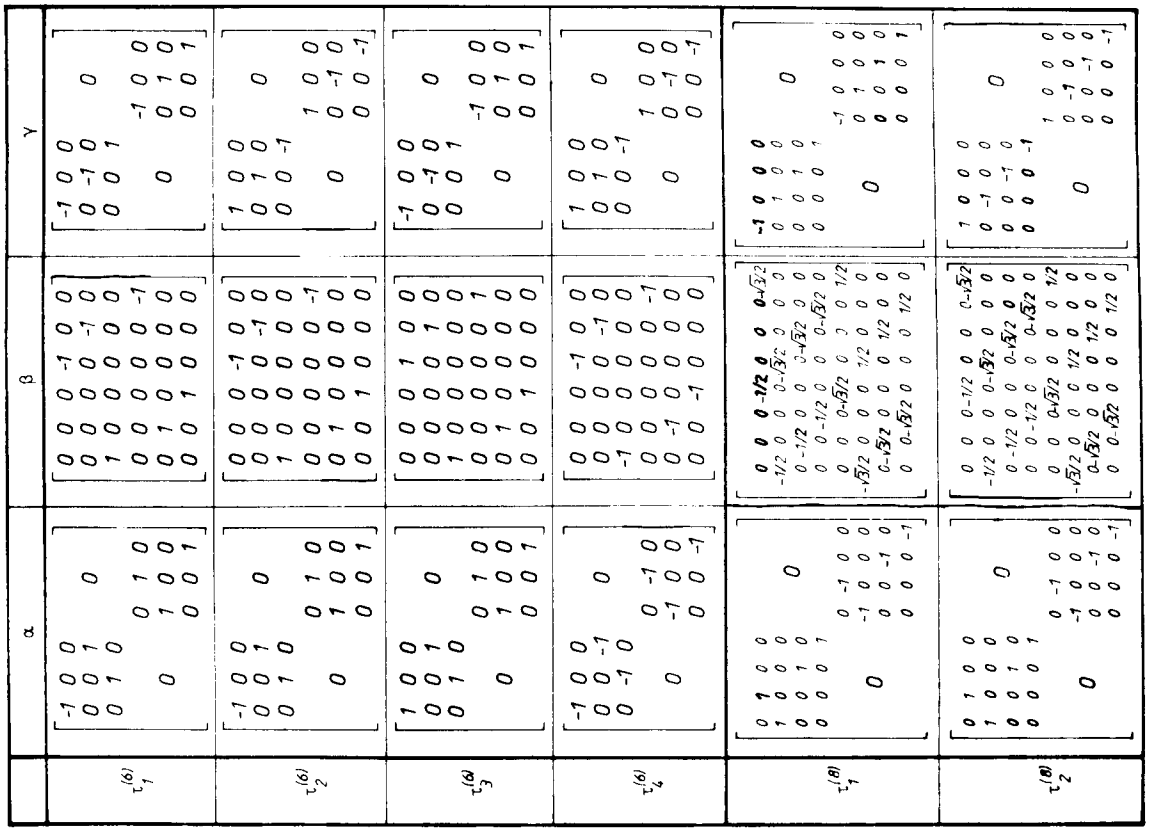
TABLE IXA. Connections between different generators of W_4.

Presentation as Generator-Product	Wreath Product	Representation Matrix	Generators in the Notation of Chen & Birman
α^2	$\left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1234\\1234\end{pmatrix}\right)$	$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	I
$(\beta\gamma)^4$	$\left(\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1234\\1234\end{pmatrix}\right)$	$\begin{bmatrix}-1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix}$	I'
$(\beta\gamma)^2\beta^2$	$\left(\begin{pmatrix}0\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1234\\1234\end{pmatrix}\right)$	$\begin{bmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&1\end{bmatrix}$	T
$(\gamma\beta)^2(\beta\gamma)^2$	$\left(\begin{pmatrix}0\\1\\0\\1\end{pmatrix}, \begin{pmatrix}1234\\1234\end{pmatrix}\right)$	$\begin{bmatrix}1&0&0&0\\0&-1&0&0\\0&0&1&0\\0&0&0&-1\end{bmatrix}$	R
$(\beta\gamma)^2\beta^2\alpha\beta$	$\left(\begin{pmatrix}0\\1\\0\\1\end{pmatrix}, \begin{pmatrix}1234\\1342\end{pmatrix}\right)$	$\begin{bmatrix}1&0&0&0\\0&0&0&-1\\0&-1&0&0\\0&0&1&0\end{bmatrix}$	S
$\beta\gamma(\beta\alpha)^2$	$\left(\begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \begin{pmatrix}1234\\1324\end{pmatrix}\right)$	$\begin{bmatrix}1&0&0&0\\0&0&-1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}$	W
$\alpha(\gamma\beta)^2\alpha$	$\left(\begin{pmatrix}0\\0\\1\\1\end{pmatrix}, \begin{pmatrix}1234\\4321\end{pmatrix}\right)$	$\begin{bmatrix}0&0&0&-1\\0&0&-1&0\\0&1&0&0\\1&0&0&0\end{bmatrix}$	A
$(\beta\gamma)^2$	$\left(\begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \begin{pmatrix}1234\\3412\end{pmatrix}\right)$	$\begin{bmatrix}0&0&1&0\\0&0&0&-1\\-1&0&0&0\\0&1&0&0\end{bmatrix}$	B
γ	$\left(\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1234\\1234\end{pmatrix}\right)$	$\begin{bmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	Y
α	$\left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1234\\2134\end{pmatrix}\right)$	$\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	$WSBAY$
β	$\left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1234\\2341\end{pmatrix}\right)$	$\begin{bmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}$	$SWAYTR$
γ	$\left(\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1234\\1234\end{pmatrix}\right)$	$\begin{bmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	Y

TABLE IXB. Representation matrices of W_4.

	α	β	γ
S_4-cycles	$1^2 2$	4	1^4
S_8-cycles	$1^4 2^2$	4^2	$1^6 2$
S_8-permutation	21346578	23415678	52341678
$\tau_1^{(1)}$	1	1	1
$\tau_2^{(1)}$	1	1	-1
$\tau_3^{(1)}$	-1	-1	1
$\tau_4^{(1)}$	-1	-1	-1
$\tau_1^{(2)}$	$\begin{bmatrix}1&0\\0&-1\end{bmatrix}$	$\begin{bmatrix}-1/2&-\sqrt{3}/2\\-\sqrt{3}/2&1/2\end{bmatrix}$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$
$\tau_2^{(2)}$	$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	$\begin{bmatrix}-1/2&-\sqrt{3}/2\\-\sqrt{3}/2&1/2\end{bmatrix}$	$\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$
$\tau_1^{(3)}$	$\begin{bmatrix}1&0&0\\0&1&0\\0&0&-1\end{bmatrix}$	$\begin{bmatrix}-1/3&\sqrt{8}/3&0\\-\sqrt{2}/3&-1/6&\sqrt{3}/2\\-\sqrt{6}/3&-\sqrt{3}/6&-1/2\end{bmatrix}$	$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$
$\tau_2^{(3)}$	$\begin{bmatrix}1&0&0\\0&1&0\\0&0&-1\end{bmatrix}$	$\begin{bmatrix}-1/3&\sqrt{8}/3&0\\-\sqrt{2}/3&-1/6&\sqrt{3}/2\\-\sqrt{6}/3&-\sqrt{3}/6&-1/2\end{bmatrix}$	$\begin{bmatrix}-1&0&0\\0&-1&0\\0&0&-1\end{bmatrix}$
$\tau_3^{(3)}$	$\begin{bmatrix}-1&0&0\\0&-1&0\\0&0&1\end{bmatrix}$	$\begin{bmatrix}1/3&-\sqrt{8}/3&0\\\sqrt{2}/3&1/6&-\sqrt{3}/2\\\sqrt{6}/3&\sqrt{3}/6&1/2\end{bmatrix}$	$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$
$\tau_4^{(3)}$	$\begin{bmatrix}-1&0&0\\0&-1&0\\0&0&1\end{bmatrix}$	$\begin{bmatrix}1/3&-\sqrt{8}/3&0\\\sqrt{2}/3&1/6&-\sqrt{3}/2\\\sqrt{6}/3&\sqrt{3}/6&1/2\end{bmatrix}$	$\begin{bmatrix}-1&0&0\\0&-1&0\\0&0&-1\end{bmatrix}$
$\tau_1^{(4)}$	$\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	$\begin{bmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}$	$\begin{bmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$
$\tau_2^{(4)}$	$\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	$\begin{bmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}$	$\begin{bmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix}$
$\tau_3^{(4)}$	$\begin{bmatrix}0&-1&0&0\\-1&0&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix}$	$\begin{bmatrix}0&0&0&-1\\-1&0&0&0\\0&-1&0&0\\0&0&-1&0\end{bmatrix}$	$\begin{bmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$
$\tau_4^{(4)}$	$\begin{bmatrix}0&-1&0&0\\-1&0&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix}$	$\begin{bmatrix}0&0&0&-1\\-1&0&0&0\\0&-1&0&0\\0&0&-1&0\end{bmatrix}$	$\begin{bmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix}$

TABLE IXC. Representation matrices of W_4.



$$\tau_2^{(6)} = \tau_1^{(6)} \otimes \tau_2^{(1)}$$
$$\tau_4^{(6)} = \tau_3^{(6)} \otimes \tau_4^{(1)}.$$

For example, we get

$$\begin{aligned}\tau_3^{(4)} \otimes \tau_4^{(6)} &= \tau_1^{(4)} \otimes \tau_3^{(1)} \otimes \tau_3^{(6)} \otimes \tau_4^{(1)} \\ &= \tau_3^{(1)} \otimes \tau_4^{(1)} \otimes \tau_1^{(4)} \otimes \tau_3^{(6)} \\ &= \tau_2^{(1)} \otimes \tau_1^{(4)} \otimes \tau_3^{(6)} \\ &= \tau_2^{(4)} \otimes \tau_3^{(6)} = \tau_1^{(4)} \otimes \tau_3^{(6)}.\end{aligned}$$

As all irreducible representations can be formed by multiplication of $\tau_1^{(2)}$, $\tau_1^{(3)}$, $\tau_1^{(4)}$, $\tau_1^{(6)}$, $\tau_3^{(6)}$, and $\tau_1^{(8)}$ with one-dimensional representations, it is sufficient to look at Kronecker products of some so-called "fundamental" representations with one another. The result is listed in Table VIII.

V. THE MATRIX REPRESENTATIONS

It is convenient to select a sufficiently small number of elements which generate the whole group. It is evident that we have to list the matrix representations only for these elements. It is not difficult to prove that W_4 is generated by the following three elements:

$$\alpha = \left(\mathbf{0}, \begin{pmatrix}1 & 2 & 3 & 4\\ 2 & 1 & 3 & 4\end{pmatrix}\right), \beta = \left(\mathbf{0}, \begin{pmatrix}1 & 2 & 3 & 4\\ 2 & 3 & 4 & 1\end{pmatrix}\right),$$
$$\gamma = \left(\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \mathrm{id}\right).$$

Furthermore, these three elements form a minimal system of generators. The connections to the nine generators used by Chen and Birman[2] are listed in Table IX A.

Since $W_4 \cong O(4,\mathbb{Z})$ and $O(4,\mathbb{Z}) \subset O(4)$, every representation of W_4 is equivalent to an orthogonal one. Thus we chose orthogonal representation matrices. The matrices for $\tau_1^{(2)}$ and $\tau_1^{(3)}$ come from the S_4 representations (of Hamermesh[17]). The construction rules for the matrices of $\tau_1^{(4)}$, $\tau_1^{(6)}$, and $\tau_3^{(6)}$ are explained above, and for $\tau_1^{(8)}$ they are received by $\tau_1^{(8)} = \tau_1^{(2)} \otimes \tau_1^{(4)}$. The rest, again, is obtained by multiplication with one-dimensional representations (see Tables IXB,IXC).

VI. STRUCTURE OF SW_4

Let us now consider the following subgroup of W_4:

$$SW_4 := \{(\mathbf{a},\pi) \in W_4 | \det(\mathbf{a},\pi) = +1\},$$

where

$$\det(\mathbf{a},\pi) := \det(T(\mathbf{a},\pi)) = (-1)^{\sum_{i=1}^{4} a_i} \operatorname{sgn}\pi.$$

SW_4 is a normal subgroup of W_4 of order 192. Simultaneously we have $SW_4 \cong SO(4,\mathbb{Z})$ with $SO(4,\mathbb{Z}) = O(4,\mathbb{Z}) \cap SO_4$.

Restricting the conjugacy classes of W_4 to the subgroup SW_4—in doing so one has to take eleven classes of W_4 into consideration—one finds that two classes decompose, each of them in two parts of equal order. Thus we get 13 conjugacy classes in SW_4 (see Table X).

As we did for W_4, we list the number of elements of SW_4 of a given order (Table XI), the number of unequal cyclic subgroups of SW_4 (Table XII), and the number of axes of SW_4 (Table XIII).

TABLE X. Classes of conjugate elements of SW_4.

S_8-cycles	1^8	$1^4 4$	$1^2 2^3$	$1^2 2^3$	$1^2 3^2$	8		4^2		$2^2 4$	26	2^4	
S_4-cycles	1^4	$1^2 2$	1^4	$1^2 2$	13	4	4	2^2	2^2	$1^2 2$	13	2^2	1^4
order	1	12	6	24	32	24	24	6	6	12	32	12	1

TABLE XI. Number of elements of a given order for the group SW_4.

order	1	2	3	4	5	6	7	8
number	1	43	32	36	0	32	0	48

TABLE XII. Number of cyclic subgroups of a given order for the group SW_4.

order	1	2	3	4	5	6	7	8
number	1	43	16	18	0	16	0	12

TABLE XIII. Number of axes of a given order for the group SW_4.

order	1	2	3	4	5	6	7	8
number	0	36	0	12	0	16	0	12

TABLE XIV. Characters of SW_4.

S_8-cycles	1^8	$1^4 4$	$1^2 2^2$	$1^2 3^2$	$1^3 2^3$	8		4^2		$2^2 4$	26	2^4	
S_4-cycles	1^4	$1^2 2$	1^4	$1^2 2$	13	4	4	2^2	2^2	$1^2 2$	13	2^2	1^4
order	1	12	6	24	32	24	24	6	6	12	32	12	1
${}^s\chi_1^{(1)}$	1	1	1	1	1	1	1	1	1	1	1	1	1
${}^s\chi_2^{(1)}$	1	-1	1	-1	1	-1	-1	1	1	-1	1	1	1
${}^s\chi^{(2)}$	2	0	2	0	-1	0	0	2	2	0	-1	2	2
${}^s\chi_1^{(3)}$	3	1	3	1	0	-1	-1	-1	-1	1	0	-1	3
${}^s\chi_2^{(3)}$	3	-1	3	-1	0	1	1	-1	-1	-1	0	-1	3
$*\,{}^s\chi_3^{(3)}$	3	1	-1	-1	0	1	-1	3	-1	1	0	-1	3
$*\,{}^s\chi_4^{(3)}$	3	1	-1	-1	0	-1	1	-1	3	1	0	-1	3
$*\,{}^s\chi_5^{(3)}$	3	-1	-1	1	0	-1	1	3	-1	-1	0	-1	3
$*\,{}^s\chi_6^{(3)}$	3	-1	-1	1	0	1	-1	-1	3	-1	0	-1	3
${}^s\chi_1^{(4)}$	4	2	0	0	1	0	0	0	0	-2	-1	0	-4
${}^s\chi_2^{(4)}$	4	-2	0	0	1	0	0	0	0	2	-1	0	-4
${}^s\chi^{(6)}$	6	0	-2	0	0	0	0	-2	-2	0	0	2	6
${}^s\chi^{(8)}$	8	0	0	0	-1	0	0	0	0	0	1	0	-8

TABLE XV. Multiplicity of irreducible representations in n-fold Kronecker products of T with itself, $1<n<10$, for the group SW_4.

	${}^s T_1^{(1)}$	${}^s T_2^{(1)}$	${}^s T^{(2)}$	${}^s T_1^{(3)}$	${}^s T_2^{(3)}$	${}^s T_3^{(3)}$	${}^s T_4^{(3)}$	${}^s T_5^{(3)}$	${}^s T_6^{(3)}$	${}^s T_1^{(4)}$	${}^s T_2^{(4)}$	${}^s T^{(6)}$	${}^s T^{(8)}$
${}^s T_1^{(4)} = T$	0	0	0	0	0	0	0	0	0	1	0	0	0
$\otimes^2 T$	1	0	0	1	0	1	1	0	0	0	0	1	0
$\otimes^3 T$	0	0	0	0	0	0	0	0	0	5	1	0	5
$\otimes^4 T$	5	1	5	10	6	10	10	6	6	0	0	16	0
$\otimes^5 T$	0	0	0	0	0	0	0	0	0	51	35	0	85
$\otimes^6 T$	51	35	85	136	120	136	136	120	120	0	0	256	0
$\otimes^7 T$	0	0	0	0	0	0	0	0	0	715	651	0	1365
$\otimes^8 T$	715	651	1365	2080	2016	2080	2080	2016	2016	0	0	4096	0
$\otimes^9 T$	0	0	0	0	0	0	0	0	0	11051	10795	0	21845
$\otimes^{10} T$	11051	10795	21845	32896	32640	32896	32896	32640	32640	0	0	65536	0

VII. REPRESENTATIONS OF SW_4

Since SW_4 is a normal subgroup of W_4 of index 2, we get the complete system of irreducible representations of SW_4 by application of Clifford's theorem.

Except for $\chi_1^{(6)}$, and $\chi_2^{(6)}$ all irreducible W_4 representations stay irreducible after restriction to SW_4. The W_4 representations $\tau_1^{(6)}$ and $\tau_2^{(6)}$ split, respectively, into two three-dimensional representations. It should be noted that all representations of SW_4 are equivalent to orthogonal ones. The

isomorphism between SW_4 and $SO(4,\mathbf{Z})$ is still yielded by the restriction of the canonical W_4 representation T. Thus we have all information needed for the character table (see Table XIV). Note that all representations and characters of SW_4 are marked with the left upper index "s".

Let us shortly explain the character table:

${}^s\chi_1^{(1)} = \chi_1^{(1)}|_{SW_4}$, "identity";

${}^s\chi_2^{(1)} = \chi_2^{(1)}|_{SW_4}$, "signum S_8";

${}^s\chi^{(2)} = \chi_1^{(2)}|_{SW_4}$;

${}^s\chi_1^{(3)} = \chi_1^{(3)}|_{SW_4}$;

${}^s\chi_2^{(3)} = {}^s\chi_1^{(3)} \cdot {}^s\chi_2^{(1)} = \chi_2^{(3)}|_{SW_4}$;

${}^s\chi_3^{(3)}, {}^s\chi_4^{(3)}$ are irreducible parts of the restriction of the W_4-character $\chi_1^{(6)}$;

${}^s\chi_5^{(3)}, {}^s\chi_6^{(3)}$ are irreducible parts of the restriction of the W_4 character $\chi_2^{(6)}$;

${}^s\chi_5^{(3)} = {}^s\chi_3^{(3)} \cdot {}^s\chi_2^{(1)}$; ${}^s\chi_6^{(3)} = {}^s\chi_4^{(3)} \cdot {}^s\chi_2^{(1)}$;

${}^s\chi_1^{(4)} = \chi_1^{(4)}|_{SW_4}$;

${}^s\chi_2^{(4)} = \chi_2^{(4)}|_{SW_4}$;

${}^s\chi^{(6)} = \chi_3^{(6)}|_{SW_4}$;

${}^s\chi^{(8)} = \chi_1^{(8)}|_{SW_4}$.

TABLE XVI. Decomposition of mixed Kronecker products for the group SW_4.

	${}^sT_1^{(1)}$	${}^sT_2^{(1)}$	${}^sT_1^{(2)}$	${}^sT_1^{(3)}$	${}^sT_2^{(3)}$	${}^sT_3^{(3)}$	${}^sT_4^{(3)}$	${}^sT_5^{(3)}$	${}^sT_6^{(3)}$	${}^sT_1^{(4)}$	${}^sT_2^{(4)}$	${}^sT^{(6)}$	${}^sT^{(8)}$
${}^sT^{(2)} \otimes {}^sT^{(2)}$	1	1	1	0	0	0	0	0	0	0	0	0	0
${}^sT^{(2)} \otimes {}^sT_1^{(3)}$	0	0	0	1	1	0	0	0	0	0	0	0	0
${}^sT^{(2)} \otimes {}^sT_3^{(3)}$	0	0	0	0	0	1	0	1	0	0	0	0	0
${}^sT^{(2)} \otimes {}^sT_4^{(3)}$	0	0	0	0	0	0	1	0	1	0	0	0	0
${}^sT^{(2)} \otimes {}^sT_1^{(4)}$	0	0	0	0	0	0	0	0	0	0	0	0	1
${}^sT^{(2)} \otimes {}^sT^{(6)}$	0	0	0	0	0	0	0	0	0	0	0	2	0
${}^sT^{(2)} \otimes {}^sT^{(8)}$	0	0	0	0	0	0	0	0	0	1	1	0	1
${}^sT_1^{(3)} \otimes {}^sT_1^{(3)}$	1	0	1	1	1	0	0	0	0	0	0	0	0
${}^sT_1^{(3)} \otimes {}^sT_3^{(3)}$	0	0	0	0	0	0	1	0	0	0	0	1	0
${}^sT_1^{(3)} \otimes {}^sT_4^{(3)}$	0	0	0	0	0	1	0	0	0	0	0	1	0
${}^sT_1^{(3)} \otimes {}^sT_1^{(4)}$	0	0	0	0	0	0	0	0	0	1	0	0	1
${}^sT_1^{(3)} \otimes {}^sT^{(6)}$	0	0	0	0	0	1	1	1	1	0	0	1	0
${}^sT_1^{(3)} \otimes {}^sT^{(8)}$	0	0	0	0	0	0	0	0	0	1	1	0	2
${}^sT_3^{(3)} \otimes {}^sT_3^{(3)}$	1	0	1	0	0	1	0	1	0	0	0	0	0
${}^sT_3^{(3)} \otimes {}^sT_4^{(3)}$	0	0	0	1	0	0	0	0	0	0	0	1	0
${}^sT_3^{(3)} \otimes {}^sT_1^{(4)}$	0	0	0	0	0	0	0	0	0	1	0	0	1
${}^sT_3^{(3)} \otimes {}^sT^{(6)}$	0	0	0	1	1	0	1	0	1	0	0	1	0
${}^sT_3^{(3)} \otimes {}^sT^{(8)}$	0	0	0	0	0	0	0	0	0	1	1	0	2
${}^sT_4^{(3)} \otimes {}^sT_4^{(3)}$	1	0	1	0	0	0	1	0	1	0	0	0	0
${}^sT_4^{(3)} \otimes {}^sT_1^{(4)}$	0	0	0	0	0	0	0	0	0	1	0	0	1
${}^sT_4^{(3)} \otimes {}^sT^{(6)}$	0	0	0	1	1	1	0	1	0	0	0	1	0
${}^sT_4^{(3)} \otimes {}^sT^{(8)}$	0	0	0	0	0	0	0	0	0	1	1	0	2
${}^sT_1^{(4)} \otimes {}^sT_1^{(4)}$	1	0	0	1	0	1	1	0	0	0	0	1	0
${}^sT_1^{(4)} \otimes {}^sT^{(6)}$	0	0	0	0	0	0	0	0	0	1	1	0	2
${}^sT_1^{(4)} \otimes {}^sT^{(8)}$	0	0	1	1	1	1	1	1	1	0	0	2	0
${}^sT^{(6)} \otimes {}^sT^{(6)}$	1	1	2	1	1	1	1	1	1	0	0	2	0
${}^sT^{(6)} \otimes {}^sT^{(8)}$	0	0	0	0	0	0	0	0	0	2	2	0	4
${}^sT^{(8)} \otimes {}^sT^{(8)}$	1	1	1	2	2	2	2	2	2	0	0	4	0

TABLE XVIIA. Representation matrices of SW_4.

	$\alpha\gamma$	$\beta\gamma$
S_4-cycles	$1^2 2$	4
S_8-cycles	$1^4 4$	8
S_8-permutation	61342578	63412785
${}^s\tau_1^{(1)}$	1	1
${}^s\tau_2^{(1)}$	-1	-1
${}^s\tau^{(2)}$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix}$
${}^s\tau_1^{(3)}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1/3 & \sqrt{8}/3 & 0 \\ -\sqrt{2}/3 & -1/6 & \sqrt{3}/2 \\ -\sqrt{6}/3 & -\sqrt{3}/6 & -1/2 \end{bmatrix}$
${}^s\tau_2^{(3)}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1/3 & -\sqrt{8}/3 & 0 \\ \sqrt{2}/3 & 1/6 & -\sqrt{3}/2 \\ \sqrt{6}/3 & \sqrt{3}/6 & 1/2 \end{bmatrix}$
${}^s\tau_3^{(3)}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$
${}^s\tau_4^{(3)}$	$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$
${}^s\tau_5^{(3)}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$
${}^s\tau_6^{(3)}$	$\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$
${}^s\tau_1^{(4)}$	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
${}^s\tau_2^{(4)}$	$\begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}$
${}^s\tau^{(6)}$	$\begin{bmatrix} -1 & 0 & 0 & & & \\ 0 & 0 & 1 & & 0 & \\ 0 & -1 & 0 & & & \\ & & & 0 & 1 & 0 \\ & 0 & & -1 & 0 & 0 \\ & & & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

TABLE XVIIB. Representation matrices of SW_4.

	$\alpha\gamma$	$\beta\gamma$
${}^s\tau^{(8)}$	$\begin{bmatrix} 0 & 1 & 0 & 0 & & & & \\ -1 & 0 & 0 & 0 & & & 0 & \\ 0 & 0 & 1 & 0 & & & & \\ 0 & 0 & 0 & 1 & & & & \\ & & & & 0 & -1 & 0 & 0 \\ & & & & 1 & 0 & 0 & 0 \\ & & 0 & & 0 & 0 & -1 & 0 \\ & & & & 0 & 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & -1/2 & 0 & 0 & 0 & -\sqrt{3}/2 \\ 1/2 & 0 & 0 & 0 & \sqrt{3}/2 & 0 & 0 & 0 \\ 0 & -1/2 & 0 & 0 & 0 & -\sqrt{3}/2 & 0 & 0 \\ 0 & 0 & -1/2 & 0 & 0 & 0 & -\sqrt{3}/2 & 0 \\ 0 & 0 & 0 & -\sqrt{3}/2 & 0 & 0 & 0 & 1/2 \\ \sqrt{3}/2 & 0 & 0 & 0 & -1/2 & 0 & 0 & 0 \\ 0 & -\sqrt{3}/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & -\sqrt{3}/2 & 0 & 0 & 0 & 1/2 & 0 \end{bmatrix}$

Representations that result from the split of a W_4 representation are marked with an asterisk in the character table.

VIII. KRONECKER PRODUCTS

Once more, we can obtain all irreducible representations of SW_4 by decomposition of the multiple Kronecker product of the canonical representation $T = {}^s\tau_1^{(4)}$ with itself. This is done in Table XV.

For the purpose of decomposition of an arbitrary Kronecker product of irreducible representations, we have to know the decomposition numbers for the twofold Kronecker products of ${}^s\tau^{(2)}$, ${}^s\tau_1^{(3)}$, ${}^s\tau_3^{(3)}$, ${}^s\tau_4^{(3)}$, ${}^s\tau_1^{(4)}$, ${}^s\tau^{(6)}$, and ${}^s\tau^{(8)}$ with each other. These numbers are given in Table XVI. All other products shall be considered with the aid of the following rules:

$$
\begin{aligned}
&{}^s\tau_1^{(1)} \otimes {}^s\tau_1^{(1)} = {}^s\tau_1^{(1)};\\
&{}^s\tau_1^{(1)} \otimes {}^s\tau_2^{(1)} = {}^s\tau_2^{(1)};\\
&{}^s\tau_2^{(1)} \otimes {}^s\tau_2^{(1)} = {}^s\tau_1^{(1)};\\
&{}^s\tau_2^{(1)} \otimes {}^s\tau_1^{(3)} = {}^s\tau_2^{(3)};\\
&{}^s\tau_2^{(1)} \otimes {}^s\tau_3^{(3)} = {}^s\tau_5^{(3)};\\
&{}^s\tau_2^{(1)} \otimes {}^s\tau_4^{(3)} = {}^s\tau_6^{(3)};\\
&{}^s\tau_2^{(1)} \otimes {}^s\tau_1^{(4)} = {}^s\tau_2^{(4)}.
\end{aligned}
$$

IX. MATRIX REPRESENTATIONS

It is easy to check that two elements

$$
\alpha\cdot\gamma = \left[\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1&2&3&4\\2&1&3&4\end{pmatrix}\right]
$$

and

$$
\beta\cdot\gamma = \left[\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix}\right]
$$

generate the whole group SW_4. Thus we list the representation matrices of these two elements. In order to obtain the representation matrices of $\tau_1^{(6)}$ and $\tau_2^{(6)}$ in block-diagonal form, we use the unitary transformation matrix

$$
U = \frac{1}{2^{1/2}}\begin{pmatrix}
1&0&0&0&0&-1\\
0&1&0&0&1&0\\
0&0&1&-1&0&0\\
0&0&1&1&0&0\\
0&1&0&0&-1&0\\
1&0&0&0&0&1
\end{pmatrix}
$$

Through this matrix we get the three-dimensional irreducible constituents of the reducible representations $\tau_1^{(6)}|_{sw_4}$ and $\tau_2^{(6)}|_{sw_4}$ by means of a similarity transformation. The representation matrices are given in the Tables XVIIA and XVIIB.

ACKNOWLEDGMENT

This paper grew out of an undergraduate seminar on group theory supervised by Professor V. Rittenberg. We wish to thank him for his encouragement and support.

[1] W. Miller, *Symmetry Groups and Their Applications* (Academic, New York, 1972).
[2] J. L. Birman and Li-Ching Chen, "Irreducible vector and ray representations of some cubic crystal point groups in four dimensions," J. Math. Phys. **12**, 2454 (1971).
[3] E. Goursat, "Sur les substitutions orthogonales et les divisions régulières de l'espace," Ann. Sci. Ecole. Norm. Sup. Paris (3) **6**, 9 (1889).
[4] A. C. Hurley, "Finite Rotation Groups and Crystal Classes in Four Dimensions," Proc. Camb. Phil. Soc. **47**, 650 (1951).
[5] A. Young, "On Quantitative Substitutional Analysis," Proc. Lond. Math. Soc. **33**(1), 97 (1900);**34** 361 (1902).
[6] G. de B. Robinson, "A Geometrical Study of the Hyperoctahedral Group," Proc. Cambridge Philos. Soc. **26**, 94 (1930).
[7] J. S. Frame "Orthogonal Group Matrices of Hyperoctahedral Groups, Nagoya Math. J. **27**, 585 (1966).
[8] A. Kerber, *Representations of Permutation Groups I*, Lecture Notes in Mathematics, Vol. 240 (Springer-Verlag, Berlin, 1971).
[9] A. Kerber, "Darstellungstheorie an Kranzprodukten," Can. J. Math. **20**, 665 (1968).
[10] S. J. Maier, "On the Irreducible Characters of the Symmetric Group," Adv. Math. **15**, 127 (1975).
[11] S. J. Maier, "On the Characters of the Weyl Group of Type C, J. Algebra **33**, 59 (1975).
[12] M. Osima, "On the Representations of the Generalized Symmetric Group I," Math. J. Okayama Univ. **4**, 39 (1954).
[13] M. Osima, "On the Representations of the Generalized Symmetric Group II," Math. J. Okayama Univ. **6**, 81 (1956).
[14] M. Baake, B. Gemünden and R. Oedingen, "On Relations between Representations of the Hyperoctahedral Group and O(4) Representations," in preparation.
[15] M. Marcu, A. Regev, and V. Rittenberg, "The global symmetries of spin systems defined on abelian groups-I," J. Math. Phys. **22**, 2740 (1981).
[16] J. M. Isaacs, *Character Theory of Finite Groups* (Academic, New York, 1976).
[17] M. Hamermesh, *Group Theory and Its Application to Physical Problems*, 2nd ed., (Addison-Wesley, Reading, Mass., 1964).

Infinite-dimensional representations of the graded Lie algebra (Sp(4):4). Representation of the para-Bose operators with real order of quantization

I. Inaba, T. Maekawa, and T. Yamamoto
Department of Physics, Kumamoto University, Kumamoto, Japan

(Received 25 September 1981; accepted for publication 18 December 1981)

Infinite-dimensional representations for the system of the two para-Bose operators, which generate the graded Lie algebra (GLA) (Sp(4):4), are constructed by using the irreducible representations (*discrete series*) of the Lie algebra $Sp(4) \approx SO(3,2)$. It is shown that there are four kinds of the irreducible representations of the GLA's (Sp(4):4), i.e., cases (I), (II), (III), and (IV). Case (I) is described by the three irreducible representations of Sp(4) and corresponds to the para-Bose quantization with real order of the quantization greater than 1 and case (IV) corresponding to the ordinary Bose quantization by the two irreducible representations of Sp(4). Cases (II) and (III), which are described by the four and two irreducible representations of Sp(4), respectively, and cannot be obtained by the method of Fock space, express the representations of the graded Lie algebra (Sp(4):4).

PACS numbers: 02.20.Sv, 02.20.Qs

I. INTRODUCTION

Explicit construction of the irreducible representations of the graded Lie algebra has been studied by many authors.[1] It, however, seems that this is restricted to the simple cases and finite-dimensional representation[1] or the special case which uses the method of Fock space with the assumption of the vacuum state.[2] It will, therefore, be important to give a nontrivial example of the graded Lie algebra (GLA) and then to find the infinite-dimensional irreducible representations of the GLA in order to obtain a more general result.

Parafield quantization due to Green[3] is fruitful, and various properties are made clear.[4] Among these, it is interesting to note that the system of f parafield operators is connected with the orthogonal group $SO(2f+1)$[5] of the sympletic group $Sp(2f,R)$[6] according to the para-Fermi or the para-Bose operators, respectively. It is also known that the group $Sp(2f)$ together with the f para-Bose operators becomes the graded Lie algebra (Sp(4):4). Through the connection, the representation of the para-Fermi operators is constructed, and the order of the quantization becomes positive integer,[1,7] which may be considered to result from the compactness of the orthogonal group. On the other hand, as is well known, the sympletic group is noncompact, and explicit construction of the unitary irreducible representations seems lacking except for the simple cases.[8] Of course, the representation of $Sp(2f)$ is studied by the method of Fock space,[2] which assumes the vacuum state together with the integer order of the quantization. However, it seems that there is a possibility of the para-Bose quantization with real order of the quantization for finite f, because we know the real order of the quantization in the case of $f=1$.[9]

The purpose of this article is to construct explicitly some infinite-dimensional irreducible representations of the graded Lie algebra (Sp(4):4) generated by the two para-Bose operators and to show that the order of the para-Bose quantization is also characterized by a real number greater than one in the case of $f=2$.

In Sec. II, the discrete series of the irreducible representations of $SO(3,2) \approx Sp(4)$,[10] which are necessary for our discussion, are summarized. In Sec. III, the matrix elements of the generators of the graded Lie algebra (Sp(4):4), which consists of the para-Bose operators, are explicitly constructed for all classes. Section IV is devoted to a discussion of the result.

II. SUMMARY OF SO(3,2)

We summarize the results of the discrete series of the irreducible representation of SO(3,2),[10] which are necessary for our discussion.

The Hermitian generators ($J_{ij} = -J_{ji}$, $i,j = 1,2,\ldots,5$) of SO(3,2) satisfy the commutation relations

$$[J_{jk}, J_{lm}] = i(g_{jl}J_{km} + g_{km}J_{jl} - g_{kl}J_{jm} - g_{jm}J_{kl}), \quad (2.1)$$

where the metric tensor g_{jk} takes the values $g_{11} = g_{22} = -g_{33} = -g_{44} = -g_{55} = 1$ and $g_{jk} = 0$ for $j \neq k$. The Casimir operators of SO(3,2)[11] are given by

$$\begin{aligned} F &= \tfrac{1}{2}\sum g_{ik}g_{jl}J_{ij}J_{kl}, \\ G &= (C_1)^2 + (C_2)^2 - (C_3)^2 - (C_4)^2 - (C_5)^2, \end{aligned} \quad (2.2)$$

where

$$\begin{aligned} C_1 &= J_{23}J_{45} + J_{42}J_{35} + J_{25}J_{34}, \\ C_2 &= J_{31}J_{45} + J_{43}J_{15} + J_{35}J_{14}, \\ C_3 &= J_{12}J_{45} + J_{41}J_{25} + J_{24}J_{15}, \\ C_4 &= J_{21}J_{35} + J_{13}J_{25} + J_{32}J_{15}, \\ C_5 &= J_{23}J_{14} + J_{31}J_{24} + J_{12}J_{34}. \end{aligned}$$

We define the quantities

$$\begin{aligned} J_1^{(1)} &= \tfrac{1}{2}(J_{23} + J_{14}), \quad J_1^{(2)} = \tfrac{1}{2}(J_{23} - J_{14}), \\ J_2^{(1)} &= \tfrac{1}{2}(J_{31} + J_{24}), \quad J_2^{(2)} = \tfrac{1}{2}(J_{31} - J_{24}), \\ J_3^{(1)} &= \tfrac{1}{2}(J_{12} - J_{34}), \quad J_3^{(2)} = \tfrac{1}{2}(J_{12} + J_{34}). \end{aligned} \quad (2.3)$$

 0022-2488/82/060954-10$02.50

Then they satisfy the commutation relations

$$[J_1^{(j)},J_2^{(k)}] = -i\delta_{jk}J_3^{(j)},$$
$$[J_3^{(j)},J_1^{(k)}] = i\delta_{jk}J_2^{(j)}, \tag{2.4}$$
$$[J_2^{(j)},J_3^{(k)}] = i\delta_{jk}J_1^{(j)}.$$

It follows from (2.4) that the quantities defined in (2.3) are generators of SO(2,1). The irreducible representations of SO(2,1) are well known[12] and characterized by the eigenvalue of the Casimir operator of SO(2,1):

$$\mathbb{J}^{(i)2} = J_3^{(i)2} - J_1^{(i)2} - J_2^{(i)2} \quad (i = 1,2). \tag{2.5}$$

The unitary irreducible representations of SO(3,2) (discrete series) can be constructed on the product space of the bases of SO(2,1) $\times$ SO(2,1) [10]; that is, the following bases can be used:

$$\left.\begin{aligned} F|pq{:}j_1m_1j_2m_2\rangle &= [\,p(p-3)+q(q+1)]\,|pq{:}j_1m_1j_2m_2\rangle,\\ G|pq{:}j_1m_1j_2m_2\rangle &= -(p-1)(p-2)q(q+1)|pq{:}j_1m_1j_2m_2\rangle,\\ \mathbb{J}^{(i)2}|pq{:}j_1m_1j_2m_2\rangle &= j_i(j_i-1)|pq{:}j_1m_1j_2m_2\rangle\\ J_3^{(i)}|pq{:}j_1m_1j_2m_2\rangle &= m_i|pq{:}j_1m_1j_2m_2\rangle \end{aligned}\right\}(i=1,2), \tag{2.6}$$

where the numbers p and q specifying the eigenvalues of F and G characterize the irreducible representation of SO(3,2). The numbers j_i and m_i ($=1,2$) are related to the representation of SO(2,1). We restrict the eigenvalue m_i of $J_3^{(i)}$ to the postivie-definite value, because only the case is needed for us in the following. j_i is the minimum value of m_i.

The matrix elements of the generators with respect to the bases (2.6) are given as follows[10]:

$$J_+^{(1)}|pq{:}j_1m_1j_2m_2\rangle = [(m_1+j_1)(m_1-j_1+1)]^{1/2}|pq{:}j_1m_1+1j_2m_2\rangle,$$
$$J_+^{(2)}|pq{:}j_1m_1j_2m_2\rangle = [(m_2+j_2)(m_2-j_2+1)]^{1/2}|pq{:}j_1m_1j_2m_2+1\rangle,$$
$$\langle pq{:}j_1'm_1'j_2'm_2'|J_+^{(3)}|pq{:}j_1m_1j_2m_2\rangle = \left(\frac{\Gamma(m_1'+j_1')\Gamma(m_1'-j_1'+1)}{\Gamma(m_1+j_1)\Gamma(m_1-j_1+1)}\right)^{1/2}\left(\frac{\Gamma(m_2'+j_2')\Gamma(m_2'-j_2'+1)}{\Gamma(m_2+j_2)\Gamma(m_2-j_2+1)}\right)^{1/2}\langle pq{:}j_1'j_2'\|J_+^{(3)}\|pq{:}jj_2\rangle$$
$$\text{for } m_i' = m_i + \tfrac{1}{2},\quad j_i' = j_i \pm \tfrac{1}{2},$$
$$\langle pq{:}j_1'm_1'j_2'm_2'|J_+^{(4)}|pq{:}j_1m_1j_2m_2\rangle = \left(\frac{\Gamma(m_1'+j_1')\Gamma(m_1'-j_1'+1)}{\Gamma(m_1+j_1)\Gamma(m_1-j_1+1)}\right)^{1/2}\left(\frac{\Gamma(m_2+j_2)\Gamma(m_2-j_2+1)}{\Gamma(m_2'+j_2')\Gamma(m_2'-j_2'+1)}\right)^{1/2}\langle pq{:}j_1'j_2'\|J_+^{(4)}\|pq{:}jj_2\rangle$$
$$\text{for } m_1' = m_1 + \tfrac{1}{2},\quad m_2' = m_2 - \tfrac{1}{2},\quad j_i' = j_i \pm \tfrac{1}{2},$$
$$\langle pq{:}j_1 + \tfrac{1}{2}j_2 + \tfrac{1}{2}\|J_+^{(3)}\|pq{:}jj_2\rangle = \left(\frac{(\sigma+q)(\sigma-q-1)(\sigma-p+1)(\sigma+p-2)}{2j_1(2j_1-1)2j_2(2j_2-1)}\right)^{1/2},$$
$$\langle pq{:}j_1 + \tfrac{1}{2}j_2 - \tfrac{1}{2}\|J_+^{(3)}\|pq{:}jj_2\rangle = \left(\frac{(q+\delta+1)(q-\delta)(p-\delta-2)(p+\delta-1)}{2j_1(2j_1-1)(2j_2-1)(2j_2-2)}\right)^{1/2},$$
$$\langle pq{:}j_1j_2\|J_+^{(3)}\|pq{:}j_1 - \tfrac{1}{2}j_2'\rangle = \langle pq{:}j_1 - \tfrac{1}{2}j_2'\|J_+^{(3)}\|pq{:}j_1j_2\rangle \quad \text{for } j_2' = j_2 \pm \tfrac{1}{2},$$
$$\langle pq{:}j_1'j_2'\|J_+^{(3)}\|pq{:}jj_2\rangle = \langle pq{:}j_1'j_2'J_-^{(3)}\|pq{:}jj_2\rangle = \langle pq{:}j_1'j_2'\|J_+^{(4)}\|pq{:}jj_2\rangle = \langle pq{:}j_1'j_2'\|J_-^{(4)}\|pq{:}jj_2\rangle \quad \text{for } j_i' \pm \tfrac{1}{2},$$

where

$$J_\pm^{(j)} = J_1^{(j)} \pm iJ_2^{(j)} \quad (j=1,2),$$
$$J_\pm^{(3)} = J_{15} \pm iJ_{25}, \quad J_\pm^{(4)} = -J_{35} \mp iJ_{45},$$
$$\sigma \equiv j_1 + j_2, \qquad \delta \equiv j_1 - j_2,$$

and $\Gamma(x)$ denotes a gamma function with an argument x. The expressions for $J_-^{(i)}$ are obvious from those for $J_+^{(i)}$.

The irreducible unitary representations of SO(3,2) (discrete series) are classified in terms of p and q as follows[10]:

(a) $p > q + \frac{1}{2}$ ($q = 0,\frac{1}{2}$) and $p > q + 1$ ($q = 1,\frac{3}{2},2,\cdots$):

$j_1 = (p+n+\delta)/2$, $j_2 = (p+n-\delta)/2$,
$\delta = -q, -q+1,\ldots q-1, q$; $n = 0,1,2,\cdots$.

(b) $p = q + 1$ ($q = 1,\frac{3}{2},2,\ldots$):

$j_1 = (p+\delta)/2$, $j_2 = (p-\delta)/2$,
$\delta = -q, -q+1,\ldots,q-1,q$;
$j_1 = (p+n-q)/2$, $j_2 = (p+n+q)/2$

or

$j_1 = (p+n+q)/2$, $j_2 = (p+n-q)/2$,
$n = 1,2,\cdots$.

(c) $p = q + \frac{1}{2}$ ($q = 0$):

$j_1 = j_2 = \frac{1}{4}$ or $j_1 = j_2 = \frac{3}{4}$.

(d) $p = q + \frac{1}{2}$ ($q = \frac{1}{2}$):

$j_1 = \frac{3}{4}$, $j_2 = \frac{1}{4}$ or $j_1 = \frac{1}{4}$, $j_2 = \frac{3}{4}$.

m_i, of course, takes the values $j_i, j_i + \frac{1}{2}, j_i + 1, \cdots$ in all classes.

In the next section, we construct the irreducible representations for the system consisting of the two para-Bose operators by using the above result.

III. REPRESENTATIONS OF PARA-BOSE OPERATORS

The para-Bose operators satisfy the following relations[3]:

$$\begin{aligned}
&[\{a_i,a_j\},a_k^\dagger] = 2\delta_{ik}a_j + 2\delta_{jk}a_i,\\
&[\{a_i,a_j^\dagger\},a_k^\dagger] = 2\delta_{ik}a_j^\dagger,\\
&[\{a_i,a_j\},a_k] = 0,
\end{aligned} \tag{3.1}$$

where $a_j^\dagger$ is the Hermitian conjugate of a_j and $[A,B] = AB - BA$, $\{A,B\} = AB + BA$. The Hamiltonian H and the number operator N of the system are defined by

$$\begin{aligned}
&H = \tfrac{1}{2}\omega \sum \{a_k,a_k^\dagger\},\\
&N = \sum N_k, \quad N_k = \tfrac{1}{2}\{a_k,a_k^\dagger\} - p,
\end{aligned} \tag{3.2}$$

where $2p$ is order of the para-Bose quantization defined in terms of the vacuum state $|0\rangle$ by[3]

$$a_k|0\rangle = 0, \quad a_j a_k^\dagger|0\rangle = 2\delta_{jk}p|0\rangle. \tag{3.3}$$

We define the operators $X_{\mu\nu}$ in terms of a_j and $a_j^\dagger$ as follows:

$$\begin{aligned}
&X_{j-k} = X_{-kj} = \tfrac{1}{2}(a_j^\dagger a_k + a_k a_j^\dagger),\\
&X_{jk} = X_{kj} = \tfrac{1}{2}(a_j^\dagger a_k^\dagger + a_k^\dagger a_j^\dagger),\\
&X_{-j-k} = X_{-k-j} = \tfrac{1}{2}(a_j a_k + a_k a_j).
\end{aligned} \tag{3.4}$$

Then if follows from (3.1) that they satisfy the commutation relation of Sp(2f,R) [6]:

$$\begin{aligned}
[X_{\mu\nu},X_{\rho\sigma}] = {}& \epsilon^\rho\delta_{\nu-\rho}X_{\mu\sigma} + \epsilon^\sigma\delta_{\nu-\sigma}X_{\mu\rho} + \epsilon^\rho\delta_{\mu-\rho}X_{\nu\sigma}\\
&+ \epsilon^\sigma\delta_{\mu-\sigma}X_{\nu\rho},
\end{aligned} \tag{3.5}$$

where μ,ν,ρ,σ take $\pm j$ ($j = 1,2,\ldots,f$) and ϵ^ρ takes $+1$ for $\rho > 0$ and -1 for $\rho < 0$. We, then, see that the operators (3.4) together with a_j and $a_j^\dagger$ generate a graded Lie algebra.[1] In what follows, we restrict our discussion to the case of Sp(4) (i.e., $f = 2$), and then show that the representation of the graded Lie algebra is determined by that of the Lie algebra of Sp(4).

As is well known, the group Sp(4), is isomorphic to SO(3,2).[13] Explicitly writing the relation in terms of the generators between Sp(4) and SO(3,2), we get

$$\begin{aligned}
&J_{12} - J_{34} = X_{1-1}, \quad J_{23} + J_{14} + i(J_{31} + J_{24}) = X_{11},\\
&J_{23} + J_{14} - i(J_{31} + J_{24}) = X_{-1-1},\\
&J_{12} + J_{34} = X_{2-2}, \quad J_{23} - J_{14} + i(J_{31} - J_{24}) = X_{22},\\
&J_{23} - J_{14} - i(J_{31} - J_{24}) = X_{-2-2},\\
&J_{15} + iJ_{25} = X_{12}, \quad J_{15} - iJ_{25} = X_{-1-2},\\
&-J_{35} - iJ_{45} = X_{1-2}, \quad -J_{35} + iJ_{45} = X_{2-1},
\end{aligned} \tag{3.6}$$

where the same symbol as in Sec. II, J_{jk}, for the generators of SO(3,2) is used. Therefore, we can use the result in Sec. II for our discussion. It is convenient for us to express $J_{\pm}^{(i)}$ defined in (2.3) and below (2.6) in terms of a_j and $a_j^\dagger$:

$$\left.\begin{aligned}
&J_3^{(i)} = \tfrac{1}{4}(a_i^\dagger a_i + a_i a_i^\dagger)\\
&J_+^{(i)} = \tfrac{1}{2}a_i^\dagger a_i^\dagger, \quad J_-^{(i)} = \tfrac{1}{2}a_i a_i
\end{aligned}\right\} \quad (i = 1,2),$$

$$\begin{aligned}
&J_+^{(3)} = \tfrac{1}{2}(a_1^\dagger a_2^\dagger + a_2^\dagger a_1^\dagger), \quad J_-^{(3)} = J_+^{(3)\dagger},\\
&J_+^{(4)} = \tfrac{1}{2}(a_1^\dagger a_2 + a_2 a_1^\dagger), \quad J_-^{(4)} = J_+^{(4)\dagger}.
\end{aligned} \tag{3.7}$$

It follows from (3.1) that the following and their Hermitian conjugate relations hold:

$$\left.\begin{aligned}
&[a_j,J_3^{(k)}] = \tfrac{1}{2}\delta_{jk}a_j\\
&[a_j,J_+^{(k)}] = \delta_{jk}a_j^\dagger\\
&[a_j,J_-^{(k)}] = 0
\end{aligned}\right\} \quad (j,k = 1,2),$$

$$\begin{aligned}
&[a_1,J_+^{(3)}] = a_2^\dagger, \quad [a_1,J_+^{(4)}] = a_2,\\
&[a_1,J_-^{(3)}] = [a_1,J_-^{(4)}] = 0,\\
&[a_2,J_+^{(3)}] = a_1^\dagger, \quad [a_2,J_-^{(4)}] = a_1,\\
&[a_2,J_-^{(3)}] = [a_2,J_+^{(4)}] = 0.
\end{aligned} \tag{3.8}$$

Before solving our problem, let us explain the simple case of Sp(2)$\approx$SO(2,1), which is well known.[9] That is, we consider the quantities

$$\begin{aligned}
&J_3 = \tfrac{1}{4}(aa^\dagger + a^\dagger a), \quad J_+ = \tfrac{1}{2}a^\dagger a^\dagger, \quad J_- = J^\dagger{}_+ .\\
&[a,J_3] = \tfrac{1}{2}a, \quad [a,J_+] = a^\dagger, \quad [a,J_-] = 0.
\end{aligned} \tag{3.9}$$

As the first three operators correspond to those of SO(2,1) and the eigenvalues of J_3 must be positive as is seen from the expression of J_3, we can adopt the following bases[9]:

$$\begin{aligned}
&\mathbb{J}^2|j,m\rangle = j(j-1)|j,m\rangle,\\
&J_3|j,m\rangle = m|j,m\rangle,
\end{aligned} \tag{3.10}$$

where $\mathbb{J}^2$ means the Casimir operator (2.4) of SO(2,1), j characterizes the irreducible representation of SO(2,1), and m takes the values $j, j+1, \cdots$.

The action of J_+ on the bases is given by

$$J_+|j,m\rangle = [(m+j)(m-j+1)]^{1/2}|j,m+1\rangle. \tag{3.11}$$

Let us now determine the action of $a(a^\dagger)$ on the bases. We get from $[J_3,a] = -a/2$ and $[J_3,a^\dagger] = a^\dagger/2$

$$\begin{aligned}
&\langle j'm'|a|jm\rangle \neq 0 \quad \text{for } m' = m - \tfrac{1}{2},\\
&\langle j'm'|a^\dagger|jm\rangle \neq 0 \quad \text{for } m' = m + \tfrac{1}{2}.
\end{aligned} \tag{3.12}$$

Taking the matrix elements of $[a,J_-] = [a^\dagger,J_+] = 0$, we get

$$\begin{aligned}
&\langle j'\ m - \tfrac{1}{2}|a|jm\rangle\\
&\quad = \left(\frac{\Gamma(m-j+1)\Gamma(m+j)}{\Gamma(m-j'+\frac{1}{2})\Gamma(m+j'-\frac{1}{2})}\right)^{1/2}\langle j'\|a\|j\rangle,\\
&\langle j'\ m + \tfrac{1}{2}|a^\dagger|jm\rangle\\
&\quad = \left(\frac{\Gamma(m-j'+\frac{3}{2})\Gamma(m+j'+\frac{1}{2})}{\Gamma(m-j+1)\Gamma(m+j)}\right)^{1/2}\langle j'\|a^\dagger\|j\rangle,
\end{aligned} \tag{3.13}$$

where $\langle j'\|a\|j\rangle$ denotes the reduced matrix elements of a. Similarly, taking the matrix elements of $[a,J_+] = a^\dagger$ and $[a^\dagger,J_-] = -a$ and taking into account (3.13), we obtain

the following relation:

$$(j'-j-\tfrac{1}{2})(j'-j+\tfrac{1}{2})(j'+j-\tfrac{1}{2})(j'+j-\tfrac{3}{2})\langle j'\|a\|j\rangle = 0, \tag{3.14}$$

and the same relation for $a^\dagger$. Thus it is sufficient for us to consider the cases of $j' = j \pm \frac{1}{2}$ for which the reduced matrix elements $\langle j'\|a\|j\rangle$ and $\langle j'\|a^\dagger\|j\rangle$ are not zero. It, then, follows from $[a,J_+] = a^\dagger$ together with (3.13) that the following relation holds:

$$\langle j'\|a\|j\rangle = \langle j'\|a^\dagger\|j\rangle = (\langle j\|a^\dagger\|j'\rangle)^* \quad \text{for } j' = j \pm \tfrac{1}{2}. \tag{3.15}$$

Thus we see that the action of a on the base $|jm\rangle$ gives the bases with $j' = j \pm \frac{1}{2}$. As the number j takes a positive-definite value (trivial representation for $j = 0$ is omitted), it may be assumed that there exists a minimum . Then we can write the action of a on the bases as follows:

$$\begin{aligned} a|jm\rangle &= (m-j)^{1/2}|j+\tfrac{1}{2}\ m-\tfrac{1}{2}\rangle\langle j+\tfrac{1}{2}\|a\|j\rangle, \\ a|j+\tfrac{1}{2}\ m+\tfrac{1}{2}\rangle &= (m-j)^{1/2}|j+1\ m\rangle\langle j+1\|a\|j+\tfrac{1}{2}\rangle \\ &\quad + (m+j)^{1/2}|jm\rangle\langle j\|a\|j+\tfrac{1}{2}\rangle, \end{aligned} \tag{3.16}$$

and it is obvious for $a^\dagger$. Because the relation (3.11) must hold for any j, we get

$$\begin{aligned} &\langle j+\tfrac{1}{2}\|a\|j\rangle\langle j\|a\|j+\tfrac{1}{2}\rangle = 2, \\ &\langle j+\tfrac{1}{2}\|a\|j\rangle\langle j+1\|a\|j+\tfrac{1}{2}\rangle = 0, \\ &\langle j+\tfrac{1}{2}\|a\|j+1\rangle\langle j+1\|a\|j+\tfrac{1}{2}\rangle = 0, \\ &\langle j+\tfrac{3}{2}\|a\|j+1\rangle\langle j+1\|a\|j+\tfrac{1}{2}\rangle = 0. \end{aligned} \tag{3.17}$$

The third relation together with (3.15) gives $\langle j+1\|a\|j+\frac{1}{2}\rangle = 0$, and we obtain from the first relation

$$\langle j+\tfrac{1}{2}\|a\|j\rangle = 2^{1/2}e^{i\delta}, \quad \langle j\|a\|j+\tfrac{1}{2}\rangle = 2^{1/2}e^{-i\delta}, \tag{3.18}$$

where δ denotes a phase factor which may be fixed to zero without loss of generality. In this way, we could determine the action of a on the bases as follows:

$$\begin{aligned} a|jm\rangle &= [2(m-j)]^{1/2}|j+\tfrac{1}{2}\ m-\tfrac{1}{2}\rangle, \\ a|j+\tfrac{1}{2}\ m+\tfrac{1}{2}\rangle &= [2(m+j)]^{1/2}|jm\rangle, \end{aligned} \tag{3.19}$$

and obvious for $a^\dagger$. We, therefore, see that two irreducible representations of SO(2,1) are needed in order to describe the representation of the graded Lie algebra GLA(Sp(2):2).

It is obvious that the vacuum state $|0\rangle$ defined in (3.3) is given by $|jj\rangle$ and the following relation holds:

$$aa^\dagger|jm\rangle = 2(m+j)|jm\rangle. \tag{3.20}$$

Thus the order of quantization is given by $4j(=2p)$. We find the eigenvalue of the number operator $N = 2J_3 - 2j$:

$$\begin{aligned} N|jm\rangle &= 2(m-j)|jm\rangle, \\ N|j+\tfrac{1}{2}\ m+\tfrac{1}{2}\rangle &= 2(m-j)|j+\tfrac{1}{2}\ m+\tfrac{1}{2}\rangle, \end{aligned} \tag{3.21}$$

which agree with the known result.[4,9]

Let us now consider the case a_1, a_2 and their Hermitian conjugates. The bases (2.6) may be used for our purpose. It follows from the above discussion for a and (3.8) that the matrix elements for a_i are expressed as follows:

$$\begin{aligned} &\langle p'q':j_1+\tfrac{1}{2}\ m_1-\tfrac{1}{2}j_2m_2|a_1|pq:j_1m_1j_2m_2\rangle \\ &\quad = (m_1-j_1)^{1/2}\langle p'q':j_1+\tfrac{1}{2}j_2\|a_1\|pq:j_1j_2\rangle, \\ &\langle p'q':j_1-\tfrac{1}{2}\ m_1-\tfrac{1}{2}j_2m_2|a_1|pq:j_1m_1,j_2m_2\rangle \\ &\quad = (m_1+j_1-1)^{1/2}\langle p'q':j_1-\tfrac{1}{2}j_2\|a\|pq:j_1j_2\rangle. \end{aligned} \tag{3.22}$$

The expressions for $a_1^\dagger$ are obvious and those for $a_2(a_2^\dagger)$ are obtained by interchanging the role of j_1,m_1 and j_2,m_2. We get from the second relation of (3.8), i.e., $[a_1,J_+^{(1)}] = a_1^\dagger$, together with (3.22)

$$\begin{aligned} \langle p'q':j_1'j_2\|a_1^\dagger\|pq:j_1j_2\rangle &= \langle p'q':j_1'j_2\|a_1\|pq:j_1j_2\rangle \\ &= \langle pq:j_1j_2\|a_1\|p'q':j_1'j_2\rangle^* \quad \text{for } j_1' = j_1 \pm \tfrac{1}{2}, \end{aligned} \tag{3.23}$$

and the corresponding relations for a_2 and $a_2^\dagger$. Thus we see that the equality of the reduced matrix elements given in the last row in (2.7) is obvious, and it is sufficient for us to consider the matrix elements for one of $J_\pm^{(3)}$ and $J_\pm^{(4)}$. We, also, see from (3.8) that a_1 (a_2) commutes with $J_3^{(2)}$ and $J_\pm^{(2)}$ ($J_3^{(1)}$ and $J_\pm^{(1)}$, and thus the matrix elements of a_1 (a_2) are diagonal with respect to j_2 and m_2 (j_1 and m_1).

In order to obtain the j_1 and j_2 dependence of a_1, we take the matrix elements of the relation $[a_1,J_-^{(3)}] = 0$:

$$\langle p'q':j_1+1\ m_1-1,j_2'\ m_2-\tfrac{1}{2}|\,[a_1,J_-^{(3)}]\,|pq:j_1m_1j_2m_2\rangle = 0$$
$$\text{for } j_2' = j_2 \pm \tfrac{1}{2}.$$

It is straightforward to obtain the following relations by using (2.7):

$$\begin{aligned} &\left(\frac{\Gamma(2j_1+2)\Gamma(\sigma+q+1)\Gamma(\sigma-q)\Gamma(\sigma-p+2)\Gamma(\sigma+p-1)}{\Gamma(2j_1)\Gamma(\sigma+q'+\frac{3}{2})\Gamma(\sigma-q'+\frac{1}{2})\Gamma(\sigma-p'+\frac{5}{2})\Gamma(\sigma+p'-\frac{1}{2})}\right)^{1/2}\langle p'q':j_1+1\,j_2+\tfrac{1}{2}\|a_1\|pq:j_1+\tfrac{1}{2}\,j_2+\tfrac{1}{2}\rangle \\ &\quad = \left(\frac{\Gamma(2j_1+1)\Gamma(\sigma+q)\Gamma(\sigma-q-1)\Gamma(\sigma-p+1)\Gamma(\sigma+p-2)}{\Gamma(2j_1-1)\Gamma(\sigma+q'+\frac{1}{2})\Gamma(\sigma-q'-\frac{1}{2})\Gamma(\sigma-p'+\frac{3}{2})\Gamma(\sigma+p'-\frac{3}{2})}\right)^{1/2}\langle p'q':j_1+\tfrac{1}{2}\,j_2\|a_1\|pq:j_1j_2\rangle, \end{aligned} \tag{3.24}$$

$$\begin{aligned} &\left(\frac{\Gamma(2j_1+2)\Gamma(\delta+q+2)\Gamma(\delta-q+1)\Gamma(\delta-p+3)\Gamma(\delta+p)}{\Gamma(2j_1)\Gamma(\delta+q'+\frac{5}{2})\Gamma(\delta-q'+\frac{3}{2})\Gamma(\delta-p'+\frac{7}{2})\Gamma(\delta+p'+\frac{1}{2})}\right)^{1/2}\langle p'q':j_1+1\,j_2-\tfrac{1}{2}\|a_1\|pq:j_1+\tfrac{1}{2}\,j_2-\tfrac{1}{2}\rangle \\ &\quad = \left(\frac{\Gamma(2j_1+1)\Gamma(\delta+q+1)\Gamma(\delta-q)\Gamma(\delta-p+2)\Gamma(\delta+p-1)}{\Gamma(2j_1-1)\Gamma(\delta+q'+\frac{3}{2})\Gamma(\delta-q'+\frac{1}{2})\Gamma(\delta-p'+\frac{5}{2})\Gamma(\delta+p'-\frac{1}{2})}\right)^{1/2}\langle p'q'j_1+\tfrac{1}{2}\,j_2\|a_1\|pq:j_1j_2\rangle. \end{aligned}$$

(3.24) gives the following expression:

$$\begin{aligned} \langle p'q':j_1+\tfrac{1}{2}\,j_2\|a_1\|pqj_1j_2\rangle &= \left(\frac{\Gamma(2j_1-1)\Gamma(\sigma+q'+\frac{1}{2})\Gamma(\sigma-q'-\frac{1}{2})\Gamma(\sigma-p'+\frac{3}{2})}{\Gamma(2j_1+1)\Gamma(\sigma+q)\Gamma(\sigma-q-1)\Gamma(\sigma-p+1)\Gamma(\sigma+p-2)}\right)^{1/2} \\ &\quad\times\left(\frac{\Gamma(\sigma+p'-\frac{3}{2})\Gamma(\delta+q'+\frac{3}{2})\Gamma(\delta-q'+\frac{1}{2})\Gamma(\delta-p'+\frac{5}{2})\Gamma(\delta+p'-\frac{1}{2})}{\Gamma(\delta+q+1)\Gamma(\delta-q)\Gamma(\delta-p+2)\Gamma(\delta+p-1)}\right)^{1/2}\langle p'q'\|a_1\|pq\rangle. \end{aligned} \tag{3.25}$$

It is noted that the dependence of a_1 on j_1 and j_2 is completely determined except for the reduced matrix elements $\langle p'q'\|a_1\|pq\rangle$. Similarly, considering the matrix elements $\langle p'q':j_1-1\ m_1-1 j_2'\ m_2-\frac{1}{2}|[a_1,J_-^{(3)}]|pq:j_1m_1j_2m_2\rangle=0$ for $j_2'=j_2\pm\frac{1}{2}$, we get the following expression:

$$\langle p'q':j_1-\tfrac{1}{2}j_2\|a_1\|pq:j_1j_2\rangle=\left(\frac{\Gamma(2j_1-2)\Gamma(\sigma+q)\Gamma(\sigma-q-1)\Gamma(\sigma-p+1)}{\Gamma(2j_1)\Gamma(\sigma+q'-\frac{1}{2})\Gamma(\sigma-q'-\frac{3}{2})\Gamma(\sigma-p'+\frac{1}{2})}\right)^{1/2}$$
$$\times\left(\frac{\Gamma(\sigma+p-2)\Gamma(\delta+q+1)\Gamma(\delta-q)\Gamma(\delta-p+2)\Gamma(\delta+p-1)}{\Gamma(\sigma+p'-\frac{5}{2})\Gamma(\delta+q'+\frac{1}{2})\Gamma(\delta-q'-\frac{1}{2})\Gamma(\delta-p'+\frac{3}{2})\Gamma(\delta+p'-\frac{3}{2})}\right)^{1/2}\langle p'q'\|a_1\|pq\rangle. \quad (3.26)$$

The expressions for a_2 are obtained from (3.25) and (3.26) by interchanging the role of j_1 and j_2. The expressions for $a_1^\dagger$ and $a_2^\dagger$ are obvious from (3.23).

Let us now determine p' and q' for which the reduced matrix element $\langle p'q'\|a_1\|pq\rangle$ is not zero. Taking into account the matrix elements $\langle p'q':j_1\ m_1-1 j_2'\ m_2-\frac{1}{2}|[a_1,J_-^{(3)}]|pq:j_1m_1j_2m_2\rangle=0$ for $j_2'=j_2\pm\frac{1}{2}$ together with (2.7), (3.25), and (3.26), we obtain the following:

$$\begin{aligned}
&\{[\,j_1(\sigma+q'-\tfrac{1}{2})(\sigma-q'-\tfrac{3}{2})(\sigma-p'+\tfrac{1}{2})(\sigma+p'-\tfrac{5}{2})-(j_1-1)(\sigma+q)(\sigma-q-1)(\sigma-p+1)(\sigma+p-2)]\\
&\quad\times\frac{\Gamma(\sigma+q)\Gamma(\sigma-q-1)\Gamma(\sigma-p+1)\Gamma(\sigma+p-2)\Gamma(\delta+q+1)\Gamma(\delta-q)\Gamma(\delta-p+2)\Gamma(\delta+p-1)}{\Gamma(\sigma+q'+\frac{1}{2})\Gamma(\sigma-q'-\frac{1}{2})\Gamma(\sigma-p'+\frac{3}{2})\Gamma(\sigma+p'-\frac{3}{2})\Gamma(\delta+q'+\frac{1}{2})\Gamma(\delta-q'-\frac{1}{2})}\\
&\quad-[\Gamma(\delta-p'+\tfrac{3}{2})\Gamma(\delta+p'-\tfrac{3}{2})]^{-1}\\
&\quad\times[\,j_1(\delta+q)(\delta-q-1)(\delta-p+1)(\delta+p-2)-(j_1-1)(\delta+q'+\tfrac{1}{2})(\delta-q'-\tfrac{1}{2})(\delta-p'+\tfrac{3}{2})(\delta+p-\tfrac{3}{2})]\,)\\
&\quad\times\langle p'q'\|a_1\|pq\rangle=0.\\
&\{[\,j_1(\delta+q'+\tfrac{1}{2})(\delta-q'-\tfrac{1}{2})(\delta-p'+\tfrac{3}{2})(\delta+p'-\tfrac{3}{2})-(j_1-1)(\delta+q+1)(\delta-q)(\delta-p+2)(\delta+p-1)]\\
&\quad\times\frac{\Gamma(\sigma+q)\Gamma(\sigma-q-1)\Gamma(\sigma-p+1)\Gamma(\sigma+p-2)\Gamma(\delta+q+1)\Gamma(\delta-q)\Gamma(\delta-p+2)\Gamma(\delta+p-1)}{\Gamma(\sigma+q'-\frac{1}{2})\Gamma(\sigma-q'-\frac{3}{2})\Gamma(\sigma-p'+\frac{1}{2})\Gamma(\sigma+p'-\frac{5}{2})\Gamma(\delta+q'+\frac{3}{2})\Gamma(\delta-q'+\frac{1}{2})\Gamma(\delta-p'+\frac{5}{2})\Gamma(\delta+p'-\frac{1}{2})}\\
&\quad-[\,j_1(\sigma+q-1)(\sigma-q-2)(\sigma-p)(\sigma+p-3)-(j_1-1)(\sigma+q'-\tfrac{1}{2})(\sigma-q'-\tfrac{3}{2})(\sigma-p'+\tfrac{1}{2})(\sigma+p'-\tfrac{5}{2})]\,)\\
&\quad\times\langle p'q'\|a_1\|pq\rangle=0.
\end{aligned} \quad (3.27)$$

The reduced matrix elements $\langle p'q'\|a_1\|pq\rangle$ are not zero only when the coefficients in (3.27) vanish. Thus we get the equation

$$\begin{aligned}
&(\delta+q'+\tfrac{1}{2})(\delta-q'-\tfrac{1}{2})(\delta-p'+\tfrac{3}{2})(\delta+p'-\tfrac{3}{2})[(\sigma+\delta)(\sigma+q-1)(\sigma-q-2)(\sigma-p)(\sigma+p-3)\\
&\quad-(\sigma+\delta-2)(\sigma+q'-\tfrac{1}{2})(\sigma-q'-\tfrac{3}{2})(\sigma-p'+\tfrac{1}{2})(\sigma+p'-\tfrac{5}{2})]\ [(\sigma+\delta)(\sigma+q'-\tfrac{1}{2})(\sigma-q'-\tfrac{3}{2})\\
&\quad\times(\sigma-p'+\tfrac{1}{2})(\sigma+p'-\tfrac{5}{2})-(\sigma+\delta-2)(\sigma+q)(\sigma-q-1)(\sigma-p+1)(\sigma+p-2)]\\
&\quad-(\sigma+q'-\tfrac{1}{2})(\sigma-q'-\tfrac{3}{2})(\sigma-p'+\tfrac{1}{2})(\sigma+p'-\tfrac{5}{2})[(\sigma+\delta)(\delta+q'+\tfrac{1}{2})(\delta-q'-\tfrac{1}{2})(\delta-p'+\tfrac{3}{2})(\delta+p'-\tfrac{3}{2})\\
&\quad-(\sigma+\delta-2)(\delta+q+1)(\delta-q)(\delta-p+2)(\delta+p-1)]\,[(\sigma+\delta)(\delta+q)(\delta-q-1)\\
&\quad\times(\delta-p+1)(\delta+p-2)-(\sigma+\delta-2)(\delta+q'+\tfrac{1}{2})(\delta-q'-\tfrac{1}{2})(\delta-p'+\tfrac{3}{2})(\delta+p'-\tfrac{3}{2})]=0.
\end{aligned} \quad (3.28)$$

(3.28) must hold for any possible σ and δ. Equating the coefficients of each power of σ and δ to zero, we get

$$\begin{aligned}
&[(q'+\tfrac{1}{2})^2+(p'-\tfrac{3}{2})^2][(q'+\tfrac{1}{2})^2+(p'-\tfrac{3}{2})^2-q(q+1)-(p-1)(p-2)-1]^2+4(q'+\tfrac{1}{2})^2(p'-\tfrac{3}{2})^2-2\\
&\quad\times[(q'+\tfrac{1}{2})^2(p'-\tfrac{3}{2})^2-(p-1)(p-2)q(q+1)][(q'+\tfrac{1}{2})^2+(p'-\tfrac{3}{2})^2-q(q+1)-(p-1)(p-2)+1]\\
&\quad-[(p-1)(p-2)+q(q+1)]^2=0,\\
&\{(q'+\tfrac{1}{2})^2(p'-\tfrac{3}{2})^2-q(q+1)(p-1)(p-2)-(q'+\tfrac{1}{2})(p'-\tfrac{3}{2})[(q'+\tfrac{1}{2})^2+(p'-\tfrac{3}{2})^2-q(q+1)-(p-1)(p-2)-1]\}\\
&\quad\times\{(q'+\tfrac{1}{2})^2(p'-\tfrac{3}{2})^2-(p-1)(p-2)q(q+1)+(q'+\tfrac{1}{2})(p'-\tfrac{3}{2})[(q'+\tfrac{1}{2})^2+(p'-\tfrac{3}{2})^2-q(q+1)-(p-1)(p-2)-1]\}\\
&\quad=0.
\end{aligned} \quad (3.29)$$

It follows from the second relation of (3.29) that one of the following relations must hold

$$(q'+\tfrac{1}{2})^2(p'-\tfrac{3}{2})^2-(p-1)(p-2)q(q+1)=\pm(q'+\tfrac{1}{2})(p'-\tfrac{3}{2})[(q'+\tfrac{1}{2})^2+(p'-\tfrac{3}{2})^2-q(q+1)-(p-1)(p-2)-1]. \quad (3.30)$$

The expression of (3.30) with the negative sign is obtained from the positive one by replacing $q'\to -q'-1$ or $p'\to -p'+3$ and then the first equation of (3.29) is invariant under this replacement. Thus it is sufficient for us to find the solution for the positive sign in (3.30). From (3.30) and the first of (3.29), we find

$$\begin{aligned}
&(q'-p'+1)(q'-p'+3)[F(p,q)-F(p'+\tfrac{1}{2},q'+\tfrac{1}{2})]\\
&\quad\times[F(p,q)-F(p'-\tfrac{1}{2},q'-\tfrac{1}{2})]=0,
\end{aligned} \quad (3.31)$$

where

$$F(p,q)=p(p-3)+q(q+1).$$

As in general the first two factors in (3.31) can not satisfy

(3.30), we get

$$F(p'+\tfrac{1}{2},q'+\tfrac{1}{2})-F(p,q)=0, \tag{3.32a}$$

$$F(p'-\tfrac{1}{2},q'-\tfrac{1}{2})-F(p,q)=0. \tag{3.32b}$$

The solutions of (3.30) and (3.32a) are

$$\begin{Bmatrix}p'=p-\frac{1}{2}\\ q'=q-\frac{1}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=p-\frac{1}{2}\\ q'=-q-\frac{3}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=-p+\frac{5}{2}\\ q'=q-\frac{1}{2}\end{Bmatrix},$$
$$\begin{Bmatrix}p'=-p+\frac{5}{2}\\ q'=-q-\frac{3}{2}\end{Bmatrix}, \tag{3.33a}$$

$$\begin{Bmatrix}p'=q+\frac{3}{2}\\ q'=p-\frac{5}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=q+\frac{3}{2}\\ q'=-p+\frac{1}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=-q+\frac{1}{2}\\ q'=+p-\frac{5}{2}\end{Bmatrix},$$
$$\begin{Bmatrix}p'=-q+\frac{1}{2}\\ q'=-p+\frac{1}{2}\end{Bmatrix}. \tag{3.33b}$$

Those of (3.30) and (3.32b) are

$$\begin{Bmatrix}p'=p+\frac{1}{2}\\ q'=q+\frac{1}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=p+\frac{1}{2}\\ q'=-q-\frac{1}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=-p+\frac{5}{2}\\ q'=q+\frac{1}{2}\end{Bmatrix},$$
$$\begin{Bmatrix}p'=-p+\frac{5}{2}\\ q'=-q-\frac{1}{2}\end{Bmatrix}, \tag{3.34a}$$

$$\begin{Bmatrix}p'=q+\frac{5}{2}\\ q'=p-\frac{3}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=q+\frac{5}{2}\\ q'=-p+\frac{3}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=-q+\frac{3}{2}\\ q'=p-\frac{3}{2}\end{Bmatrix},$$
$$\begin{Bmatrix}p'=-q+\frac{3}{2}\\ q'=-p+\frac{3}{2}\end{Bmatrix}. \tag{3.34b}$$

It is easy to see that the set of solutions in (3.33b) and (3.34b) cannot make the coefficients in (3.27) vanish. Thus the solutions are given by (3.33a) and (3.34a). However, we must take into account the nonnegativeness of p' and q' as well as p and q. Taking into account the fact and the remarks below (3.30), we obtain p' and q' for which the reduced matrix elements $\langle p'q'\|a_1\|pq\rangle$ are not zero:

$$\begin{Bmatrix}p'=p+\frac{1}{2}\\ q'=q+\frac{1}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=p+\frac{1}{2}\\ q'=q-\frac{1}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=p-\frac{1}{2}\\ q'=q+\frac{1}{2}\end{Bmatrix},\quad \begin{Bmatrix}p'=p-\frac{1}{2}\\ q'=q-\frac{1}{2}\end{Bmatrix}. \tag{3.35}$$

It is obvious that the same conditions hold for the reduced matrix elements of a_2.

Let us determine the action of a_1 on the bases (2.6) in the similar way as in SO(2,1). We start with the case of $q=0$ in the class (a). The bases with $q=0$ are given by

$$|p0{:}j_1m_1j_2m_2\rangle, \tag{3.36}$$

where j_1 and j_2 take the same value, i.e., $j_1=j_2=(p+n)/2$, $n=0,1,2,\cdots$. Those with $q=\frac{1}{2}$ have the form

$$|p'\tfrac{1}{2}{:}j_1'm_1j_2'm_2\rangle, \tag{3.37}$$

where j_1' and j_2' are given by

$$j_1'=(p'+n)/2+\epsilon/4,\quad j_2'=(p'+n)/2-\epsilon/4,$$
$$\epsilon^2=1.$$

Assuming that p in (3.36) is minimum, we see from (3.35) that p' and q' after the action of a_1 on the bases (3.36) take only the values $p'=p+\frac{1}{2}$ and $q'=\frac{1}{2}$. Thus the bases after the action of a_1 on the bases (3.36) have the form $|p+\frac{1}{2}\frac{1}{2}{:}j_1'\ m_1-\frac{1}{2}j_2'm_2\rangle$ with $j_1'=j_1+(1+\epsilon)/4$, $j_2'=j_2+(1-\epsilon)/4$ and $j_1=j_2=(p+n)/2$. As is seen from (3.22), j_1 should change into $j_1\pm\frac{1}{2}$ and j_2 does not change after the action of a_1. This actually occurs in the above result as follows:

$$\begin{aligned} j_1'&=j_1+\tfrac{1}{2},\quad j_2'=j_2 \quad \text{for } \epsilon=1,\\ j_1'&=j_1-\tfrac{1}{2},\quad j_2'=j_2 \quad \text{for } \epsilon=-1.\end{aligned} \tag{3.38}$$

Thus the action of a_1 on the bases (3.36) with a minimum p can be written as follows:

$$\begin{aligned} a_1|p0{:}j_1m_1j_2m_2\rangle &= (m_1-j_1)^{1/2}|p+\tfrac{1}{2}\tfrac{1}{2}{:}j_1+\tfrac{1}{2}\ m_1-\tfrac{1}{2}j_2m_2\rangle\\ &\times\langle p+\tfrac{1}{2}\tfrac{1}{2}{:}j_1+\tfrac{1}{2}j_2\|a_1\|p0{:}j_1j_2\rangle\\ &+(m_1+j_1-1)^{1/2}|p+\tfrac{1}{2}\tfrac{1}{2}{:}j_1-\tfrac{1}{2}\ m_1-\tfrac{1}{2}j_2m_2\rangle\\ &\times\langle p+\tfrac{1}{2}\tfrac{1}{2}{:}j_1-\tfrac{1}{2}j_2\|a_1\|p0{:}j_1j_2\rangle,\end{aligned} \tag{3.39}$$

where (3.22) is taken into account. Similar consideration gives the following expressions:

$$\begin{aligned} a_1|p+\tfrac{1}{2}\tfrac{1}{2}{:}j_1+\tfrac{1}{2}\ m_1+\tfrac{1}{2}j_2m_2\rangle &= (m_1+j_1)^{1/2}|p0{:}j_1m_1j_2m_2\rangle\ f_+(p0{:}j_1j_2)\\ &+(m_1+j_1)^{1/2}|p+1\ 0{:}j_1m_1j_2m_2\rangle\ f_+(p+1\ 0{:}j_1j_2)\\ &+(m_1+j_1)^{1/2}|p1{:}j_1m_1j_2m_2\rangle\ f_+(p1{:}j_1j_2)\\ &+(m_1+j_1)^{1/2}|p+1\ 1{:}j_1m_1j_2m_2\rangle\ f_+(p+1\ 1{:}j_1j_2),\end{aligned} \tag{3.40a}$$

$$\begin{aligned} a_1|p+\tfrac{1}{2}\tfrac{1}{2}{:}j_1-\tfrac{1}{2}\ m_1+\tfrac{1}{2}j_2m_2\rangle &= (m_1-j_1+1)^{1/2}|p0{:}j_1m_1j_2m_2\rangle\ f_-(p0{:}j_1j_2)\\ &+(m_1-j_1+1)^{1/2}|p+1\ 0{:}j_1m_1j_2m_2\rangle\ f_-(p+1\ 0{:}j_1j_2)\\ &+(m_1-j_1+1)^{1/2}|p1{:}j_1m_1j_2m_2\rangle\ f_-(p1{:}j_1j_2)\\ &+(m_1-j_1+1)^{1/2}|p+1\ 1{:}j_1m_1j_2m_2\rangle\ f_-(p+1\ 1{:}j_1j_2),\end{aligned} \tag{3.40b}$$

where

$$f_\pm(p'q'{:}j_1j_2)=\langle p'q'{:}j_1j_2\|a_1\|p+\tfrac{1}{2}\tfrac{1}{2}{:}j_1\pm\tfrac{1}{2}j_2\rangle.$$

The matrix elements of the generators of SO(3,2) consisting of a_i and $a_i^\dagger$ must coincide with those in Sec. II. Thus the following relation must hold:

$$\begin{aligned} J_-^{(1)}|p0{:}j_1m_1j_2m_2\rangle &= [(m_1-j_1)(m_1+j_1-1)]^{1/2}|p0{:}j_1m_1-1,j_2m_2\rangle,\\ 2J_-^{(1)} &= a_1a_1.\end{aligned} \tag{3.41}$$

We obtain from (3.39), (3.40), and (3.41)

$$f_+^*(p0{:}j_1j_2)f_+(p0{:}j_1j_2)+f_-^*(p0{:}j_1j_2)f_-(p0{:}j_1j_2)=2, \tag{3.42}$$

$$f_+^*(p0{:}j_1j_2)f_+(p'q'{:}j_1j_2)+f_-^*(p0{:}j_1j_2)f_-(p'q'{:}j_1j_2)=0$$
$$\text{for } (p',q')=(p+1,0),(p,1), \text{ and } (p+1,1).$$

Substitution of (3.25) and (3.26) into (3.42) gives the following results:

(i) For $p>1$,

$$\langle p+\tfrac{1}{2}\,\tfrac{1}{2}||a_1||p0\rangle = (p-1)^{-1/2}e^{i\delta},$$

$$f_+(p+1\,0;j_1j_2) = i\epsilon\left(\frac{(2j_1-p)(p-1)}{2j_1-1}\right)^{1/2} \times\langle p+1\,0||a_1||p+\tfrac{1}{2}\,\tfrac{1}{2}\rangle, \tag{3.43}$$

$$f_-(p+1\,0;j_1j_2) = -i\epsilon\left(\frac{(2j_1+p-2)(p-1)}{2j_1-1}\right)^{1/2} \times\langle p+1\,0||a_1||p+\tfrac{1}{2}\,\tfrac{1}{2}\rangle,$$

$$\langle p1||a_1||p+\tfrac{1}{2}\,\tfrac{1}{2}\rangle = \langle p+1\,1||a_1||p+\tfrac{1}{2}\,\tfrac{1}{2}\rangle = 0.$$

(ii) For $\frac{1}{2} < p < 1$,

$$\langle p+\tfrac{1}{2}\,\tfrac{1}{2}||a_1||p0\rangle = (1-p)^{-1/2}e^{i\delta},$$

$$f^*_+(p0;j_1j_2) = i\epsilon\left(\frac{(2j_1+p-2)(1-p)}{2j_1-1}\right)^{1/2} \times\langle p+\tfrac{1}{2}\,\tfrac{1}{2}||a_1||p0\rangle, \tag{3.44}$$

$$f^*_-(p0;j_1j_2) = -i\epsilon\left(\frac{(2j_1-p)(1-p)}{2j_1-1}\right)^{1/2} \times\langle p+\tfrac{1}{2}\,\tfrac{1}{2}||a_1||p0\rangle,$$

$$\langle p1||a_1||p+\tfrac{1}{2}\,\tfrac{1}{2}\rangle = \langle p+1\,1||a_1||p+\tfrac{1}{2}\,\tfrac{1}{2}\rangle = 0.$$

δ is a phase factor and ϵ is an arbitrary constant taking 1 or -1. These constants may, of course, differ in (3.43) and (3.44). It is noted that the difference of the relative sign in the terms involving ϵ in (3.43) and (3.44) is due to the double valuedness of the function of the square root and then the second relation for $p' = p+1$ and $q' = 0$ in (3.42) is satisfied. Similar consideration on the bases $|p+1\,0;j_1m_1j_2m_2\rangle$ gives

(i) $p > 1$,

$$\langle p+\tfrac{1}{2}\,\tfrac{1}{2}||a_1||p+1\,0\rangle = (p-1)^{-1/2}e^{i\eta},$$
$$\langle p+\tfrac{3}{2}\,\tfrac{1}{2}||a_1||p+1\,0\rangle = 0. \tag{3.45}$$

(ii) $\frac{1}{2} < p < 1$,

$$\langle p+\tfrac{1}{2}\,\tfrac{1}{2}||a_1||p+1\,0\rangle = (1-p)^{-1/2}e^{i\eta},$$
$$\langle p+\tfrac{3}{2}\,\tfrac{1}{2}||a_1||p+1\,0\rangle = 0. \tag{3.46}$$

η is a phase factor. No further result is obtained from $J^{(1)}_-$. (3.43)–(3.46) together with (3.22) give the matrix elements of a_1 up to the phase factors. The expressions for $a_1^\dagger$ are obvious due to (3.23). Similar expressions for a_2 ($a_2^\dagger$) hold and we express them with the primed phases.

In order to obtain the relation among the phases, we consider $J^{(3)}_-$ [$= (a_1a_2 + a_2a_1)/2$]. The action of $J^{(3)}_-$ on the base $|p0;j_1m_1j_2m_2\rangle$ ($p > 1$) gives the expression

$$\begin{aligned}
2J^{(3)}_-&|p0;j_1m_1j_2m_2\rangle \\
&= (e^{i(\delta-\delta')} + e^{-i(\delta-\delta')})[(m_1-j_1)(m_2-j_2)(2j_1+p-2)(2j_1-p+1)/2j_1(2j_1-1)]^{1/2}|p0;j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2+\tfrac{1}{2}\,m_2-\tfrac{1}{2}\rangle \\
&\quad + (e^{i(\delta-\delta')} + e^{-i(\delta-\delta')})[(m_1+j_1-1)(m_2+j_2-1)(2j_1-p)(2j_1+p-3)/(2j_1-1)(2j_1-2)]^{1/2} \\
&\quad \times|p0;j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2-\tfrac{1}{2}\,m_2-\tfrac{1}{2}\rangle + i(\epsilon e^{i(\delta'-\eta)} + \epsilon' e^{i(\delta-\eta')})[(m_1+j_1-1)(m_2+j_2-1)(2j_1-p) \\
&\quad \times(2j_1-p-1)/(2j_1-1)(2j_1-2)]^{1/2}|p+1\,0;j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2-\tfrac{1}{2}\,m_2-\tfrac{1}{2}\rangle \\
&\quad - i(\epsilon e^{i(\delta'-\eta)} + \epsilon' e^{i(\delta-\eta')})[(m_1-j_1)(m_2-j_2)(2j_1+p-1)(2j_1+p-2)/2j_1(2j_1-1)]^{1/2} \\
&\quad \times|p+1\,0;j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2+\tfrac{1}{2}\,m_2-\tfrac{1}{2}\rangle.
\end{aligned} \tag{3.47}$$

Similar expression holds for $\frac{1}{2} < p < 1$. Comparison of the matrix elements for $J^{(3)}_-$ in Sec. II and (3.47) gives

(i) $p > 1$,

$$e^{i\delta} = e^{i\delta'}, \quad \epsilon e^{i\eta} + \epsilon' e^{i\eta'} = 0. \tag{3.48}$$

(ii) $\frac{1}{2} < p < 1$,

$$e^{i\eta} = e^{i\eta'}, \quad \epsilon e^{i\delta} + \epsilon' e^{i\delta'} = 0. \tag{3.49}$$

No further condition except for (3.48) or (3.49) exists. Thus we may fix the arbitrary phases without loss of generality in the form; i.e., $\delta = 0$, $\epsilon e^{i\eta} = -i$ for $p > 1$ and $\eta = 0$, $\epsilon e^{i\delta} = -i$ for $\frac{1}{2} < p < 1$. Thus we could determine the action of a_i and $a_i^\dagger$ on the bases, and they are summarized as follows:

(I) The case with $q = 0$ in (a):

$$\begin{aligned}
a_1|p0;j_1m_1j_2m_2\rangle &= [(m_1-j_1)(2j_1+p-2)/(2j_1-1)]^{1/2}|p+\tfrac{1}{2}\,\tfrac{1}{2};j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle \\
&\quad + \epsilon[(m_1+j_1-1)(2j_1-p)/(2j_1-1)]^{1/2}|p+\tfrac{1}{2}\,\tfrac{1}{2};j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle,
\end{aligned}$$

a_2: $(1 \rightleftarrows \epsilon)$,

$$\begin{aligned}
a_1|p+1\,0;j_1m_1j_2m_2\rangle &= -\epsilon[(m_1-j_1)(2j_1-p)/(2j_1-1)]^{1/2}|p+\tfrac{1}{2}\,\tfrac{1}{2};j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle \\
&\quad + [(m_1+j_1-1)(2j_1+p-2)/(2j_1-1)]^{1/2}|p+\tfrac{1}{2}\,\tfrac{1}{2};j_1-\tfrac{1}{2}m_1-\tfrac{1}{2}j_2m_2\rangle,
\end{aligned}$$

a_2: $(-\epsilon \rightleftarrows 1)$,

$$\begin{aligned}
a_1|p+\tfrac{1}{2}\,\tfrac{1}{2};j_1+\tfrac{1}{2}\,m_1+\tfrac{1}{2}j_2m_2\rangle &= [(m_1+j_1)(2j_1+p-2)/(2j_1-1)]^{1/2}|p0;j_1m_1j_2m_2\rangle \\
&\quad - \epsilon[(m_1+j_1)(2j_1-p)/(2j_1-1)]^{1/2}|p+1\,0;j_1m_1j_2m_2\rangle,
\end{aligned}$$

a_2: $(1 \rightarrow \epsilon, -\epsilon \rightarrow 1)$,

$$a_1|p+\tfrac{1}{2}\tfrac{1}{2}:j_1-\tfrac{1}{2}\,m_1+\tfrac{1}{2}j_2m_2\rangle=\epsilon[(m_1-j_1+1)(2j_1-p)/(2j_1-1)]^{1/2}|p0:j_1m_1j_2m_2\rangle,$$
$$+[(m_1-j_1+1)(2j_1+p-2)/(2j_1-1)]^{1/2}|p+1\,0:j_1m_1j_2m_2\rangle,$$
$$a_2\text{: }(\epsilon\to 1,\ 1\to-\epsilon), \tag{3.50}$$

where ϵ is equal to 1 for $p>1$ and -1 for $\frac{1}{2}<p<1$, and the notation such as a_2:$(1\rightleftarrows\epsilon)$ after each expression for a_1 means that the expression for a_2 is obtained from that corresponding to a_1 by interchanging the role of j_1,m_1 and j_2,m_2 and further the coefficients 1 and ϵ on the right-hand side. For instance, the last notation a_1:$(\epsilon\to 1,\ 1\to-\epsilon)$ denotes the following:

$$a_2|p+\tfrac{1}{2}\tfrac{1}{2}:j_1m_1j_2-\tfrac{1}{2}\,m_2+\tfrac{1}{2}\rangle=[(m_2-j_2+1)(2j_2-p)/(2j_2-1)]^{1/2}|p0:j_1m_1j_2m_2\rangle$$
$$-\epsilon[(m_2-j_2+1)(2j_2+p-2)/(2j_2-1)]^{1/2}|p+1\,0:j_1m_1j_2m_2\rangle. \tag{3.51}$$

It is noted that j_1 and j_2 in the above expressions take the same values $(p+n)/2$ $(n=0,1,2,\cdots)$. The expressions for $p=1$ are obtained from those for $p>1$ as the limit of $p\to 1$. For instance, we get

$$a_1|1\,0:j_1m_1j_2m_2\rangle=(m_1-j_1)^{1/2}|\tfrac{3}{2}\tfrac{1}{2}:j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle+(m_1+j_1-1)^{1/2}|\tfrac{3}{2}\tfrac{1}{2}:j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle,$$
$$j_1=j_2=\tfrac{1}{2}(1+n). \tag{3.52}$$

It is straightforward to see that they are valid and give the irreducible representations of SO(3,2).

The representations of a_i and $a_i^\dagger$ in the other classes are treated in a similar way. We give only the results below.

(II) The cases with $q\neq 0$ in (a): This class is characterized by $p>q+\frac{1}{2}$ $(q=\frac{1}{2})$ and $p>q+1$ $(q=1,\frac{3}{2},2,\cdots)$. j_1 and j_2 take the values $j_1=(p+n+\delta)/2$ and $j_2=(p+n-\delta)/2$ with $\delta=-q,-q+1,\ldots,q-1,q$. The action of a_1 is given as follows:

$$a_1|pq:j_1m_1j_2m_2\rangle=\left(\frac{(m_1-j_1)(\sigma+q)(\sigma+p-2)(q+\delta+1)(p+\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q+\tfrac{1}{2}:j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle$$
$$+\left(\frac{(m_1+j_1-1)(\sigma-q-2)(\sigma-p)(q-\delta-1)(p-\sigma-1)}{(2j_1-1)(2j_1-2)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q+\tfrac{1}{2}:j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle$$
$$+\epsilon\left(\frac{(m_1-j_1)(\sigma-q-1)(\sigma+p-2)(q-\delta)(p+\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q-\tfrac{1}{2}:j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle$$
$$-\epsilon\left(\frac{(m_1+j_1-1)(\sigma+q-1)(\sigma-p)(q+\delta)(p-\delta-1)}{(2j_1-1)(2j_1-2)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q-\tfrac{1}{2}:j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle,$$

$$a_1|p+1q:j_1m_1j_2m_2\rangle=\epsilon\left(\frac{(m_1-j_1)(\sigma+q)(\sigma-p)(q+\delta+1)(p-\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q+\tfrac{1}{2}:j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle$$
$$-\epsilon\left(\frac{(m_1+j_1-1)(\sigma-q-2)(\sigma+p-2)(q-\delta+1)(p+\delta-1)}{(2j_1-1)(2j_1-2)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q+\tfrac{1}{2}:j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle$$
$$+\left(\frac{(m_1-j_1)(\sigma-q-1)(\sigma-p)(q-\delta)(p-\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q-\tfrac{1}{2}:j_1+\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle$$
$$+\left(\frac{(m_1+j_1-1)(\sigma+q-1)(\sigma+p-2)(q+\delta)(p+\delta-1)}{(2j_1-1)(2j_1-2)(2q+1)(p-1)}\right)^{1/2}|p+\tfrac{1}{2}\,q-\tfrac{1}{2}:j_1-\tfrac{1}{2}\,m_1-\tfrac{1}{2}j_2m_2\rangle,$$

$$a_1|p+\tfrac{1}{2}\,q+\tfrac{1}{2}:j_1+\tfrac{1}{2}\,m_1+\tfrac{1}{2}j_2m_2\rangle=\left(\frac{(m_1-j_1)(\sigma-q-1)(\sigma-p+1)(q-\delta)(p-\delta-2)}{2j_1(2j_1+1)(2q+1)(p-1)}\right)^{1/2}|pq:j_1+1m_1j_2m_2\rangle$$
$$+\left(\frac{(m_1+j_1)(\sigma+q)(\sigma+p-2)(q+\delta+1)(p+\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|pq:j_1m_1j_2m_2\rangle$$
$$-\epsilon\left(\frac{(m_1-j_1)(\sigma-q-1)(\sigma+p-1)(q-\delta)(p+\delta)}{2j_1(2j_1+1)(2q+1)(p-1)}\right)^{1/2}|p+1q:j_1+1m_1j_2m_2\rangle$$
$$+\epsilon\left(\frac{(m_1+j_1)(\sigma+q)(\sigma-p)(q+\delta+1)(p-\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|p+1\,q:j_1m_1j_2m_2\rangle,$$

$$a_1|p+\tfrac{1}{2}\,q-\tfrac{1}{2}:j_1+\tfrac{1}{2}\,m_1+\tfrac{1}{2}j_2m_2\rangle=-\epsilon\left(\frac{(m_1-j_1)(\sigma+q)(\sigma-p+1)(q+\delta+1)(p-\delta)}{2j_1(2j_1+1)(2q+1)(p-1)}\right)^{1/2}|pq:j_1+1\,m_1j_2m_2\rangle$$
$$+\epsilon\left(\frac{(m_1+j_1)(\sigma-q-1)(\sigma+p-2)(q-\delta)(p+\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|pq:j_1m_1j_2m_2\rangle$$
$$+\left(\frac{(m_1-j_1)(\sigma+q)(\sigma+p-1)(q+\delta+1)(p+\delta)}{2j_1(2j_1+1)(2q+1)(p-1)}\right)^{1/2}|p+1\,q:j_1+1m_1j_2m_2\rangle$$
$$+\left(\frac{(m_1+j_1)(\sigma-q-1)(\sigma-p)(q-\delta)(p-\delta-1)}{2j_1(2j_1-1)(2q+1)(p-1)}\right)^{1/2}|p+1\,q:j_1m_1j_2m_2\rangle, \tag{3.53}$$

where $\epsilon^2=1$ as well as $\sigma=j_1+j_2$ and $\delta=j_1-j_2$. The expressions for $a_1^\dagger$ are easily obtained by taking into account (3.23).

Those for a_2 are given by interchanging j_1,m_1 and j_2,m_2 and substituting $-\epsilon$ for ϵ in each expression for a_1. It is noted that case (I) with $p>1$ is obtained from the above result as the limit of $q\to 0$.

(III) The case in (b): This class is characterized by $p=q+1$ $(q=1,\frac{3}{2},2,\cdots)$. There are three sets of j_1 and j_2, i.e., (A) $j_1=(q+1+\delta)/2, j_2=(q+1-\delta)/2, \delta=-q,-q+1,\ldots,q-1,q$; (B) $j_1=q+(n+1)/2, j_2=(n+1)/2, n=1,2,\cdots$; (C) $j_1=(n+1)/2, j_2=q+(n+1)/2, n=1,2,\cdots$. The action of a_1 and a_2 has the form in the three cases (A), (B), and (C) as follows;

$$\begin{aligned}
\text{(A)}\quad & a_1|q+1q{:}j_1m_1j_2m_2\rangle=[2(m_1-j_1)]^{1/2}|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1+\tfrac{1}{2}m_1-\tfrac{1}{2}j_2m_2\rangle,\\
& a_1|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1+\tfrac{1}{2}m_1+\tfrac{1}{2}j_2m_2\rangle=[2(m_1+j_1)]^{1/2}|q+1\,q{:}j_1m_1j_2m_2\rangle,\\
& a_2|q+1\,q{:}j_1m_1j_2m_2\rangle=[2(m_2-j_2)]^{1/2}|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1\ m_1,j_2j_2+\tfrac{1}{2}m_2-\tfrac{1}{2}\rangle,\\
& a_2|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1m_1j_2+\tfrac{1}{2}m_2+\tfrac{1}{2}\rangle=[2(m_2+j_2)]^{1/2}|q+1q{:}j_1m_1j_2m_2\rangle.
\end{aligned}\tag{3.54a}$$

$$\begin{aligned}
\text{(B)}\quad & a_1|q+1\,q{:}j_1m_1j_2m_2\rangle=[2(m_1-j_1)]^{1/2}|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1+\tfrac{1}{2}m_1-\tfrac{1}{2}j_2m_2\rangle,\\
& a_1|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1+\tfrac{1}{2}m_1+\tfrac{1}{2}j_2m_2\rangle=[2(m_1+j_1)]^{1/2}|q+1\,q{:}j_1m_1j_2m_2\rangle,\\
& a_2|q+1\,q{:}j_1m_1j_2m_2\rangle=[2(m_2+j_2-1)]^{1/2}|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1m_1j_2-\tfrac{1}{2}m_2-\tfrac{1}{2}\rangle,\\
& a_2|q+\tfrac{3}{2}q+\tfrac{1}{2}{:}j_1m_1j_2-\tfrac{1}{2}m_2+\tfrac{1}{2}\rangle=[2(m_2-j_2+1)]^{1/2}|q+1\,q{:}j_1m_1j_2m_2\rangle.
\end{aligned}\tag{3.54b}$$

(C). This case is clear from (B) because only the role of j_1 and j_2 changes. The expressions for $a_1^\dagger$ and $a_2^\dagger$ in each case are obvious.

(IV) The case in (c) and (d): The class (c) is characterized by $p=\frac{1}{2}$ and $q=0$. j_1 and j_2 in (c) take the same value $\frac{1}{4}$ or $\frac{3}{4}$. Similarly, (d) is characterized by $p=1$ and $q=\frac{1}{2}$. j_1 and j_2 in (d) take $j_1=\frac{1}{4}j_2=\frac{3}{4}$ or $j_1=\frac{3}{4}j_2=\frac{1}{4}$. Class (c) together with (d) gives the expressions for a_i. The results are

$$\begin{aligned}
& a_1|\tfrac{1}{2}\,0{:}\tfrac{1}{4}\,m_1,\tfrac{1}{4}\,m_2\rangle=[2(m_1-\tfrac{1}{4})]^{1/2}|1\,\tfrac{1}{2}{:}\tfrac{3}{4}\,m_1-\tfrac{1}{2},\tfrac{1}{4}\,m_2\rangle,\quad a_1|\tfrac{1}{2}\,0{:}\tfrac{3}{4}\,m_1,\tfrac{3}{4}\,m_2\rangle=[2(m_1-\tfrac{1}{4})]^{1/2}|1\,\tfrac{1}{2}{:}\tfrac{1}{4}\,m_1-\tfrac{1}{2},\tfrac{3}{4}\,m_2\rangle,\\
& a_1|1\,\tfrac{1}{2}{:}\tfrac{1}{4}\,m_1,\tfrac{3}{4}\,m_2\rangle=[2(m_1-\tfrac{1}{4})]^{1/2}|\tfrac{1}{2}\,0{:}\tfrac{3}{4}\,m_1-\tfrac{1}{2},\tfrac{3}{4}\,m_2\rangle,\quad a_1|1\,\tfrac{1}{2}{:}\tfrac{3}{4}\,m_1,\tfrac{1}{4}\,m_2\rangle=[2(m_1-\tfrac{1}{4})]^{1/2}|\tfrac{1}{2}\,0{:}\tfrac{1}{4}\,m_1-\tfrac{1}{2},\tfrac{1}{4}\,m_2\rangle.
\end{aligned}\tag{3.55}$$

The expressions for other a's are obvious. It is noted that the results in (III) and (IV) can be considered as the limit of those in (II). (IV) will be easily generalized to an arbitrary number of the Bose system.

IV. DISCUSSION

We have constructed the infinite-dimensional representations for the system of the two para-Bose operators which induce the graded Lie algebra written as GLA(Sp(4):4). We have known that in order to describe the representations of GLA(Sp(4):4) three irreducible representations of $\mathrm{Sp}(4)\approx\mathrm{SO}(3,2)$ are needed in case (I), four in case (II), and two in cases (III) and (IV).

The vacuum state in (3.3) is unique and must satisfy the conditions

$$a_1|0\rangle=a_2|0\rangle=0.\tag{4.1}$$

It, therefore, follows that the vacuum state is contained in cases (I) and (IV), because there is a state with $q=0$ and the minimum $j_1=j_2=m_1=m_2$. Indeed, $|0\rangle$ is given by $|0\rangle=|\,p0{:}p/2\,p/2,p/2\,p/2\rangle$ in case (I) and $|0\rangle=|\frac{1}{2}\,0{:}\frac{11}{44},\frac{11}{44}\rangle$ in case (IV), and they evidently satisfy (4.1). However, it is easy to see that case (IV) corresponds to the ordinary Bose quantization. We, therefore, see that case (I) corresponds to the para-Bose quantization.

It follows from (4.1) and (3.50) that the following holds:

$$a_ia_j^\dagger|0\rangle=2p\delta_{ij}|0\rangle.\tag{4.2}$$

Thus order of quantization in our case is $2p$ which takes any real value greater than 1. The action of the commutators $[a_1,a_2]$ etc. on the bases is easily calculated and a few of them in case (I) are given below:

$$\begin{aligned}
&\tfrac{1}{2}[a_1^\dagger,a_2^\dagger]|p0{:}j_1m_1j_2m_2\rangle\\
&=\epsilon\left(\frac{(m_1+j_1)(m_2+j_2)(2j_1+p-1)(2j_1+p-2)}{2j_1(2j_1-1)}\right)^{1/2}\\
&\quad\times|p+1\,0{:}j_1+\tfrac{1}{2}\,m_1+\tfrac{1}{2}j_2+\tfrac{1}{2}\,m_2+\tfrac{1}{2}\rangle\\
&\quad-\epsilon\left(\frac{(m_1-j_1+1)(m_2-j_2+1)(2j_1-p)(2j_1-p-1)}{(2j_1-1)(2j_1-2)}\right)^{1/2}\\
&\quad\times|p+1\,0{:}j_1-\tfrac{1}{2}\,m_1+\tfrac{1}{2}j_2-\tfrac{1}{2}\,m_2+\tfrac{1}{2}\rangle,
\end{aligned}\tag{4.3}$$

$$\begin{aligned}
&\tfrac{1}{2}[a_1,a_1^\dagger]|p0{:}j_1m_1j_2m_2\rangle=p|p0{:}j_1m_1j_2m_2\rangle\\
&\quad-\epsilon[(2j_1-p)(2j_1+p-2)]^{1/2}|p+1\,0{:}j_1m_1j_2m_2\rangle.
\end{aligned}$$

It follows that the commutator $[a_1^\dagger,a_2^\dagger]$ has only the matrix elements between different irreducible representations of Sp(4). It is noted that the anticommutator $\{a_1^\dagger,a_2^\dagger\}$ $(=2J_+^{(3)})$ has only the matrix elements in the same irreducible representation of Sp(4). The second of (4.3) gives $2p(2p-1)$ as the square of the magnitude of $[a_1^\dagger,a_2^\dagger]|0\rangle/2$, which agrees with the known result.[3]

The bases of Alabiso *et al.*[2] are given by

$$\begin{aligned}
&|nlm\rangle=N_{nlm}(J_+^{(4)})^m(a_2^\dagger)^l(J^{(+)})^n|0\rangle,\\
&J^{(+)}\equiv\tfrac{1}{2}[a_1^\dagger,a_2^\dagger],
\end{aligned}\tag{4.4}$$

where n,l,m are nonnegative integers and N_{nlm} is a normalization constant. Their bases are given for the integer order of the quantization, i.e., $2p=\text{integer}\gtrsim 2$. It is easy to express the right-hand side of (4.4) in terms of our bases, because the

action of $J^{(+)}$, $a_2^\dagger$, and $J_+^{(4)}$ on our bases is known. We give the result for a special case, for the expressions in the general case are too lengthy:

For l = even,

$$|0l\,0\rangle = N_{0l0}\times 2^{l/2}\left(\frac{\Gamma(p+l/2)\Gamma(1+l/2)}{\Gamma(p)}\right)^{1/2} \times|\,p0{:}p/2\;p/2{,}p/2\;(p+l)/2\rangle. \tag{4.5}$$

For l = odd,

$$|0l\,0\rangle = \epsilon N_{0l0}\times 2^{l/2}\left(\frac{\Gamma[p+(l+1)/2]\Gamma[(l+1)/2]}{\Gamma(p)}\right)^{1/2} \times\left|p+\frac{1}{2}\,\frac{1}{2}{:}\frac{p}{2}\,\frac{p}{2}{,}(p+1)/2\;(p+l)/2\right\rangle, \tag{4.6}$$

$$N_{0l0}^2 = \begin{cases} \dfrac{(2p-2)!}{l!!(2p+l-2)!!(2p-3)!!} & \text{for } l = \text{even}, \\ \dfrac{(2p-2)!}{(l-1)!!(2p+l-1)!!(2p-3)!!} & \text{for } l = \text{odd}. \end{cases}$$

The Hamiltonian of our system is given by $H = 2\omega(J_3^{(1)} + J_3^{(2)})$. Thus on making use of $m_i = j_i + n_i$, with non-negative integer n_i, the eigenvalues of H in the case (I) are given by $2(p+n+n_1+n_2)$ for the base $|p0{:}j_1m_1j_2m_2\rangle$, $2(p+n+n_1+n_2+1)$ for the base $|\,p+10{:}j_1m_1,j_2m_2\rangle$ and $2(p+n+n_1+n_2)+1$ for the bases $|p+\frac{1}{2}\,\frac{1}{2}{:}j_1+\frac{1}{2}m_1+\frac{1}{2}j_2m_2\rangle$ and $|p+\frac{1}{2}\,\frac{1}{2}{:}j_1m_1j_2+\frac{1}{2}\,m_2+\frac{1}{2}\rangle$, where n is nonnegative integer. Thus the degeneracy[14] of the eigenvalue of H is given by $(N/2+1)^2$ for even N and $(N+1)/2\times[(N+1)/2+1]$ for odd N, respectively.

[1] V. G. Kac, Commun. Math. Phys. **53**, 31 (1977); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146,155 (1977); M. Bednař and V. F. Šachl, *ibid.* **20**, 367 (1979).
[2] C. Alabiso, F. Duimio, and J. L. Redondo, Nuovo Cimento A **61**, 266 (1969); C. Alabiso and F. Duimio, *ibid.* **7**, 63 (1973); **10**, 499 (1972); T. D. Palev, J. Math. Phys. **21**, 1293 (1980).
[3] H. S. Green, Phys. Rev. **90**, 270 (1953).
[4] O. W. Greenberg and A. M. L. Messiach, Phys. Rev. **136**, B 248 (1964); Y. Ohnuki and S. Kamefuchi, Soryushiron Kenkyu (Kyoto) **53**, 130 (1976).
[5] C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[6] S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1962).
[7] A. J. Bracken and H. S. Green, Nuovo Cimento A **9**, 349 (1972); T. Maekawa and N. Noguchi, J. Math. Phys. **21**, 1659 (1980).
[8] M. K. F. Wong, J. Math. Phys. **21**, 630 (1980).
[9] E. P. Wigner, Phys. Rev. **77**, 711 (1950); T. F. Jordon, N. Mukunda, and S. V. Pepper, J. Math. Phys. **4**, 1089 (1963).
[10] N. T. Evans. J. Math. Phys. **8**, 170 (1967).
[11] T. Maekawa, Math. Jpn. **18**, 129 (1973).
[12] V. Bargmann, Ann. Math. **48**, 586 (1947).
[13] G. Racah, *Group Theory and Spectroscopy, Lecture Notes* (Princeton, 1951).
[14] I. Inaba, T. Maekawa, and T. Yamamoto, Lett. Nuovo Cimento (to be published).

A family of sums of products of Legendre functions

H. van Haeringen
Department of Mathematics, Delft University of Technology, Delft, The Netherlands

(Received 15 December 1981; accepted for publication 5 February 1982)

We present simple analytic expressions for a few sums of products of Legendre functions, of the type $\Sigma_{n=0}^{\infty}(2n+1)\mathfrak{P}_n^{\alpha}(x)\mathfrak{P}_n^{\beta}(y)\mathfrak{P}_n^{\gamma}(z)\mathfrak{Q}_n^{\mu}(u)$.

PACS numbers: 02.30. + g, 02.90. + p, 03.65. − w, 03.65.Nk

In Ref. 1 we derived a simple closed expression for the sum $\Sigma_{n=0}^{\infty}(2n+1)\mathfrak{P}_n(x)\mathfrak{Q}_n^{-\mu}(y)\mathfrak{Q}_n^{\mu}(z)$. This expression followed from the unitarity relation for the Coulomb T matrix. In this paper we shall evaluate some related series.

We define the family of sums $S_{3\,1}$ by

$$S_{3\,1}\left(\begin{matrix}\alpha,\beta,\gamma;\mu\\ x,y,z;u\end{matrix}\right) := \sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n^{\alpha}(x)\mathfrak{P}_n^{\beta}(y)\mathfrak{P}_n^{\gamma}(z)\mathfrak{Q}_n^{\mu}(u), \tag{1}$$

where $\mathfrak{P}_n^{\alpha}$ and $\mathfrak{Q}_n^{\mu}$ are Legendre functions of the first and second kind, respectively.[2] The variables $x, y, z, u, \alpha, \beta, \gamma$, and μ are in general complex. The infinite series in Eq. (1) is convergent when

$$|x+(x^2-1)^{1/2}|\cdot|y+(y^2-1)^{1/2}|\cdot|z+(z^2-1)^{1/2}| < |u+(u^2-1)^{1/2}|. \tag{2}$$

We shall assume throughout this paper that this condition is satisfied. To avoid ambiguities in the definition of complex powers we shall take the real part of x, y, z, and u positive. E.g., in Ref. 2, p. 123, the definition

$$(z^2-1)^a := (z-1)^a(z+1)^a, |\arg(z\pm 1)|<\pi, |\arg z|<\pi,$$

is used which means that $(z^2-1)^a$ is different from $\exp(a\ln(z^2-1))$ when $\operatorname{Re} z<0$.

It is interesting to note that the equation

$$|z+(z^2-1)^{1/2}| = R, \quad R>1,$$

represents in the complex z plane an ellipse with foci at $+1$ and -1, with major axis equal to $R+R^{-1}$ and with minor axis equal to $R-R^{-1}$.

In this paper we present the simple analytic expressions given by the right-hand members of Eqs. (3)–(7). In order to avoid problems related to branch cuts we take, in these equations, the real part of z [and of u in Eq. (6)] sufficiently large and positive.

In the first place we have obtained

$$\sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n(x)\mathfrak{P}_n^{-\mu}(y)\mathfrak{Q}_n^{\mu}(z) = e^{i\pi\mu}W^{-1/2}(yz-x-W^{1/2})^{\mu/2}(yz-x+W^{1/2})^{-\mu/2} \tag{3a}$$

$$= e^{i\pi\mu}W^{-1/2}(y^2-1)^{\mu/2}(z^2-1)^{\mu/2}(yz-x+W^{1/2})^{-\mu}, \tag{3b}$$

where

$$W := W(x,y,z) := x^2+y^2+z^2-1-2xyz.$$

It is interesting to note that the closed formula for the sum $\Sigma(2n+1)\mathfrak{P}_n(x)\mathfrak{Q}_n^{-\mu}(y)\mathfrak{Q}_n^{\mu}(z)$ obtained in Ref. 1 can be derived from Eq. (3), and vice versa. We have obtained a new, independent proof of Eq. (3), which is given in Ref. 3. Furthermore, we have found

$$\sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n(x)\mathfrak{P}_n(y)\mathfrak{Q}_n^{\mu}(z) = e^{i\pi\mu}\Gamma(1+\mu)(z^2-1)^{\mu/2}W^{-1/2-\mu/2}\mathfrak{P}_{\mu}((z-xy)W^{-1/2}) \tag{4a}$$

$$= e^{i\pi\mu}\Gamma(1+\mu)(z^2-1)^{\mu/2}W^{-1/2-\mu/2}\,{}_2F_1 \times(-\tfrac{1}{2}\mu,\tfrac{1}{2}+\tfrac{1}{2}\mu;1;(1-x^2)(y^2-1)W^{-1}) \tag{4b}$$

$$= e^{i\pi\mu}\Gamma(1+\mu)(z^2-1)^{\mu/2}(z-xy)^{-\mu-1}\,{}_2F_1 \times(1+\tfrac{1}{2}\mu,\tfrac{1}{2}+\tfrac{1}{2}\mu;1;(x^2-1)(y^2-1) \times(z-xy)^{-2}), \tag{4c}$$

and

$$\sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n^{\mu}(x)\mathfrak{P}_n^{-\mu}(y)\mathfrak{Q}_n^{\mu}(z) = e^{i\pi\mu}W^{-1/2-\mu/2}[\tfrac{1}{2}(x+1)(y-1)(z+1)]^{\mu/2} \times\mathfrak{P}_{\mu}^{\mu}((x+z-y-1)W^{-1/2}) \tag{5a}$$

$$= \frac{e^{i\pi\mu}}{\Gamma(1-\mu)}\left(\frac{x+1}{x-1}\frac{y-1}{y+1}\frac{z+1}{z-1}\right)^{\mu/2} \times W^{-1/2}\,{}_2F_1(1,-2\mu;1-\mu;-\tfrac{1}{2}AW^{-1/2}) \tag{5b}$$

$$= \frac{e^{i\pi\mu}}{\Gamma(1-\mu)}\left(\frac{x+1}{x-1}\frac{y-1}{y+1}\frac{z+1}{z-1}\right)^{\mu/2}W^{-1/2}\,{}_2F_1 \times(\tfrac{1}{2},-\mu;1-\mu;-2W^{-1}(x-1)(y+1)(z-1)), \tag{5c}$$

with W as before and $A := x+z-y-1-W^{1/2}$.

We have evaluated one member of the family defined by Eq. (1) containing products of *four* Legendre functions,

$$\sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n(x)\mathfrak{P}_n(y)\mathfrak{P}_n(z)\mathfrak{Q}_n(u) = (W^2-T)^{-1/4}\mathfrak{P}_{-1/2}(W(W^2-T)^{-1/2}) \tag{6a}$$

$$= 2^{1/2}\pi^{-1}T^{-1/4}\mathfrak{Q}_{-1/2}(WT^{-1/2}) \tag{6b}$$

$$= W^{-1/2}\,{}_2F_1(\tfrac{1}{4},\tfrac{3}{4};1;TW^{-2}), \tag{6c}$$

where

$$W := W(x,y,z,u) := x^2+y^2+z^2+u^2-2-2xyzu,$$

$$T := T(x,y,z,u) := 4(1-x^2)(1-y^2)(1-z^2)(1-u^2).$$

There exist simple relations between $\mathfrak{P}_n^{-\alpha}$ and the Jacobi polynomial $P_n^{(\alpha,-\alpha)}$, and similarly between $\mathfrak{Q}_n^{\mu}$ and the so-called Jacobi function of the second kind, $Q_n^{(\mu,-\mu)}$, see, e.g., Szegö.[4]

We have obtained the following interesting relation:

 0022-2488/82/060964-03$02.50

$$\sum_{n=0}^{\infty}(2n+1)\frac{\Gamma^2(n+1)}{\Gamma(n+\alpha+1)\Gamma(n-\mu+1)}$$

$$\times P_n^{(\alpha,-\alpha)}(y)Q_n^{(-\mu,\mu)}(z)$$

$$=\frac{1}{z-1}\frac{\Gamma(\mu+1)}{\Gamma(\alpha+1)}\,{}_2F_1\left(1,1+\mu;1+\alpha;\frac{y-1}{z-1}\right). \quad (7)$$

This result suggests that interesting generalizations may exist for the more general case of Jacobi polynomials $P_n^{(\alpha,\beta)}$ and Jacobi functions $Q_n^{(\mu,\nu)}$.

In the special case $\mu=\alpha$, Eq. (7) reduces to the simple relation

$$\sum_{n=0}^{\infty}(2n+1)\frac{\Gamma^2(n+1)}{\Gamma(n+\alpha+1)\Gamma(n-\alpha+1)}$$

$$\times P_n^{(\alpha,-\alpha)}(y)Q_n^{(-\alpha,\alpha)}(z)=(z-y)^{-1}, \quad (8)$$

which is well known.[4,5]

For the proof of Eqs. (3), (5), and (7) we refer to Ref. 3, where many other interesting sums of products of Legendre functions are also given. The remaining part of this paper consists of the proof of Eqs. (4) and (6).

For the proof of Eq. (4) we start with Heine's formula[2]

$$\sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n(x)\mathfrak{Q}_n(z)=(z-x)^{-1}. \quad (9)$$

For convenience we take here $z>1$, and x and y between 0 and 1. Later these conditions can be relaxed by analytic continuation. We introduce the notation

$$\bar{x}:=(1-x^2)^{1/2},\quad \bar{y}:=(1-y^2)^{1/2}.$$

By using

$$\pi^{-1}\int_0^{\pi}\mathfrak{P}_n(xy+\bar{x}\bar{y}\cos\varphi)\,d\varphi=\mathfrak{P}_n(x)\mathfrak{P}_n(y) \quad (10)$$

and

$$\pi^{-1}\int_0^{\pi}(t-\cos\varphi)^{-1}\,d\varphi=(t^2-1)^{-1/2},\mathrm{Re}\,t>0,$$

we obtain from Eq. (9),

$$\sum_{n=0}^{\infty}(2n+1)\,\mathfrak{P}_n(x)\,\mathfrak{P}_n(y)\mathfrak{Q}_n(z)$$

$$=\pi^{-1}\int_0^{\pi}(z-xy-\bar{x}\bar{y}\cos\varphi)^{-1}\,d\varphi$$

$$=W^{-1/2}=((z-xy)^2-(\bar{x}\bar{y})^2)^{-1/2}. \quad (11)$$

Note that Eq. (11) is just Eq. (4) with $\mu=0$. By using, furthermore,[2]

$$\Gamma(-\mu)\mathfrak{Q}_n^{\mu}(z)$$

$$=e^{i\pi\mu}(z^2-1)^{\mu/2}\int_z^{\infty}\mathfrak{Q}_n(t)(t-z)^{-\mu-1}\,dt,\quad -1<\mathrm{Re}\,\mu<0,$$

we obtain from Eq. (11)

$$\Gamma(-\mu)\sum_{n=0}^{\infty}(2n+1)\mathfrak{P}_n(x)\mathfrak{P}_n(y)\mathfrak{Q}_n^{\mu}(z)$$

$$=e^{i\pi\mu}(z^2-1)^{\mu/2}\int_z^{\infty}(t-z)^{-\mu-1}$$

$$\times(x^2+y^2+t^2-1-2xyt)^{-1/2}\,dt$$

$$=e^{i\pi\mu}(z^2-1)^{\mu/2}W^{-1/2-\mu/2}$$

$$\times\int_0^{\infty}\tau^{-\mu-1}(\tau^2+2\tau W^{-1/2}(z-xy)+1)^{-1/2}\,d\tau$$

$$=e^{i\pi\mu}(z^2-1)^{\mu/2}W^{-1/2-\mu/2}$$

$$\times B(1+\mu,-\mu)\mathfrak{P}_{\mu}(W^{-1/2}(z-xy)),$$

which completes the proof of Eq. (4). Here we have used (cf. Ref. 6, Formula 3.252.11),

$$\int_0^{\infty}\tau^{-\mu-1}(\tau^2+2\beta\tau+1)^{-1/2}\,d\tau$$

$$=B(1+\mu,-\mu)\mathfrak{P}_{\mu}(\beta),\quad -1<\mathrm{Re}\mu<0. \quad (12)$$

For the proof of Eq. (6) we combine Eqs. (10)–(12) and get

$$\sum_{n=0}^{\infty}(2n+1)\,\mathfrak{P}_n(x)\,\mathfrak{P}_n(y)\,\mathfrak{P}_n(z)\mathfrak{Q}_n(u)$$

$$=\pi^{-1}\int_0^{\pi}[(xy+\bar{x}\bar{y}\cos\varphi)^2$$

$$+z^2+u^2-1-2uz(xy+\bar{x}\bar{y}\cos\varphi)]^{-1/2}\,d\varphi$$

$$=\pi^{-1}\int_0^{\pi}(a+2b\cos\varphi+c\cos^2\varphi)^{-1/2}\,d\varphi$$

$$=\pi^{-1}(a+c+2b)^{-1/4}(a+c-2b)^{-1/4}$$

$$\times\int_0^{\infty}\tau^{-1/2}(\tau^2+2\beta\tau+1)^{-1/2}\,d\tau$$

$$=[(a+c)^2-4b^2]^{-1/4}\mathfrak{P}_{-1/2}(\beta).$$

Here

$$a:=x^2y^2+z^2+u^2-1-2xyzu,$$

$$b:=\bar{x}\bar{y}(xy-zu),$$

$$c:=\bar{x}^2\bar{y}^2,$$

$$\beta:=(a-c)[(a+c)^2-4b^2]^{-1/2},$$

which implies $a-c=W$, $(a+c)^2-4b^2$ $=(a-c)^2-4(b^2-ac)=W^2-T$, and we have used the substitution

$$\tau:=(a+c-2b)^{1/2}(a+c+2b)^{-1/2}\tan^2\varphi/2.$$

Using, finally, the well-known relations[2,7]

$$\mathfrak{P}_{-1/2}(z)=z^{-1/2}\,{}_2F_1(\tfrac{1}{4},\tfrac{3}{4};1;1-z^{-2}),\quad \mathrm{Re}\,z>0,$$

$$=2^{1/2}\pi^{-1}(z^2-1)^{-1/4}\mathfrak{Q}_{-1/2}(z(z^2-1)^{-1/2}),$$

$$\mathrm{Re}\,z>0,$$

the proof of Eq. (6) is completed.

ACKNOWLEDGMENT

I am grateful to R. Askey for informing me of the fact that M. Rahman [SIAM J. Math. Anal. 8, 1014 (1977)] has evaluated related sums of products of Jacobi functions and polynomials.

[1]H. van Haeringen and L. P. Kok, J. Math. Phys. **22**, 2482 (1981).
[2]Bateman Manuscript Project, A. Erdélyi, Editor, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I. Unfortunately, there are misprints and errors in the five volumes of Bateman's Manuscript Project: (i) *Higher Transcendental Functions* (McGraw-Hill, New York), Vols. I and II (1953), and Vol. III (1955); (ii) *Table of Integral Transforms* (McGraw-Hill, New York), Vols. I and II (1954). About a hundred of these, excluding those reported in the later printings of these volumes, have been collected in H. van Haeringen and L. P. Kok, Corrigenda Report 82 04 (Delft-Groningen, 1982).
[3]H. van Haeringen, "Sums of Products of Legendre Functions," preprint (1981).
[4]G. Szegö, *Orthogonal Polynomials, 4th ed.* (American Mathematical Society, Providence, R.I., 1975), Eq. (9.2.1).
[5]E. R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, N. J. 1975), Eq. (45.6.1).
[6]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products, Corrected and Enlarged Edition* (Academic, New York, 1980). Unfortunately, this work contains misprints and errors. A few hundred of these have been collected in H. van Haeringen and L. P. Kok, Corrigenda Report 82 01 (Delft-Groningen, 1982).
[7]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966). Unfortunately, this work contains misprints and errors. More than a hundred of these have been collected in H. van Haeringen and L. P. Kok, Corrigenda Report 82 03 (Delft-Groningen, 1982).

Variational calculation of the multipole potential from an arbitrary localized charge distribution

P. Csavinszky
Department of Physics, University of Maine, Orono, Maine 04469

(Received 22 October 1981; accepted for publication 26 February 1982)

Very recently Painter has developed a method for solving Poisson's equation as a set of finite-difference equations for an arbitrary localized charge distribution $\rho(\mathbf{r})$ that is expanded in a partial-wave representation as $\rho(\mathbf{r}) = \Sigma_L\, \rho_L(r)Y_L(\hat{r})$, where L denotes l and m. In the present work a variational principle is established, and a possible approach is outlined, for obtaining approximate partial-wave coefficients $V_L(r)$ of the potential $V(\mathbf{r}) = \Sigma_L\, V_L(r)Y_L(\hat{r})$.

PACS numbers: 02.30. + g, 41.10. − j

I. INTRODUCTION

Painter's method[1] for calculating the potential of an arbitrary localized charge distribution is a generalization of Loucks' finite-difference method[2] that is applicable to spherically-symmetric charge distributions. The basis of Painter's method is the representation of the nonspherically-symmetric localized charge density by a rapidly convergent partial-wave expansion

$$\rho(\mathbf{r}) = \sum_L \rho_L(r)Y_L(\hat{r}), \tag{1}$$

where L denotes l and m. In this case, the potential can also be represented as the partial-wave expansion

$$V(\mathbf{r}) = \sum_L V_L(r)Y_L(\hat{r})\,. \tag{2}$$

With $\rho(\mathbf{r})$ expressed by Eq. (1), and $V(\mathbf{r})$ by Eq. (2), Poisson's equation (in atomic units),

$$\nabla^2 V(\mathbf{r}) = -8\pi\rho(\mathbf{r})\,, \tag{3}$$

leads to the differential equation

$$\frac{1}{r^2}\,d\left(r^2\frac{dV_L(r)}{dr}\right)/dr - \frac{l(l+1)}{r^2}\,V_L(r) = -8\pi\rho_L(r) \tag{4}$$

for each partial-wave component V_L belonging to the partial-wave component ρ_L.

With the transformation

$$x = \ln r\,, \tag{5}$$

Equation (4) is transformed to

$$\frac{d^2V_L}{dx^2} + \frac{dV_L}{dx} = -8\pi e^{2x}\rho_L + l(l+1)V_L\,. \tag{6}$$

The first-derivative term in Eq. (6) can be eliminated by the transformation

$$W_L(x) = e^{x/2}V_L(x)\,, \tag{7}$$

leading to the second-order inhomogeneous differential equation

$$\frac{d^2W_L}{dx^2} = (l+\tfrac{1}{2})^2 W_L - 8\pi e^{5x/2}\rho_L\,. \tag{8}$$

This is the differential equation that should be solved for each partial-wave component W_L. The potential of the charge distribution is then specified by Eq. (2), upon consideration of Eqs. (5) and (7).

II. THE VARIATIONAL APPROACH

In order to solve Eq. (8) by an equivalent variational principle, one considers the functional

$$L\{W_L\} = \int_{-\infty}^{+\infty} F(W_L, W_L', x)dx\,, \tag{9}$$

where F has to be so chosen that upon its substitution into the Euler–Lagrange equation,[3]

$$\frac{\partial}{\partial W_L}F - \frac{d}{dx}\frac{\partial}{\partial W_L'}F = 0\,, \tag{10}$$

Eq. (8) is recovered. It is easy to show that the expression

$$F = -\tfrac{1}{2}(W_L')^2 - \tfrac{1}{2}(l+\tfrac{1}{2})^2 W_L^2 + 8\pi e^{5x/2}\rho_L W_L \tag{11}$$

satisfies the above requirement.

What remains now is a specification of the boundary conditions at $r\to 0$ (corresponding to $x\to -\infty$), and at $r\to +\infty$ (corresponding to $x\to +\infty$). (In order to keep a possible model in mind, one can assume that an originally spherically-symmetric atom is placed in a crystal where its electron cloud undergoes a distortion).

The boundary condition at $r\to +\infty$ is l-dependent. The potential of the nonspherically-symmetric charge distribution must approach the multipole expansion[4] as r increases. The multipole potential is

$$V_m(\mathbf{r}) = 8\pi\sum_L \frac{q_L}{2l+1}\frac{Y_L(\hat{r})}{r^{l+1}}\,, \tag{12}$$

where the multipole moment, corresponding to the L th partial-wave component of the charge distribution [Eq. (1)] is[1]

$$\rho_L = \int_0^r (r')^{l+2}\rho_L(r')dr'\,. \tag{13}$$

In the limit of $r\to +\infty$, only the monopole term ($l = 0, m = 0$) of Eq. (12) is important. In this limit

$$V_M(r\to +\infty) = \frac{2(4\pi)^{1/2}q_{00}}{r}\,, \tag{14}$$

where[4]

$$q_{00} = \frac{1}{(4\pi)^{1/2}}\,q\,, \tag{15}$$

 0022-2488/82/060967-02$02.50

with q denoting the total charge of the distribution. For a neutral atom $q = Z$, where Z is the atomic number.

The boundary condition at $r \to +\infty$ can serve as a check on the "goodness" of the approximate W_L's: the variational W_L's are good, at least at large distances from the atom, if $V(\mathrm{r})$ agrees well with $V_M(\mathrm{r})$.

To establish the boundary condition as $r \to 0$, one considers that, for small values of r, Eq. (8) becomes

$$\frac{d^2 W_L}{dx^2} - (l + \tfrac{1}{2})^2 W_L = 0 . \tag{16}$$

The solution of Eq. (16) is

$$W_L = A e^{(l+1/2)x} , \tag{17}$$

where A is a constant. In view of this result, the boundary condition at $r \to 0$ can serve as a guide in selecting a variational W_L.

The problem that remains now is the determination of the constant A. This problem did not arise in Painter's theory,[1] since in the finite-difference method it was only necessary to relate values of W_L at adjacent mesh points.

To determine A, one considers that the charge distribution is finite. This means that $\rho \to 0$ as $r \to +\infty$. The vanishing of the charge density ρ is tantamount to demanding that $\rho_L \to 0$ as $r \to +\infty$. In this case, the term $8\pi e^{5x/2}\rho_L$ in Eq. (8) goes to zero as r goes to infinity and one again has Eq. (16), whose solution is Eq. (17). Combining Eq. (7) with Eq. (14), one finds that

$$A = 2q_{00} . \tag{18}$$

APPENDIX: OUTLINE OF A VARIATIONAL PROCEDURE

One may assume that, for the interval $(-\infty, +\infty)$, W_L can be written as

$$W_L(x) = C W_L(x \to -\infty) P + D W_L(x \to +\infty) Q , \tag{A1}$$

where, from Eq. (17),

$$W_L(x \to -\infty) = A e^{(l+1/2)x} \tag{A2}$$

and, from Eqs. (7) and (14),

$$W_L(x \to +\infty) = B e^{-x/2} \tag{A3}$$

with $A = 2q/(4\pi)^{1/2}$, and $B = 2q$, as follows from Eqs. (15), (18), and (14), (15), respectively.

In Eq. (A1), C and D are constants, and the quantities P and Q must be such that

$$\left.\begin{matrix} Q(x) \to 0 \\ P(x) \to 1 \end{matrix}\right\} \text{ as } x \to -\infty , \tag{A4}$$

and

$$\left.\begin{matrix} Q(x) \to 1 \\ P(x) \to 0 \end{matrix}\right\} \text{ as } x \to +\infty . \tag{A5}$$

The problem is to find a $Q(x)$ and a $P(x)$ with such behavior.

It is easily seen that the functions

$$P(x) = b/[b + e^{(l+1)x}] , \tag{A6}$$

and

$$Q(x) = a/[a + e^{-(l+1)x}] , \tag{A7}$$

where the quantities a and b are considered as positive variational parameters, satisfy Eqs. (A4) and (A5).

With Eqs. (A6) and (A7), it follows from Eq. (A1) that

$$W_L(x \to +\infty) = (CAb + DB) e^{-x/2} , \tag{A8}$$

which is the x-behavior required by Eq. (A3). In obtaining Eq. (A8), the constant b has been dropped in the denominator of the first term in Eq. (A1), and the term $e^{-(l+1)x}$ has been dropped in the second term of Eq. (A1).

Again, with Eqs. (A6) and (A7), it follows from Eq. (A1) that

$$W_L(x \to -\infty) = (CA + DBa) e^{(l+1/2)x} , \tag{A9}$$

which is the x-behavior required by Eq. (A2). In obtaining Eq. (A9), the term $e^{(l+1)x}$ has been dropped in the denominator of the first term in Eq. (A1), and the constant a has been dropped in the denominator of the second term in Eq. (A1).

What remains now is the determination of the constants C and D. It follows from Eqs. (A3) and (A8) that

$$CAb + DB = B , \tag{A10}$$

while Eqs. (A2) and (A9) require that

$$CA + DBa = A . \tag{A11}$$

Solving Eqs. (A10) and (A11) for C and D, one obtains

$$C = \frac{Ba - A}{Aab - A} ; \qquad D = \frac{Ab - B}{Bab - B} . \tag{A12}$$

Using Eqs. (A2), (A3), (A6), and (A7), one finds that Eq. (A1) becomes

$$W_L(x) = \frac{CAbe^{(l+1/2)x}}{b + e^{(l+1)x}} + \frac{DBae^{-x/2}}{a + e^{-(l+1)x}} . \tag{A13}$$

The variational procedure can now be outlined by adopting a model density. One can, for instance, choose the same model density as that of Painter,[1] namely

$$\rho_L(r) = K_l r^l e^{-\alpha_l r} , \tag{A14}$$

where $K_l = \alpha_l^{l+3}/[\sqrt{4\pi}(l+2)!]$ and the constants α_l are assumed to be known. Using Eq. (5), one can express Eq. (A14) as

$$\rho_L(x) = K_l e^{lx} e^{-\alpha_l e^x} . \tag{A15}$$

With Eqs. (A13) and (A15), the quantity F in Eq. (11) is specified. A little reflection shows that the complicated term arising from Eq. (A15) is well-behaved both at $x = +\infty$ and at $x = -\infty$. Upon choosing values for the variational parameters a and b, the integrations required by Eq. (9) can be carried out by numerical techniques. This concludes the outline of an approach to the variational procedure.

[1] G. S. Painter, Phys. Rev. B **23**, 1624 (1981).
[2] T. L. Loucks, *Augmented Plane Wave Method* (Benjamin, New York, 1967), Appendix 1.
[3] R. Courant, *Differential and Integral Calculus* (Interscience, New York, 1968), Vol. II, p. 491.
[4] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962), p. 99.

A nonstandard infinite dimensional vector space approach to Gaussian functional measures

Louis M. Pecora
Naval Research Laboratory, Washington, D. C. 20375

(Received 21 December 1981; accepted for publication 5 February 1982)

Nonstandard analysis is used to apply the usual concepts of finite dimensional vector spaces in order to define various Gaussian functional measures without using a limiting process. The approach is to define matrices on the infinite dimensional space and use straightforward techniques to determine the properties of the functions the measures are concentrated on. The power of nonstandard analysis allows one to work directly with infinite and infinitesimal quantities and "visualize" certain sets upon which the basic Gaussian form is or is not infinitesimal. The relation of the choice of Gaussian to the function (process) properties remains heuristic since a proof of the Holder continuity of the Weiner paths is not complete, but remains at a local (infinitesimal) level. However, given the Holder continuity the nonbounded variation property easily follows. Higher derivative Gaussian measures are also easily developed and their analytic properties displayed along with their covariances. By using Fourier analysis on finite abelian groups transferred to the nonstandard universe and applied to hyperfinite abelian groups a rigorous transformation from the discrete form of the Weiner measure to a Fourier series form is accomplished. It is shown here that the functions (processes) are infinitesimally close to those of the discrete version of the measure. The Fourier approach is also extended to more general measures. Finally, some speculations show directions in which this direct approach to Gaussian functional measures can be extended and generalized.

PACS numbers: 02.30.Cj

I. INTRODUCTION

The usual physicist's approach to functional integration[1-6] over some Gaussian measure (for example, an exponentiated action integral) is to select some equal-spaced time slicing and write the functional as though it were a functional on an n-dimensional vector space $\mathbb{R}^n$. Then the attempt would be to integrate over each "vector component" and after doing the integration take the limit as $n \to \infty$ and hope it exists. Very often the correct answer is obtained. Because nonstandard analysis is well suited for rigorously handling infinite dimensional spaces, it seems a natural choice to see how far the simple vector space approach can be taken with regard to certain Gaussian measures.

The attempt here is to derive many of the usual path properties of the Weiner measure by using infinite dimensional spaces and to use these results to define other Gaussian measures which are shown to be related to the Weiner paths. The attempt falls short because the proof that the Weiner paths are Holder continuous of order $< \frac{1}{2}$ has only been accomplished at a "local" level. Hence on the whole the development must remain heuristic. However, given Holder continuity the nonbounded variation of the paths can be shown by examining the region of ${}^*\mathbb{R}^{\Omega}$ (where Ω is an infinite number) on which the measure is concentrated. Other geometric notions arise here and because of the power of nonstandard analysis a clearer vision of the paths as vector quantities emerges.

On the whole the word measure is used loosely and simply refers to the Gaussian form on ${}^*\mathbb{R}^{\Omega}$. The question of whether or not this is a true measure has been deferred. Questions about the measure algebras of sets have been neglected. The approach here is a direct one and somewhat along naive lines but, thanks to nonstandard analysis, the results are more rigorous than would be expected and more geometric in meaning. A short attempt is made in the section on integration in ${}^*\mathbb{R}^{\Omega}$ to show how a true measure over an algebra of cylinder sets can be related to some Gaussian forms. In particular, the connection to the Weiner measure is shown. The emphasis throughout the paper is on the vector space operations and quantities, their relationship to standard entities, and the properties selected by various Gaussian forms.

Physicists also use Fourier representations for functions and integrate over the coefficients to do functional integration.[3,5-9] The direct connection between the time-slicing vector-space representation of the Weiner function integral and its Fourier representation is shown here. It is shown that the transformation to Fourier space (also an ${}^*\mathbb{R}^{\Omega}$ space) yields functions which are infinitesimally close to the Weiner functions. The Fourier representation also allows the usual generalization to fractional derivatives.

Some concluding remarks also suggest ways to extend this explicit representation of Gaussian measures to more general cases.

Overall, the power of nonstandard analysis will become apparent in the ability to treat infinite and infintesimal quantities directly as standard numbers are treated in classical analysis. This often gives more meaning to a result than simply saying, for example, that the limit of some variable is infinite. In fact, having "concrete" values for certain infinite or infinitesimal quantities allows the "visualization" of certain geometric aspects of the problem.

 0022-2488/82/060969-14$02.50

II. NONSTANDARD ANALYSIS

A. Theory

In recent times Robinson[10] put the Liebnitzian concepts of infinite and infinitesimal numbers on a sound mathematical foundation while retaining their intuitive aspects. The general advantage of this scheme, called nonstandard analysis, is that the usual roundabout "$\delta\epsilon$" methods can be replaced by the direct use of infinitely large and small quantities. This does not mean the $\delta\epsilon$ techniques have to be altogether abandoned but, instead, that they become a smaller part of the general scheme of mathematical analysis. Below, a simplified heuristic introduction to the formulation of nonstandard analysis is given which is a combination of the approaches in Refs. 10–16. This is not meant to be a complete introduction to the subject, but only a guide for this article.

Classical analysis can be thought of as a system of statements about various n-ary relations based on an infinite set $\mathbb{R}$ of individuals. For example, the sum c of two numbers a and b can be denoted by $S(c,a,b)$, a ternary relation, where $c = a + b$. Then the statement "for any $a,b \in \mathbb{R}$ there exists a $c \in \mathbb{R}$ such that $S(c,a,b)$" is true. Similarly, there are relations for "$>$" and "$\cdot$" so that $a > b$ is written as a binary relation $Q(a,b)$ and the product of two numbers, $c = a \cdot b$, is written as the ternary relation $P(c,a,b)$. There also exist relations between sets and individuals, i.e., members of $\mathbb{R}$, between sets and sets, etc. In general, one can think of the relations as subsets of the Cartesian products $\mathbb{R}\times\mathbb{R}\times\cdots\times\mathbb{R}$ (n times for an n-ary relation), subsets of the Cartesian products of these subsets, etc., producing an entire heirarchy of Cartesian products and their subsets. Sets can be thought of as singulary relations. These relations and statements about them [as in the statement above about $S(c,a,b)$] constitute the *standard* system of classical analysis and will be denoted by U. In other words, this is just the set of relations, lemmas, definitions, and theorems usually called classical mathematics.

An abstract language L can be established in which all the relations of U can be written as abstract relations Φ in L. For example, $S(c,a,b)$ in U becomes $\Phi(s,c',a',b')$ in L, where s stands for sum, i.e., it names the relation. The individuals of L, the set A, are denoted by c',a',b' here, and are abstract entities which merely name the individuals ($\mathbb{R}$) of U. All the statements in U can be written in L in terms of the relations $\Phi(\cdots)$ and a set of logical symbols: $\exists,\forall,\neg,\vee,\wedge,\Rightarrow$, [, and]. These have the usual logical meanings. Then in L the former statement in U about the existence of a sum of two numbers becomes the sentence

$$\forall a,b\,[\Phi(A,a)\wedge\Phi(A,b)\exists c[\Phi(A,c)\wedge\Phi(s,c,a,b)]],$$

where $\Phi(A,x)$ is the relation $x \in A$. This can be read "for each a and b in A there exists a c in A such that c is the sum (in the sum relation) of a and b." The entire set of sentences in L which are the abstract counterparts of U is denoted by K. In fact one can think of L (and therefore K) as given and U as being a particular concrete model for K. It is emphasized here that L and K are sets of abstract individuals, relations, symbols, and sentences and do not necessarily represent anything concrete. There may in fact be other models for K besides U. The intention here, however, is not to find new models of K but to extend K to a new set of sentences of L, K_c, in such a way that a model of K_c will include the desired infinite and infintesimal numbers. The relation between the model of K_c and the model of $K(U)$ will be given as the now famous *Transfer Theorem*. This will provide the connection allowing statements in one system to be interpreted and used to develop new results in the other.

A *concurrent relation* is any binary relation, say $\Phi(x, a, b)$, in K where a and b can be individuals, relations, sets, etc., and x is the name of the relation with the following property: For any finite set of entities G in the domain of the relation there exists an entity in the range of the relation for which the relation holds for all the entities of the finite set G. Symbolically, given any finite set $G \subset$ domain $[\Phi(x, a, b)]$, $G = \{g_1,\ldots, g_n\}$ say, there exists $y \in$ range $[\Phi(x, a, b)]$ such that

$$\Phi(x, g_1, y)\wedge\cdots\wedge\Phi(x, g_n, y)$$

is true. An example of this is the abstract relation for "$<$," written $\Phi(l, a, b)$, which means $a < b$ (or, more acurately, the images of a and b in L are in this relation). Then it is well known that for any finite set of numbers there always exists another number which is greater than all those in the finite set, i.e., "$<$" or $\Phi(l, a, b)$ is a concurrent relation. Now, using the set of concurrent relations, the set of sentences K can be enlarged.

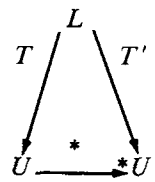
Take the set of all concurrent relations of K and add to K entities which satisfy the concurrent relations for all (i.e., infinite) sets of entities of the domains of the relations. In other words, for any concurrent relation $\Phi(x, a, b)$ and *any* subset of the domain of this relation, say $Q = \{a_i | i \in \mathbb{N}\}$, add to K from other available and unused entities of L, always assumed to exist, entities b such that $\Phi(x, a_i, b)$ is true $\forall i \in \mathbb{N}$, and all other formal sentences which were true still hold true. Call this new set of sentences so generated K_c. Note $K \subset K_c$ and K_c still retains formally all the qualities of K (sum relations, inequalities, set inclusion, theorems, lemmas, etc.). The great accomplishment of Robinson[10,12] was in showing that K_c also possess a model called a *nonstandard* universe ${}^*U$ based on an extension of $\mathbb{R}$, ${}^*\mathbb{R}$. Obviously $\mathbb{R} \subset {}^*\mathbb{R}$, but ${}^*\mathbb{R} - \mathbb{R} \neq \emptyset$. This can be seen by considering the concurrent relation "$<$". There exists, by construction, in ${}^*\mathbb{R}$ a number ω such that $n < \omega\,\forall n \in \mathbb{N}$. In fact, there exists an infinite number of such ω's since $\omega + x (x \in \mathbb{R})$, ω^2, $\frac{1}{2}\omega$, etc., also satisfy the concurrent relation "$<$" $\forall x \in \mathbb{R}$. These are the infinite numbers. Similarily, using "$>$" one obtains "$\exists \eta \in {}^*\mathbb{R}$ such that $(1/n) > \eta$, $n \in \mathbb{N}$." The number η is a member of the infinitesimal numbers. In general, standard sets are always subsets of their nonstandard extensions (denoted with a star prescript), but they are not necessarily equal to their nonstandard extensions. In fact, if the standard set is of infinite cardinality the nonstandard set will be larger, i.e., will contain nonstandard entities.

The question can now be asked: How does one go from the standard to the nonstandard universe and what rules apply to the new entities? At first one might guess that since both universes model similar sets of sentences in a formal abstract language all sets of individuals and all relations in ${}^*U$ have the same properties as those of U. That this is not so

can be seen by considering the set $\mathbb{R} \subset {}^*\mathbb{R}$. $\mathbb{R}$ is bounded above by ω, an infinite number, but $\mathbb{R}$ does not have a least upper bound (lub). If $\omega_1 = \text{lub}(\mathbb{R})$ then $\omega_1 - 1$ is an upper bound to $\mathbb{R}$ which is infinite also but is less than ω_1. From this it is seen that there exist sets in ${}^*U$ which do not have the usual standard properties (actually the logical formula of U have been presented rather loosely concerning the quantifier $\forall$; only bounded quantifiers can be transferred to ${}^*U$ and be true[15]). The answer to this situation lies in the realization that ${}^*U$ obeys all the *formal* properties of U as exemplified by the set of sentences K_c generated from K. Thus, one must first couch all entities from U in the formal language of L and only these statements will get through to ${}^*U$ and remain true. The reverse is also true, that is, only statements in ${}^*U$ which can be translated into a formal sentence of K_c and hence of K will be true (formally) in U. If the "*" prescript represents the map (transform) of entities and formal statements from U to ${}^*U$, then the above is a rephrasing of Robinson's

Transfer Theorem: All formal logical sentences are true in U if and only if the * transform of the sentence is true in ${}^*U$.

If T represents the 1–1 map which associates the model U with $K \subset L$ and T' represents the analogous map between ${}^*U$ and $K_c \subset L$, then the following diagram illustrates the *-map between U and ${}^*U$:



Thus $\mathbb{R}$ in ${}^*U$ is not the *-image of anything in U. ${}^*\mathbb{R}$ is the image of $\mathbb{R}$ and ${}^*\mathbb{R}$ has all the formal properties in ${}^*U$ of $\mathbb{R}$ in U. These considerations lead to the definition of the properties internal and external which can be assigned to the entities of ${}^*U$. An entity in ${}^*U$ is *internal* if and only if it is the *-image of a formal entity in U (e.g., ${}^*\mathbb{R}$ is internal) or if it is defined using only other internal quantities. If an entity is not internal it is *external* ($\mathbb{R}$ is external in ${}^*U$ as are the set of infinite numbers and the set of infinitesimal numbers). The transfer theorem basically states that only the internal quantities of ${}^*U$ obey the formal rules of U. In the above diagram the map T' is not onto since ${}^*U$ contains external quantities.

B. Some nonstandard applications and nomenclature

The *-transform applies to all standard quantities (e.g., $f \to {}^*f$, $2 \to {}^*2$, $\lim_{n\to\infty} \to {}^*\lim_{n\to{}^*\infty}$, etc.), but will be dropped when no confusion will arise. This can be done for many standard entities since their behavior is essentially the same in ${}^*U$ as in U.

The numbers ${}^*1, {}^*2, \ldots \in {}^*\mathbb{R}$ can be written and treated as one does the numbers $1, 2, \ldots$. However, there are other quantities in ${}^*\mathbb{R}$; for example, the set of infinite numbers ${}^*\mathbb{R}_\infty = \{x \in {}^*\mathbb{R} \mid |x| > |a|, a \in \mathbb{R}\}$ or the set of infinite integers ${}^*\mathbb{N}_\infty = {}^*\mathbb{N} - \mathbb{N} \subset {}^*\mathbb{R}_\infty$. There is also the set of infinitesimal numbers ${}^*\mathbb{R}_0 = \{x \in {}^*\mathbb{R} \mid |x| < |a|, a \in \mathbb{R}\}$. Numbers like $45 + \gamma$, where $\gamma \in {}^*\mathbb{R}_0$, are also in ${}^*\mathbb{R}$ since they belong to the sum relation. This number differs from 45 by an infinitesimal amount. The relation $\simeq$ is used to denote any two quantities which differ only by an infinitesimal amount. Thus $45 \simeq 45 + \gamma$ and ${}^*\mathbb{R}_0 = \{x \in {}^*\mathbb{R} \mid x \simeq 0\}$. Whenever a number c is finite ($c \notin {}^*\mathbb{R}_\infty$) then it can be shown that $c \simeq b$, where b is some standard number ($b \in \mathbb{R}$). This relation is written $b = {}^\sigma(c)$, where σ is the *standard part map*. Thus $45 = {}^\sigma(45 + \gamma)$. A number which has a standard part is said to be *near standard*. When two numbers differ by a finite amount they are said to be in the same *galaxy* or that one is in the galaxy of the other. That is, if $a, b \in {}^*\mathbb{R}$ and $a - b \notin {}^*\mathbb{R}_\infty$ then $a \in$ galaxy of b. For example, the galaxy of 0 (or any finite number) is just the set of finite numbers. The galaxy of an infinite number ω is just the set of other infinite numbers which differ from ω by a finite amount. It can be seen that $\simeq$ and "galaxy" set up equivalence classes in ${}^*\mathbb{R}$.

The word infinite can become confusing since it has many definitions here. It might refer to a number in ${}^*\mathbb{R}_\infty$ or to a $\lim_{n\to{}^*\infty}$, which is a *-transform of a standard definition of infinity. If one examines the quantities in ${}^*\mathbb{R}_\infty$ in this light they are *-finite in the sense that the elements of $\mathbb{R}$ are finite; however, when there will be no confusion one can say that the elements of ${}^*\mathbb{R}_\infty$ are infinite. The difference between the use of a *-transformed quantity and a standard quantity which is examined in ${}^*U$ comes out in the concept of continuity.

If a standard real valued function $f(x)$ on $[a, b]$ is continuous at $x = y$ then one knows that $\forall \epsilon > 0 \exists \delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. Thus picking a fixed standard ϵ and transforming this statement to ${}^*U$ it can be seen that for some positive $\delta \simeq 0$ this is always true $\forall$ standard $\epsilon > 0$. Thus one must have $f(x) \simeq f(y)$ whenever $x \simeq y$ and $x \in [a, b]$. This is the intuitive notion of standard continuity. Robinson[12] has shown that it can be taken as the starting point and the standard definition can be derived from it. However, in ${}^*U$ one also has the definition of *-continuity: An ${}^*\mathbb{R}$ valued function f on $[a, b]$, $a, b \in {}^*\mathbb{R}$, is *-continuous at $y \in [a, b]$ iff for each $\epsilon > 0 \exists \delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. In this last case ϵ and δ can be infinitesimal. For example, the function $f(x) = x^2$ is standard continuous only on the finite numbers, is not standard continuous on ${}^*\mathbb{R}_\infty$, but is *-continuous on all ${}^*\mathbb{R}$. Another example is the nonstandard internal function $\sin \omega x$, where $x \in {}^*\mathbb{R}$ and $\omega \in {}^*\mathbb{R}_\infty$. This is *-continuous but not standard continuous.

If $f(x)$ is standard and continuous on $[a, b]$ then a concept which is intuitively useful and thoroughly rigorous in ${}^*U$ is that of *vectorization*. For any infinite $\Omega \in {}^*\mathbb{N}_\infty$ define a vector $\tilde{f}$ in the Cartesian space ${}^*\mathbb{R}^\Omega = {}^*\mathbb{R} \times \cdots \times {}^*\mathbb{R}$ (Ω times) by $f_j \equiv (\tilde{f})_j = f(a + [(j-1)/\Omega]b)$.

Differentiation is intuitively defined also (but now the definition is rigorous). Given $f(x)$ a standard function then $f(x)$ has a derivative at $x = y$ iff there exists a unique near standard number $f'(y)$ such that

$$f'(y) \simeq \frac{f(y+\delta) - f(y)}{\delta} \quad \text{for all } \delta \simeq 0$$

or, equivalently, $f'(x) = {}^\sigma([f(y+\delta) - f(y)]/\delta)$ for all $\delta \simeq 0$. Another definition for the existence of $f'(y)$ is that $f(x)$ is "linear" in $\delta \simeq 0$ at $x = y$, which means $f(y + \delta) = f(y)$

$+ f'(y)\delta + \eta\delta$, where $\delta \simeq 0$ and $\eta \simeq 0$. If a standard function has a derivative everywhere on some interval then one can simply use the vectorization method and show that $f'(y) \simeq (f_j - f_{j-1})/\mu$, where $\mu = 1/\Omega$ and $j\mu \simeq y$. As in the case of continuity there are functions in $^*U$ which are not standard (but are internal) and possess a *-derivative (which is a *-transform of the standard definition of the derivative); for example, $e^{-\omega x}$ for some $\omega \in {}^*\mathbb{R}_\infty$.

The definition of integration of a continuous function is very direct along the usual intuitive lines. If $f(x)$ is a standard continuous function on $[a,b]$ then

$$\int_a^b f(x)dx \simeq \sum_{j=1}^{\Omega} f_j \mu,$$

where Ω, μ, and f_j are as before. The concept of integral can be star transformed to $^*U$ and one can write $^*\int_{\mathscr{S}} g(x)dx$, where g is some internal integrable function and $\mathscr{S}$ is a measurable set. For a standard function $h(x)$ the improper integral $\int_\infty^\infty h(x)dx$ can be examined in $^*U$. If the standard integral exists in U then

$$^*\int_{-\eta}^{\eta} h(x)dx \simeq \int_{-\infty}^{\infty} h(x)dx \left(\text{or } {}^\sigma\left[\int_{-\eta}^{\eta} h(x)dx\right] = \int_{-\infty}^{\infty} h(x)dx\right)$$

$$\text{and} \int_{\eta}^{\infty} h(x)dx \simeq 0$$

for all infinite $\eta (\eta \in {}^*\mathbb{R}_\infty)$. Note that for a nonstandard function, "$p(x) = 1$ if $x < \omega$ for ω fixed, positive, and infinite and e^{-x} for all other x;" for example, the integral $\int_0^\infty p(x)\,dx$ exists (in $^*U$) because of the *-transformed limit definition of an improper integral.

Overall, one can see that various operations and concepts for standard quantities can be defined on $^*U$ and mapped back to U (by σ, for example) and these concepts can be *-transformed back to $^*U$ to operate on quantities in $^*U$! While this may at first seem confusing very little perplexity emerges so long as one realizes that the * concepts apply only to *internal* quantities. Looking back at the three nonstandard functions defined in the preceding paragraphs one can see that they are internal. An example of the external function is "$f(x) = 1$ if $x \simeq 0$ and 0 otherwise," since it involves the set $^*\mathbb{R}_0$ which is external, i.e., it has no image in $K_c \subset L$] and, of course, has no counterpart in U. Similarly, $^*\mathbb{R}_\infty$ and $^*\mathbb{N}_\infty$ are external. Internal objects are those which are generated by a *-transform or through a formal definition in K_c involving other internal objects. For example, $\{x | x < \omega\}$ is internal for any fixed ω even if it is infinite. The concepts of internal and external are both used to develop new results. It is their interplay which is important.

If an internal quantity $Q(n) \in {}^*\mathbb{R}$ for $n \in {}^*\mathbb{N}$ is infinitesimal for all $n \in \mathbb{N}$ then it is infinitesimal for some infinite $n_1 \in {}^*\mathbb{N}_\infty$. This is the infinitesimal prolongation theorem.[13] The proof is quite straightforward and illustrates very well the concepts of internal and external. Consider the set $A = \{n \in {}^*\mathbb{N} | \, |nQ(n)| < 1\}$. A is internal (it is defined in terms of internal quantities and relations). $A \neq \mathbb{N}$ since $\mathbb{N}$ is external, but from the definition of Q it is true that $|nQ(n)| < 1$ for all $n \in \mathbb{N}$ (because infinitesimal x finite = infinitesimal). Thus, there must exist an $n_1 \in {}^*\mathbb{N}_\infty$ such that $|n_1 Q(n_1)| < 1$ and hence $|Q(n_1)| < |1/n_1| \simeq 0$.

The concepts in this section and other *-transformed properties will be of use in the following development.

III. THE SPACE $^*\mathbb{R}^\Omega$ AND GAUSSIAN MEASURES

Since, intuitively, one wants to integrate over an infinite dimensional space one can introduce the space $^*\mathbb{R}^\Omega$, where $\Omega \in {}^*\mathbb{N}_\infty$. Although no direct attempt is made here to rigorously relate integrals in $^*\mathbb{R}^\Omega$ with a measure on a particular algebra of sets, it can be shown that in certain cases it is possible to define $^*\mathbb{R}^\Omega$ consistently with such an algebra.

For example, on the space of continuous functions $\mathscr{C}[0,1]$ a definiton of an algebra of cylinder sets is as follows. Given a finite set of points $(t_1,\ldots,t_n)$, $t_1 < t_2 < \cdots < t_n$, in $(0,1)$ define the map $z_t: \mathscr{C} \to \mathbb{R}^n$ by $z_t(f) = (f(t_1),\ldots,f(t_n))$. Then the sets $C_t^B = z_t^{-1}(B)$, where B is a Borel set in $\mathbb{R}^n$, are cylinder sets in $\mathscr{C}$. A Gaussian measure may be defined on $\mathscr{C}$ by using these sets as the basis for a σ-algebra S on $\mathscr{C}$ and defining a measure on the C_t^B's by an integration over a Gaussian in $\mathbb{R}_n$. For example, a conventional definition of the Weiner measure is the following (in the next section a more heuristic definition will be given). Define the function

$$w(s,x,y) = \frac{1}{\sqrt{2\pi s}} e^{-(x-y)^2/2s}.$$

Then define the measure λ on C_t^B by

$$\lambda(C_t^B) \equiv \int_B db_1 \cdots db_n \, w(t_1,0,b) w(t_2,b_1,b_2) \cdots w(t_n, b_{n-1}, b_n).$$

This induces a measure on S, the Weiner measure. In order to use an infinite dimensional space in this case note that given any set of sequences $\{t^j | j = 1,\ldots,n\}$, where $t^j = (t_1^j,\ldots,t_{m_j}^j) \in \mathbb{R}^{m_j}$, one can define an ascending set of sequences of ordered points $u^k = \cup_{j=1}^k t^j$ so that $u^k \subset u^{k+1}$ and u^k represents a point in $\mathbb{R}^{l_k}$ where $l_k = \Sigma_1^k m_j$. A concurrent relation has been set up here which states that for any finite set of t^j's there is a new set u^n which includes all the points of the t^j's and has associated with it a new space $\mathbb{R}^{l_n}$. This means that in the nonstandard universe there exists a sequence of Ω elements t, say, and an associated space $^*\mathbb{R}^\Omega$ for some $\Omega \in {}^*\mathbb{N}_\infty$ for which $t^j \subset t$ for all standard t^j's. Note that t contains all standard points in $(0,1)$ as well as infinitely many nonstandard ones. A Weiner measure for a cylinder set is now easily defined on this space because the Borel set B which defines C_u^B has an image $^*B$ in $^*\mathbb{R}^n \subset {}^*\mathbb{R}^\Omega$ and one merely integrates from $-\infty$ to $+\infty$ over all variables not in $^*B$. Put more rigorously, if π is the projection from $^*\mathbb{R}^\Omega$ to $^*\mathbb{R}^n$, where $^*B \subset {}^*\mathbb{R}^n$, then define the measure

$$\lambda(C_u^B) = {}^\sigma\left(\int_{\pi^{-1}(^*B)} db^1 \cdots db_\Omega \; W(b^1,\cdots,b_\Omega)\right), \tag{1}$$

where W is the appropriate infinite product of the w functions (Ω of them). In general, it will be assumed that something of this nature can be done for each measure in this article.

In the previous example the criteria that all standard elements of $(0,1)$ be in t can be relaxed to just making $t_i = i/\Omega$, $i = 1,\ldots,\Omega - 1$, for the components of t. The integral in Eq. (1) can then be defined the same way except taking as the variables in $^*B$ a set of n points which are infinitesimal-

ly close to those of u. Since the integrand in (1) is a continuous standard function and the elements of u are all different by standard amounts, the two integrands will differ only by an infinitesimal amount and thus on taking standard parts the measure will remain the same for all C_u^B. The definition of the Gaussian in terms of equal spacings such as this will be done in the next section. A general outline of this process is presented below.

Using the space ${}^*\mathbb{R}^\Omega$ as a starting point a general Gaussian measure can be defined on it. In fact, the method throughout shall be to first define the Gaussian measure and then determine what standard entities it is concentrated on. The Gaussian measures here will be defined using square matrices on ${}^*\mathbb{R}^\Omega$ with nonzero real eigenvalues (although in general they may be infinite or infinitesimal). If A is such a matrix on ${}^*\mathbb{R}^\Omega$ then, letting $G = A^T A$, the Gaussian measure related to A can be defined using

$$\exp[-(\tfrac{1}{2})(\tilde{f}, G\tilde{f})], \tag{2}$$

where $\tilde{f} \in {}^*\mathbb{R}^\Omega$ and $(,)$ is the natural inner product on ${}^*\mathbb{R}^\Omega$. Thus G is symmetric and has positive definite eigenvalues. Then, for example, if this is to be a measure on the cylinder sets of $\mathscr{C}[0,1], C_u^B$, say, pick the components of f according to the previous prescription involving equally spaced t_j's and define the measure by

$$\lambda(C_u^B) = {}^\sigma\left(\int_{\pi^{-1}({}^*B)} \frac{df_1 \cdots df_\Omega}{(2\pi)^{\Omega/2}\sqrt{\det G}} e^{-(1/2)(\tilde{f}, G\tilde{f})}\right). \tag{3}$$

The exact form of G will be shown later. All integrals of this nature will be written with the integrand

$$d\tilde{f}\, e^{-(1/2)(\tilde{f}, G\tilde{f})}, \tag{4}$$

where $d\tilde{f}$ includes all the normalization factors like π and $\det G$. However, all functional integrals in this article will be handled by the simple approach of assuming the functional to be integrated is of the form in which its arguments can be vectorized and one can then integrate over each component. For any functional F this will be expressed by the symbolic form $\int d\lambda_G\, F$, where $d\lambda_G = d\tilde{f} \exp[-(\tfrac{1}{2})(\tilde{f}, G\tilde{f})]$.

Using the variables $\tilde{x} = A\tilde{f}$, where $G = A^T A$, (4) becomes

$$d\tilde{x}\, e^{-(1/2)(\tilde{x}, \tilde{x})} = d\tilde{x} \exp\left(-(\tfrac{1}{2}) \sum_1^\Omega x_m^2\right). \tag{5}$$

An important heuristic concept in what follows will be that of the external set on which the measure is concentrated. That is, in order to visualize the subset of vectors of ${}^*\mathbb{R}^\Omega$ which contribute to the standard part of the measure in (5) a definition will be offered which produces a set E which is maximum in some sense (to be defined) on which the integral of (5) is infinitesimal. Then the integral is concentrated on $H = {}^*\mathbb{R}^\Omega - E$. The set E is external since its definition involves the concept "infinitesimal" and thus H is also external. Hence, integration over E of H cannot be seriously considered since in general integrals exist only over internal sets. But the E and H concepts will be useful to help visualize the region of ${}^*\mathbb{R}^\Omega$ in which the representatives $\tilde{f}$ of standard entities are concentrated by the measure.

The entire contribution to the measure comes on the set $H = {}^*\mathbb{R}^\Omega - E$. This means that the $\tilde{f}$ which contribute to the standard part of the integral are contained in $A^{-1}(H)$. It is these vectors in ${}^*\mathbb{R}^\Omega$ which will relate to standard entities. The heuristic sets E and H will be referred to throughout. Although E and H cannot be exhibited explicitly, several "approximations" to them will be shown in the next paragraphs.

Several theorems about the properties of the vectors $\tilde{x}$ in Eq. (5) can be proven. Since the quantities of interest $\tilde{f}$ are related to $\tilde{x}$ through a transformation A, knowing $\tilde{x}$ will immediately shed light on properties of $\tilde{f}$. The first property is derived in

Theorem 1: At least one component of $\tilde{x}$ can be infinite, i.e., for each $\tilde{x} \in H$ one x_i can be an infinite number. An "upper bound" on this number is $\sqrt{\ln(8\Omega^2/\pi\gamma^2)}$, where γ is an infinitesimal number.

Proof: Start with a more exact definition for a set encompassing E, namely the internal set $E_c \equiv \{\tilde{x} \mid |x_i| > c$ for some i, all other components are arbitrary$\}$. The object now could be to find a value of c such that $\lambda(E_c) \simeq 0$. However, it is easier to calculate the measure of $H_c = {}^*\mathbb{R}^\Omega - E_c$ $= \{\tilde{x} \mid |x_i| < c$ for all $i\}$ and determine for what c values it is $\simeq 1$. Then

$$\lambda(H_c) = \prod_{i=1}^{\Omega} 2 \int_0^c \frac{dx_i}{\sqrt{2\pi}} e^{-(1/2)x_i^2}. \tag{6}$$

Letting $2\int_0^c e^{-(1/2)x^2} dx/\sqrt{2\pi} = (1 - \xi)$, where $0 \le \xi \le 1$, then $\lambda(H_c) = (1 - \xi)^\Omega$. Now although $\xi \simeq 0$ whenever $c \in {}^*\mathbb{R}_\infty$, $(1 - \xi)^\Omega \not\simeq 1$ for certain infinitesimal ξ. Thus defining $H' = \cap\{H_c \mid \lambda(H_c) \simeq 1\}$ it is seen that the measure is "concentrated" on H' and H' contains some $\tilde{x}$ for which at least one component can be an infinite number. An "upper bound" can be placed on these numbers by considering the cases for which $(1 - \xi)^\Omega \simeq 1$. ξ is given by $\xi = (2/\sqrt{\pi}) \int_{c/\sqrt{2}}^\infty dx\, e^{-x^2}$, which follows from the properties of the error integral transferred to ${}^*U$. In order to see what values of c will yield $(1 - \xi)^\Omega \simeq 1$, note that for all infinitesimal γ, $e^{-\gamma} \simeq 1$. Then for some positive $\gamma \simeq 0$ set $(1 - \xi)^\Omega > e^{-\gamma}$ to insure the result. Thus $\xi < 1 - e^{-\gamma/\Omega}$. Since ξ is monotonically decreasing in c this relation determins a c such that for any sets which contain $\tilde{x}$'s with any component larger that this c the measure of the set will be infinitesimal. An estimate of c (in terms of γ) can be obtained by using the approximation

$$\xi = \left(\frac{2}{\sqrt{\pi}}\right) \int_{c/\sqrt{2}}^\infty dx\, e^{-x^2} < \left(\frac{2}{\sqrt{\pi}}\right) \int_{c/\sqrt{2}}^\infty dx\, \frac{\sqrt{2}x}{c} e^{-x^2}$$
$$= \frac{2}{c}\left(\frac{2}{\pi}\right)^{1/2} e^{-c^2/2}. \tag{7}$$

Then setting

$$\frac{2}{c}\left(\frac{2}{\pi}\right)^{1/2} e^{-c^2/2} < 1 - e^{-\gamma/\Omega} \tag{8}$$

insures that $\lambda(H_c) \simeq 1$. Solving (8) for c gives $c > \sqrt{\ln(8\Omega^2/\gamma^2\pi)}$. ■

Note that this depends on an external relation, $\simeq$, and so one cannot get a true upper bound for x_1. However, it does provide a geometric picture of an approximation to H, name-

ly the external set $H' = \{\tilde{x} | |x_1| < \sqrt{\ln(8\Omega^2/\gamma^2\pi)}$ for all positive $\gamma \simeq 0\}$. This is just an Ω-dimensional cube.

Because of the form of the Gaussian it is natural to ask whether there are any bounds on $r^2 = \Sigma_1^\Omega x_j^2$ when $\tilde{x} \in H$. From Theorem 1 a bound would be $r^2 < \Omega \ln(8\Omega^2/\gamma^2\pi)$. In fact, a better set of bounds on r^2 is given by

Theorem 2: For all $\tilde{x} \in H$ $r^2 = \Sigma x_j^2 \in$ galaxy of Ω.

Proof: Using hyperspherical coordinates for an n-dimensional space $\mathbb{R}^n$

$$x_1 = r\cos\theta_1,$$
$$x_2 = r\sin\theta_1\cos\theta_2,$$
$$\vdots$$
$$x_{n-1} = r\sin\theta_1\cdots\sin\theta_{n-2}\cos\theta_{n-1},$$
$$x_n = r\sin\theta_1\cdots\sin\theta_{n-1},$$

where $r^2 = \Sigma x_i^2$, yields a volume element $dV = r^{n-1}dr dA_n$, where dA_n = solid angle $= \sin^{n-2}\theta_1\cdots\sin^2\theta_{n-2}\sin\theta_{n-1}d\theta_1\cdots d\theta_{n-1}$.
For simplicity choose n even. Then

$$\int_{\substack{\text{Unit}\\ \text{hypersphere}}} dA_n = \frac{2\pi^{n/2}}{(n/2-1)!}.$$

Transfer these results to $^*U$ and apply them to the measure in Eq. (5) and the set $H(c_1,c_2) = \{\tilde{x} | \sqrt{2}c_1 < |\Sigma_{i=1}^\Omega x_i^2| < \sqrt{2}c_2\}$. This gives

$$\lambda(H(c_1,c_2)) = \frac{2}{(\Omega/2-1)!}\int_{c_1}^{c_2} r^{\Omega-1}e^{-r^2}dr. \tag{9}$$

To see for what values of c_1 and c_2 $\lambda(H(c_1,c_2)) \simeq 1$ let $k = \Omega/2$ and examine first the integral

$$\lambda_2 = \frac{2}{(k-1)}\int_{c_2}^{\infty} r^{\Omega-1}e^{-r^2}dr.$$

The integrand has a maximum at $r_{\max}^2 = (\Omega-1)/2$ and is an increasing function of r for $r < r_{\max}$ and a decreasing function of r for $r > r_{\max}$. So it is expected that the integral will be concentrated in the region of $r_{\max}$.

Changing variables to $y = r^2$,

$$\lambda_2 = \frac{1}{(k-1)!}\int_{c_2^2}^{\infty} y^{k-1}e^{-y}dy.$$

Transferring the standard integral formula[17] to $^*U$ and doing some algebra yields

$$\lambda_2 = e^{-c_2^2}\sum_{n=0}^{k-1}\frac{c_2^{2n}}{n!}.$$

Since it is expected that the c_2^2 for which $\lambda_2 \simeq 0$ will be greater than $(\Omega-1)/2$ pick $c_2^2 = a(\Omega-1)$ with $a > 1$. Then

$$\lambda_2 < e^{-a(\Omega-1)}a^k\sum_{n=0}^{k-1}\frac{(\Omega-1)^n}{n!} < e^{-a(\Omega-1)}a^k e^{(\Omega-1)}.$$

Rewriting this as an exponential and substituting for k,

$$\lambda_2 < e^{(1-a+(1/2)\ln a)\Omega-(1-a)}.$$

Obviously, if a is such that $^\sigma(a) > 1$ then $\lambda_2 \simeq 0$, since the exponent is negative and infinite [this follows from $a > \frac{1}{2}\ln a$]. A closer estimate on a can be made by taking $a = 1 + \eta/\Omega$, where $\eta > 0$ and $\eta/\Omega \simeq 0$. Note that η can still be infinite. This gives for the exponent

$$\left(-\frac{\eta}{\Omega} + \frac{1}{2}\ln(1+\eta/\Omega)\right)\Omega + \frac{\eta}{\Omega}.$$

Using $\ln(1+\eta/\Omega) < \eta/\Omega$ this gives for λ_2,

$$\lambda_2 < e^{-(1/2)\eta+\eta/\Omega}.$$

Thus $\lambda_2 \simeq 0$ whenever η is infinite or, equivalently, whenever $c_2^2 \geqslant \Omega + \eta$ for some infinite η.

The case for a similarly defined λ_1 as a function of c_1 is the mirror image of that for λ_2. Take $a < 1$ and $c_1^2 = a(\Omega-1)$. Then the same formulas result as before for the exponential. Picking $a = 1 - \eta/\Omega$, where $\eta > 0$ and $\eta/\Omega \simeq 0$, gives the condition that $\lambda_1 \simeq 0$ whenever $c_1^2 \leqslant \Omega - \eta$ for some infinite $\eta < \Omega$. Thus $H(c_1,c_2)$ is concentrated in the region in which $\Sigma x_1^2 = \Omega + \eta$ for some finite η. This is just the galaxy of Ω.■

This presents the geometric picture of H being contained in a spherical shell in $^*\mathbb{R}^\Omega$ of nominal radius Ω and "thickness" the external set $^*\mathbb{R} - {}^*\mathbb{R}_\infty$, the set of finite numbers. This result will be particularly useful in showing the property of nonbounded variation for functions of the Weiner measure.

One final theorem of this section, while not so general, will be of use in plausibility arguments for the finiteness and continuity properties of the Weiner functions.

Theorem 3: For $\tilde{x} \in H$ $\frac{1}{\sqrt{\Omega}}\Sigma_{i=1}^\Omega x_1$ is near standard.

Proof: In order to prove this the full use of $^*\mathbb{R}^\Omega$ as a Euclidean vector space is made. Let $\{\hat{e}_j\}$ be an orthonormal basis for $^*\mathbb{R}^\Omega$ so that $\tilde{x} = \Sigma_1^\Omega x_1\hat{e}_i$. Let $\{\hat{a}_i\}$ be a new basis related to $\hat{e}_i$ by an orthogonal transformation O: $\hat{e}_j = \Sigma_1^\Omega O_{ji}\hat{a}_i$. Pick O such that $\hat{a}_1 = (1/\sqrt{\Omega})\Sigma_1^\Omega\hat{e}_i$. This can be done since $\hat{a}_1$ determines one direction and it is always possible to define the rest of the transformation on the subspace $\{\tilde{v} \in {}^*\mathbb{R}^\Omega | \tilde{v}\cdot\hat{a}_1 = 0\}$ so that it is orthogonal. This is assumed to be done and the detailed form of O on this subspace will not be of concern here. The factor of $1/\sqrt{\Omega}$ in the definition of $\hat{a}_1$ normalizes $\hat{a}_1$. Then the component of $\tilde{x}$ along $\hat{a}_1$ is given by $y_1 \equiv \tilde{x}\cdot\hat{a}_1 = x_j O_{j1}$. From the properties of orthogonal transformations transferred to $^*U$, $O_{j1} = O_{1j} = 1/\sqrt{\Omega}$ so that $y_1 = (1/\sqrt{\Omega})\Sigma_1^\Omega x_i$. Because O is orthogonal $\Sigma_1^\Omega x_1^2 = \Sigma_1^\Omega y_i^2$. Let $E_c = \{\tilde{x} | (1/\sqrt{\Omega})|\Sigma_1^\Omega x_i| > c\}$. Then transforming the integral over E_c using O gives

$$\lambda(E_c) = 2\int_c^\infty \frac{dy_1}{\sqrt{2\pi}}e^{-(1/2)y_1^2}.$$

Thus for all infinite c $\lambda(E_c) \simeq 0$ and so for all $\tilde{x} \in H$ $(1/\sqrt{\Omega})\Sigma_1^\Omega x_i$ is near standard.■

IV. DISCRETE FUNCTIONAL MEASURES

A. First order or Weiner Gaussian measures

The first Gaussian functional measure considered here is the usual Weiner measure,[18] which was historically the first functional measure studied. The path spaces will be taken throughout to be the paths with domain [0,1]. The Weiner

paths form the set of paths of Brownian motion of a point particle.[18] This measure will be the foundation from which several other Gaussian measures will be built. The path properties are already well known[18] and those necessary to this article will be stated with only a few nonstandard proofs shown.

The Weiner measure and several subsequent measures are defined here by considering the following matrix D. Let $\Omega \in {}^*\mathbf{N}_\infty$ and let $\mu = 1/\Omega$, then on the space ${}^*\mathbf{R}^\Omega$ define the matrix

$$D = \frac{1}{\mu}\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}. \tag{10}$$

Let $G_1 = \mu D^T D$. Note that the definition of G_1 differs slightly here from the previous section. This defines a Gaussian measure using Eq. (3). To put the exponential into more familiar form use Eqs. (3) and (1) to write (using $f_0 = 0$)

$$\exp[-(1/2)(\tilde{f},G\tilde{f})] = \exp\left[-\frac{1}{2}\sum_{j=1}^{\Omega}(f_j - f_{j-1})^2/\mu\right]. \tag{11}$$

Then, using the nonstandard expression for the integral of a function and the derivative of a function, (11) becomes

$$\exp\left[-\frac{1}{2}\int_0^1 \left(\frac{df}{dt}\right)^2 dt\right], \tag{12}$$

the form more familiar to physicists as the measure in a path integral. The diffusivity has been taken to be one-half here. Also one should note the derivative df/dt is a weak derivative (a distribution) since it will be seen that the paths f are not differentiable.

It is well known that the Weiner measure is concentrated on a set of continuous functions $\mathscr{W}_1$ with certain properties. The external set in ${}^*\mathbf{R}^\Omega$ on which the measure is concentrated is $(1/\sqrt{\mu})D^{-1}(H)$ in the basis of the vectors $\tilde{f}$. The assumption here (which makes the subsequent development heuristic and not entirely rigorous) is that the vectors $\tilde{f} \in (1/\sqrt{\mu})D^{-1}(H)$ represent the Weiner functions in that they are infinitesimally close to them. Thus, assume that if $\tilde{f} \in (1/\sqrt{\mu})D^{-1}(H)$ then $f_j \simeq \alpha(j/\Omega)$, where $\alpha \in \mathscr{W}_1$. Some plausibility arguments for this are given below, but so far it has not been possible to fully show this using these nonstandard vector-space methods.

Let $F = (1/\sqrt{\mu})D^{-1}(H)$. Then the following arguments show that the above associations of $\tilde{f}$'s and α's is plausible.

1. $\tilde{f} \in F$ are all near standard. From Theorem 3 it follows that f_Ω is finite. For any $j < \Omega$ Theorem 3 can be generalized by applying it to the subspace generated by $\{\hat{e}_i \mid i = 1,\ldots,j\}$ to show that $(1/\sqrt{j})\Sigma_1^j x_i$ is finite. Now for $j\mu \simeq t$, where t is any standard point in $(0,1)$, $j = a\Omega$, where ${}^\sigma a \in (0,1)$ and therefore $(1/\sqrt{\Omega})\Sigma_1^j x_i$ is finite. Since the sum to j (with the $1/\sqrt{\Omega}$ factor) is just f_j, this is highly suggestive that f_j is finite. But the actual set to measure is the set
$H_c = \{\tilde{f} \mid\mid f_j| = \sqrt{\mu}|\Sigma_1^j x_i| < c \text{ for all } j\}$ to see whether $\lambda_{G_1}(H_c) \simeq 1$ for all infinite c. This is analogous to finding the allowed values of the separate x_i in Theorem 1. A similar plausibility argument could have been made there that the $\tilde{x} \in H$ are all finite, but as shown this is not necessarily true. That f_j is finite is no doubt true since the Weiner measure is well known to be concentrated on finite functions, but integration over H_c is not a simple matter.

2. $\tilde{f} \in F$ correspond to continuous standard functions. This can be proven given 1. Let $E_c = \{\mid|f_j - f_{j-1}| > c \text{ for some } j\}$. Obviously, if $\tilde{f}$ is to represent some $\alpha \in \mathscr{W}_1$ it must be true that $\lambda(E_c) \simeq 0$ for all standard c. Let $\sqrt{\mu}x_j = f_j - f_{j-1}$, then by a simple change of variables $\lambda_{G_1}(E_c) = \lambda(K_c)$, where $K_c = \{\tilde{x} \mid |x_j| > c/\sqrt{\mu} \text{ for some } j\}$. Let $K_c^j = \{\tilde{x} \mid |x_j| > c/\sqrt{\mu}$, all other components arbitrary$\}$. Then $K_c = \cup_j K_c^j$ and $\lambda(K_c) < \Sigma_1^\Omega \lambda(K_c^j)$. Now

$$\lambda(K_c^j) = 2\int_{c/\sqrt{\mu}}^{\infty} \frac{dx_j}{(2\pi)^{1/2}} e^{-(1/2)x_j^2} < \frac{2}{c}\left(\frac{\mu}{\pi}\right)^{1/2} e^{-c^2/\mu} \quad \forall j;$$

then using the definition of μ,

$$\lambda(K_c) < \frac{2}{c}\left(\frac{\Omega}{\pi}\right)^{1/2} e^{-c^2\Omega} \simeq 0 \quad \text{for all finite } c,$$

since the exponential dominates the Ω factor. By the infinitesimal prolongation theorem $\lambda(K_c) \simeq 0$ for certain infinitesimal c's, which suggests that the continuity properties of $\alpha \in \mathscr{W}_1$ can be restricted.

3. The $\tilde{f} \in F$ are "locally" Holder continuous of order $\gamma < \frac{1}{2}$. The sense of local will be made clear in the following. Consider the approach of Theorem 3 applied only to the subspace generated by $\{e_l, e_{l+1}, \ldots, e_k\}\, k > l$. Make an orthogonal transformation to $y_1 = (1/\sqrt{k-l})\,\Sigma_{j=l}^k x_j$. Then $\sqrt{\mu(k-l)}\, y_1 = f_k - f_l$ and the restriction $|f_k - f_l| > c|(k-l)\mu|^\gamma$ becomes $|y_1| > c|(k-l)\mu|^{\gamma - 1/2}$. Now define the sets $E_{kl}^c = \{\tilde{f} \mid |f_k - f_l| > c|(k-l)\mu|^\gamma\}$. Let $L \in {}^*\mathbf{N}$ and $L < \Omega$. Then the set $E_L^c = \cup_{k-l<L} E_{kl}$ is the collection of vectors $\tilde{f}$ whose components f_k, f_l, say, differ by values larger than $c|(k-l)\mu|^\gamma$ whenever $L > (k-l)$. This permits investigation of the continuity on a "local" level determined by the size of L. Now $\lambda_{G_1}(E_L^c) < \Sigma_{k-l<L}\lambda_{G_1}(E_{kl}^c)$ and

$$\lambda_{G_1}(E_{kl}^c) = 2\int_{c/|(k-l)\mu|^{(1/2)-\gamma}}^{\infty} \frac{dy_1}{(2\pi)^{1/2}} e^{-(1/2)y_1^2}$$
$$< \frac{2|(k-l)\mu|^{(1/2)-\gamma}}{(2\pi)^{1/2}c} \exp(-c^2|(k-l)\mu|^{2\gamma-1}),$$

so that

$$\lambda_{G_1}(E_L^c) < \sum_{(k-l)<L} \frac{2|(k-l)\mu|^{(1/2)-\gamma}}{(2\pi)^{1/2}c} \exp(-c^2|(k-l)\mu|^{2\gamma-1}).$$

Letting $j = k - l$ the sum becomes

$$\lambda_{G_1}(E_L^c) < \sum_{j=1}^{L} \frac{(\Omega - j)|j\mu|^{(1/2)-\gamma}}{(2\pi)^{1/2}c} \exp(-c^2|j\mu|^{2\gamma-1}),$$

since for each j there are $(\Omega - j)$ equal terms. Noting that $|j\mu|^{(1/2)-\gamma} < 1$ and $|j\mu|^{2\gamma-1} \geqslant |L\mu|^{2\gamma-1}$, the previous expression becomes

$$\lambda_{G_1}(E_L^c) < \frac{2}{(2\pi)^{1/2}c} \exp(-c^2|L\mu|^{2\gamma-1}) \sum_{j=1}^{L} (\Omega - j).$$

Replacing the sum by its (transferred) inductive formula,

$$\lambda_{G_1}(E_L^c) < \frac{L(2\Omega + L + 1)}{(2\pi)^{1/2}c} \exp(-c^2|L\mu|^{2\gamma-1}).$$

This will not be infinitesimal in general for all infinite c because of the explicit presence of the Ω factor. However, if $|L\mu| \lesssim \Omega^{-b}$, where b is standard and $b \in (0,1)$, the following expression results:

$$\lambda_{G_1}(E_L^c) < \frac{L(2\Omega + L + 1)}{(2\pi)^{1/2}c} e^{-c^2\Omega^b} < \frac{3\Omega^2}{(2\pi)^{1/2}c} e^{-c^2\Omega^b},$$

where the fact that $L \lesssim \Omega - 1$ has been used. The crucial question is whether $(\Omega^2/c)\exp(-c^2\Omega^b)$ is infinitesimal. Rewrite this as a total exponential:
$\exp(-c^2\Omega^b + 2\ln\Omega - \ln c)$. Now for standard $b \in (0,1)$ and any noninfinitesimal c, $c^2\Omega^b > 2\ln\Omega$ by an infinite amount. Thus, $\lambda_{G_1}(E_L^c) \simeq 0$ and the $\tilde{f} \in H$ are locally Holder continuous of order $\gamma < \frac{1}{2}$, the locality being determined by L.■

The general result for Weiner functions is that they are truly Holder continuous of order $< \frac{1}{2}$.[18] If an attempt is made to obtain this result above by choosing $\gamma < \frac{1}{2}$ and c infinite it is noted that for certain values of L (now restricted only by $L \lesssim \Omega - 1$) $L\mu$ is not infinitesimal and there is no guarantee that the total exponential expression in the theorem will be negative infinite unless c is on the order of Ω^b for b as given in the theorem. The problem is that the inequality $\lambda_{G_1}(E_L^c) < \Sigma_{(k-l)<L}\lambda(E_{kl}^c)$ is too crude and a lower upper-bound is needed. Of course, neither the local nor the total result holds if $\gamma \geqslant \frac{1}{2}$. The local result will be useful for converting the present discretely defined measure to a measure on Fourier series coefficients.

A property which follows from the Holder continuity property is that the paths are nowhere differentiable. This can be seen by recalling that for a standard function $g(x)$ to possess a derivative at x_0 there must exist a standard number, $g'(x_0)$ say, and an infinitesimal ρ such that[13]

$$g(x_0 + \mu) - g(x_0) = g'(x_0)\mu + \rho\mu$$

whenever $\mu \simeq 0$. That is, $g(x_0 + \mu) - g(x_0)$ is linear "to first order" in μ. But if $g \in \mathscr{W}_1$, then g is at best Holder continuous; this means $g(x_0 + \mu) - g(x_0)$ goes as μ^γ, where $\gamma \in [0, 1/2)$, which is "less than linear" and therefore $g'(x_0)$ is not necessarily standard or near standard.

Finally, the property of nonbounded variation of Weiner paths can be derived directly by nonstandard means. A function h on $[0,1]$ is said to be of bounded variation if

$$\sup \sum_{i=1}^{n} |h(y_i) - (y_{i-1})| < \infty$$

for all partitions $\{y_j j = 1, \ldots, n\}$ of $[0,1]$. It can be shown that all vectors in the set on which the measure is concentrated, F, represent functions which are of *nonbounded* variation. Use the following notation again, $x_j = (f_j - f_{j-1})/\sqrt{\mu}$. Then the nonstandard Weiner measure becomes

$$e^{-(1/2)(\tilde{f},G\tilde{f})}d\tilde{f} = e^{-(1/2)(\tilde{x},\tilde{x})}d\tilde{x}. \tag{13}$$

In the previous section is was shown that if $\tilde{x} \in H$ then $\Sigma_1^\Omega x_i^2 = \Omega + a$, where a is near standard, contribute to the integral. Now, recalling the definition of x_j and μ gives the relation

$$\sum_1^\Omega (f_i - f_{i-1})^2 \simeq 1,$$

which must be true for all f which are present a function $\alpha \in \mathscr{W}_1$ and so it must be true for the vectorization of α [that is, for $\tilde{\alpha}$ where $\alpha_j = \alpha(j\mu)$]. Geometrically, the result states that the differentials of the members of $\mathscr{W}_1$ must lie on a spherical shell of nominal radius 1 and with thickness of the set of infinitesimals.

Let $\tilde{\alpha}$ be the vectorization of $\alpha \in \mathscr{W}_1$. Then, with

$$\Delta\alpha = \max_{i=1,\ldots,\Omega} |\alpha_i - \alpha_{i-1}|,$$

$$1 \simeq \sum_1^\Omega (\alpha_i - \alpha_{i-1})^2 < \Delta\alpha \sum_1^\Omega |\alpha_i - \alpha_{i-1}|.$$

But since $\Delta\alpha \simeq 0$ this means that the variation of α must be infinite and hence each $\alpha \in \mathscr{W}_1$ is of nonbounded variation.

The covariance of the Weiner measure can also be calculated using the vectors of ${}^*\mathbb{R}^\Omega$ in a straightforward way. The covariance is defined as the expectation value of the product $\alpha(t)\alpha(s)$ for $t,s \in [0,1]$. This can be evaluated by an integration on ${}^*\mathbb{R}^\Omega$. Let the functional $C(\alpha) = \alpha(t)\alpha(s)$ for $\alpha \in \mathscr{W}_1$. The covariance is $\int d\lambda_{G_1} C$. Choose j and k such that $j\mu \simeq t$ and $k\mu \simeq s$, then write

$$\text{covariance} = \text{Cov}\, G_1(t,s) = {}^\sigma\left(\int_{{}^*\mathbb{R}^\Omega} f_j f_k e^{-1/2)(\tilde{f},G\tilde{f})} d\tilde{f}\right) \equiv {}^\sigma(J_{G_1}). \tag{14}$$

This is easily calculated by first noting that $\int_{{}^*\mathbb{R}^\Omega} x_j x_k \times \exp(-\Sigma x_m^2) d\tilde{x}$ reduces to δ_{jk}. Using $I = D^{-1}$ the integral in (14) becomes

$$J_{G_1} = (1/\mu) I_{jl} I_{km} \delta_{lm}$$

and

$$\text{Cov}\, G_1(t,s) = {}^\sigma((1/\mu)(II^T)_{jk}).$$

The matrix product in this is easily shown to be

$$(1/\mu)II^T = \mu \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots \\ 1 & 2 & 2 & 2 & \cdots \\ 1 & 2 & 3 & 3 & \cdots \\ 1 & 2 & 3 & 4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

and thus Cov $G_1 = {}^\sigma(\mu \min(j,k)) = \min(s,t)$. This is the usual form of the covariance of the Weiner measure.[18]

B. Higher order discrete Gaussian measures

The quantities introduced for the Weiner measure can be used to define new functional measures merely by iteration. A Gaussian measure on a space of once differentiable functions whose derivatives are continuous but nowhere differentiable can be defined by using the matrix D of Eq. (1). Define $G_2 = \mu(D^2)^T D^2$. A general component of $D^2\tilde{f}$ will look like $(f_j - 2f_{j-1} + f_{j-2})/\mu^2$, which is a discrete version of the second derivative of a function. As before, the exponential part of the measure can be translated into an action integral form which becomes

$$\exp\left[-\frac{1}{2}\int_0^1 dt \left(\frac{d^2 f}{dt^2}\right)^2\right].$$

The above-mentioned properties of the paths of this measure are easily derived. Again use the variable $x = (\sqrt{\mu})D^2\tilde{f}$. It is known that the integral is concentrated on the set $\tilde{x}\in H$. From the last section the quantities $\tilde{h} = (1/\sqrt{\mu})I\tilde{x}$ are Weiner paths, where as before $I = D^{-1}$. Then the $\tilde{f}$'s are given by $\tilde{f} = I\tilde{h}$ or by component

$$f_j = \sum_{i=1}^{j} h_i\mu\,. \tag{15}$$

Now since the $\tilde{h}$'s represent continuous functions the sum in (15) represents an integral. Taking standard parts of (15) yields the relation

$$\alpha(t) = \int_0^1 ds\,\beta(s), \tag{16}$$

where $\alpha(t)$ and $\beta(t)$ are the standard functions corresponding to $\tilde{f}$ and $\tilde{h}$, respectively, and $t\simeq j\mu$. Obviously from (16) α has a first derivative everywhere on [0,1]. To see that α is independent of which representation of β is used in (15) let g be another representation of β. Then $g_j\simeq h_j$ for all j. Write $g_j - h_j = \eta_j$ and let $\eta = \max(|\eta_j|)$. Then $\Sigma\, h_j\mu = \Sigma\, g_j\mu - \Sigma\,\eta_j\mu \simeq \Sigma\, g_j\mu$ because $|\Sigma\,\eta_j\mu| < \eta \simeq 0$ and therefore the standard part of both sums is the same. Define $\mathscr{W}_2$ as the space of paths of the measure G_2. Then $\alpha\in\mathscr{W}_2$ is a path whose derivative is a Weiner path, i.e., $d\alpha/dt\in\mathscr{W}_1$.

Obviously this process can be extended to any finite order n by defining $G_n = \mu(D^n)^T D^n$ as the bilinear form. This induces a measure $d\lambda_{G_n}$ on the space $\mathscr{C}[0,1]$ and is concentrated on the set $\mathscr{W}_n$. Define $\tilde{h}^j = D^j\tilde{f}$ in analogy with the G_2 case. It follows immediately that $\tilde{h}^{n-1}$ represents a Weiner path. Using the same argument used for Eq. (16) the vector $\tilde{h}^{n-2}$ represents a path in $\mathscr{W}_2$. Continuing this yields the vector $\tilde{h}^{n-m}$, $m > 1$, which represents a path in $\mathscr{W}_m$ and is related to $\tilde{h}^{n-m+1}$ by an integration. Let $\tilde{h}^{n-j}, j = 1,\ldots,n$, be a series of vectors generated by this procedure. Let $\alpha^{n-j} = {}^{\circ}(\tilde{h}^{n-j})$. Then α^k is a function in $\mathscr{W}_{n-k}$ which possesses $k-1$ derivatives and whose $(k-1)$ derivative is a Weiner path. In particular the $\tilde{f}$ represent paths α^0, which possess $n-1$ derivatives.

The covariances of the measures $d\lambda_{G_n}$ can be calculated. It is possible to calculate them as was done for the Weiner measure, but this becomes cumbersome. A better method is derived as follows. Note that the operator $I = D^{-1}$ acts componentwise as an integral [see Eq. (10)]. This will be useful in this derivation. Define

$$\theta(i-j) = \begin{cases} 1 & \text{if } i\leqslant j\,, \\ 0 & \text{if } j > i\,. \end{cases} \tag{17}$$

Then $I_{ij} = \theta(i-j)\mu$ and

$$\frac{1}{\mu} I_{ik}I_{kj} = \sum_{k=1}^{\Omega} \theta(i-k)\theta(k-j)\mu. \tag{18}$$

Since in (18) there exist near standard numbers t, s, p such that $t\simeq i\mu$, $s\simeq k\mu$, and $p\simeq j\mu$ this means

$$\frac{1}{\mu} I_{ik}I_{kj} \simeq \int_0^1 \theta(t-s)\theta(s-p)ds = \theta(t-p)\int_p^t ds. \tag{19}$$

A similar reduction is possible for the transposes

$$\frac{1}{\mu} I^T_{ik}I^T_{kj} = \theta(p-t)\int_t^p ds. \tag{20}$$

Now let $t\simeq i\mu$ and $s\simeq j\mu$. Then

$$\text{Cov}\; G_n(t,s) \simeq (1/\mu^{n-1})[I^n(I^T)^n]_{ij}. \tag{21}$$

Use (19) and (20) to turn all the sums in (21) into integrals to obtain

$$\text{Cov}\; G_n(t,s) = \int_0^1 dy_1 \cdots \int_0^1 dy_n \cdots \int_0^1 dy_{2n-1}\,\theta(t-y_1)\cdots\theta \times (y_{n-1} - y_n)\theta(y_{n+1} - y_n)\cdots\theta(s - y_{2n-1}). \tag{22}$$

The $(n-1)$ integrals from y_1 to y_{n-1} can be reduced using (19) to

$$\int_{y_n}^t dy_{n-1}\int_{y_n}^t dy_{n-2}\cdots\int_{y_2}^t dy_1. \tag{23}$$

This is related to a volume integral over an $n-1$ dimensional rectangle by

$$\int_{y_n}^t dy_{n-1}\cdots\int_{y_2}^t dy_1 = \frac{1}{(n-1)!}\int_{y_n}^t\cdots\int_{y_n}^t dy_{n-1}\cdots dy_1 = \frac{(t-y_n)^{n-1}}{(n-1)!}.$$

In a similar fashion the integrals from y_{n+1} to y_{2n-1} can be reduced to $(s-y_n)^{n-1}/(n-1)!$. Combining these results in the last integral over y_n gives

$$\text{Cov}\; G_n(t,s) = \frac{1}{[(n-1)!]^2}\int_0^{\min(t,s)} dy(t-y)^{n-1}(s-y)^{n-1}. \tag{24}$$

Using the binomial expansion, letting $m = n-1$, and taking $t < s$ this can be written as

$$\text{Cov}\; G_{m+1}(t,s) = \sum_{l=0}^{m} \frac{t^{m+l+1}s^{m-l}}{l!(m-l)!}\sum_{k=0}^{m}\frac{(-1)^{k+l}}{k!(m-k)!(k+l+1)}. \tag{25}$$

For the case $s < t$ simply interchange s with t. Note that when $m = 0$ this is the covariance of the Weiner measure and (25) gives the correct result. The covariances of several low order derivative measures are listed below.

Order	*Bilinear form*	*Function space*	*Covariance* $(t<s)$
2	$\mu(D^2)^TD^2$	$\mathscr{W}_2$	$-\frac{1}{6}t^3 + \frac{1}{2}t^2s$
3	$\mu(D^3)^TD^3$	$\mathscr{W}_3$	$\frac{3}{80}t^5 - \frac{1}{24}t^4s + \frac{1}{12}t^3s^2$
4	$\mu(D^4)^TD^4$	$\mathscr{W}_4$	$-\frac{1}{2520}t^7 + \frac{1}{720}t^6s - \frac{1}{240}t^5s^2 + \frac{1}{144}t^4s^3$

It is possible to extend the discrete form further and define a measure on the space of infinitely differentiable functions on $[0,1] = \mathscr{C}^\infty[0,1]$. Let $\nu\in{}^*\mathbf{N}_\infty$ and $\nu < \Omega$. Define the matrix

$$G_\nu = \mu(D^\nu)^TD^\nu \tag{26}$$

and use this to define a measure $d\lambda_{G_\infty}$ on a space $\mathscr{W}_\infty$. The action integral may be represented formally by $\int_0^1 dt\,(d^\infty f/dt^\infty)^2$. It will be shown that $\mathscr{W}_\infty = \mathscr{C}^\infty[0,1]$ by

showing that the vectors in ${}^*\mathbb{R}^\Omega$ on which the measure is concentrated represent functions which possess any (standard) order derivative. As before let $\tilde{h}^j = D^j\tilde{f}$. Then $\tilde{h}^{\nu-1}$ represents a continuous Weiner path and in general $\tilde{h}^j$ represents a path which possesses $\nu - j - 1$ derivatives and whose $(\nu - j - 1)$th derivative is a Weiner path. To show this merely use the same argument leading to Eq. (11), recursively starting with the "lowest" path $\tilde{h}^{\nu-1}$ which represents the continuous function $\beta_{\nu-1}(t)$, say. Then build up to the general path $(n < \nu)$

$$\beta_n(t) = \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{\nu-n-2}} ds_{\nu-n-1}\beta_{\nu-1}(t). \tag{27}$$

Then β_0 are the paths of $\mathscr{W}_\infty$. It can be seen from (27) that for any $n\in\mathbb{N}$ $d^n B_0/dt^n$ exists everywhere on [0,1].

V. MEASURES IN FOURIER SPACE

A. Fourier measures on Weiner paths

On a simple level the Weiner measure can be thought of as measure based on an action integral $\int_0^1 dt\,(d\alpha/dt)^2$. If one naively writes the derivative as a Fourier series $d\alpha/dt = \Sigma\, b_n e^{2\pi i n t}$, then the action becomes $\Sigma\, b_n^2$ and there is the possibility of doing the functional integration by integrating over the coefficients b_n, since by an integration in t these determine $\alpha(t)$. This has been an approach used in the past for a number of simple functional integrals[3,5-9]. It will be seen below that the conversion of the Weiner measure to something of this nature is possible, but that the proof that the same functions are associated with each measure is not straightforward.

Before proceeding some nonstandard versions of Fourier series formula will be presented. These results come from the very good article on a nonstandard approach to Fourier series by Luxemburg.[19] The basic approach is to use the results from harmonic analysis on finite groups.

For $m\in\mathbb{N}$ consider the additive group $T(m) = \{0,1/m,2/m,\ldots,(m-1)/m \bmod 1\}$ of $m+1$ elements. That is, if $0\leqslant k < m$ and $n\in\mathbb{N}$ the points k/m and $(k + nm)m$ are identified. Then define an algebra of functions L_m on $T(m)$, $L_m \ni s: T(m)\to\mathbb{C}$, where $\mathbb{C}$ are the complex numbers. The muliplication $\circ$ on L_m is defined by the convolution sum

$$(s\circ r)(z) = \sum_{k=0}^{m} s(z - z_k)r(z_k)$$

for $z\in T(m) = \{z_k\}$ and $s,r\in L_m$. In the above sense the elements of L_m act like periodic functions on the set of rational numbers $\{0,1,m,\ldots\}$ with period = 1. It is possible to define a discrete Fourier representation for the elements of L_m in terms of a sum over $m+1$ terms. The relations between an element s of L_m and its Fourier coefficients a_n are[19]

$$a_n = \frac{1}{m+1}\sum_{k=0}^{m} s(z_k)e^{-2\pi i n k/(m+1)}, \tag{28}$$

$$s(z_j) = \sum_{n=0}^{m} a_n e^{2\pi i n j/(m+1)}. \tag{29}$$

Note that Eq. (29) is exact, that is, the Fourier representation of s is not just an approximation to each $s(z_j)$. If one thinks of $\{a_n\}$ and $\{s(z_j)\}$ as $m+1$ dimensional vectors then, apart from the factor of $1/(m+1)$ in Eq. (28), Eqs. (28) and (29) appear as unitary transformations using the matrix $M_{nk} = e^{2\pi i n k/(m+1)}$. One other relation between the members of L_m and their Fourier coefficients which is of use here is the equality[19]

$$\frac{1}{m+1}\sum_{k=0}^{m} s^2(z_k) = \sum_{n=0}^{m} a_n^2. \tag{30}$$

To extend these results to the usual Fourier series of continuous functions simply put $m + 1 = \omega\in{}^*\mathbb{N}_\infty$ and transfer all results to ${}^*U$. Then Eq. (28) is related to the usual integral and Eq. (30) becomes the form of Parseval's formula familiar to most physicists. The point is that using the formal properties of (28), (29), and (30) one can work with continuous functions and other entities as discrete quantities, vectors, in a Euclidean space ${}^*\mathbb{R}^\omega$.

It can be seen that Eq. (30) is especially well suited for application to the Weiner measure with $s(z_j) = g_j = (f_j - f_{j-1})/\mu$, which would yield the simple exponent $\Sigma\, b_n^2$ of Fourier coefficients of g_j. This will turn out to be true in general, but requires some study in detail to exhibit the correct subspace of Fourier coefficients over which to integrate. Later a nontrivial theorem will relate the f_j to the direct integration of the Fourier series of g_j. First the properties of the Fourier series of f_j and g_j are derived.

To do Fourier analysis $\tilde{f}$ and $\tilde{g}$, vectors in ${}^*\mathbb{R}^\Omega$ from before, must first be changed to cyclic functions. One way to do this is to extend the domain from [0,1] to $(-1,1]$ in a systematic way. The method chosen here is consistent with the requirement that $f_0 = 0$. Define $f_{-j} = -f_j$ for $j = 0,1,\ldots,\Omega$ and $f_{j+2\Omega} = f_j$. Let $\omega = 2\Omega$. Then the new vector space becomes ${}^*\mathbb{R}^\omega$ and $\tilde{f} = (f_{-\Omega+1},\ldots f_0,\ldots f_\Omega)$ with $f_{-\Omega} = +f_\Omega$, etc. Define g, the "derivative", by the above formula, where $\mu = 2/\omega$. The $\tilde{g}$'s are almost symmetric, $g_{-j} = g_{j+1}$, and are cyclic mod ω.

The exponent in the Weiner measure becomes $(\mu/4)\,\Sigma_{-\Omega+1}^{\Omega} g_j^2$ and in terms of the Fourier coefficients b_n of $\tilde{g}$ this transforms to $\frac{1}{2}\,\Sigma_{-\Omega+1}^{\Omega} b_n^2$. But the b_n are complex in general and $\tilde{g}$ is a real vector, hence not all b_n's are independent, just as not all of the components of $\tilde{g}\in{}^*\mathbb{R}$ are independent. To see which b_n will be integration variables these relations need to be explored. These will also be of use in showing the relation between f and the integral of the Fourier series of $\tilde{g}$.

The notation used in the following will be to write $\overline{\Sigma}_j$ for $\Sigma_{j=-\Omega+1}^{\Omega}$ and $\underline{\Sigma}_j$ for $\Sigma_{j=0}^{\Omega}$. When the index of summation is clear (e.g., a repeated index) it will be omitted from the summation symbol.

The Fourier coefficients a_n of the f's determined by using Eq. (28) behave almost like Fourier coefficients of antisymmetric functions. Using the definition of a cyclic $\tilde{f}\in{}^*\mathbb{R}^\omega$ the requirement that the a_n represent a real function $(a_n{}^* = a_{-n})$, and being careful with indices leads to the following relations for a_n:

$$\begin{aligned} &\mathrm{Re}(a_0) = f_\Omega/\omega, \quad \mathrm{Im}(a_0) = 0, \quad \mathrm{Im}(a_\Omega) = 0, \\ &\mathrm{Re}(a_n) = (-1)^n a_0, \quad \mathrm{Im}(a_{-n}) = \mathrm{Im}(a_n), \\ &n = -\Omega + 1, \ldots, \Omega - 1. \end{aligned} \tag{31}$$

There are initially 4Ω coefficients, but there are 3Ω relations leaving Ω independent coefficients, the same number as there are independent components in $\tilde{f}$. Note that the behavior is almost like an antisymmetric function in that $\mathrm{Re}(a_n) \simeq 0$ for all near standard $\tilde{f}$.

The Fourier coefficients b_n of g_i can likewise be reduced by similiar methods using the following obtained relations:

$$b_0 = 0,$$

$$b_n^* = b_{-n} = e^{2i\pi n/\omega} b_n = \frac{2(-1)^n}{\Omega} [e^{2\pi i n/\omega} - 1] f_{\Omega - 1}. \tag{32}$$

The factor of $e^{2\pi i n/\omega}$ comes from the relation $g_{-j} = g_{j+1}$. Again there are only Ω independent coefficients.

A relationship between a_n and b_n can be obtained by using $b_n = (1/\omega) \overline{\Sigma} g_j e^{-2\pi i j n/\omega} = \frac{1}{2} \overline{\Sigma} (f_j - f_{j-1}) e^{-2\pi i j n/\omega}$ and rearranging terms. This gives

$$b_n = i\pi n e^{\pi i n/\omega} \left(\frac{\sin n\pi/\omega}{n\pi/\omega} \right) a_n. \tag{33}$$

This is almost what one would expect for a derivative relation. In fact for $n/\Omega \simeq 0$, $b_n \simeq i\pi n\, a_n$, but this breaks down for larger n. However, Eq. (33) will be useful later in relating a_n to an auxillary function which approximates $\tilde{f}$.

In order to transform the Weiner measure $d\lambda_{G_1}$ into a measure on Fourier space first write

$$D = \frac{1}{\mu} \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & & -1 \\ -1 & 1 & 0 & 0 & \cdots & & 0 \\ 0 & -1 & 1 & 0 & \cdots & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & & -1 & 1 \end{pmatrix}$$

as a matrix on ${}^*\mathbb{R}^\omega$ or ${}^*\mathbb{C}^\omega$ (see the following comment) and take $G_1 = \mu D^T D$. Define the measure as

$$d\lambda_{G_1} = \frac{e^{-(1/4)(\tilde{f}, G_1 \tilde{f})}}{\sqrt{\det\, G_1}} \prod_{j=-\Omega+1}^{\Omega} \frac{df_j}{2\sqrt{\pi}}. \tag{34}$$

In order to facilitate the transformation to a general complex Fourier representation $\tilde{f}$ will be treated as a vector in ${}^*\mathbb{C}^\omega$ and after the transformation to a Fourier space the proper restrictions on the coefficients will be imposed. With this in mind Eq. (34) transforms to

$$d\lambda_{G_1} = e^{-(1/4)\mu \overline{\Sigma} g_j^2} d\tilde{g}. \tag{35}$$

To transform to the Fourier coefficients use the fact that $\{g_j\}$ and $\{b_n\}$ are related by a nearly unitary transformation with the extra ω factor being canceled by the μ which appears in (35). This yields a Jacobian which is just 1 and a transformed measure

$$d\lambda_{G_1} = e^{-(1/2)\overline{\Sigma} b_n^2} d\tilde{b}. \tag{36}$$

Here $d\tilde{b} = \Pi_{n=-\Omega+1}^{\Omega} db_n^r\, db_n^i$, where $b_n^r = \mathrm{Re}(b_n)$ and $b_n^i = \mathrm{Im}(b_n)$. Now restrictions can be placed on $\tilde{f}$ to make it conform to the previous case of a real antisymmetric function on $(-1,1]$. Then $\tilde{g}$ has the appropriate restrictions and Eq. (36) can remain the same, but the domains of integration for the b_n will be constrained by Eq. (32).

Equation (36) shows that the Fourier representation is a natural one for the Weiner measure. This is because it represents an inner product on a certain function space which is best represented by Fourier series.[20] Equation (36) can also be changed to an integration over a_n, but this is not as clean as integration over b_n. Also, the possibility of using the Fourier representation as a continuous representation instead of a discrete one presents itself.

This last possibility is an interesting one since it would be particularly nice to be able to use the transferred definition of derivative and integral to operate on the simple functions $e^{2\pi i n t}$, $t \in (-1,1]$. This would allow the direct calculation of $\tilde{f}$ and $\tilde{g}$ in a *-continuous fashion and avoid going through a "time-slicing" process. However, the question remains whether one can obtain the f's by direct integration of the Fourier series for the g's. That is, can it be shown that

$$f_j \simeq \int_0^t dt\, \overline{\sum}\, b_n\, e^{2\pi i n t} \tag{37}$$

whenever $t \simeq j\mu$? It is this last relation which would allow a direct link from the integration variables b_n to the function values f_j. But because of Eq. (33) it is not clear that this will work. The extra factor $e^{\pi i n/\omega}$ (sin $n\pi/\omega$), which will not appear in Eq. (37), casts the doubt.

In order to show that Eq. (37) is valid it will be necessary to introduce an auxillary function $F(t)$ for each $\tilde{f}$ in the manner of Luxemburg.[19] Define

$$F(t) = f_j \text{ when } -1/\omega < t - j\mu \leqslant 1/\omega. \tag{38}$$

Thus, $F(t)$ is a step function infinitesimally approximating $\tilde{f}$. When $\tilde{f}$ represents a Weiner function then $f_j \simeq F(t)$ whenever $t \simeq j\mu$.

The necessity of using $F(t)$ becomes more apparent when one calculates the Fourier coefficients of $F(t)$ using the *transferred* Fourier series results from U. That is, calculate[19]

$$F_n = \frac{1}{2} \int_{-1}^{1} F(t) e^{-2\pi i n t} = a_n \frac{(\sin n\pi/\omega)}{(n\pi/\omega)}; \tag{39}$$

this shows that it is the true Fourier coefficients of $F(t)$ which have a relation to the Fourier coefficients of $\tilde{g}$ more like the relation expected between a function and its derivative, namely

$$b_n = in\pi e^{-\pi i n/\omega} F_n. \tag{40}$$

The factor $e^{-\pi i n/\omega}$ will only shift the sum by an infinitesimal amount which does not matter for continuous functions. Thus, integrating $\overline{\Sigma} b_n e^{2\pi i n t}$ according to Eq. (37) will lead to an approximation to $F(t)$. Since $F(t) \simeq f_j$ when $j\mu \simeq t$ it remains to be shown that the sum $\overline{\Sigma} F_n e^{2\pi i n t}$ is infinitesimally close to $F(t)$ whenever $F(t)$ is an auxillary function for a vector $\tilde{f}$ which represents a Weiner function.

In order to show this it will be necessary to use the version of the Riemann localization theorem developed by Hille and Klein:[21]

Theorem: If $S_N(t) = \Sigma_{n=-N}^{N} F_n e^{2\pi i n t}$, where F_n are the Fourier coefficients of a function $F(t)$ on $(-1,1]$, then

$$R_{N\delta} \equiv \left| S_N(t) - \frac{1}{\pi}\int_{-\delta}^{\delta} F(t+x)\frac{\sin Nx}{x}dx \right| \leqslant \frac{K}{\delta}(\|F\| + 1)M_1(1/N),$$

where

$$\|F\| \equiv \frac{1}{2}\int_{-1}^{1} |F(t)|dt$$

and

$$M_1(1/N) \equiv \text{Integral Modulus of Continuity} \equiv \max_{t<1/N}\int_{-1}^{1} |F(x+t) - F(x)|dx$$

and K is a finite constant. This shows how the partial Fourier sum is related locally to the function.

Now transfer this theorem to $^*U$ and apply it to the auxillary function $F(t)$. Set $N = \omega$ (or $\omega - 1$ if a symmetric partial sum is desired as in the theorem since the elimination of one infinitesimal term from the partial sum will not matter). It can be seen that $\|F\| < F_{\max}$, which is near standard. Since $|F(t+x) - F(x)| \leqslant$[the largest step when $t \leqslant 1/\omega$] and since $\tilde{f}$ is locally Holder continuous, it follows that for some standard c and $\beta \in (0,1)$ $M_1(1/\omega) < 2\pi c(1/\omega)^{\beta}$. The question is, can δ be chosen so that $R_{\omega\delta} \simeq 0$ for the Hille–Klein theorem? The answer is yes.

Lemma 1: There exists an infinitesimal δ such that $R_{\omega\delta} \simeq 0$ for $F(t)$, an auxillary function to a Weiner vector $\tilde{f}$ locally Holder continuous of order β.

Proof: From before, $R_{\omega\delta} \leqslant (2\pi K/\delta)(F_{\max} + 1)c(1/\omega)^{\beta}$, where $\beta \in (0,1)$ and is standard. With the exception of $(1/\omega)^{\beta}$ and δ all terms are near standard so that all that matters is the factor $(1/\delta)(1/\omega)^{\beta}$. Pick $\gamma \in (0,1)$ and standard such that $\gamma < \beta$. Then set $\delta = (1/\omega)^{\gamma}$ so that $(1/\delta)(1/\omega)^{\beta} = (1/\omega)^{\beta-\gamma} \simeq 0$ and $R_{\omega\delta} \simeq 0$.■

This means the partial sum of Fourier coefficients of $F(t)$, $S_{\omega}(t)$ approximates $\int_{-\delta}^{\delta} F(t)[\sin(\omega t)/t\,]dt$ infinitesimally closely. The main theorem now goes as follows.

Theorem 5: If $\tilde{f}$ is a Weiner vector in $^*\mathbb{R}^{\Omega}$ then the partial sum $S_{\omega}(t)$ of the auxillary function to $\tilde{f}$, $F(t)$, is related to $\tilde{f}$ by $S_{\omega}(t) \simeq f_j$ whenever $j\mu \simeq t$, where $\mu = 2/\omega$.

Proof: If it can be shown that $S_{\omega}(t) \simeq F(t)$, then from the definition of $F(t)$ it automatically follows that $S_{\omega}(t) \simeq f_j$ when $j\mu \simeq t$. The proof comes down to showing that $(1/\pi)\int_{-\delta}^{\delta} dx\, F(t+x)(\sin \omega x)/x \simeq F(t)$ for an appropriate δ and applying the previous lemma. To show this last result pick $\gamma \in (0,1)$ and $\gamma < \beta$, where β is the order of $\tilde{f}$ Holder continuity as before. Then set $\delta = (1/\omega)^{\gamma}$. Write $F(t+x) = F(t) + \alpha_t(x)$ on $x \in (t-\delta, t+\delta)$. From the local Holder continuity

$$|\alpha_t(x)| < c\delta^{\beta} = c(1/\omega)^{\beta\gamma} \simeq 0.$$

Then

$$\frac{1}{\pi}\int_{-\delta}^{\delta} F(t+x)\frac{\sin \omega x}{x}dx = \frac{F(t)}{\pi}\int_{-\delta}^{\delta} \alpha_t(x)\frac{\sin \omega x}{x}dx + \frac{1}{\pi}\int_{-\delta}^{\delta} \alpha_t(x)\frac{\sin \omega x}{x}dx.$$

The first integral becomes

$$\frac{F(t)}{\pi}\int_{-\delta\omega}^{\delta\omega} \frac{\sin \zeta}{\zeta}d\zeta \simeq \frac{F(t)}{\pi}\int_{-\infty}^{\infty} \frac{\sin \zeta}{\zeta}d\zeta = F(t)$$

since $\delta\omega$ is infinite. The second integral is more subtle:

$$\left|\frac{1}{\pi}\int_{-\delta\omega}^{\delta\omega} \alpha_t(\zeta/\omega)\frac{\sin \zeta}{\zeta}d\zeta\right| \leqslant \frac{1}{\pi}\int_{-\delta\omega}^{\delta\omega} \left|\alpha_t(\zeta/\omega)\frac{\sin \zeta}{\zeta}\right|d\zeta < \frac{c\delta^{\beta}}{\pi}\int_{-\delta\omega}^{\delta\omega} \left|\frac{\sin \zeta}{\zeta}\right|d\zeta \simeq 4c(1/\omega)^{\gamma\beta}\ln(\omega^{1-\gamma}),$$

where the fact that in U $\int_{-N}^{N} |\sin \zeta/\zeta|\,d\zeta \to 4\pi \ln N$ asymptotically[22] has been transferred to $^*U$. But $(1/\omega)^{\gamma\beta}\ln(\omega^{1-\gamma}) \simeq 0$ for the present choices of γ and β. Now, because $\gamma < \beta$ and γ is standard and $\gamma \in (0,1)$ use the previous lemma to show that $S_{\omega}(t) \simeq f_j$ whenever $j\mu \simeq t$.■

It should be remarked here that contained in these results is the following standard theorem:

Theorem (standard): If f is a Holder continuous function on an interval $(-1,1]$ and $F^N(t)$ is a step function approximation to $f(t)$ with N steps,

$F^N(t) = f((2j/N) - 1)$ when
$t - [(2j/N) - 1] \in (-1/2N, 1/2N]$
for $j = 1,2,\ldots,N$.

Then the Fourier coefficients of $F^N(t)$, F_n^N are such that

$\lim_{N\to\infty} S_N(t) = f(t)$ pointwise, where
$S_N(t) = \sum_{n=-N}^{N} F_n^N e^{2\pi i n t}$.

It may be possible to weaken the assumptions of this theorem by introducing a standard definition of locally Holder continuous.

These results complete the circle of correspondence of $\tilde{f}, g$ and their Fourier coefficients $\tilde{a}$ and $\tilde{b}$. Not only is it assured that the Fourier coefficients b_n are related via Eqs. (28) and (29) to a Holder continuous function, but that the function can be approximated infinitesimally well by merely integrating the Fourier series in b_n via Eq. (37) and using the theorem relating $\Sigma F_n e^{\pi i n t}$ and $F(t)$. This means

$$\int_0^t \overline{\sum} b_n e^{\pi i n t}dt \simeq F(t - 1/\omega) - F(1/\omega) \simeq F(t) \simeq f_j \text{ for } j\mu \simeq t.$$

The internal *-continuous functions $g(t) = \overline{\Sigma}\, b_n e^{\pi i n t}$ and $f(t) = \overline{\Sigma}\, a_n e^{\pi i n t}$ can be defined and the usual calculus operations transferred to $^*U$ can now be applied to them to generate new internal functions, some of which are near standard, as well be seen.

B. Fourier measures in other function spaces

The results of the last section show that one could alternately define the Weiner measure by an integration over the variables a_n, the coefficients of f. In this case one would use

the matrix

$$K=\begin{pmatrix}1 & & & & \\ & 2 & & & \\ & & 3 & & 0 \\ & 0 & & \ddots & \\ & & & & \Omega\end{pmatrix} \tag{41}$$

on ${}^*\mathbb{R}^{\Omega}$ to define the Gaussian generated by $G=K^2$. The functions $\overline{\Sigma}(b_n/n)e^{\pi int}$ (with the appropriate conditions on the b_n) are infinitesimally close to the usual Weiner functions in $\mathscr{W}_1$.

In a similar way one can define the Gaussian generated by $G=K^4$. This yields Fourier coefficients like b_n/n^2, where the b_n take on the same values as the b_n in the previous section. The Fourier series generated by these coefficients yield paths which can be shown to be infinitesimally close to a double integral of the derivative paths $g(t)$ of part A (which is what one would expect intuitively):

$$\int_0^s dt\int_0^t dx\, g(x)\simeq \overline{\sum}\frac{b_n}{i\pi n}\int_0^s e^{\pi int}dt-\int_0^s F(1/\omega)dt$$
$$\simeq \overline{\sum}\frac{b_n}{-\pi^2n^2}e^{\pi ins},$$

since the second integral is infinitesimal. Thus, by a redefinition of b_n to absorb the $-1/\pi^2$ the path $h(s)=\overline{\Sigma}(b_n/n^2)e^{\pi ins}$ is infinitesimally close to paths in $\mathscr{W}_2$.

The generalization of this using $G=K^{2n}$ for paths in $\mathscr{W}_n$ is obvious, but the use of Fourier series allows a wider generalization. Since K is diagonal, the matrix K^{α}, where α is standard and $\alpha\in(0,1)$, is well defined. Then one can define the Gaussian generator $G=K^{2(n+\alpha)}$. This can be said to generate the space of paths $\mathscr{W}_{n+\alpha}$. One might think of this in terms of paths with a fractional derivative. A further step is simply to use a monotonically increasing function ρ and define $K_{nm}=\delta_{nm}\rho(n)$ and $G=K^2$. A number of interesting mathematical questions come up here regarding the form of ρ required to yield quantities which are continuous functions, functions with various derivatives, etc. Also of interest is the "minimal" ρ which still yields meaningful results in terms of the integral still being concentrated on a space of standard functions. It would appear that many of the forms of allowable ρ's are determinable by what is presently known about trigonometric series,[22] although this is by no means clear.

VI. CONCLUDING REMARKS

It is possible to "go the other way" in deriving new Weiner spaces $\mathscr{W}_{\alpha}$ from $\mathscr{W}_1$ by using the inverse of the matrix D given by Eq. (10), I. Define $G_{-1}=\mu I^TI$. The properties of the vectors which emerge from this definition are easily derived. Let $\tilde{h}\in(1/\mu)I^{-1}(H)$ and write $h_j=(g_j-g_{j-1})/\mu$, where $\tilde{g}$ is defined as in Eq. (10). Now examine the quantity $\Sigma_{\Omega}^1h_j\alpha_j\mu$ for a standard continuous function α on [0,1] for which $\alpha(1)=0$. This becomes $-\Sigma_{\Omega}^1(\alpha_{j+1}-\alpha_j)g_j$. If α has a first derivative everywhere in [0,1], then $\alpha_{j+1}-\alpha_j=\beta_j\mu$, where β_i is near standard. Then $|\Sigma_{\Omega}^1h_j\alpha_j\mu|=|-\Sigma_1^{\Omega-1}\beta_jg_j\mu|<\max(\beta_j)|\Sigma_1^{\Omega-1}g_j\mu|$, which is near standard (see the section on the Weiner measure). Thus $\forall\tilde{h}\in(1/\mu)I^{-1}(H)$ the sum $\Sigma h_j\alpha_j\mu$ is near standard so long as α has the above-stated properties. Each h associates a number with each α in a linear fashion and so the h's represent distributions (linear functionals) on a subset B of $\mathscr{C}^1[0,1]$, which consists of the above α functions. Equivalently, the quantities $\gamma_i=h_i\mu$ can be thought of as a measure on B. Notationally, it can be written that $\tilde{h}$ represents a quantity in $\mathscr{W}_{-1}$ which is at least a subset of the set of distributions on B.

The generalization to the case $\mathscr{W}_{\pm n}$ is immediate and corresponds to $G_{\pm n}=\mu(D^{\pm n})^TD^{\pm n}$. The case of $n=0$ simply yields the vectors g of Sec. IV as the representatives which are at least distributions on the space of continuous functions and are known to represent white noise.[18] A further step is to use a Fourier series to describe the paths. Then using K as defined by Eq. (41) define $G_{\alpha}=K^{2\alpha}$ and presume that the entities which emerge are in the space $\mathscr{W}_{\alpha}$. Or, using a function ρ as before, generate $\mathscr{W}_{\rho(\alpha)}$. Whether this is possible for all $\alpha\in\mathbb{R}$ and all $\rho\neq0$ is not clear.

Another generalization follows from an attempt to get away from the simple method of vectorization and come closer to a measure theoretic construction. One can define a hyperfinite partition $\mathscr{P}_{\omega}$ on [0,1] following the ideas of Loeb.[23] Consider a finite partition of [0,1], say $\mathscr{P}_n=\{A_1,\ldots,A_n\}$, where $\cup_{i=1}^nA_i=[0,1]$ and $A_i\cap A_j=\varnothing$ if $i\neq j$. Then it is known that for each finite partition $\mathscr{P}_n$ (or for a finite set of $\mathscr{P}_{n_i}$'s) there is a refining partition $\mathscr{P}_m$ such that each member of $\mathscr{P}_m$ is a subset of one set of $\mathscr{P}_n$ (or one set of each $\mathscr{P}_{n_i}$). This is a concurrent relation and so in ${}^*U\,\exists$ a hyperfinite partition $\mathscr{P}_{\omega}$ $\omega\in{}^*\mathbb{R}_{\infty}$ which refines all standard finite partitions of [0,1]. Let $A_i\in\mathscr{P}_{\omega}$ for $i=1,\ldots,\omega$ and let μ be a measure on [0,1] which can now be transferred to ${}^*U$. One can assign the numbers $\mu_i=\mu(A_i)$ to each member of $\mathscr{P}_{\omega}$. Loeb[23] has shown that by choosing the values of an integrable function f by the following prescription, $f_i=f(x)$ for a fixed $x\in A_i$ $\forall i=1,\ldots,\omega$, then the integral of $f(x)$ is given by $\int_0^1 d\mu f(x)\simeq\Sigma_1^{\omega}f_i\mu_i$. This resembles the vectorization methods, but for certain A_i one could have $\mu_i=0$ (e.g., if A_i is one point and μ is the Lebesgue measure). If μ is a regular Borel measure then Shilov *et al.*[24] have shown that the derivative of a set function ψ on an algebra of subsets of [0,1] can be defined. Naively, this would be $\psi'(x)=\lim_{\mu(A)\to0}\psi(A)/\mu(A)$ for all A of nonzero measure such that $x\in A$. This concept could be merged with that of Loeb and more general measure theoretic derivatives and integrals could be introduced into the definition of a Gaussian generating matrix G. For example use the set $M=\{i|\mu_i\neq0\}$ of *-finite cardinality Ω, say, to define a Gaussian $\exp[-\frac{1}{2}\Sigma(\psi(A_i)/\mu_i)^2\mu_i]$ to generate new spaces of internal quantities $\psi(A_i)$ or $\Sigma\psi(A_i)\mu_i$, etc., which also depend on the measure μ.

[1] R. P. Feynman, Rev. Mod. Phys. **20**, 367–387 (1948).
[2] I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
[3] B. Davison, Proc. R. Soc. London, Ser. A **225**, 252 (1954).

[4]J. B. Keller and D. W. McLaughlin, Amer. Math. Monthly **82**, 451 (1975).
[5]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[6]S. G. Brush, Rev. Mod. Phys. **33**, 79 (1961).
[7]W. K. Burton and A. H. DeBorde, Nuovo Cimento **2**, 197 (1955).
[8]H. Davies, Proc. Cambridge Philos. Soc. **53**, pt. 1, 199 (1957).
[9]S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977).
[10]A. Robinson, Proc. R. Acad. Amsterdam, Ser. A **64**, 432–440 (1961).
[11]K. D. Stroyan and W. A. J. Luxemburg, *Introduction to the Theory of Infinitesimals* (Academic, New York, 1976).
[12]A. Robinson, *Non-Standard Analysis* (North-Holland/Elsevier, New York, 1974).
[13]M. Davis, *Applied Non-Standard Analysis* (Wiley, New York, 1977).
[14]A. Voros, J. Math. Phys. **14**, 292 (1973).
[15]K. D. Stroyan, "Infinitesimal Analysis of Curves and Surfaces," in *Handbook of Mathetical Logic*, edited by J. Barwise (North-Holland/Elsevier, New York, 1977).
[16]P. J. Keleman and A. Robinson, J. Math. Phys. **13**, 1875 (1972).
[17]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), 4th ed.
[18]N. Wiener, A. Seigel, B. Rankin, and W. T. Martin, *Differential Space, Quantum Systems and Prediction* (MIT, Cambridge, Mass., 1966).
[19]W. A. J. Luxemberg, "A Non-Standard Analysis Approach to Fourier Analysis," in *Contributions to Non-Standard Analysis*, edited by W. A. J. Luxemberg and A. Robinson (North-Holland, Amsterdam, 1972).
[20]A. Truman, J. Math. Phys. **17**, 1852 (1976).
[21]E. Hille and G. Klein, Duke Math. J. **21**, 587 (1954).
[22]N. K. Bary, *A Treatise on Trigonometric Series* (Pergammon, New York, 1964), Vols. I and II.
[23]P. A. Loeb, "An Introduction to Non-Standard Analysis and Hyperfinite Probability Theory," in *Probabilistic Analysis and Related Topics*, edited by Bharucha-Reid (Academic, New York, 1979), Vol. 2.
[24]G. E. Shilov and B. L. Gurevich, *Integral, Measure and Derivative, a Unified Approach* (Prentice-Hall, Englewood Cliffs, N. J., 1966).

Asymptotic behavior of the nonlinear diffusion equation $n_t = (n^{-1}n_x)_x$

James G. Berryman [a)]
University of California, Lawrence Livermore National Laboratory, Livermore, California 94550

Charles J. Holland
Office of Naval Research, Code 411, Arlington, Virginia 22217

(Received 10 November 1981; accepted for publication 23 December 1981)

The asymptotic behavior of the equation $n_t = (\ln n)_{xx}$ is studied on the finite interval $0 \leqslant x \leqslant 1$ with the boundary conditions $n(0,t) = n(1,t) = n_0$ and initial data $n(x,0) \geqslant n_0$. We prove that asymptotically $\ln[n(x,t)/n_0] \to A \exp(-\pi^2 t/n_0) 2^{1/2} \sin \pi x$ and also provide rigorous upper and lower bounds on the asymptotic amplitude A in terms of integrals of nonlinear functions of the initial data. The rigorous bounds are compared to values of A obtained from computer experiments. The lower bound $L = (2^{3/2}/\pi)\exp[\mathrm{li}(1+Q) - \gamma]$, where li is the logarithmic integral, γ is Euler's constant, and $Q = (\pi/2)\int[n(x,0)/n_0 - 1]\sin \pi x \, dx$, is found to be the best known estimate of A.

PACS numbers: 02.30.Jr

I. INTRODUCTION

A nonlinear diffusion process of the form

$$\frac{\partial n}{\partial t} = D \frac{\partial}{\partial x}\left(\frac{1}{n}\frac{\partial n}{\partial x}\right), \tag{1}$$

where $n(x,t)$ is the density and x,t are the space, time coordinates, has been predicted for cross-field convective diffusion of plasma including mirror effects[1] and has been observed during experiments using the Wisconsin toroidal octupole plasma containment device.[2–4] The same equation describes the expansion of a thermalized electron cloud.[5] The equation also arises in studies[6,7] of the central limit approximation to Carleman's model[8] of the Boltzmann equation. In this latter application, Eq. (1) serves to control a kind of nonlinear Brownian motion arising as the central limit approximation to Carleman's equation.

We wish to study the behavior of (1) on the finite spatial interval $0 \leqslant x \leqslant 1$ so that (1) will serve as an idealized model of the situation found in the experiments at Wisconsin.[2–4,9] Based upon physical considerations,[10] we will require finite values of n at the boundaries. For simplicity, we choose to set $n(0,t) = n(1,t) = n_0$ at the boundaries. The constant n_0 is a small density to be thought of as a background value. This background value must be sufficiently small so that the initial density profile satisfies $n(x,0) \geqslant n_0$. Having made these stipulations, we note that, since

$$\frac{d}{dt}\int_0^1 n(x,t)dx = D\left(\frac{1}{n}\frac{\partial n}{\partial x}\right)\Big|_0^1, \tag{2}$$

the flux [i.e., the right-hand side of (2)] will be finite and the mathematical problem will be well posed. Furthermore, it proves useful to transform (1) into

$$(e^m)_t = m_{xx} \quad \text{on } 0 \leqslant x \leqslant 1, \tag{3}$$

with $m(0,t) = m(1,t) = 0$, where $m(x,t) = \ln(n/n_0)$, m is nonnegative, and the new time scale differs from the old by a factor of n_0/D.

First, the derivation of rigorous inequalities on various integrals of functions of m and its derivatives is presented. Next these inequalities are used to prove that the solution $m(x,t)$ asymptotically approaches

$$m(x,t) \to A \exp(-\pi^2 t)\phi_1(x), \tag{4}$$

where the

$$\phi_k(x) = 2^{1/2} \sin k\pi x \tag{5}$$

are the normalized eigenfunctions vanishing at the boundaries satisfying

$$\phi_{k,xx} + k^2\pi^2\phi_k = 0 \tag{6}$$

and the asymptotic amplitude A is a constant dependent on the initial data. Then further inequalities are derived to place bounds on A and, finally, these estimates are compared to the results of computer experiments.

II. INEQUALITIES

The simplest integral bound for problem (3) is derived in the following sequence:

$$\begin{aligned} -\frac{d}{dt}\int_0^1 (e^m - 1)\phi_1 dx &= -\int_0^1 m_{xx}\phi_1 \, dx \\ &= -\int_0^1 m\phi_{1xx}\, dx = \pi^2 \int_0^1 m\phi_1 \, dx \\ &\leqslant \pi^2 \int_0^1 (e^m - 1)\phi_1 \, dx. \end{aligned} \tag{7}$$

The steps in (7) include using (3), twice integrating by parts, using (6) to substitute for ϕ_{1xx}, and finally using the well-known inequality $m \leqslant (e^m - 1)$. With the definition

$$Q(t) = \frac{\pi}{2^{3/2}}\int_0^1 [e^{m(x,t)} - 1]\phi_1(x)\, dx, \tag{8}$$

inequality (7) can be integrated to yield

$$Q(t) \geqslant Q(0)e^{-\pi^2 t} \equiv Q_0 e^{-\pi^2 t}. \tag{9}$$

Thus, we have found an elementary lower bound for $Q(t)$.

To obtain an upper bound on some relevant integral requires a more sophisticated argument. First note that, for any differentiable function $f(x)$ such that $f(0) = f(1) = 0$,

[a)]Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract No. W-7410-ENG-48.

 0022-2488/82/060983-05$02.50

$$f(x) = \int_0^x f_x\, dx = -\int_x^1 f_x\, dx. \tag{10}$$

The Cauchy–Schwarz inequality then implies

$$f^2(x) \leqslant x \int_0^x f_x^2\, dx \quad \text{and} \quad f^2(x) \leqslant (1-x)\int_x^1 f_x^2\, dx. \tag{11}$$

From (11), it follows that

$$\left(\frac{1}{x} + \frac{1}{1-x}\right) f^2 = \frac{f^2}{x(1-x)} \leqslant \int_0^1 f_x^2\, dx. \tag{12}$$

In particular,

$$m^2(x,t) \leqslant x(1-x)\int_0^1 m_x^2 dx \leqslant \frac{1}{4}\int_0^1 m_x^2\, dx \equiv z^2(t). \tag{13}$$

[Note that the one-dimensional character of (3) is essential here because no such bound on sup m in terms of an integral of the squared gradient of m is available in higher dimensions.]

Next recall from standard variational arguments that

$$\pi^2 \leqslant \frac{\int m_x^2 dx}{\int m^2 dx}. \tag{14}$$

From (13), $\exp(m-z) \leqslant 1$ so

$$\int m^2 dx \geqslant e^{-z}\int m^2 e^m dx \tag{15}$$

and therefore

$$\pi^2 \leqslant e^z \frac{\int m_x^2 dx}{\int m^2 e^m dx}. \tag{16}$$

Then again use Cauchy–Schwarz at the end of the following sequence of manipulations:

$$\begin{aligned}\left(\int m_x^2 dx\right)^2 &= \left(-\int m m_{xx} dx\right)^2 \\ &= \left[\int (m e^{m/2})(m_{xx} e^{-m/2})\, dx\right]^2 \\ &\leqslant \int m^2 e^m dx \int m_{xx}^2 e^{-m} dx.\end{aligned} \tag{17}$$

Combining (16) and (17) yields

$$\pi^2 \leqslant e^z \frac{\int m_{xx}^2 e^{-m} dx}{\int m_x^2 dx} = -\frac{e^z}{z}\frac{dz}{dt}, \tag{18}$$

where we have used the fact that

$$\frac{d}{dt} z(t) = -\int m_{xx}^2 e^{-m} dx/4z \tag{19}$$

in the final equality of (18).

Inequality (18) can be integrated[11] to yield

$$\mathrm{Ei}(z) \leqslant \mathrm{Ei}(z_0) - \pi^2 t, \tag{20}$$

where $z_0 = z(0)$ and $\mathrm{Ei}(\cdot)$ is the exponential integral.[12] Furthermore, using the identity[11,12]

$$\mathrm{Ei}(y) = \gamma + \ln|y| + \sum_{k=1}^{\infty} \frac{y^k}{k \cdot k!}, \tag{21}$$

where γ is Euler's constant, and also the fact that $z \geqslant 0$, it is straightforward to show that

$$z \leqslant \exp[\mathrm{Ei}(z) - \gamma] \leqslant \exp[\mathrm{Ei}(z_0) - \gamma - \pi^2 t] \equiv z_B\, e^{-\pi^2 t}. \tag{22}$$

Thus, (22) provides an upper bound on z and, together with (13), it therefore bounds sup m.

The single most important characteristic of both bounds (9) and (22) is the fact that the time dependence is $\exp(-\pi^2 t)$ for both the upper and lower bounds. This fact is crucial for the latter developments.

III. ASYMPTOTIC BEHAVIOR

To determine the asymptotic behavior of the solution (3), consider the function

$$u(x,t) = e^{\pi^2 t} m(x,t), \tag{23}$$

which satisfies

$$u_t = e^{-m} u_{xx} + \pi^2 u. \tag{24}$$

We know from (13) and (22) that $m \to 0$ as $t \to \infty$. Therefore, if we could show, in some sense yet to be defined, that $|u_t| \to 0$, then in some corresponding sense the solution of (24) approaches the solution of

$$0 = u_{xx} + \pi^2 u. \tag{25}$$

What is required then is an estimate of $|u_t|$ showing that it must tend to vanish. The required estimates can be obtained by studying the Lyapunov functional[13,14]

$$I(u) = \int_0^1 u_x^2 dx - \pi^2 \int_0^1 e^m u^2\, dx. \tag{26}$$

First, from the definitions (13), (22), and (23), it follows that

$$0 \leqslant \int u_x^2 dx = 4e^{2\pi^2 t} z^2 \leqslant 4 z_B^2 \tag{27}$$

and that

$$0 \leqslant \int e^m u^2 dx \leqslant e^z \int u^2 dx \leqslant \frac{e^z}{\pi^2}\int u_x^2\, dx. \tag{28}$$

The last inequality in (28) follows from (14). Thus, both terms of (26) are bounded for all time and therefore $I(u)$ is bounded.

Using (23) and (24), the time derivative of the Lyapunov functional is

$$\frac{d}{dt} I = -2\int e^m u_t^2 dx + 2\pi^2 e^{-\pi^2 t}\int u u_x^2\, dx. \tag{29}$$

Again using (12), we find

$$0 \leqslant \int u u_x^2\, dx \leqslant \frac{1}{2}\left(\int u_x^2\, dx\right)^{3/2} \leqslant 4 z_B^3 \tag{30}$$

so the second integral in (29) is bounded above by a positive constant (say $C \equiv 4 z_B^3$) as determined by (27). Now *suppose* that $|u_t|$ does not tend to vanish. Then there exists an ϵ such that

$$2\int e^m u_t^2 dx \geqslant \epsilon > 0. \tag{31}$$

From (29), it follows that

$$\frac{d}{dt} I \leqslant -\epsilon + 2\pi^2 C e^{-\pi^2 t}. \tag{32}$$

Integrating (32) yields

$$I \leqslant 2C(1 - e^{-\pi^2 t}) - \epsilon t \leqslant 2C - \epsilon t. \tag{33}$$

The right-hand side of (33) tends to $-\infty$ as $t\to+\infty$. Thus, if (31) were true, $I(u)$ could not be bounded below as we have shown previously. Hence, (31) is false and there must exist a sequence of times t_i for which

$$\int e^m u_t^2 dx \to 0 \quad \text{as} \quad t_i \to \infty. \tag{34}$$

It follows, as in Ref. 15, that there exists a function R such that $u(\cdot,t_i)\to R(\cdot)$ as $t_i\to\infty$ as in the space H_0^1 of Sobolev.[16,17] We must still show that R is a solution of the linear equation for ϕ_1.

Now, if (24) is multiplied by e^m and also by any C^∞-function P vanishing at the boundaries, then after integrating we have

$$\int Pe^m u_t\, dx = -\int P_x u_x dx + \pi^2 \int Pe^m u\, dx. \tag{35}$$

Using the Cauchy–Schwarz inequality again, we have

$$\left(\int Pe^m u_t\, dx\right)^2 \leqslant \int e^m u_t^2\, dx \int e^m P^2\, dx. \tag{36}$$

From the boundedness of P and m and from the result (34), we find that the left-hand side of (35) tends to zero for the sequence of times t_i. Since $e^m\to 1$ as $t_i\to\infty$, it follows that $u\to R$, where the function $R(x)$ satisfies

$$-\int P_x R_x\, dx + \pi^2 \int PR\, dx = 0. \tag{37}$$

Thus, $R(x)$ is a weak solution of (25). From the existence of the lower bound (9), we infer that R is not identically zero. Then, by using the same arguments used in Ref. 15, we conclude further that R must be a classical solution of (25).

To show that $u\to R$ for all t requires the further information that $\int u^2 dx\to$const as $t\to\infty$. Consider the time derivative of this integral

$$\frac{d}{dt}\int u^2\, dx = 2\pi^2 \int u^2\, dx + 2\int e^{-m} u u_{xx}\, dx$$

$$= 2\pi^2 \int u^2\, dx - 2\int (1-m)e^{-m} u_x^2\, dx, \tag{38}$$

where we have used (23) and (24) and integrated once by parts to obtain the second equality in (38).

Again using (14), we find that (38) implies

$$\frac{d}{dt}\int u^2\, dx \leqslant 2\int [1-(1-m)e^{-m}] u_x^2\, dx$$

$$\leqslant 4\int m u_x^2 dx = 4e^{-\pi^2 t}\int u u_x^2\, dx. \tag{39}$$

The second inequality in (39) follows from the fact that $[1-(1-m)e^{-m}]\leqslant 2m$ for any m, which can be proved using elementary calculus.[18] Equations (30) and (39) show that

$$\frac{d}{dt}\int u^2\, dx \leqslant 16 z_B^3 e^{-\pi^2 t}. \tag{40}$$

Now *suppose* that $\int u^2 dx$ does not converge to a constant for all t and consider two time sequences s_i and t_i for which the integral $\int u^2 dx$ converges to the distinct constants α and β, respectively, satisfying $\alpha<\beta$. Using the estimate (40) together with standard arguments, it is not difficult to show that both sequences must converge to the lower constant α—thus contradicting the hypothesis. Therefore, $\int u^2 dx$ converges to a constant for all t. It then follows, using the same arguments as in Ref. 15, that u converges to R for all t.

We have shown that

$$u(x,t)\to R(x) = A\phi_1(x) \tag{41}$$

or equivalently

$$m(x,t)\to Ae^{-\pi^2 t}\phi_1(x), \tag{42}$$

uniformly in x and for all time. The asymptotic amplitude A appearing in (41) and (42) is a definite constant whose magnitude will be studied in the next section.

IV. AMPLITUDE ESTIMATES

We have established the general asymptotic behavior of the solution of (3) but we still need to find estimates of the asymptotic amplitude A. Bounds on A can be obtained from the inequalities (9) and (22). To illustrate, note

$$\lim_{t\to\infty} e^{\pi^2 t} Q(t) = \frac{\pi}{2^{3/2}}\int_0^1 R(x)\phi_1(x)\, dx = \frac{\pi}{2^{3/2}} A \geqslant Q_0 \tag{43}$$

and

$$\lim_{t\to\infty} e^{\pi^2 t} z(t) = \frac{1}{2}\left[\int_0^1 R^2(x)\, dx\right]^{1/2} = \frac{\pi}{2} A \leqslant z_B. \tag{44}$$

We define the resulting bounds on the amplitude to be

$$L_1 \equiv (2^{3/2}/\pi) Q_0 \quad \text{and} \quad U_1 \equiv (2/\pi) z_B. \tag{45}$$

Other integral inequalities can also be derived and a few of the more interesting ones will be obtained in the following paragraphs.

The most significant new bound can be derived by a slight modification of (7) using the properties of convex functions.[19] If f is a convex function and $p(x)\geqslant 0$, then in general

$$f\left[\frac{\int p(x)v(x)\, dx}{\int p(x)\, dx}\right] \leqslant \frac{\int p(x) f[v(x)]\, dx}{\int p(x)\, dx}. \tag{46}$$

The exponential function is convex so we have

$$\exp\left[\frac{\int m\phi_1 dx}{\int \phi_1 dx}\right] \leqslant \frac{\int e^m \phi_1 dx}{\int \phi_1 dx} \tag{47}$$

or equivalently, using the definition of Q from (8),

$$\int m\phi_1 dx \leqslant (\pi/2^{3/2}) \ln(1+Q), \tag{48}$$

since $\int \phi_1 dx = (2^{3/2}/\pi)$. Replacing the last step in (7) with (47), we have

$$-\frac{d}{dt} Q \leqslant \pi^2 \ln(1+Q), \tag{49}$$

which can be integrated[11] to yield

$$\mathrm{li}(1+Q) \geqslant \mathrm{li}(1+Q_0) - \pi^2 t \tag{50}$$

in terms of the logarithmic integral li or, equivalently,

$$\mathrm{Ei}[\ln(1+Q)] \geqslant \mathrm{Ei}[\ln(1+Q_0)] - \pi^2 t \tag{51}$$

in terms of the exponential integral, since $\mathrm{li}(y) = \mathrm{Ei}[\ln(y)]$. Checking the limiting forms as in (43) and (44), we find

$$A \geqslant (2^{3/2}/\pi)\exp[\mathrm{li}(1+Q_0) - \gamma] \equiv L_2. \tag{52}$$

Although upper bounds other than (22) exist, none have been found which generally give a significant improvement

over (22). However, one set of bounds has been found to be slightly better when the initial data are small. To derive these bounds, first consider the following inequalities, valid for $m \geqslant 0$:

$$m^2 \leqslant 2[1-(1-m)e^m] \leqslant m^2 e^m \leqslant (e^m-1)^2. \tag{53}$$

These inequalities can be derived using elementary calculus.[18] Combining (15) and (53) gives

$$\int m^2 dx \geqslant e^{-z} \int 2[1-(1-m)e^m]\, dx, \tag{54}$$

from which it follows that

$$\pi^2 \leqslant \frac{\int m_x^2 dx}{\int m^2\, dx} \leqslant e^z \frac{\int m_x^2 dx}{\int 2[1-(1-m)e^m]\, dx}. \tag{55}$$

Similarly, using (15) and $m \geqslant 0$, we have

$$\pi^2 \leqslant \frac{\int m_x^2 dx}{\int m^2 dx} \leqslant e^z \frac{\int m_x^2 dx}{\int m^2 e^m dx} \leqslant e^z \frac{\int (1+m) m_x^2 dx}{\int m^2 e^m dx}. \tag{56}$$

Then, since (e^m-1) is a function that vanishes at $x=0$ and 1, the same variational argument leading to (14) also shows that

$$\pi^2 \leqslant \frac{\int e^{2m} m_x^2\, dx}{\int (e^m-1)^2 dx} \leqslant e^z \frac{\int e^m m_x^2 dx}{\int (e^m-1)^2 dx}. \tag{57}$$

In all three cases (55)–(57), the integral in the numerator on the right is one-half the negative of the time derivative of the integral in the denominator. For example, with the definition

$$W(t) = 2\int [1-(1-m)e^m]\, dx, \tag{58}$$

we have

$$-\frac{1}{2}\frac{d}{dt} W = -\int m e^m m_t\, dx = -\int m m_{xx}\, dx = \int m_x^2\, dx. \tag{59}$$

From (55) and (59), it follows that

$$-\frac{1}{2}\frac{d}{dt} \ln W \geqslant \pi^2 e^{-z}. \tag{60}$$

To integrate (60), we need to study the right-hand side of the equation. Using the inequality (22) on z, note that

$$\begin{aligned} \int_0^t e^{-z} dt &= t + \int_0^t (e^{-z}-1)\, dt \\ &\geqslant t + \int_0^t [\exp(-z_B\, e^{-\pi^2 t}) - 1]\, dt \\ &= t + \int_{z_B e^{-\pi^2 t}}^{z_B} (e^{-y}-1)\frac{dy}{\pi^2 y} \\ &\geqslant t + \frac{1}{\pi^2}[\mathrm{Ei}(-z_B) - \mathrm{Ei}(z_0)], \end{aligned} \tag{61}$$

where in the last step of (61) we used the definition[11]

$$\mathrm{Ei}(-z_B) = \gamma + \ln z_B + \int_0^{z_B} (e^{-y}-1)\frac{dy}{y} \tag{62}$$

and the definition (22) of z_B.

So, integrating (60) using (61), we find

$$\ln[W(t)/W(0)]^{1/2} \leqslant -\pi^2 t + \mathrm{Ei}(z_0) - \mathrm{Ei}(-z_B). \tag{63}$$

TABLE I. Values of the asymptotic amplitude A and the rigorous upper and lower bounds on A obtained in seven computer experiments. The initial data for the first six cases are given by $m(x,0) = \Sigma_{k=1}^4 \alpha_k \sin k\pi x$, where $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ equal (i) (1,0.4,0,0), (ii) (1,0,0.3,0), (iii) (1,0, −0.3,0), (iv) (1,0,0,0.225), (v) (0.1,0.04,0,0), and (vi) (0.01,0.004,0,0). For case (vii), $m(x,0) = 1$ for $0 \leqslant x \leqslant 1$. Formulas for L_1, U_1, L_2, and U_2 are given in Eqs. (45), (52), and (64), respectively.

Case	L_1	L_2	A	U_1	U_1
i	1.202	2.222	2.375	7.715	23.51
ii	1.089	1.906	1.987	9.532	26.04
iii	1.243	2.343	2.508	9.532	28.83
iv	1.146	2.061	2.172	9.527	26.54
v	0.074 2	0.077 3	0.077 6	0.104 9	0.092 3
vi	0.007 11	0.007 13	0.007 13	0.009 18	0.007 75
vii	1.547	3.363	3.558	∞	∞

Applying the same reasoning which lead to (43), (44), and (52) gives

$$A \leqslant W^{1/2}(0)\exp[\mathrm{Ei}(z_0) - \mathrm{Ei}(-z_B)] \equiv U_2. \tag{64}$$

Similar bounds can be obtained from (56) and (57) but (64) is the best of this set of bounds because of the relation (53).

From the definition (22) of z_B and the identity (21), it follows that

$$z_B = z_0 \exp\left[\sum_{k=1}^{\infty} \frac{z_0^k}{k\cdot k!}\right]. \tag{65}$$

If z_0 is small, then the exponential factor in (64) can approach unity. If z_0 is large, the exponential factor can become enormous, making this set of bounds useless. In the former case, U_2 can give a slight improvement over U_1. In the latter case, U_2 is a much worse estimate of A than U_1.

To check our results, a series of computer experiments were performed. The results are presented in Table I. Equation (3) was integrated numerically using a linear implicit three-level difference scheme developed by Lees[20,21] for quasilinear parabolic equations. The initial values for $m(x,0)$ were chosen in six out of seven cases to be of the form

$$m(x,0) = \sum_{k=1}^{4} \alpha_k \sin k\pi x \tag{66}$$

using different sets of values for the $\{\alpha_k\}$. The six cases tested of this form $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ were: (i) (1, 0.4, 0, 0), (ii) (1, 0, 0.3, 0), (iii) (1, 0, −0.3, 0), (iv) (1, 0, 0, 0.225), (v) (0.1, 0.04, 0, 0), and (vi) (0.01, 0.004, 0, 0). The exceptional case was (vii) for which $m(x,0) = 1$ for the 99 grid points in the interval $0 \leqslant x \leqslant 1$ but $m(0,0) = m(1,0) = 0$ at the boundaries. The amplitude A can be determined during the course of the computer experiment by monitoring the upper and lower bounds until the difference $(U-L) \leqslant$ a predetermined error. The values of A quoted in Table I are correct ± a unit in the last significant figure quoted.

The results show that L_2 is always the closest estimate of A and that neither of the upper bounds is a very good estimate of A.

V. COMMENTS

The results obtained here differ from the results of our previous work[15] in a fundamental way. To obtain the proper

time dependence of the upper bounds (22) and (64), we required an estimate of sup m in terms of the integral of the squared gradient of m. Such estimates are available in one dimension but not in higher dimensions. Therefore, the methods used in this paper to determine the asymptotic behavior of (3) cannot be generalized to higher dimensions. In our previous work,[15] estimates such as (13) were not required in the one-dimensional case so the generalization to higher dimensions was relatively straightforward. The mathematical consequences of this difficulty have not yet been adequately explored.

[1]T. Kamimura and J. M. Dawson, Phys. Rev. Lett. **36**, 313 (1976).
[2]J. R. Greenwood, Ph.D. thesis, Univ. of Wisconsin, 1975 (unpublished).
[3]J. R. Drake, D. W. Kerst, G. A. Navratil, R. S. Post, S. Ejima, R. LaHaye, C. P. Moeller, T. Ohkawa, P. J. Petersen, R. Prater, and S. K. Wong, in *Plasma Physics and Controlled Nuclear Fusion Research* (International Atomic Energy Agency, Vienna, 1977), Vol. II, p. 333.
[4]A. B. Ehrhardt, H. R. Garner, G. A. Navratil, and R. S. Post, Phys. Fluids **24**, 1859 (1981).
[5]K. E. Lonngren and A. Hirose, Phys. Lett. A **59**, 285 (1976).
[6]T. G. Kurtz, Trans. Am. Math. Soc. **186**, 259 (1973).
[7]H. P. McKean, Israel J. Math. **21**, 54 (1975).
[8]T. Carleman, *Problèmes Mathématiques dans la Théorie Cinétique des Gas* (Almquist-Wiksells, Uppsala, 1957).
[9]J. R. Drake and J. G. Berryman, Phys. Fluids **20**, 851 (1977).
[10]J. G. Berryman, UCRL-86783 (preprint) (unpublished).
[11]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, (Academic, New York, 1965), pp. 93, 204, 925–930.
[12]W. Gautschi and W. F. Cahill, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972), Chap. 5, pp. 227–252.
[13]J. G. Berryman and C. J. Holland, Phys. Rev. Lett. **40**, 1720 (1978).
[14]G. Rosen, Phys. Rev. B **23**, 3093 (1981).
[15]J. G. Berryman and C. J. Holland, Arch. Rational Mech. Anal. **74**, 379 (1980).
[16]The Sobolev space H_0^1 (or $W_{1,2}$) assigns the norm $||f||^2 = \int (f^2 + f'^2) dx$ to a function f which vanishes on the boundaries.
[17]Note that, in one dimension, convergence in H_0^1 implies uniform convergence in space for the sequence of times t_i.
[18]G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities* (Cambridge U.P., Cambridge, 1952), Chap. IV, pp. 106–107.
[19]Reference 18, Chap. III, pp. 74–78.
[20]M. Lees, Math. Comp. **20**, 516 (1966).
[21]A. R. Mitchell, *Computational Methods in Partial Differential Equations* (Wiley, London, 1969), pp. 95–99.

New Stokes' line in WKB theory

H. L. Berk
Institute for Fusion Studies, University of Texas, Austin, Texas 78712

William McCay Nevins
Lawrence Livermore National Laboratory, Livermore, California 94550

K. V. Roberts
Euratom/UKAEA Fusion Association, Culham Laboratory, Abingdon, Oxon OX 14 3DB, United Kingdom

(Received 18 August 1981; accepted for publication 23 December 1981)

The WKB theory for differential equations of arbitrary order or integral equations in one dimension is investigated. The rules previously stated for the construction of Stokes' lines for N th-order differential equations, $N \geqslant 3$, or integral equations are found to be incomplete because these rules lead to asymptotic forms of the solutions that depend on path. This paradox is resolved by the demonstration that new Stokes' lines can arise when previously defined Stokes' lines cross. A new formulation of the WKB problem is given to justify the new Stokes' lines. With the new Stokes' lines, the asymptotic forms can be shown to be independent of path. In addition, the WKB eigenvalue problem is formulated, and the global dispersion relation is shown to be a functional of loop integrals of the action.

PACS numbers: 02.30.Jn, 02.30.Rz, 03.65.Sq

1. INTRODUCTION

Phase integral methods have been used for many years as a means of obtaining approximate solutions to linear ordinary differential equations in the limit that the solution is rapidly varying in comparison with the variation in the coefficients of the equation. The development of quantum mechanics led to a great deal of interest in the application of phase integral methods to the solution of Schrödinger's equation. A comprehensive theory of the application of phase integral methods to second-order differential equations, WKB theory, was developed.[1-3]

In addition to the work on second-order differential equations, there has been a continuing interest in the application of phase integral methods to the solution of higher-order systems of equations. Such equations are of particular interest in the study of wave propagation in inhomogeneous media. In this paper we examine the application of phase integral methods to the solution of higher-order equations. Following previous workers,[4-6] we consider equations of the form

$$\int dx'\, G(x - x', x + x'/2, \omega)\phi(x') = 0. \tag{1}$$

Differential equations of arbitrary order are a subclass of Eq. (1) in which the kernel G involves a sum of delta functions and their derivatives, with the singularities located at $x = x'$. Our results may be extended to vector systems of integral (or differential) equations by using the methods of Berk and Pfirsch.[6]

For the purposes of WKB theory, it is sufficient to consider the characteristic equation

$$\epsilon(k, x, \omega) = \int dy\, G(y, x, \omega)\exp(-iky) = 0. \tag{2}$$

Early work on higher-order differential equations[7-9] and integral equations[10] demonstrates that Eq. (1) has asymptotic solutions of the form

$$\phi(x) \sim \exp\left[i\int^x k_j(z)\, dz\right], \tag{3}$$

where $k_j(z)$ is a particular branch of the function $k(z)$ defined by

$$\epsilon[k(z), z, \omega] = 0. \tag{4}$$

In an analogy to the WKB theory of second-order differential equations, it is found that the asymptotic solutions, (3), are coupled at turning points where two branches, $k_i(x)$ and $k_j(x)$, of the function $k(z)$ merge.[4-6,11-18] This coupling of the asymptotic solutions, which has been particularly well studied in certain fourth-order differential equations, leads to linear mode conversion and wave reflection.[11-16] Multiple wave reflections can lead to the formation of linear normal modes.[17,18]

In addition to this work on fourth-order equations, the quasiclassical theory of mechanics has been developed by applying phase integral methods to the solution of higher-order equations in which the differential operator is Hermitian.[19-22] While this restriction to Hermitian operators does not raise any difficulty in examining the classical limit of quantum mechanics, it is a serious restriction when studying the propagation of waves in a dissipative medium. The differential operator describing wave propagation in a dissipative medium is never Hermitian, as dissipation is associated with the anti-Hermitian part of an operator. In the limit of weak dissipation, progress can be made by using the quasiclassical approximation while treating the anti-Hermitian part of the operator as a perturbation.[23]

In the general case in which the dissipation is not weak, such perturbation expansions must be abandoned in favor of a method in which the Hermitian and anti-Hermitian parts of the differential operator are treated on an equal footing. Such a generalization has been successfully made in the

 0022-2488/82/060988-15$02.50

WKB theory of second-order differential equations.[1] The major complication that arises in second-order differential equations when the operator is not Hermitian is that the turning points do not, in general, lie on the real axis. Hence, it is necessary to consider the behavior of the solution in the complex plane, rather than just on the real axis. It is convenient to introduce Stokes' lines[24] when examining the behavior of solutions in the complex plane.

Stokes' lines also play a fundamental role in the WKB theory of integral equations. Past work[4–6,25] indicates that Stokes' lines emerge from turning points of the characteristic equation where

$$\epsilon(k_{ij}^T, x_{ij}^T, \omega) = \epsilon_k(k_{ij}^T, x_{ij}^T, \omega) = 0. \tag{5}$$

The subscript on ϵ denotes a partial derivative with respect to the indicated variable. Equation (5) is the condition for the merging of two branches, $k_i(x)$ and $k_j(x)$, of the function $k(z)$. The Stokes' lines satisfy[6]

$$\mathcal{S}_{ij}(x) \ni \mathrm{Re} \int_{x_{ij}^T}^{x} [k_i(z) - k_j(z)]\, dz = 0. \tag{6}$$

Note that for second-order differential equations in standard form (i.e., the coefficient of the first derivative vanishes), $k_i(x) = -k_j(x)$ and $k_{ij}^T = 0$. Hence, condition (6) reduces to the usual definition of a Stokes' line.[1–3]

Stokes' lines are significant because they delimit regions of the complex x plane in which the asymptotic approximations to solutions of Eq. (1) have a particular form. These asymptotic approximations change discontinuously upon crossing Stokes' lines.[24]

The object of WKB theory is to develop a set of rules for continuing asymptotic solutions to all regions of the complex x plane. Past workers[4–6,10–18] have focused on the behavior of solutions in the neighborhood of an isolated turning point. The integral equation may be reduced to an Airy equation in this region.[25] The familiar connection rules of the WKB theory of second-order differential equations[1–3] are then recovered. Since these rules were first put forward by Furry[1] for a general linear ordinary differential equation of second order, we call them "Furry's rules."

Furry's rules are complete for second-order differential equations. However, for Nth-order differential equations (with $N > 2$), it is possible for Stokes' lines [as defined in Eq. (6)] to cross. When such a crossing occurs, Furry's rules can lead to contradictions in the asymptotic properties of the solution. This problem is discussed in Sec. 2, where we present a particular example of an equation that leads to crossed Stokes' lines.

We resolve this difficulty with Furry's rules by reformulating the problem of obtaining the WKB connection formula. The breakdown of Furry's rules indicates that there are phenomena that occur in differential equations of third and higher order that have no analog in second-order differential equations. To study these phenomena, we extend the methods of quasiclassical mechanics[22] in using the Fourier space representation of the equation to obtain a contour integral representation of the solution to Eq. (1) that is valid in the neighborhood of one or more turning points x_{ij}^T (Ref. 6). The asymptotic solutions and connection rules are then deduced from this integral representation.

This formulation of the problem has several advantages. It resolves the philosophical difficulties surrounding the retention in the asymptotic solution of a subdominant term in the presence of a dominant term. Stokes' lines appear naturally in this formulation. Hence, the ambiguity in the definition of Stokes' lines for higher-order equations (compare Refs. 5, 6, and 26) is resolved. In the process new classes of Stokes' lines are discovered. When these new Stokes' lines are considered, the difficulties in the asymptotic properties of the solution are resolved. Since this formulation does not make use of comparison equations, it is applicable to both differential equations of all orders and integral equations. Since our formulation of the connection problem is new, a careful presentation is given where we shall rederive known properties of Stokes' lines as well as obtain new results.

In this paper we restrict our attention to simple turning points. That is, we assume

(i) $k(x)$ is finite for $|x| < \infty$,

(ii) $\epsilon_{kk}(k_{ij}^T, x_{ij}^T, \omega) \neq 0$ for any i, j,

and

(iii) $\epsilon_x(k_{ij}^T, x_{ij}^T, \omega) \neq 0$ for any i, j.

An extension of this formulation to include singular turning points, where (i) is violated, and multiple turning points, where (ii) and/or (iii) are violated, is reserved for future considerations. In addition, we do not consider systems in which two turning points are joined by a Stokes' line. This singular configuration is encountered in quantum mechanics when studying tunneling through potential barriers and has led to a great deal of controversy[2,3] regarding the proper statement of the WKB connection rules for second-order differential equations.

The plan of this paper is as follows. In Sec. 2 we examine a particular third-order differential equation to show that the rules put forward by Berk and Pfirsch[6] lead to inconsistencies in the asymptotic properties of the solution to this equation. Our contour integral representation of the solution is developed in Sec. 3, where we point out that the usual WKB wavelets may be recovered when the contour integral is evaluated by the method of steepest descent. In Sec. 4 Stokes' lines are discussed in the context of this contour integral representation. We demonstrate that in addition to the Stokes' lines emerging from turning points, there are additional classes of Stokes' lines. In Sec. 5 we evaluate the Stokes' multipliers associated with each class of Stokes' line, and follow this in Sec. 6 by a review of rules for continuing asymptotic solutions to all regions of the complex plane. In Sec. 7 we show that the global dispersion relation is a functional of loop integrals in the complex plane. Finally, in Sec. 8 several unresolved problems are discussed.

2. A PARADOX INVOLVING CROSSED STOKES' LINES

In studying higher-order differential equations we find that Stokes' lines, as defined by Eq. (6), can cross. In fact, experience[27,28] with Nth-order systems of differential equations ($N > 2$) indicates that the Stokes' lines often do cross. This is to be expected because Eq. (6) defines a Stokes' line

$\mathscr{S}_{ij}$ through two branches, $k_i(x)$ and $k_j(x)$, of the multivalued function $k(z)$. When the system of equations is of third or higher order, then $k(z)$ has more than two branches. Hence, there are several "kinds" of Stokes' lines.

Consider the Stokes' lines $\mathscr{S}_{ij}$ which emerge from turning points x_{ij}^T, together with the Stokes' lines $\mathscr{S}_{mn}$ which emerge from $x_{mn}^T(\{i,j\} \neq \{m,n\})$. Since the elements of $\mathscr{S}_{ij}$ obey an equation that is independent of the equation obeyed by the elements of $\mathscr{S}_{mn}$, we expect that these Stokes' lines will go off to $|x| = \infty$ in the complex x plane at independent angles $\theta = \arg(x)$. Hence, the generic situation is that at least one of the triplet of Stokes' lines $\mathscr{S}_{ij}$ that emerge from x_{ij}^T will intersect a Stokes' line $\mathscr{S}_{mn}$ from x_{mn}^T.

Such a crossing can lead to difficulties with the rules put forward by Berk and Pfirsch[6] for continuing the asymptotic solutions to all regions of the complex x plane. We illustrate these difficulties by considering the simplest equation that exhibits crossed Stokes' lines:

$$(id^3/dx^3 + 3id/dx + x)\phi = 0. \tag{7}$$

The characteristic equation is given by

$$\epsilon(k, x) = k^3 - 3k + x = 0. \tag{8}$$

Setting $\epsilon = \epsilon_k = 0$, we find two turning points:

$$x_{\alpha\beta}^T = -2, \quad k_{\alpha\beta}^T = -1$$

and

$$x_{\beta\gamma}^T = 2, \quad k_{\beta\gamma}^T = 1.$$

Since the characteristic equation is cubic in k, there are three branches to the function $k(x)$, which we label α, β, and γ. Hence, the two turning points must share at least one common branch. A numerical solution of Eq. (8) for $k_i(x)$ reveals that the two turning points share exactly one branch, which we have labeled β.

The Stokes' structure, which was obtained by numerically solving Eq. (6), is shown in Fig. 1. The Stokes' line $(x_{\alpha\beta}^T, C)$ is seen to cross the Stoke' line $(x_{\beta\gamma}^T, A)$ at x_q. The numerical solution of the characteristic equation shows that the WKB wavelet

$$\langle k_\alpha | x_{\alpha\beta}^T, x\rangle \sim \exp\left(i\int_{x_{\alpha\beta}^T}^{x} k_\alpha \, dx\right)$$

is exponentially large compared to the wavelet

$$\langle k_\beta | x_{\alpha\beta}^T, x\rangle \sim \exp\left(i\int_{x_{\alpha\beta}^T}^{x} k_\beta \, dx\right)$$

on $(x_{\alpha\beta}^T, C)$. Hence, $\langle k_\alpha \rangle$ is dominant with respect to $\langle k_\beta \rangle$ on $(x_{\alpha\beta}^T, C)$. Similarly, we find that $\langle k_\beta \rangle$ is dominant with respect to

$$\langle k_\gamma | x_{\beta\gamma}^T, x\rangle \sim \exp\left(i\int_{x_{\beta\gamma}^T}^{x} k_\gamma \, dx\right)$$

on $(x_{\beta\gamma}^T, A)$. We note that the crossing of two Stokes' lines illustrated in Fig. 1 is an example of an *ordered crossing* of Stokes' lines because the relative dominance of all three WKB wavelets is determined at x_q. Namely, $\langle k_\alpha \rangle$ is dominant with respect to $\langle k_\beta \rangle$, which is in turn dominant with respect to $\langle k_\gamma \rangle$. We will discuss the importance of ordered crossings of Stokes' lines in Sec. 4.

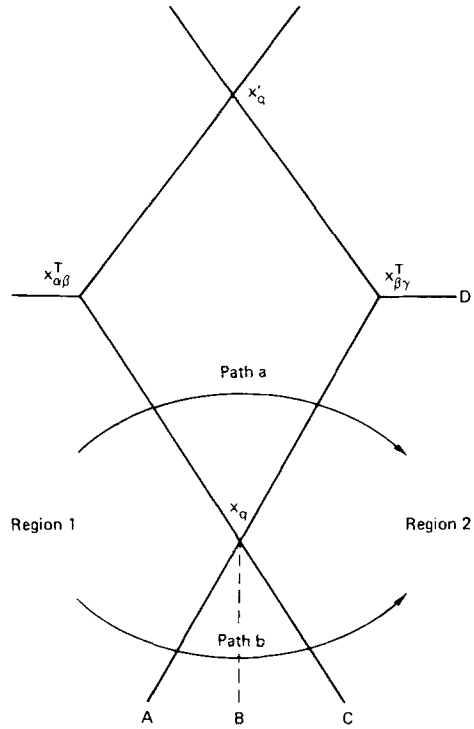


FIG 1. Example of an ordered crossing of Stokes' lines.

Suppose that we have a solution $\phi(x)$ that can be approximated by $\langle k_\alpha \rangle$ in region 1, and we attempt to continue this solution to region 2 by following path a. Path a crosses the Stokes' line $(x_{\alpha\beta}^T, C)$, and the subdominant wavelet must be added according to the rules delineated by Berk and Pfirsch.[6] Then $\phi(x)$ becomes

$$\phi(x) \sim \langle k_\alpha \rangle + P_1 \langle k_\beta \rangle,$$

where P_1 is a constant that will be as determined precisely in Sec. 4. When path a crosses the Stokes' line $(x_{\beta\gamma}^T, A)$, the wavelet $\langle k_\beta \rangle$ induces a $\langle k_\gamma \rangle$ wavelet, so that $\phi(x)$ becomes

$$\phi(x) \sim \langle k_\alpha \rangle + P_1 \langle k_\beta \rangle + P_2 \langle k_\gamma \rangle. \tag{9}$$

If, on the other hand, $\phi(x)$ is continued along path b, we do *not* induce a wavelet on crossing the Stokes' line $(x_{\beta\gamma}^T, A)$ because no $\langle k_\beta \rangle$ wavelet is yet present. The wavelet $\langle k_\beta \rangle$ is induced on crossing the Stokes' line $(x_{\alpha\beta}^T, C)$. Hence, we obtain

$$\phi(x) \sim \langle k_\alpha \rangle + P_1 \langle k_\beta \rangle$$

as the asymptotic solution in region 2. These two versions of the asymptotic solution in region 2 differ by a term in $\langle k_\gamma \rangle$. This term is exponentially large compared to the term in $\langle k_\beta \rangle$ on the Stokes' line $(x_{\beta\gamma}^T, D)$ that bounds region 2. Clearly, only one of these asymptotic approximations to the solution of Eq. (7) can be correct. Hence, previous rules do not provide a complete prescription for continuing asymptotic solutions to all regions of the complex plane.

We anticipate the results of Secs. 4 and 5 by noting the existence of an additional Stokes' line (x_q, B) on which the wavelet $\langle k_\alpha \rangle$ is dominant with respect to $\langle k_\gamma \rangle$. When this new Stokes' line is included, the wavelet $\langle k_\gamma \rangle$ is induced upon crossing (x_q, B), while the $\langle k_\beta \rangle$ wavelet is induced on

crossing the Stokes' line $(x_{\alpha\beta}^T, C)$. Hence, we obtain

$$\phi(x) \sim \langle k_\alpha \rangle + P_1 \langle k_\beta \rangle + P_2 \langle k_\alpha \rangle \tag{9'}$$

as the asymptotic solution in region 2. We see that with the introduction of a new Stokes' line it will be possible to both establish rules for connecting asymptotic solutions to all regions of the complex x plane, and obtain results that are independent of the path used in continuing these solutions.

3. AN INTEGRAL REPRESENTATION OF THE SOLUTION

Equation (1) has asymptotic (WKB) solutions of the form[4-6]

$$\phi_i(x) \sim \frac{A}{\epsilon_k^{1/2}} \exp\left[i \int_{x_0}^{x} k_i(z) dz \right].$$

These solutions break down in the neighborhood of x-space turning points; i.e., those points where $\epsilon_k [k(x_{ij}^T), x_{ij}^T] = 0$. To derive WKB connection rules, we require an approximate solution to Eq. (1) that is valid in the neighborhood of these x-space turning points. Following quasiclassical theory[19-22] and previous workers,[5,6] we obtain such an approximate solution by considering Fourier space representation of Eq. (1). This Fourier representation also has an approximate WKB solution,

$$\phi(k) \sim \frac{A}{\epsilon_x^{1/2}} \exp\left[-i \int_{k_0}^{k} x(k) dk \right],$$

where $x(k)$ is determined from the equation $\epsilon[k, x(k)] = 0$. This k-space WKB solution breaks down in the neighborhood of the k-space turning points, where $\epsilon_x [k_T^{mn}, x(k_T^{mn})] = 0$. When the k-space and x-space turning points are well separated, i.e., when

$$\left[\frac{\epsilon_x(k_{ij}^T, x_{ij}^T)}{\epsilon_{kk}(k_{ij}^T, x_{ij}^T)} \right]^{1/2} |x_{ij}^T - x(k_T^{mn})|^{3/2} \gg 1 \tag{10}$$

for all (m,n), then an approximate representation of the solutions to Eq. (1), valid in the neighborhood of x_{ij}^T, is given by the Fourier transform of $\phi(k)$; viz.,

$$\phi_i(x) = \int_{C_i(k)} \frac{dk}{\epsilon_x^{1/2}(k,z)} \exp\left[-i \int_{k_0}^{k} x(k')\, dk' + ikx \right]. \tag{11}$$

The contour $C_i(k)$ avoids regions where $\epsilon_x(k,x) = 0$ and terminates where the integrand vanishes.

It is convenient to transform to the variable $z = x(k)$. The z manifold is chosen such that the function $k(z)$ is single-valued. Hence, this manifold consists of N sheets—one for each branch of the multivalued function $k_j(x)$. We then find

$$\begin{aligned} \phi_i(x) &= \int_{C_i(z)} dz \frac{dk(z)}{dz} \frac{1}{\epsilon_x^{1/2}[k(z), z]} \\ &\quad \times \exp\left[-i \int_{z_0}^{z} dz' z' \frac{dk(z')}{dz} + ik(z)x \right] \\ &= -\int_{C_i(z)} dz \frac{\epsilon_x^{1/2}}{\epsilon_k} \exp\left[i \int_{z_0}^{x_j} dz' k(z') + i\Phi_j(z,x) + ik_0 z_0 \right], \end{aligned} \tag{12}$$

where

$$\Phi_j(z,x) = -k(z)(z-x) + \int_{x_j}^{z} dz' k(z'), \tag{13}$$

$z_0 = x(k_0)$, and $dk/dz = -\epsilon_k/\epsilon_x$. The $\Phi_j(z,x)$ term in Eq. (12) is obtained from an integration by parts. The subscript j on x_j labels the sheet of the z manifold on which x, the end point of the z' integrals, resides. The contour $C_i(z)$ is the mapping of the contour $C_i(k)$ from the k plane into the z manifold under the transformation $z = x(k)$.

A given independent solution $\phi_i(x)$ is defined by a contour in the z manifold. Allowable contours must begin and end in a remote region at infinite $|z|$, where the integrand of Eq. (12) vanishes. Two contours are equivalent when they connect the same two remote regions; they are independent when they do not connect the same two remote regions. An N th-order differential equation [$\epsilon(k,x)$ is then an N th-order polynomial in k] will possess N independent contours corresponding to the N independent solutions. An integral equation will have an infinite number of independent contours. We assume that the contours are denumerable. The number of independent contours is denoted by N, with N allowed to go to infinity.

In some region of the x plane it is always possible to choose N independent contours $\{C_i\}$, such that the N independent solutions $\{\phi_i\}$—defined by these contours together with Eq. (12)— can each be approximated by a single WKB wavelet,

$$\phi_i(x) \sim \frac{A}{\epsilon_k^{1/2}} \exp\left[i \int_{x_0}^{x} k_i(z) dz \right].$$

These contours may be constructed as follows. At some point $x = \tilde{x}$ let us find stationary points with respect to z of the integrand in Eq. (12). We will assume that $\exp[i\Phi(z,x)]$ is rapidly varying when compared to ϵ_z. The stationary phase condition is then

$$-(z-\tilde{x}) \frac{dk(z=\tilde{x})}{dz} = \frac{\epsilon_z[k(z), z]}{\epsilon_k[k(z), z]} (z-\tilde{x}) = 0. \tag{14}$$

Hence, stationary phase points occur where $z = \tilde{x}$, $\epsilon_z[k(z), z] = 0$, or $\epsilon_k[k(z), z] = \infty$. Only the first condition is allowed. The second condition is ruled out since neither Eq. (11) or (12) is an accurate representation of the solution to the integral equation near points where $\epsilon_z = 0$, and the third condition implies a singular $\epsilon(k,z)$ for finite z, a condition that we have assumed not to exist. Thus, the stationary phase points of Eq. (14) are

$$z = \tilde{x}.$$

Since the z manifold has N sheets, there are N stationary phase points—one on each sheet—where $z = \tilde{x}$. These stationary phase points are saddle points of the function $\exp[i\Phi(z,x)]$. From each of the N saddle points we generate a contour C_α by following the two paths of steepest descent from the saddle point to $|z| = \infty$. The change in the phase

Φ_α along the steepest descent path is

$$d\Phi_\alpha \equiv i d\beta = (dz_R + i dz_I)\left[\frac{\partial \Phi_\alpha}{\partial z}\right]$$

$$= \frac{\epsilon_z(k_\alpha,z)}{\epsilon_k(k_\alpha,z)}(z-\tilde{x})(dz_R + i dz_I), \tag{15}$$

where $z = z_R + i z_I$. Note that $d\Phi$ must be positive imaginary on a steepest descent path or, equivalently, $d\beta$ is positive and real. Hence, the real part of the right-hand side of (15) vanishes so that the steepest descent contour satisfies the first-order differential equation,

$$\frac{dz_R}{dz_I} = \frac{\mathrm{Im}[\epsilon_z(k_\alpha,z)(z-\tilde{x})/\epsilon_k(k_\alpha,z)]}{\mathrm{Re}[\epsilon_z(k_\alpha,z)(z-\tilde{x})/\epsilon_k(k_\alpha,z)]}. \tag{16}$$

In the vicinity of $z = \tilde{x}$ Eq. (16) is singular. However, near $z = \tilde{x}$ one can expand $\Phi_\alpha(\tilde{x} + \delta z, \tilde{x})$ to second order in δz and find

$$\delta z = \pm (2i\delta\beta\epsilon_k/\epsilon_z)^{1/2}, \tag{17}$$

where $\delta\beta$ is a small positive real number (negative $\delta\beta$ would describe the initial direction of a steepest ascent path). Starting from $z = \tilde{x} + \delta z$, and using Eq. (16), we can then generate the two steepest descent paths, one for each branch of the square root in Eq. (17); these paths are precisely defined provided that

$$\frac{\partial \Phi_\alpha}{\partial z}(z,\tilde{x}) \neq 0 \tag{18}$$

anywhere along the steepest descent paths other than at the original point $z = \tilde{x}$ on the αth sheet, where $k = k_\alpha(\tilde{x})$. The contour C_α is then defined to be the union of the two steepest descent paths from the saddle point at $z = \tilde{x}$ on the αth sheet. Hence, for each root $k_\alpha(\tilde{x})$, a contour C_α is generated. These N contours $\{C_\alpha\}$, together with Eq. (12), define N independent solutions to the equation $\{\Phi_\alpha(x)\}$.

We now show that if $\partial\Phi_\alpha/\partial z \neq 0$ anywhere on the αth steepest descent contour (except for the original point $z = x$ on the αth sheet), then this contour must deform continously as x is varied. Consider the steepest descent path of the function $\exp[i\Phi_\alpha(z,x)]$ from the saddle point at $z = \tilde{x}$. We may label points on this path by

$$\beta = \mathrm{Im}\ \Phi_\alpha(z,\tilde{x}). \tag{19}$$

β monotonically increases from zero as you move away from the saddle point along the steepest descent path. Hence, Eq. (19) may be inverted to write an equation for the steepest descent path in the form

$$z = Z_\alpha^{(\pm)}(\beta,\tilde{x}), \tag{20}$$

where the plus/minus sign in Eq. (20) corresponds to the plus/minus sign in Eq. (17).

Similarly, we may evaluate the phase function at a nearby point, $\tilde{x} + \delta x$. The steepest descent path now satisfies

$$z = Z_\alpha^{(\pm)}(\beta,\tilde{x} + \delta x). \tag{21}$$

Note that Re Φ_α vanishes on both of the steepest descent paths, (20) and (21). Hence, the separation δz between corresponding points (i.e., points labeled by the same value of β) on these two steepest descent paths is determined by the condition that $\delta\Phi_\alpha$ vanish. When δx is small, this gives

$$\delta z \equiv Z_\alpha^{(\pm)}(\beta,\tilde{x} + \delta x) - Z_\alpha^{(\pm)}(\beta,\tilde{x})$$

$$= -\frac{\partial\Phi_\alpha[Z_\alpha^{(\pm)}(\beta,\tilde{x}),\tilde{x}]/\partial x}{\partial\Phi_\alpha[Z_\alpha^{(\pm)}(\beta,\tilde{x}),x]/\partial z}\delta x.$$

If $\partial\Phi/\partial z$ does not vanish anywhere on the steepest descent path $Z_\alpha^{(\pm)}(\beta,\tilde{x})$, then δz will be small when δx is small, and the steepest descent path must deform continuously as x is varied. In particular, the end points of each steepest descent path must remain in the same remote regions. Hence, the N steepest descent contours define N independent solutions for all x in some region about $\tilde{x}$ such that (18) is satisfied.

An approximate representation of these independent solutions $\{\phi_i(x)\}$ can be obtained by evaluating the contour integral in Eq. (12) by the saddle point method.[29] Expanding $\Phi(z + \delta z, x)$ to second order in Eq. (12) yields

$$\phi_j(x) \sim -\exp\left[i\int_{z_0}^{x_j} k(z')dz'\right]\frac{\epsilon_x^{1/2}}{\epsilon_k}$$

$$\times \int_{C_j} dz \exp\left[i\frac{(z-x)^2}{2}\frac{\epsilon_z}{\epsilon_k}\right]$$

$$= -\frac{(2\pi i)^{1/2}}{\epsilon_k^{1/2}}\exp\left[i\int_{x_0}^{x} k_j(z')dz'\right]. \tag{22}$$

This is precisely the x-space WKB representation[6] for $\phi(x)$. Exactly one WKB wavelet is needed to represent each of the independent solutions $\phi_i(x)$ in the vicinity of $\tilde{x}$. The choice of the branch of $\epsilon_k^{1/2}$ is implicitly determined by the direction of the integration path C_j. This choice will be discussed more fully in Sec. 5.

4. STOKES' LINES

As we move away from $x = \tilde{x}$, we can reach a region where more than one WKB wavelet is needed to approximate a particular solution, $\phi_\alpha(x)$. This happens as follows. As point x (at which the solution ϕ_α is to be evaluated) is varied, we may reach a curve $\mathcal{S}$ where Eq. (18) fails; i.e., for $x \in \mathcal{S}$, $\partial\Phi_\alpha(z,x)/\partial z = 0$ at some point along the contour C_α. When this occurs, the steepest descent prescription will no longer define a unique contour passing through the point $z = x_\alpha$, where we use the notation x_α to denote the point $z = x$ on the αth sheet of the z manifold.

We shall refer to the region containing $\tilde{x}$ as region 1. The curve $\mathcal{S}$ is a boundary separating region 1 from region 2. We see that $\mathcal{S}$ is the locus of points x on which a steepest descent path from $z = x_\alpha$ intersects a point where $\partial\Phi(z,x_\alpha)/\partial z = 0$. However, this is just the saddle point condition, and from the discussion following Eq. (14) such a point can only occur when $z = x$. Hence, on the boundary separating region 1 from region 2 a steepest descent path from the saddle at x_α must return to $z = x$, but on a different sheet of the z manifold; e.g., at $z = x_\beta$, where $k(z) = k_\beta(x)$, another root of the equation $\epsilon(k,x) = 0$. From the point $z = x_\beta$ the two steepest descent paths can be drawn in accordance with Eq. (17). Hence, the steepest descent contour from $z = x_\alpha$ bifurcates when x lies on a boundary line $\mathcal{S}$. We will find that this bifurcation of the steepest descent paths

leads to a discontinuous change in the asymptotic representation of ϕ_α. Hence, we call the boundary curves $\mathscr{S}$ *Stokes' lines*. This terminology is equivalent to that employed in the theory of asymptotic expansions.[30]

We have seen that x lies on a Stokes' line $\mathscr{S}$ when a pair of saddle points x_α and x_β are connected to each other by a steepest descent path of the function $\exp[i\Phi(z,x_\alpha)]$. At one of these saddle points, say x_α, the magnitude of the integrand in Eq. (12) is exponentially large when compared to its magnitude at x_β because $i\Phi(x_\beta,x_\alpha)$ is real and negative on a steepest descent path. Hence, we call x_α the dominant saddle and the corresponding value of k, $k(x_\alpha) = k_\alpha(x)$, the dominant root on $\mathscr{S}$. Similarly, $k(x_\beta) = k_\beta(x)$ is called the subdominant root on $\mathscr{S}$.

We define the quantity $g_{\alpha\beta}(x)$ as

$$g_{\alpha\beta}(x) = i\int_{x_\alpha}^{x_\beta} dz'\, k(z')\,. \tag{23}$$

We have from Eq. (13)

$$g_{\alpha\beta}(x) = i\Phi(x_\beta,x_\alpha)\,. \tag{24}$$

When x is on a Stokes' line, x_α and x_β are connected by steepest descent path of $\exp i[\Phi(z,x_\alpha)]$, so that $g_{\alpha\beta}(x)$ is both real and negative. The direction of the Stokes' line can be determined from the differential of Eq. (23) together with the condition that the change in g be real. We then have

$$dg = i[k_\beta(x) - k_\alpha(x)]dx = \text{real},$$

which yields a differential equation for the Stokes' line,

$$\frac{dx_R}{dx_I} = \frac{\mathrm{Im}[k_\beta(x) - k_\alpha(x)]}{\mathrm{Re}[k_\beta(x) - k_\alpha(x)]}, \tag{25}$$

where x_R and x_I are real numbers satisfying $x = x_R + ix_I$.

Once a single point $x_{\mathscr{S}}$ on a Stokes' line $\mathscr{S}$ has been located, the remaining points on $\mathscr{S}$ can be found by integrating Eq. (25), using $x_{\mathscr{S}}$ as an initial condition. These points may be labeled by the value of $\mathrm{Re}\, g_{\alpha\beta}(x)$. If we follow the Stokes' line in the direction of increasing $g_{\alpha\beta}$, at least one of three things must occur. We will reach a

(i) point where $g_{\alpha\beta} = 0$ (recall that $g_{\alpha\beta}$ must be negative real)

or

(ii) maximum of $\mathrm{Re}\, g_{\alpha\beta}$, where $dg_{\alpha\beta}/dx = 0$

or

(iii) point x at which the steepest descent path from x_α breaks and fails to return to x_β.

It follows from our proof of the continuity of steepest descent paths that (iii) can only occur if there is some $x\in\mathscr{S}$ such that the steepest descent path from x_α returns to $z = x$ on some third sheet, say $z = x_\gamma$, before returning to x_β. Such a path must encircle at least two branch points (e.g., one branch point that connects sheets α and γ as well as a second that connects sheets γ and β).

We shall return to this important case later. For the present we restrict our attention to the steepest descent paths that encircle a single branch point. For these paths either (or both) of conditions (i) and (ii) must be satisfied. Condition (ii) implies [cf., Eq. (23)] that $k_\alpha(x) = k_\beta(x)$; i.e., the Stokes' line has reached a branch point connecting the α and β sheets of the z manifold. Recalling that the integrand of Eq. (23) is negative definite on the steepest descent path, we see that condition (i) can only be satisfied when the arc length along the steepest descent path between x_α and x_β vanishes. Since this path must encircle a branch point connecting sheet α to sheet β, it is clear that condition (i) can only be satisfied when the Stokes' line reaches this branch point. Hence, we conclude that the Stokes' lines associated with the steepest descent paths from x_α that encircle a single turning point before returning to x_β must pass through a branch point $x^T_{\alpha\beta}$ that connects sheets α and β of the z manifold. These branch points may be found by solving the simultaneous equations

$$\epsilon(x^T_{\alpha\beta},\, k^{\,T}_{\alpha\beta}) = 0$$

and

$$\epsilon_k(x^T_{\alpha\beta},\, k^{\,T}_{\alpha\beta}) = 0 \tag{26}$$

for $x^T_{\alpha\beta}$ and

$$k^{\,T}_{\alpha\beta} \equiv k(x^T_\alpha) = k(x^T_\beta).$$

In the vicinity of this branch point the two roots $k_\alpha(x)$ and $k_\beta(x)$ may be approximated by

$$k_{\alpha\beta}(x) \approx k^{\,T}_{\alpha\beta} \pm i \times [2(z - x^T_{\alpha\beta})\epsilon_x(k^{\,T}_{\alpha\beta}, x^T_{\alpha\beta})/\epsilon_{kk}(k^{\,T}_{\alpha\beta}, x^T_{\alpha\beta})]^{1/2}, \tag{27}$$

which directly exhibits the branch point in the function $k(z)$ at $z = x^T_{\alpha\beta}$. The two roots in (27) correspond to the dominant α root and the subdominant β root that merge at $x^T_{\alpha\beta}$, $k^{\,T}_{\alpha\beta}$. Although Eq. (25) is singular at $x^T_{\alpha\beta}$, the behavior of the Stokes' lines in the neighborhood of $x^T_{\alpha\beta}$ can be analyzed by expanding $g_{\alpha\beta}(x)$ for x near $x^T_{\alpha\beta}$.

It can readily be shown that three Stokes' lines emerge from $x^T_{\alpha\beta}$ as in Fig. 2. Along any given Stokes' line one root is dominant and the other is subdominant. If we would analytically continue the roots in an arc around the turning point in the counterclockwise direction as shown in Fig. 2, the root $k_\alpha(k_\beta)$ that is dominant (subdominant) on the Stokes' line $(x^T_{\alpha\beta}, a)$ becomes subdominant (dominant) on the adjacent Stokes' line $(x^T_{\alpha\beta}, b)$. If the arc makes a complete circuit, the analytic continuation of the original dominant (subdomin-

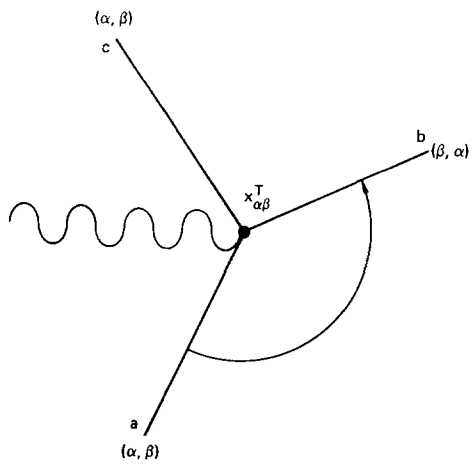


FIG. 2. Structure of Stokes' line about a turning point.

ant) root will be subdominant (dominant) when we return to the original Stokes' line. Hence, the need for the branch cut emerging from the turning point as shown. These turning points are a generalization of the simple turning points found in the WKB theory of second- order ordinary differential equations that has been discussed in previous works.[5,6,25] We will call the Stokes' lines emerging from these simple turning points "primary Stokes' lines."

Figure 3 indicates the topology of steepest descent contours in the vicinity of a primary Stokes' line. When x is immediately to the left of the Stokes' lines $(x^T_{\alpha\beta}, a)$, the steepest descent contour through the dominant saddle at $z = x_\alpha$ [which is shown as a dashed line in Fig. 3(a)], connects remote region I to remote region II. The dotted line in Fig. 3(a) represents the steepest descent contour through the subdominant saddle at $z = x_\beta$. This contour connects remote region III to remote region II. Figure 3(b) shows the configuration of the steepest descent path when x lies on the Stokes' line, whereas just to the right of Stokes' line $(x^T_{\alpha\beta}, a)$, Fig. 3(c) shows that the steepest descent path through $x = x_\alpha$ connects remote region I to remote region III. However, the steepest descent contour of the subdominant solution still connects remote region III to remote region II and is only slightly different than that shown in Fig. 3(a). For x on the Stokes' line, as in Fig. 3(b), the steepest descent path from the dominant saddle point returns to the original x point $(z = x_\beta)$, and for z beyond x_β the steepest descent path bifurcates and follows the subdominant descent contour to either remote region II or remote region III.

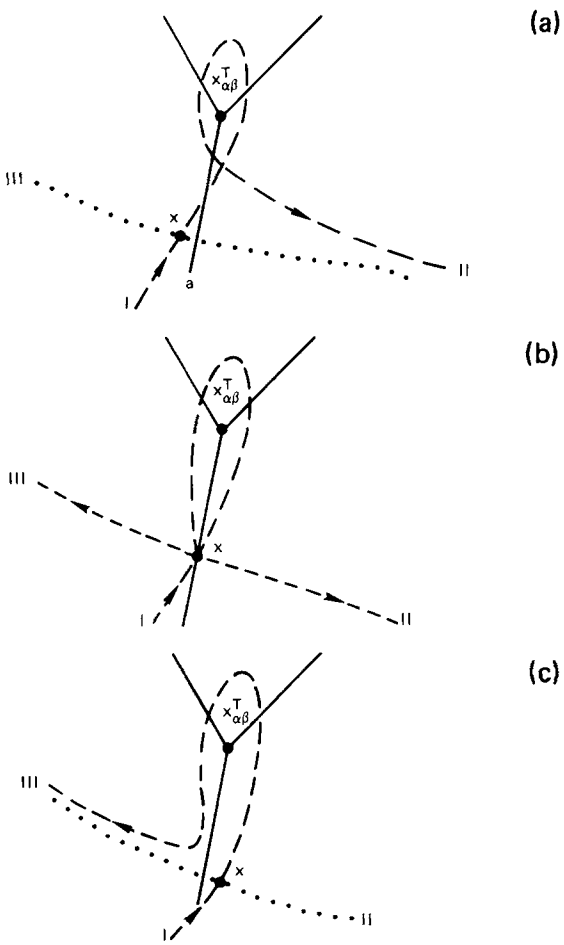


FIG. 3. Schematic diagram of steepest descent contours for a point x near a Stokes' line: (a) to the clockwise side of the line, (b) on the Stokes' line, and (c) to the counterclockwise side.

An additional class of Stokes' lines can be generated from the crossing of primary Stokes' lines. Such crossings can occur if the characteristic function $\epsilon(k,x)$ is cubic or higher order in k. It is only when $\epsilon(k,x)$ is quadratic, corresponding to second-order differential equations, that crossings are ruled out. Consider the Stokes' lines of Fig. 4(a), where $x^T_{\alpha\beta}$ is the turning point of roots k_α and k_β, and where k_α is dominant with respect to k_β. Suppose the Stokes' line $\mathcal{S}_{\alpha\beta}$ crosses a Stokes' line $\mathcal{S}_{\beta\gamma}$ emerging from $x^T_{\beta\gamma}$, where k_β is dominant with respect to k_γ. Such a crossing will be referred to as an ordered crossing. In contrast, an unordered crossing would arise if, at the crossing point, both the α and γ roots were dominant (or subdominant) with respect to β on their respective Stokes' lines. Note that for an ordered crossing we can define the relative dominance of the three modes: α dominant with respect to β, dominant with respect to γ. In an unordered crossing there is no relative dominance of all three modes. Crossings in which the Stokes' lines do not share a common root are always unordered crossings.

In the case of ordered crossings we find that a new Stokes' line (a secondary Stokes' line) $\mathcal{S}_{\alpha\gamma}$—indicated by the line x_q,c) in Fig. 4(a)—emerges from the crossing point x_q. This is shown by first considering the saddle points $x(1)$ and $x(2)$ in Fig. 4(a). The point $x(1)$ is the saddle point of a steepest descent contour that emerges from remote region I, goes through the saddle point at $z = x_\alpha(1)$, loops the turning point $x^T_{\alpha\beta}$, and returns to $z = x_\beta(1)$. Then the steepest descent path bifurcates to either remote region II or remote region III. Similarly, the point $x(2)$ is the saddle point of a steepest descent contour that emerges from remote region III, goes through the saddle point $z = x_\beta(2)$, loops the turning point $x^T_{\beta\gamma}$, returns to $z = x_\gamma(2)$, and then bifurcates to either remote region II or remote region IV.

Now let us follow this procedure on the point $x = x_q$, where $x(1) = x(2)$. In Fig. 4(b) we observe that starting from remote region I the steepest descent path goes through $(x_q)_\alpha$, loops $x^T_{\alpha\beta}$, and returns to $z = (x_q)_\beta$. There the steepest descent path bifurcates to either remote region III or to a path that must loop $x^T_{\beta\gamma}$, since x_q is on the (β,γ) Stokes' line. This path loops $x^T_{\beta\gamma}$ and returns to $(x_q)_\gamma$, where it bifurcates to either remote region II or remote region IV. Hence, we have shown that the point x_q is an element of a Stokes' line $\mathcal{S}_{\alpha\gamma}$, since a steepest descent path connects the dominant saddle $z = (x_q)_\alpha$ to $z = (x_q)_\gamma$.

Other points on this Stokes' line may be obtained by noting that

$$g_{\alpha\gamma}(x) = i\Phi(x_\alpha, x_\gamma)$$

must be real on $\mathcal{S}_{\alpha\gamma}$. Hence, the remaining elements of $\mathcal{S}_{\alpha\gamma}$ can be obtained by integrating a first-order ordinary differential equation in the form of Eq. (25)—with $k_\beta(x)$ replaced by $k_\gamma(x)$—using x_q as the initial condition and integrating in the direction of decreasing Re $g_{\alpha\gamma}$. Note that at the crossing

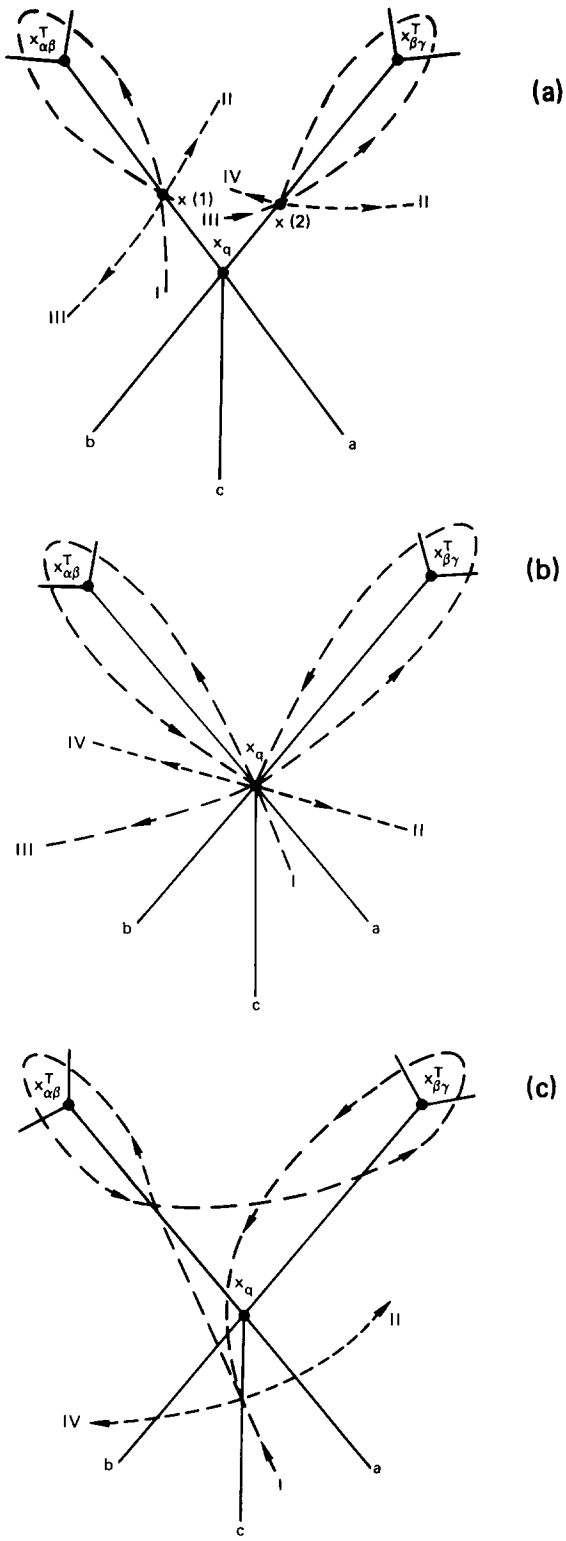


FIG. 4. Schematic diagram of steepest descent contours in the case of an ordered crossing of Stokes' lines: (a) above crossing point, (b) at crossing point, and (c) below crossing point.

point

$$g_{\alpha\gamma}(x_q) = g_{\alpha\beta}(x_q) + g_{\beta\gamma}(x_q)$$

is real and negative as it must be on a Stokes' line. The steepest descent path through a point x on $\mathcal{S}_{\alpha\gamma}$ is shown in Fig. 4(c). This path loops both $x^T_{\alpha\beta}$ and $x^T_{\beta\gamma}$.

In Appendix A we show that the Stokes' line generated by such a differential equation has the topology shown in Figs. 1 and 4 for the line (x_q, c); i.e., the direction of decreasing $g_{\alpha\gamma}$ for the Stokes' line is contained within the arc formed by the intersection of $\mathcal{S}_{\alpha\beta}$ and $\mathcal{S}_{\beta\gamma}$; the last two lines being directed towards decreasing g.

Equation (25) could also be used to generate a curve involving the α and γ roots in the direction of increasing $g_{\alpha\gamma}$ from x_q. However, in the next section we shall show that such a curve is *not* a Stokes' line because the steepest descent path from a point $z = (x_s)_\alpha$ on this curve does not return to $z = (x_s)_\gamma$.

5. CONNECTION RULES AND STOKES' MULTIPLIERS

In this section we discuss the form that the asymptotic solution takes upon passing a Stokes' line. Connection rules for continuing asymptotic solutions across Stokes' lines are derived, and the Stokes' constants are evaluated. We demonstrate that upon crossing primary Stokes' lines we recover previously obtained connection formulas. Upon crossing higher-order Stokes' lines we obtain the connection rules necessary to achieve asymptotic representations that are independent of path.

Suppose that in region 1 of the complex x plane we are examining a solution $\phi_\alpha(x)$ whose asymptotic representation is a single WKB wavelet of wave number $k_\alpha(x)$. In this region the contour integral in Eq. (12) is evaluated along the steepest descent path of the function $\exp[i\Phi_\alpha(z,x)]$ through the saddle point $z = x_\alpha$ from remote region I to remote region II. Recall that the WKB wavelet is obtained by using the method of steepest descent to evaluate the contour integral representation of $\phi_\alpha(x)$,

$$\phi_\alpha(x) = A \exp\left[i \int_{x_0}^{x} dz'\, k_\alpha(z') + ik_0 x_0 \right] \times \int_{C_\alpha(z)} \frac{dz}{(2\pi)^{1/2}} \frac{\epsilon_x^{1/2}(k,z)}{\epsilon_k(k,z)} \exp[i\Phi(z,x)] \,. \quad (12')$$

To establish a sign convention for $\epsilon_k^{1/2}$, let us first evaluate Eq. (12') near a Stokes' line that borders region 1, where α is the dominant wavelet on Stokes' line. Near such a line we know that one of the legs of the steepest descent contour loops the various branch points and then nearly returns to x_β. Let us assume that the contour $C_\alpha(a)$ starts in remote region I, goes through x_α, loops the various branch points, returns to $Z = x_\beta$, and then proceeds to remote region II.

Now, if we evaluate Eq. (12') by the steepest descent method, we find

$$\phi_\alpha(x) = A \exp\left[i \int_{x_0}^{x} dz' k_\alpha(z') + ik_0 x_0 \right] \times \frac{\exp(i\pi/4)}{\epsilon_k^{1/2}[k_\alpha(x),x]} \int_{-\infty}^{\infty} \frac{dV_\alpha}{(2\pi)^{1/2}} \exp\left(-\frac{V_\alpha^2}{2} \right), \quad (28)$$

where we have used

$$\frac{\partial^2 \Phi}{\partial z^2}(x,z)|_{z=x} = -\frac{dk_\alpha}{dx}(x) = \frac{\epsilon_x(k_\alpha,x)}{\epsilon_k(k_\alpha,x)}$$

and we made the substitution

$$V_\alpha = \exp(-i\pi/4)\frac{\epsilon_x^{1/2}[k_\alpha(x),x_\alpha]}{\epsilon_k^{1/2}[k_\alpha(x),x_\alpha]}(z-x).$$

The convention used in choosing the branch of $\epsilon_k^{1/2}(k_\alpha,x_\alpha)$ for a dominant mode is that V_α increases as the path goes through the saddle and heads towards the branch points, which it loops. As x_α varies away from the Stokes' line, the branch of $\epsilon_k^{1/2}(k,x)$ is analytically continued; there will be a one-to-one correspondence of this branch and the *continuously* changing steepest descent contour through x_α.

For convenience we define

$$A = \frac{\exp(-i\pi/4)\exp(-ik_0x_0)}{\epsilon_k^{1/2}(k_\alpha,x_0)},$$

so that

$$\phi_\alpha(x) = \frac{\epsilon_k^{1/2}(k_0,x_0)}{\epsilon_k^{1/2}(k,x)}\exp\left[i\int_{x_0}^{x}k_\alpha dz\right] \equiv\langle k_\alpha|x_0,x\rangle, \tag{29}$$

where the branch of $\epsilon_k^{1/2}(k_\alpha,x_0)$ is the analytic continuation of $\epsilon_k^{1/2}(k,x)$. Thus,

$$\lim_{x\to x_0}\left[\frac{\epsilon_k^{1/2}(k_\alpha,x_0)}{\epsilon_k^{1/2}(k_\alpha,x)}\right] = 1, \tag{30}$$

and the limit is taken along a path that is contained in region 1. We observe that the WKB wavelet $\langle k_\alpha|x_0,x\rangle$ satisfies a propagation relation

$$\langle k_\alpha|x_0,x\rangle = \langle k_\alpha|x_0,x_1\rangle\langle k_\alpha|x_1,x\rangle.$$

We can evaluate the asymptotic form of $\phi_\alpha(x)$ on the other side of the Stokes' line. To evaluate the subdominant contribution to the asymptotic form, it is necessary to study in more detail the behavior of the steepest descent path from the saddle point of the function $\exp[i\Phi(z,x_\alpha)]$ when x is in the neighborhood of $\mathcal{S}_{\alpha\beta}$. Note that for $x\in\mathcal{S}_{\alpha\beta}$ we have from Eq. (24) that $\Phi(x_\beta,x_\alpha) = -ig_{\alpha\beta}(x)$, with $g_{\alpha\beta}$ real and negative. Recall that the direction of the Stokes' line is the direction along which $g_{\alpha\beta}$ decreases (i.e., $|g_{\alpha\beta}|$ increases). Then, facing in the direction of the Stokes' line, the region to the right (left) is the clockwise (counterclockwise) side of the Stokes' line. Since $\mathrm{Im}\, g_{\alpha\beta} = 0$ on the Stokes' line, it follows from analyticity that $\mathrm{Im}\, g_{\alpha\beta}$ is greater than (less than) zero on the clockwise (counterclockwise) side of the Stokes' line.

We now consider the steepest descent path needed to evaluate ϕ_α at a nearby point $x+\delta x$, where $x\in\mathcal{S}_{\alpha\beta}$. The dominant saddle is now at $x_\alpha+\delta x$. When $x\in\mathcal{S}_{\alpha\beta}$ and δx is small, the steepest descent path from the dominant saddle must return to the neighborhood of x_β along the curve δz such that

$$\mathrm{Re}\,\Phi(x_\beta+\delta z,x_\alpha+\delta x) = 0.$$

Expanding in a Taylor series and keeping only the leading terms yields

$$\Phi(x_\beta+\delta z,x_\alpha+\delta x) = -ig_{\alpha\beta}(x+\delta x) + i/2V_\beta^2, \tag{31}$$

where

$$V_\beta = \exp(-i\pi/4)\frac{\epsilon_x^{1/2}(k_\beta,x_\beta)}{\epsilon_k^{1/2}(k_\beta,x_\beta)}(\delta z-\delta x) = \exp(-i\pi/4)\frac{\epsilon_x^{1/2}(k_\beta,x_\beta+\delta x)}{\epsilon_k^{1/2}(k_\beta,x_\beta+\delta x)}(z-x_\beta-\delta x). \tag{32}$$

Thus, the subdominant saddle on sheet β is at $V_\beta = 0$, and the real V axis is the steepest descent path passing through the subdominant saddle at $x_\beta+\delta x$ [see Fig. 5(a)]. The contour of the subdominant saddle connects remote region II to remote region III. Therefore, we define the branch $\epsilon_k^{1/2}(k_\beta,x_\beta)$ for the *subdominant* β mode to be the one that allows the contour to go from remote region III to remote region II, as V_β goes from $-\infty$ to ∞.

Now, in the neighborhood of x_β (a point on Stokes' line $\mathcal{S}_{\alpha\beta}$), the steepest descent path from the dominant saddle at $x_\alpha+\delta x$ must satisfy $\mathrm{Re}\, d\Phi = 0$, from which Eq. (31) gives

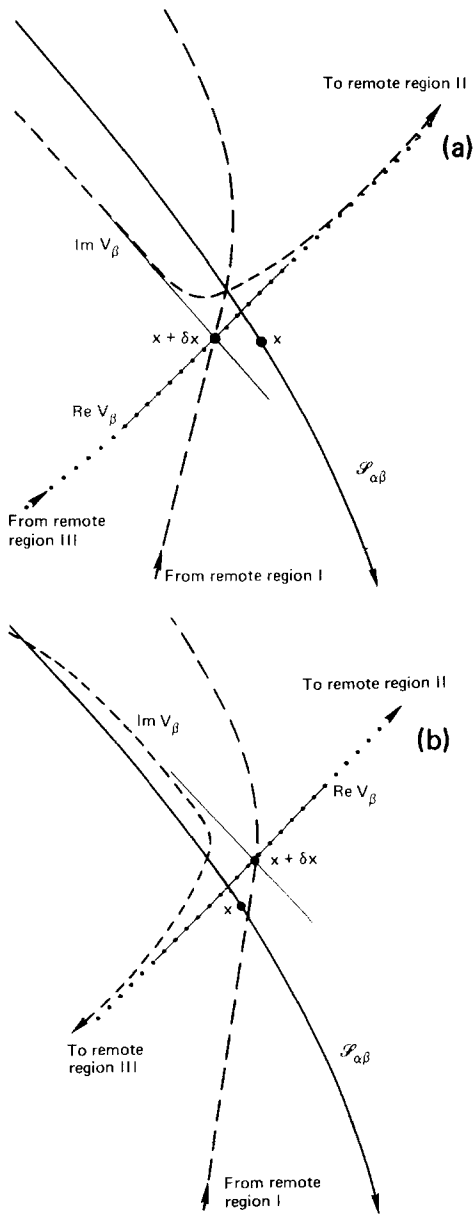


FIG. 5. Local transformation of coordinates near a Stokes' line: (a) to the clockwise side of the line and (b) to the counterclockwise side.

the hyperbola

$$(V_\beta)_R(V_\beta)_I = \operatorname{Im} g_{\alpha\beta}(x+\delta x), \tag{33}$$

where

$$(V_\beta)_R = \operatorname{Re} V_\beta \quad \text{and} \quad (V_\beta)_I = \operatorname{Im} V_\beta.$$

Suppose that the region on the clockwise (counterclockwise) side of the Stokes' line is region 1. We first consider δx such that $x+\delta x$ lies in region 1. Since $g_{\alpha\beta}(x)$ is a negative real on $\mathcal{S}_{\alpha\beta}$, and $g_{\alpha\beta}$ is an analytic function of x, it follows that $\operatorname{Im} g_{\alpha\beta}$ is greater than (less than) zero on the clockwise (counterclockwise) side of the Stokes' line. Hence, the right-hand side of Eq. (33) is positive (negative) in region 1. As in region 1 the steepest descent contour from x_α must ultimately head to remote region II, then by our convention, V_R must be positive. Thus, the steepest descent contour approaches the subdominant saddle (before it veers away) along the branch that asymptotes to the positive (negative) imaginary V axis [as shown in Fig. 5(a)]. As the steepest descent contour is directed from remote region I to remote region II and misses the saddle x_β, we only pick up the saddle point contribution at $z = x_\alpha$ in the asymptotic evaluation of $\phi_\alpha(x)$ in region 1.

We now follow the solution $\phi_\alpha(x)$ across the Stokes' line $\mathcal{S}_{\alpha\beta}$ from region 1 to region 2; i.e., we continue the solution $\phi_\alpha(x)$ across $\mathcal{S}_{\alpha\beta}$ in the counterclockwise (clockwise) sense. The right-hand side of Eq. (33) is now negative (positive) so that the steepest descent path from the dominant saddle [which still must approach the subdominant saddle along the positive (negative) imaginary V_β axis] now follows the *negative* $\operatorname{Re} V_\beta$ axis to remote region III, as shown in Fig. 5(b).

The independent solution $\phi_\alpha(x)$ in region 1 was defined by an integration contour that connected remote regions I and II. Hence, if we are to continue this solution across $\mathcal{S}_{\alpha\beta}$ into region 2, we may choose an integration contour that first follows the steepest descent path over the dominant saddle from remote region I to remote region III and then returns to remote region II by following the $\operatorname{Re} V$ axis across the subdominant saddle at $x_\beta + \delta x$. Our integration contour now crosses two saddles, and an evaluation of this solution by the method of steepest descents now yields two WKB wavelets:

$$\phi_\alpha(x) = \langle k_\alpha | x_0, x\rangle \left[1 + \frac{\epsilon_k^{1/2}(k_\alpha, x)}{\epsilon_k^{1/2}(k_\beta, x)} \exp\left(i\int_{x_\alpha}^{x_\beta} k\,dz\right)\right]. \tag{34}$$

Equation (34) gives the appropriate asymptotic form of the solution $\phi_\alpha(x)$ in region 2. It is convenient to rewrite this asymptotic solution in terms of WKB wavelets $\langle k | x_1, x_2\rangle$ and Stokes' multipliers. We first assume that $\mathcal{S}_{\alpha\beta}$ is a primary Stokes' line. The steepest descent path connecting x_α and x_β then encloses a single turning point $x_{\alpha\beta}^T$, and the asymptotic form of $\phi_\alpha(x)$ in region 2 may be rewritten as

$$\begin{aligned}\phi_\alpha(x) &= \langle k_\alpha | x_0, x\rangle \left[1 + \frac{\epsilon_k^{1/2}(k_\alpha, x)}{\epsilon_k^{1/2}(k_\alpha, x_{\alpha\beta}^T)} \exp\left(i\int_{x_\alpha}^{x_{\alpha\beta}^T} k\,dz\right)\right] \\ &\quad \times \frac{\epsilon_k^{1/2}(k_\alpha, x_{\alpha\beta}^T)}{\epsilon_k^{1/2}(k_\beta, x_{\alpha\beta}^T)} \exp\left(i\int_{x_{\alpha\beta}^T}^{x_\beta} k\,dz\right) \frac{\epsilon_k^{1/2}(k_\beta, x_{\alpha\beta}^T)}{\epsilon_k^{1/2}(k_\beta, x)} \\ &= \langle k_\alpha | x_0, x\rangle + \langle k_\alpha | x_0, x_{\alpha\beta}^T\rangle P_1 \langle k_\beta | x_{\alpha\beta}^T, x\rangle.\end{aligned} \tag{35}$$

The Stokes' multiplier P_1 is given by

$$P_1 = \lim_{\delta\to 0}\left[\epsilon_k^{1/2}(k_\alpha, x_{\alpha\beta}^T + \delta)/\epsilon_k^{1/2}(k_\beta, x_{\alpha\beta}^T + \delta)\right], \tag{36}$$

where δ is chosen such that $x_{\alpha\beta}^T + \delta$ lies on $\mathcal{S}_{\alpha\beta}$. The branches of $\epsilon_k^{1/2}(k, x_{\alpha\beta}^T + \delta)$ have been chosen such that

$$\lim_{x\to x_{\alpha\beta}^T + \delta}\left[\epsilon_k^{1/2}(k_\alpha, x)/\epsilon_k^{1/2}(k_\alpha, x_{\alpha\beta}^T + \delta)\right] = 1 \tag{37}$$

and

$$\lim_{x\to x_{\alpha\beta}^T + \delta}\left[\epsilon_k^{1/2}(k_\beta, x)/\epsilon_k^{1/2}(k_\beta, x_{\alpha\beta}^T + \delta)\right] = 1. \tag{38}$$

In order to avoid encircling the branch points of $k(x)$, the limit in Eqs. (37) and (38) is taken along a path that passes through region 2 from x to the Stokes' line $\mathcal{S}_{\alpha\beta}$ and then follows $\mathcal{S}_{\alpha\beta}$ to $x_{\alpha\beta}^T + \delta$.

In Appendix B we show that

$$P_1 = \pm i, \tag{39}$$

where the $+$ ($-$) sign is to be used when crossing a primary Stokes' line in the counterclockwise (clockwise) sense. Notice that the Stokes' multiplier P_1 depends only on parameters local to the turning point. Previous analysis (e.g., that of Ref. 6) obtained P_1 by using either the properties of the Airy equation or the Furry method.[1]

The asymptotic solution in region 2 is then

$$\phi_\alpha(x) = \langle k_\alpha | x_0, x\rangle \pm i\langle k_\alpha | x_0, x_{\alpha\beta}^T\rangle\ \langle k_\beta | x_{\alpha\beta}^T, x\rangle. \tag{40}$$

It has been observed by Miller[31] (for second-order differential equations) and Watson[27] (for integral equations) that the subdominant wavelet in Eq. (40) can also be obtained by analytically continuing the wavelet $\langle k_\alpha | x_0, x\rangle$ around the turning point $x_{\alpha\beta}^T$ in the clockwise (counterclockwise) sense. The factor $+i(-1)$ then arises from continuing $\epsilon_k^{1/2}$ [which appears in $\langle k_\alpha | x_0, x\rangle$, cf. Eq. (39)] around the branch point at $x_{\alpha\beta}^T$. Hence, a pseudophysical interpretation of Eq. (40) is that the asymptotic form of $\phi_\alpha(x)$ in region 2 is a superposition of both the wavelet propagating with wave number k_α from x_0 directly across $\mathcal{S}_{\alpha\beta}$ to x, and the wavelet that has propagated from x_0 around the turning point at $x_{\alpha\beta}^T$ and on to x in region 2.

We now need to evaluate the Stokes' multiplier P_2 on a secondary Stokes' line. For definiteness let the crossing of the primary Stokes' lines have the sense of Fig. 4; i.e., region 1 is on the clockwise side of $\mathcal{S}_{\alpha\beta}$, while region 2 is on the counterclockwise side of $\mathcal{S}_{\alpha\beta}$ and on the clockwise side of $\mathcal{S}_{\beta\gamma}$. In Fig. 6 these two regions are illustrated, as well as regions 3, 4, and 5, which we will discuss later. The reader can duplicate the following arguments if the sense of the crossing is reversed.

Now we note that a steepest descent path through a point x in the vicinity of x_q and in region 1 will come from remote region I through x_α, loop $x_{\alpha\beta}^T$, and approach x_β but veer off to the *counterclockwise* side of $\mathcal{S}_{\alpha\beta}$ along the positive V_β axis. After veering off in the counterclockwise direction, the contour has to asymptote to the steepest descent contour through $x_{q\beta}$. It is clear from the topology of Fig. 4(b) that the leg of the steepest descent contour heading in the *counterclockwise* sense past $x_{\alpha\beta}^T$ is the leg that loops $x_{\beta\gamma}^T$.

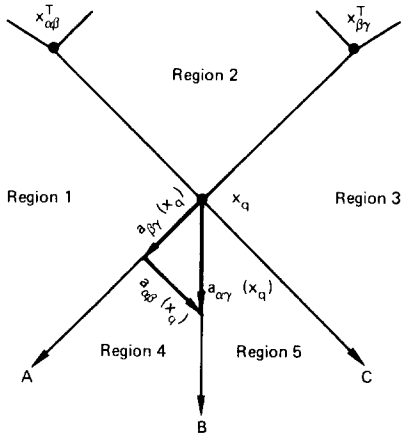


FIG. 6. Topology of secondary Stokes' lines.

Hence, in region 1, within the vicinity of x_q, the steepest descent contour heading towards remote region II will also loop $x^T_{\beta\gamma}$. Therefore, near an ordered crossing, the branch $\epsilon_k^{1/2}(k_\beta,x)$, defined from the subdominant convention, is identical to the branch of $\epsilon_R^{1/2}(k_\beta,x)$, defined from the dominant convention. By using analytic continuation the branch of $\epsilon_k^{1/2}(k_\beta,x)$ is defined unambiguously.

If we now continue x from region 1 to region 2 (still in the vicinity of x_q), the steepest descent contour from x_α will not quite return to x_β, but will veer off to the clockwise side to remote region III. Thus, the steepest contour in this region cannot return to x. We can conclude there cannot be a Stokes' line in region 2. On the other hand, x_q must lie on $\mathscr{S}_{\alpha\beta}$ because a steepest descent contour looping $x^T_{\alpha\beta}$ and $x^T_{\beta\gamma}$ exists there. We have derived the differential equation in the previous section that generates a line from x_q. Hence, x_q must be the initial condition for this line, and the line must have the topology shown in Fig. 6, where we see $\mathscr{S}_{\alpha\beta}$ separates regions 4 and 5 but does not enter region 2.

The beginning of a Stokes' line at x_q is natural from another point of view. Suppose we consider a point x on $\mathscr{S}_{\alpha\gamma}$, which is shown in Fig. 4(c) below x_q. If we follow the steepest descent path from $x = x_\alpha$, we loop both $x^T_{\alpha\beta}$ and $x^T_{\beta\gamma}$ and then return to x_γ. As the point x moves towards x_q, along $\mathscr{S}_{\alpha\gamma}$, we observe that the segment interval between x_α and x_γ of the steepest descent curve deforms continuously as x is varied. However, when we reach x_q, a bifurcation of the contour occurs at $x_{q\beta}$ after looping only $x^T_{\alpha\beta}$. Above x_q (along virtual $\mathscr{S}_{\alpha\gamma}$) the steepest descent contour through x_α will loop $x^T_{\alpha\beta}$ but then veer away from x_β and head towards remote region III.

We are now ready to evaluate the asymptotic form of $\phi_\alpha(x)$ when x crosses a secondary Stokes' line in the counterclockwise (clockwise) sense. (We now resume the convention of using parentheses to quote results for x crossing $\mathscr{S}_{\alpha\beta}$ in the clockwise sense.)

Up to a point, the evaluations of the asymptotic forms for primary Stokes' lines carry over to secondary Stokes' lines. Contour C_α connects remote region I to remote region II on the clockwise (counterclockwise) side of $\mathscr{S}_{\alpha\gamma}$. If we start with C_α, we pass through x_α, then loop $x^T_{\alpha\beta}$ and $x^T_{\beta\gamma}$, do not quite return to x_γ but veer to the counterclockwise (clockwise) side, and finally head to remote region II. The saddle-point evaluation of the integral on this contour gives Eq. (29).

We now define V_γ as

$$V_\gamma = \exp(-i\pi/4)\frac{\epsilon_x^{1/2}(k_\alpha,x_\gamma)}{\epsilon_k^{1/2}(k_\gamma,x_\gamma)}(z-x_\gamma).$$

The branch of $\epsilon_k^{1/2}(k_\gamma,x_\gamma)$ is chosen such that $V_\gamma \to +\infty$ as the subdominant steepest descent contour from remote region IV heads through x_α to remote region II. We note that in the vicinity of x_q, all steepest descent contours passing through x_α will head from remote region IV to remote region II because no bifurcation occurs on the steepest descent path of the most subdominant mode.

Now when x is on the other side of $\mathscr{S}_{\alpha\gamma}$, the steepest descent path veers away from x_γ towards the clockwise (counterclockwise) direction towards remote region IV. To recover the original solution, we must return on the steepest descent path between remote regions IV and II, where V_γ goes from $-\infty$ to ∞. The new contribution yields an asymptotic form just like Eq. (34):

$$\phi_\alpha(x) = \langle k_\alpha | x_0,x\rangle \times\left[1+\frac{\epsilon_k^{1/2}(k_\alpha,x)}{\epsilon_k^{1/2}(k_\gamma,x)}\exp\left(i\int_{x_\alpha}^{x_\gamma}k(z')dz'\right)\right]. \tag{41}$$

In the integral $\int_{x_\gamma}^{x_\alpha}k(z')dz'$ the path has to loop both $x^T_{\alpha\beta}$ and $x^T_{\beta\gamma}$.

Equation (41) can be rewritten in the form

$$\phi_\alpha(x) = \langle k_\alpha | x_0,x\rangle + P_2\langle k_\alpha | x_0,x^T_{\alpha\beta}\rangle \times\langle k_\beta | x^T_{\alpha\beta},x^T_{\beta\gamma}\rangle\ \langle k_\alpha | x^T_{\beta\gamma},x\rangle\ , \tag{42}$$

where

$$P_2 = \lim_{\delta_1\to 0}\left[\frac{\epsilon_k^{1/2}(k_\alpha,x^T_{\alpha\beta}+\delta_1)}{\epsilon_k^{1/2}(k_\beta,x^T_{\alpha\beta}+\delta_1}\right] \times\lim_{\delta_2\to 0}\left[\frac{\epsilon_k^{1/2}(k_\beta,x^T_{\beta\gamma}+\delta_2)}{\epsilon_k^{1/2}(k_\gamma,x^T_{\beta\gamma}+\delta_2)}\right]. \tag{43}$$

From Eq. (36) we have

$$P_2 = P_1\lim_{\delta_2\to 0}\left[\frac{\epsilon_k^{1/2}(k_\beta,x^T_{\beta\gamma}+\delta_2)}{\epsilon_k^{1/2}(k_\gamma,x^T_{\beta\gamma}+\delta_2)}\right]. \tag{44}$$

The ratio

$$\lim_{\delta_2\to 0}\left[\frac{\epsilon_k^{1/2}(k_\beta,x^T_{\beta\gamma}+\delta_2)}{\epsilon_k^{1/2}(k_\gamma,x^T_{\beta\gamma}+\delta_2)}\right] \tag{45}$$

looks like the Stokes' multiplier P_1 of the subdominant γ mode that is generated when a β mode crosses $\mathscr{S}_{\beta\gamma}$. However, we have to ascertain that the branches of $\epsilon_k^{1/2}$ are the proper ones. We have already noted that $\epsilon_k^{1/2}(k_\beta,x)$ is the same branch. It also follows that $\epsilon_k^{1/2}(k_\gamma,x_\gamma+\delta_2)$ will be defined as the same function whether it comes from x crossing $\mathscr{S}_{\alpha\beta}$ or $\mathscr{S}_{\beta\gamma}$. This follows because in the vicinity of $x = x_q$, where both $\mathscr{S}_{\alpha\gamma}$ and $\mathscr{S}_{\beta\gamma}$ can be crossed, $\epsilon_k^{1/2}(k_\gamma,x)$ is the branch associated with V_γ varying from $-\infty$ to ∞ as the contour integral heads from remote region IV to remote region II. Thus, $P_2 = (P_1)^2 = -1$.

The generalization to the Stokes' multiplier P_n for Stokes' lines whose steepest descent path loops n simple turning points is now apparent. The intermediate branches of $\epsilon_k^{1/2}$ may always be chosen such that P_n will involve n terms of the form (36), and hence,

$$P_n = (P_1)^n = (\pm i)^n, \tag{46}$$

where the $+(-)$ sign is to be used when crossing the nth-order Stokes' line in the counterclockwise (clockwise) sense.

6. UNIQUENESS OF ASYMPTOTIC REPRESENTATIONS

The rules for adding subdominant solutions on crossing Stokes' lines[1] ensure that the asymptotic properties of a function $\phi_\alpha(x)$ in a given region are independent of the path to that region. This uniqueness is in fact the basis of the Furry method[1] in determining the Stokes' multipliers on passing a primary Stokes' line. It should be emphasized that because of the branch cuts emanating from the turning points, the labeling of a WKB wavelet along two different paths that are separated by a turning point is not unique, even though the asymptotic properties of a particular solution are unique. The following examples demonstrate these points.

Consider Fig. 7, where three Stokes' lines emerging from x_{12}^T are shown and two paths from points x_a to x_c are shown: path $a(x_a \to x_b \to x_c)$ and path $b\,(x_a \to x_e \to x_d \to x_c)$. If

$$\phi_\alpha(x) \sim C\,\langle k_1|x_a,x\rangle \tag{47}$$

in the vicinity of x_a, and if $\langle k_1\rangle$ is dominant with respect to $\langle k_2\rangle$ along Stokes' line (x_{12}^T,x_b) and we follow path a, then we find that $\phi_\alpha(x)$ in the vicinity of x_c is of the form

$$\phi_\alpha(x) \approx C\,\langle k_1|x_a,x\rangle + i\langle k_1|x_a,x_{12}^T\rangle\,\langle k_2|x_{12}^T,x\rangle\,, \tag{48}$$

where we picked up the second WKB wavelet with the Stokes' multiplier i on crossing the Stokes' line (x_{12}^T,x_b).

On path b we do not obtain an additional WKB wavelet on crossing Stokes' line (x_{12}^T,x_e), as $\langle k_1\rangle$ must be subdominant with respect to $\langle k_2\rangle$ on this line. However, on crossing x_d we pick up a subdominant wavelet with a Stokes' multiplier $-i$, so that near x_c we have

$$\phi_\alpha(x) \sim C\,[\langle k_1|x_a,x\rangle - i\langle k_1|x_a,x_{12}^T\rangle\,\langle k_2|x_{12},x\rangle\,]\,. \tag{49}$$

The two forms of $\phi_\alpha(x)$ are identical because it can be shown[4] from the analytical properties of $\epsilon^{1/2}(k,x)$ that, in the vicinity of x_c,

$$\begin{aligned}&\langle k_1|x_a,x\rangle\,|_{\text{path 1}} = -i\langle k_1|x_a,x_{12}^T\rangle\,\langle k_2|x_{12}^T,x\rangle\,|_{\text{path 2}}\,,\\ &i\langle k_1|x_a,x_{12}^T\rangle\,\langle k_2|x_{12}^T,x\rangle\,|_{\text{path 1}} = \langle k_1|x_a,x\rangle\,|_{\text{path 2}}\,.\end{aligned} \tag{50}$$

Thus, we observe that although solutions are independent of path, the wavelet labeling depends on path. If two paths are not separated by turning points, then the same labels map to the same wavelets; if they are separated by turning points, it may be necessary to interchange labels in order to match the same wavelets.

For another example, suppose in Fig. 7 we start with a function

$$\phi_\beta(x) \sim C\,\langle k_2|x_a,x\rangle \tag{51}$$

in the vicinity of x_a. This wavelet is subdominant with respect to $\langle k_1\rangle$ on passing x_b, and no new wavelet is induced. Hence, in the vicinity of $x = x_c$ we have

$$\phi_\beta(x) \sim C\,\langle k_2|x_a,x\rangle. \tag{52}$$

However, along the upper path we find just to the right of x_e,

$$\phi_\beta(x) \sim C\,[\langle k_2|x_a,x\rangle - i\langle k_2|x_a,x_{12}^T\rangle\,\langle k_1|x_{12}^T,x\rangle\,]\,. \tag{53}$$

As we pass x_d, $\langle k_1\rangle$ induces a second $\langle k_2\rangle$ wavelet to yield

$$\begin{aligned}\phi_\beta(x) \sim C\,[&\langle k_2|x_a,x\rangle - i\langle k_2|x_a,x_{12}^T\rangle\,\langle k_1|x_{12}^T,x\rangle\\ &- \langle k_2|x_a,x_{12}^T\rangle\,\langle k_1|x_{12}^T,x_{12}^T\rangle\,\langle k_2|x_{12}^T,x\rangle\,]\\ = -iC\,&[\langle k_2|x_a,x_{12}^T\rangle\,\langle k_1|x_{12}^T,x\rangle\,]\,,\end{aligned} \tag{54}$$

where $\langle k_2\rangle$ has cancelled, and we are left only with $\langle k_1\rangle$. Using Eq. (50) we do see that $\phi_\beta(x)$ is independent of path and only the labeling of the two wavelets changes because of path.

It is apparent that bookkeeping is easier along path a because only one of the triplet of Stokes' lines emerging from a turning point is crossed and thus annihilation of wavelets cannot take place.

Now we address the original paradox mentioned in Sec. 2. We examine the asymptotic representation along two paths shown in Fig. 8. The Stokes' lines (x_{12}^T,x_g) and (x_{12}^T,x_e) are assumed to have an ordered crossing with $\langle k_1\rangle$ dominant with respect to $\langle k_2\rangle$ on the first line, $\langle k_2\rangle$ dominant with respect to $\langle k_3\rangle$ on the other line. Once the new Stokes' line emanating from x_q is recognized, the uniqueness of asymptotic representation is quite transparent. Suppose that in the vicinity of point x_a

$$\phi_\alpha(x) \approx C_\alpha\,\langle k_1|x_1,x\rangle\,. \tag{55}$$

Then along the upper path $(x_a \to x_b \to x_c \to x_d)$ we have just to the right of x_b

$$\phi_\alpha(x) \approx C_\alpha\,[\langle k_1|x_1,x\rangle + i\langle k_1|x_1,x_{12}^T\rangle\,\langle k_2|x_{12}^T,x\rangle\,]\,, \tag{56}$$

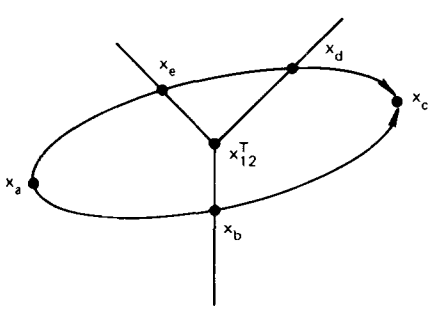


FIG. 7. Tracking of asymptotic solutions around turning points to various regions.

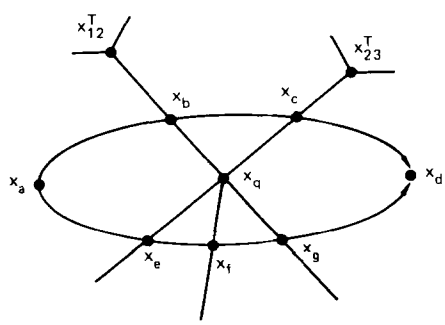


FIG. 8. Tracking of asymptotic solution around crossed ordered Stokes' lines to various regions.

and just to the right of x_c

$$\phi_\alpha(x) \approx C_\alpha [\langle k_1|x_1,x\rangle + i\langle k_1|x_1,x_{12}^T\rangle \langle k_2|x_{12}^T,x\rangle - \langle k_1|x_1,x_{12}^T\rangle \langle k_2|x_{12}^T,x_{23}^T\rangle \langle k_3|x_{23}^T,x\rangle] . \quad (57)$$

Similarly, along the lower path ($x_a \to x_e \to x_g \to x_d$) we have, just to the right of x_e,

$$\phi_\alpha(x) \approx C_\alpha \langle k_1|x_1,x\rangle ,$$

as $\langle k_3\rangle$ has to be induced by $\langle k_2\rangle$, and we do not yet have a $\langle k_2\rangle$. At the same time just to the right of x_f we have, by the rule of crossing a secondary Stokes' line,

$$\phi_\alpha(x) = C_\alpha [\langle k_1|x_1,x\rangle - \langle k_1|x_1,x_{12}^T\rangle \times \langle k_2|x_{12}^T,x_{23}^T\rangle \langle k_3|x_{23}^T,x\rangle] , \quad (58)$$

and just to the right of x_g we have

$$\phi_\alpha(x) = C_\alpha [\langle k_1|x_1,x\rangle - \langle k_1|x_1,x_{12}^T\rangle \langle k_2|x_{12}^T,x_{23}^T\rangle \langle k_3|x_{23}^T,x\rangle + i\langle k_1|x_1,x_{12}^T\rangle \langle k_2|x_{12}^T,x\rangle] . \quad (59)$$

Equations (57) and (59) have identical forms, and they are identical functions as their paths are not separated by turning points, the branch points of $\epsilon_k^{1/2}[k(x),x]$. Thus, the uniqueness of the asymptotic representations in different regions is established.

7. CONSTRUCTION OF GLOBAL DISPERSION RELATIONS

We now address the general problem of constructing the global dispersion relation. Suppose we have a $2N$ th-order (N can approach ∞) differential equation defined on the real axis from $-\infty < x < \infty$ and that the local dispersion relation (i.e., the characteristic equation) is a function of k^2. For a dispersion relation of $2N$ th order in k, there are N WKB wavelets that converge as $x \to -\infty$ on the real axis, i.e.,

$$\exp\left(i\int^x k_j\, dz\right) \to 0 \quad \text{as } x \to -\infty .$$

These converging wavelets may be labeled with the indices $j = 1\text{–}N$.

The other N wavelets have k's that are the negatives of the well-behaved group so that these WKB wavelets diverge as $x \to -\infty$. In the event that any root pairs have Im $k = 0$, then a physical criterion, such as outgoing wave energy at infinity, determines which root of the pair is allowable at infinity. In some physical problems where Im $k \ll$ Re k, the outgoing wave condition can take precedence over subdominance.[32]

As $x \to -\infty$, the eigenmodes may be represented as a linear combination of the N WKB wavelets that are well behaved at $-\infty$,

$$\phi(x) \sim \sum_{i=1}^{N} a_i \langle k_i|x_0,x\rangle .$$

The dispersion relation is obtained by first using our connection rules to continue this solution to $x = +\infty$, and then requiring that $\phi(x)$ be well behaved at $x = +\infty$.

For definiteness in this discussion we shall continue the solution along the real axis, although other paths may prove more convenient in a particular problem. This path then uniquely defines the branches of WKB wavelets as

$$\langle k_i|x_0,x\rangle \equiv \frac{\epsilon_k^{1/2}[k_i(x),x_0]}{\epsilon_k^{1/2}[k_i(x),x]} \exp\left[i\int_{x_0}^{x} k_i(x)\, dx\right] . \quad (60)$$

It is convenient to choose the phase reference point x_0 to lie on the real x axis at a large *positive* value at x, such that no Stokes' lines cross the real x axis between x_0 and ∞.

Upon continuing our solution from $x = -\infty$ to $x = +\infty$ along the real x axis, we find that the original superposition of N well-behaved wavelets at $x = -\infty$ becomes

$$\phi(x) \sim \sum_{i=1}^{N} \sum_{j=1}^{2N} a_i M_{ij}(\omega) \langle k_j|x_0,x\rangle . \quad (61)$$

The elements of the matrix M are determined from the WKB connection formulas. These matrix elements may be written in the form

$$M_{ij} = \sum_\alpha \left[\frac{\epsilon_k^{1/2}(k_j,x_0)}{\epsilon_k^{1/2}(k_i,x_0)}\right]_\alpha \exp\left(i\int_{C_\alpha} k(z)\, dz\right), \quad (62)$$

where $\{C_\alpha\}$ are contours in the z manifold that start at $(x_0)_i$ and terminate at $(x_0)_j$. The branch of $\epsilon_k^{1/2}$ is chosen such that $\epsilon_k^{1/2}(k_j,x_0)$ is the analytic continuation of $\epsilon_k^{1/2}(k_i,x_0)$ from $(x_0)_i$ to $(x_0)_j$ along C_α.

There are N wavelets that diverge as $x \to +\infty$. The requirement that our solution be well behaved at $+\infty$ is

$$\sum_{i=1}^{N} a_i M_{ij}(\omega) = 0 \quad (63)$$

for each value of j that labels a diverging wavelet at $x \to +\infty$. If we order the N diverging wavelets at $+\infty$ in one-to-one correspondence with the initial N wavelets at $-\infty$, we find that the eigenvalues of the system are determined by

$$|M_{ij}(\omega)| = 0, \quad (64)$$

where $|M_{ij}(\omega)|$ is the determinant of M_{ij}.

It follows from the definition of the determinant, together with the form of the individual matrix elements M_{ij} [cf. Eq. (62)], that this dispersion relation may be written as

$$0 = |M_{ij}(\omega)| = \sum_\beta \sigma_\beta \exp\left(i\oint C_\beta k\, dz\right), \quad (65)$$

where $\sigma_\beta = \pm 1$ (in every specific example we have investigated $\sigma_\beta = +1$). The $\{C_\beta\}$ are closed contours that weave around a particular set of branch points in the z manifold. The contours that are to be included in this sum, together with the sign of σ_β, are determined implicitly by the WKB connection rules. Note that a contour that encloses no branch points is allowed.

We have shown that the global dispersion relation is a functional of loop integrals, $\oint k\, dx$, in the complex plane. This is an important requirement for a WKB theory of higher-order equations. The global dispersion relation must be independent of the representation of the governing integral equation. One might have chosen the Fourier space representation of this equation to obtain the global dispersion relation by the WKB methods. An argument analogous to that presented here then leads to the conclusion that the global dispersion relation is a functional of loop integrals of the

form $\oint x(k)\,dk$. Since

$$\oint k(x)\,dx = -\oint x(k)\,dk, \tag{66}$$

these two dispersion relations are both functionals of the same action integral, and hence it is possible that they are identical, as a consistent theory demands.

8. UNRESOLVED PROBLEMS AND SPECULATIONS

Our presentation leaves several unresolved points of difficulty and leads to several speculative conclusions.

The first difficulty relates to Stokes' lines. In Sec. 3 we indicated that if we follow the steepest descent path from a point x_s on the Stokes' line $\mathscr{S}_{12}$ for the dominant mode k_1, we will eventually return to x_s for a subdominant value k_2. We can ask whether this will always occur as we follow the Stokes' line Re $\int^x (k_1 - k_2)dz$ away from the turning point (or points if it is a higher-order Stokes' line).

In the text we only considered saddle points that arise when $z = x$. However, we did show that another saddle point can occur if the steepest descent contour leads to a point where $\epsilon_x(k,x) = 0$. If we limit ourselves to problems where $\epsilon_k(k,x) = 0$ and $\epsilon_x(k,x) = 0$ are well separated, then such a bifurcation in the steepest contour cannot occur when the $x = z$ saddle is on the Stokes' line and x is sufficiently close to the turning point.

However, far from the turning point it is possible that the steepest descent contour from a point x_a on a Stokes' line can lead to a point $\epsilon_x(k,x) = 0$, in which case an additional bifurcation of the steepest descent path arises. Then the steepest descent path for points on the Stokes' line [as defined by Re $\int_{x_{12}^T}^x (k_1 - k_2)\,dz = 0$] beyond x_a (such as $x_a + \delta x$) will not return to $x_a + \delta x$.

We then ask whether a subdominant solution should be added when a dominant solution is followed past $x_a + \delta x$. The speculative answer is that the Stokes' line retains its character. The argument is to note that the steepest descent contour fails to return to $x_a + \delta x$ because of a bifurcation in the steepest descent path at $\epsilon_x(k,x) = 0$, precisely the point where the approximate representation of the integrand of Eq. (11) is bad, so that the bifurcation cannot be trusted.

We note that we may modify our rules slightly and choose a path for evaluating the integral

$$\int dz \frac{\epsilon_x^{1/2}}{\epsilon_k} \exp[i\Phi(z,x)]$$

from $z = x_a + \delta x_a$ and $k = k_1(x_a + \delta x)$ to $z = x_a + \delta x_a$, $k = k_2(x_a + \delta x_a)$ which avoids encircling $\epsilon_x(k,x) = 0$ but follows steepest descent contours locally near $z = x_a + \delta x_a$. We would then pick up local saddle point contributions to the integral from $z = x_a + \delta x$, $k = k_1$, and $z = x_a + \delta x$, $k = k_2$, and we also would head to the correct remote regions at infinity. Then we would have rules identical to those stated. Further, these rules would still allow unique representations of the asymptotic forms.

The second problem is that since the global dispersion relation is a functional of the loop integral of the action, $\oint k(x)\,dx$, we have the possibility that the global dispersion relation is invariant as to whether the problem is solved in x-space or k-space since

$$\oint k\,dx = -\oint x(k)\,dk, \tag{67}$$

the latter form being the loop integral of the action in k-space. However, we have not been able to prove that the functional dependences derived by starting from x-space or k-space are indeed identical. A consistent theory demands that they are identical. We believe our rules should have this identity built into them, but the proof remains to be demonstrated.

A final difficulty is that we need to develop efficient algorithms to select the correct combination of loop integrals in the dispersion relation. The correct loop integrals are implicit functions of the boundary conditions. The prescription for obtaining the correct loop integrals by explicitly continuing solutions for $-\infty$ to $+\infty$ is correct but tedious. More efficient algorithms should be developed. Knoll and Schaefer[29] indicated a procedure for second-order differential equations where one chooses a path that crosses only zero or one Stokes' line from a given turning point. With higher-order differential equations this prescription appears to be just a guide and not a rule, as complications arise due to the crossing of Stokes' lines. Further study into this problem is needed.

ACKNOWLEDGMENTS

Fruitful discussions with William Sharp are gratefully acknowledged. This work was performed under the auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

APPENDIX A: SECONDARY STOKES' LINES

In Appendix A we discuss the topology of secondary Stokes' lines. Consider Fig. 6. We will assume that mode α is dominant with respect to mode β on the Stokes' line $(x_{\alpha\beta}^T, C)$. The crossing at x_q is assumed to be ordered. Hence, mode β is dominant with respect to mode γ on the Stokes' line $(x_{\beta\gamma}^T, x)$.

We consider the quantity $g_{\alpha\beta}(x)$ defined in Eq. (23). Note that Im $g_{\alpha\beta}$ vanishes on the Stokes' line $(x_{\alpha\beta}^T, C)$, while Re $g_{\alpha\beta}$ is monotonically decreasing. Since $g_{\alpha\beta}(x)$ is an analytic function, it follows that Im $g_{\alpha\beta}$ is greater than zero in regions 2 and 3 above the curve $(x_{\alpha\beta}^T, C)$ and that Im $g_{\alpha\beta}$ is less than zero in regions 1, 4, and 5 below this curve. Similarly, Im $g_{\beta\gamma}$ is greater than zero in regions 1 and 2 above the Stokes' line $(x_{\beta\gamma}^T, A)$, while Im $g_{\beta\gamma}$ is less than zero in regions 3, 4, and 5 below $(x_{\beta\gamma}^T, A)$.

Note that $g_{\alpha\gamma} = g_{\alpha\beta} + g_{\beta\gamma}$. Hence, Im $g_{\alpha\gamma}$ can vanish only when Im $G_{\alpha\beta}$ and Im $g_{\beta\gamma}$ have opposite signs; i.e., in regions 2 or 4 and 5 of Fig. 8. A necessary condition for an (α,γ) Stokes' line is that Im $g_{\alpha\gamma} = 0$. Hence, the secondary Stokes' line must lie in regions 2 or 4 and 5.

We can determine the direction of decreasing $g_{\alpha\gamma}$ on the (α,γ) Stokes' line at x_α by considering the vectors $\mathbf{a}_{\alpha\beta}$, $\mathbf{a}_{\beta\gamma}$, and $\mathbf{a}_{\alpha\gamma}$, where

$$\mathbf{a}_{\alpha\beta} = \mathrm{Im}(k_\alpha - k_\beta)\hat{\mathbf{x}} + \mathrm{Re}(k_\alpha - k_\beta)\,\hat{\mathbf{y}}$$

and similarly for $\mathbf{a}_{\beta\gamma}$ and $\mathbf{a}_{\alpha\gamma}$. Note from Eq. (25) that when

$x \in \mathcal{S}_{\alpha\beta}$, the vector $\mathbf{a}_{\alpha\beta}$ is tangent to the (α,β) Stokes' line. A similar condition exists for $\mathbf{a}_{\beta\gamma}$, $\mathbf{a}_{\alpha\beta}$. All of these vectors point in the direction of decreasing g; that is, $\mathbf{a}_{\alpha\beta}$ points away from $x^T_{\alpha\beta}$ while $\mathbf{a}_{\beta\gamma}$ points away from $x^T_{\beta\gamma}$.

It follows from the definition of the $\mathbf{a}$ vectors that at any point x,

$$\mathbf{a}_{\alpha\gamma}(x) = \mathbf{a}_{\alpha\beta}(x) + \mathbf{a}_{\beta\gamma}(x).$$

Evaluating these vectors at x_q, we see that $\mathbf{a}_{\alpha\gamma}(x_q)$ points downward from x_q toward region 4. Hence, $g_{\alpha\gamma}$ decreases on $\mathcal{S}_{\alpha\gamma}$ as one moves down from x_q into region 4, as indicated by the arrow in Fig. 6.

APPENDIX B: EVALUATION OF P_1

The Stokes' multiplier P_1 is given by [cf. Eq. (36)]

$$P_1 = \lim_{\delta\to 0}\left[\epsilon_k^{1/2}(k_\alpha, x^T_{\alpha\beta} + \delta)/\epsilon_k^{1/2}(k_\beta, x^T_{\alpha\beta} + \delta)\right], \tag{B1}$$

where the branches of $\epsilon_k^{1/2}$ must be chosen in accordance with Eqs. (37) and (38). It is evident from (B1) that the Stokes' multiplier P_1 depends only on properties local to the turning point. Near the turning point we may expand $\epsilon(k,x)$ as

$$\epsilon(k^T_{\alpha\beta} + \delta k, x^T_{\alpha\beta} + \delta) = \tfrac{1}{2}\epsilon_{kk}\,\delta k^2 + \epsilon_x\,\delta. \tag{B2}$$

Setting $\epsilon(k,x) = 0$, we find

$$\delta k = \pm i\,(2\epsilon_x/\epsilon_{kk})^{1/2}\,\delta^{1/2}. \tag{B3}$$

Without loss of generality, we may identify the dominant root k_α with the negative sign in Eq. (B3). The requirement that $x^T_{\alpha\beta} + \delta$ lie on $\mathcal{S}_{\alpha\beta}$ then gives

$$(2\epsilon_x/\epsilon_{kk})^{1/2} = |\tfrac{3}{4}\,g_{\alpha\beta}(x^T_{\alpha\beta} + \delta)|\delta^{-3/2}, \tag{B4}$$

which was defined in Eq. (23), where $g_{\alpha\beta}$ is always a negative real number on a Stokes' line.

The appropriate branches of $\epsilon_k^{1/2}(k, x^T_{\alpha\beta} + \delta)$ may be determined by examining the steepest descent paths from saddle points at $x = x^T_{\alpha\beta} = \delta$. On the αth sheet we find

$$(\epsilon_x/\epsilon_k)_\alpha = i|\tfrac{3}{8}\,g_{\alpha\beta}|\delta^{-2} \tag{B5}$$

and

$$(\epsilon_x/\epsilon_k)_\alpha^{1/2} = \pm\,\exp(i\pi/4)|\tfrac{3}{8}\,g_{\alpha\beta}|^{1/2}/\delta. \tag{B6}$$

Hence, V_α [cf. the discussion following Eq. (28)] is given by

$$V_\alpha = \pm\,|\tfrac{3}{8}\,g_{\alpha\beta}|^{1/2}\left(\frac{z - x_\alpha}{\delta}\right). \tag{B7}$$

We must choose the sign of the square root such that increasing V_α takes us from the saddle on the αth sheet toward $x^T_{\alpha\beta}$. If the minus sign is chosen in Eq. (B7), then V_α increases when $(z - x_\alpha)$ is antiparallel to δ, i.e., movement along the steepest descent path from x_α in the direction of increasing V_α takes us toward $x^T_{\alpha\beta}$ along the Stokes' line $\mathcal{S}_{\alpha\beta}$. Hence, we choose the minus sign in Eq. (B7) and therefore in Eq. (B6) as well.

Having evaluated $(\epsilon_x/\epsilon_k)^{1/2}$ on sheet α, we turn our attention to sheet β. On sheet β we have

$$(\epsilon_x/\epsilon_k) = -i|\tfrac{3}{8}\,g_{\alpha\beta}|/\delta^2 \tag{B8}$$

and

$$(\epsilon_x/\epsilon_k)^{1/2} = \pm\,\exp(-i\pi/4)|\tfrac{3}{8}\,g_{\alpha\beta}|^{1/2}/\delta. \tag{B9}$$

Hence, V_β [cf. Eq. (32)] is given by

$$V_\beta = \mp i|\tfrac{3}{8}\,g_{\alpha\beta}|^{1/2}\left(\frac{z - x_\beta}{\delta}\right). \tag{B10}$$

The direction of the positive imaginary V_β axis is antiparallel (parallel) to $\mathcal{S}_{\alpha\beta}$ if the upper (lower) sign is chosen in both Eqs. (B9) and (B10). This is the appropriate choice of sign when crossing $\mathcal{S}_{\alpha\beta}$ in the counterclockwise (clockwise) sense because the steepest descent path will be directed down (up) the positive (negative) imaginary V_β axis as it passes from $x^T_{\alpha\beta}$ towards the saddle on sheet β, as we required in the discussion following Eq. (32). Noting that

$$\frac{\epsilon_k^{1/2}(k_\alpha, x^T_{\alpha\beta} + \delta)}{\epsilon_k^{1/2}(k_\beta, x^T_{\alpha\beta} + \delta)} = \frac{(\epsilon_x/\epsilon_k)_\beta^{1/2}}{(\epsilon_x/\epsilon_k)_\alpha^{1/2}}, \tag{B11}$$

we have

$$P_1 = \pm i,$$

where the $+\,(-)$ sign is to be used when crossing a primary Stokes' line in the counterclockwise (clockwise) sense.

[1] W. H. Furry, Phys. Rev. **71**, 360 (1947).
[2] J. Heading, *An Introduction to Phase Integral Methods* (Wiley, New York, 1962).
[3] N. Froman and P. O. Froman, *JWKB Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965).
[4] H. L. Berk and D. L. Book, Phys. Fluids **12**, 649 (1969).
[5] T. Watanabe, H. Sanuki, and M. Watanabe, J. Phys. Soc. Jpn. **47**, 286 (1979).
[6] H. L. Berk and D. Pfirsch, J. Math. Phys. **21**, 2054 (1980).
[7] G. D. Birkhoff, Trans. Am. Math. Soc. **9**, 219 (1908).
[8] J. Tamarkin, Math. Z. **27**, 1 (1928).
[9] H. L. Turrittin, Am. J. Math. **58**, 364 (1936).
[10] H. L. Berk, M. N. Rosenbluth, and R. N. Sudan, Phys. Fluids **9**, 1606 (1966).
[11] O. E. H. Rydbeck, Commun. Pure Appl. Math. **4**, 129 (1951).
[12] W. Pfister, J. Geophys. Res. **58**, 29 (1953).
[13] K. G. Budden, *Radio Waves in the Ionosphere* (Cambridge U. P., Cambridge, 1961).
[14] T. H. Stix, Phys. Rev. Lett. **15**, 878 (1965).
[15] D. Gorman, Phys. Fluids **9**, 1262 (1966).
[16] R. E. Aamodt and D. L. Book, Phys. Fluids **9**, 143 (1966).
[17] R. Z. Sagdeev, *Plasma Physics* (IAEA, Vienna, 1965), p. 625.
[18] L. D. Pearlstein and D. Bhadra, Phys. Fluids **12**, 213 (1969).
[19] M. L. Brillouin, J. Phys. (Paris) **7**, 353 (1926).
[20] J. B. Keller, Ann. Phys. **4**, 180 (1958).
[21] V. Maslov, *Theorie des Perturbations* (Dunod, Paris, 1972).
[22] I. C. Percival, Adv. Chem. Phys. **36**, 1 (1977).
[23] H. Sanuki, T. Watanabe, and M. Watanabe, Phys. Fluids **23**, 158 (1980).
[24] G. G. Stokes, Trans. Cambridge Philos. Soc. **10**, 106 (1857).
[25] H. L. Berk *et al.*, *Plasma Physics and Controlled Nuclear Fusion Research* (IAEA, Vienna, 1969), Vol. 2, p. 151.
[26] J. Heading, Proc. Cambridge Philos. Soc. **53**, 419 (1957).
[27] D. Watson, Phys. Fluids **23**, 2485 (1981).
[28] W. M. Sharp (private communication).
[29] J. Knoll and R. Schaeffer, Ann. Phys. **97**, 307 (1976).
[30] R. B. Dingle, *Asymptotic Expansions* (Academic, New York, 1973).
[31] W. H. Miller, Adv. Chem. Phys. **25**, 69 (1974).
[32] H. L. Berk, L. D. Pearlstein, J. D. Callen, C. W. Horton, and M. N. Rosenbluth, Phys. Rev. Lett. **22**, 876 (1969).

Some more developments on operators in Krein space. The exponential map

E. Behnam
Mashhad University, Department of Mathematics, Mashad, Iran

(Received 7 July 1978; accepted for publication 27 February 1981)

This paper generalizes the work of J. L. B. Cooper on symmetric operators in a Hilbert space to Pontrjagin and Krein spaces. Existence (and uniqueness) of a solution for Schrödinger's equation $d\psi(t)/dt = iA\psi(t), \psi(0) = \phi \in \Pi\ [= \Pi^+(+)\bar{\Pi}]$ for the bounded decomposable symmetric operator A (the self-adjoint operator A) is studied. Also, the existence of a solution for $(1/i)\partial\psi(t)/\partial t = A^*\psi(t),\ \psi(0) = \phi \in \Pi$, where A is a cross-bounded symmetric operator in Π, is discussed. Finally, existence and uniqueness of the equation $(1/i)\,\partial\psi(t)/\partial t = A\psi(t), \psi(0) = \phi \in \Pi_k$, where A is a self-adjoint operator in the Pontrjagin space Π_k, is studied, and the fact that the maps $\phi \to \psi(t)$ form a group of unitary operators on Π_k is discussed.

PACS numbers: 02.30.Tb

INTRODUCTION

J. L. B. Cooper in his paper "Symmetric operators in Hilbert space"[1] has discusssed symmetric, maximal symmetric, and self-adjoint operators in Hilbert spaces. He has provided a proof for the existence of a solution of the Schrödinger equation for the adjoint operator and for all symmetric operators with any element of the domain of the operator as initial value. For self-adjoint operators, he has shown that the solution of the Schrödinger equation appears as the transform of the inital value by a unitary operator: The totality of these unitary operators forms a one-parameter continuous group.

In this paper we study the very important type of linear operator in Krein spaces with decomposition

$$\Pi = \Pi^+(+)\bar{\Pi},$$

namely, the cross-boundedly decomposabe operators

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

and we find a necessary condition for such operators to be symmetric. Also in the case that A_{11} is a self-adjoint operator we find a necessary and sufficient condition for A to be self-adjoint. For bounded symmetric cross-bounded operators we discuss the existence and uniqueness of the Schrödinger equation

$$\frac{d\psi(t)}{dt} = iA\psi(t)$$

with the initial condition $\psi(0) = \phi \in \Pi$ in a bounded interval $[0,T]$. Also we study the existence of a solution of the equation

$$\frac{1}{i}\frac{\partial\psi(t)}{\partial t} = A^*\psi(t)$$

with the initial condition $\psi(0) = \phi \in \Pi$ when A is a cross-bounded symmetric operator. Finally we discuss the existence and uniqueness of the solution of the equation

$$\frac{1}{i}\frac{\partial\psi(t)}{\partial t} = A\psi(t), \quad \psi(0) = \phi \in \Pi_k,$$

where A is a self-adjoint operator in the Pontrjagin space and we deduce that the maps

$$\phi \to \psi(t)$$

form a group of unitary operators in Π_k.

SOME MORE DEVELOPMENTS ON OPERATORS IN KREIN SPACES—THE EXPONENTIAL MAP

A vector-valued function $\psi(t)$ mapping the interval (α,β) into a Krein space Π is said to be *weakly (strongly) differentiable* at point $t = t_0$ if there exists an element $\psi'(t_0) \in \Pi$ such that the difference quotient

$$h^{-1}[\psi(t_0 + h) - \psi(t_0)]$$

tends weakly (strongly) to $\psi'(t_0)$ as $h \to 0$. We shall call $\psi'(t_0)$ the weak (strong) derivative of $\psi(t)$ at t_0 and denote it by $[\delta\psi(t)/\delta t][d_\psi(t)/dt]$.

Now let Π be a Krein space with the fundamental decomposition

$$\Pi = \Pi^+(+)\Pi^-, \quad \Pi^+ \subset B^{++}, \quad \Pi^- \subset B^{--},$$

and let p_+, p_-, and J be projectors of Π on Π^+, Π^- and the corresponding fundamental symmetry respectively. Clearly we have

$$P_+ + P_- = I, \quad P_+P_- = P_-P_+ = 0,$$
$$P_+^2 = P_+, \quad P_-^2 = P_-, \quad J = P_+ - P_-.$$

Now let A be a closed operator in Π with $\overline{D(A)} = \Pi$. The operator A is said to be *cross-boundedly decomposable* if for some fundamental decomposition of Π it can be written in the following form:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where

$$A_{11}: \Pi_+ \to \Pi_+, \quad A_{12}: \Pi_- \to \Pi_+,$$
$$A_{21}: \Pi_+ \to \Pi_-, \quad A_{22}: \Pi_- \to \Pi_-,$$

i.e.,

$$A_{11} = P_+AP_+, \quad A_{12} = \overline{P_+AP_-},$$
$$A_{21} = \overline{P_-AP_+}, \quad A_{22} = P_-AP_- \qquad (1)$$

 0022-2488/82/061003-08$02.50

and such that A_{12} and A_{21} are bounded. Let $\chi \in \Pi$. Then $\chi = \chi_+ (+) \chi_-$, where $\chi_+ = P_+\chi$, $\chi_- = P_-\chi$.

Theorem 1: If A has decomposition (1) (not necessarily bounded), then for A to be a symmetric operator in Π it is necessary that A_{11} and A_{22} be symmetric operators and

$$A_{12} \subset A_{21}^*, \quad A_{21} \subset A_{12}^*.$$

If $\Pi_+ \subseteq D(A)$, then A_{11} must be self-adjoint and bounded; in this case A_{12} is bounded and so is A_{21}.

Corolary 2: All symmetric operators on a Pontrjagin space Π_K are cross-boundedly decomposable.

For Π_K can be written in the form $\Pi_K = \Pi_+ (+) \Pi_-$, where Π_+ is a finite K-dimensional positive space, so A_{11} and A_{21} are necessarily bounded.

Theorem 3: Let A be a bounded symmetric operator with decomposition (1) in the Krein space Π. Then for any $\phi \in D(A)$ the equation

$$\frac{d\psi}{dt} = iA\psi, \quad \psi(0) = \phi, \tag{2}$$

has a unique solution, and for any bounded interval $[0,T]$ there is a constant $K_T(A_{12})$ whose size depends only on T and $\|A_{12}\|_J$ such that for all $t \in T$, $\|\psi(t)\|_J \leqslant K_T \|\phi\|_J$.

Theorem 4: Let A be a cross-bounded symmetric operator in the Krein space Π with decomposition

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{3}$$

corresponding to the decomposition $\Pi = \Pi^+ (+) \Pi^-$, $\Pi^+ \subset B^{++}, \Pi^- \subset B^{--}$. Then for any $\phi \in D(A)$ the equation

$$\frac{1}{i}\frac{\delta\psi}{\delta t} = A^*\psi(t), \quad \psi(0) = \phi, \tag{4}$$

has a solution.

Theorem 5: Let A be a self-adjoint operator in the Krein space with the cross-bounded decomposition

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

corresponding to the fundamental decomposition $\Pi = \Pi^+ (+) \Pi^-$.

Then for any $\phi \in D(A)$ the equation

$$\frac{1}{i}\frac{\delta\psi}{\delta t} = A\psi, \quad \text{with } \psi(0) = \phi,$$

has a unique solution and for any bounded interval $[-T,T]$ there exists a constant K whose size depends only on T and $\|A_{12}\|_J$ such that for all $t \in [-T,T]$ we have

$$\|\psi(t)\|_J \leqslant K_T(\|A_{12}\|) \cdot \|\phi\|_J.$$

Furthermore, the maps $\phi \to \psi(t)$ form a group of unitary operators on Π.

Theorem 6: Let H be a cross-bounded symmetric operator in the Krein space Π with the fundamental decompositions.

$$\Pi = \Pi^+ (+) \Pi^-, \tag{5}$$

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}.$$

If H_{11} is a self-adjoint operator in Π^+, then H_{22} to densely defined and closed. Furthermore, H_{22} is self-adjoint in Π^- if and only if H is self-adjoint in Π.

Before we continue with the proof of the theorems, we state some lemmas.

Lemma A: Let $\{e_n\}$ be a complete orthonormal system in a separable Hilbert space H, and let $\psi(t) \in H$ be a continuous vector-valued function defined in the interval (a,b) with the property that for each n there exist an element $\phi(t) \in H$ such that, for each $t \in (a,b)$,

$$\frac{d}{dt}(\psi(t),e_n) = (\phi(t),e_n), \quad \|\phi(t)\| < M,$$

where M is a constant. Then $(d/dt)(\psi(t),g) = (\phi(t),g)$ for all $g \in H$.

Proof: Let $g \in H$. Then $g = \Sigma_{n=1}^{\infty} (g,e_n)e_n$. Set $g_N = \Sigma_{n=1}^{N} (g,e_n)e_n$; then we have $\|g - g_N\| \to 0$ as $N \to \infty$.

Let $t \in (a,b)$. Then there exists a closed interval $[a',b'] \subset (a,b)$ such that $t \in [a',b']$.

Now let us define the following complex valued functions on $[a',b']$:

$$f_N(t) = (\psi(t),g_N),$$
$$f(t) = (\psi(t),g),$$
$$h_N(t) = (\phi(t),g_N),$$

and

$$h(t) = (\phi(t),g).$$

Clearly for all $t \in [a',b']$ we have

$$|f_N(t) - f(t)| = |(\psi(t),g_N - g)| \leqslant \|\psi(t)\| \|g_N - g\| \to 0$$

as $N \to \infty$. Hence $f_N \to f$ uniformly on $[a',b']$. By the same reasoning $h_N \to h$ uniformly, and by linearity we have $f'_N = h_N$. Hence f is differentiable on $[a',b']$. Now

$$\begin{aligned} f(t) - f(a) &= \lim_{N \to \infty} [f_N(t) - f_N(a)] \\ &= \lim_{N \to \infty} \int_a^t h_N(\tau)\, d\tau \\ &= \int_a^t h(\tau)\, d\tau. \end{aligned}$$

This gives that $f' = h$, i.e., $(d/dt)(\psi(t),g) = (\phi(t),g)$ for all $t \in (a,b)$ and all $g \in H$.

Lemma B: Now let Π be a Pontrjagen space with the fundamental decomposition $\Pi = \Pi^+ (+) \Pi^-$ and let $\dim \Pi^- = K < \infty$. Let $\{e_N\}$ be a complete orthonormal system in Π with $\{e_n\}_{n=1}^K \subset \Pi^-$.

A necessary and sufficient condition for the weak differentiability of a vector-valued function $\psi(t) \in \Pi$ at a point $t \in (a,b)$ is that for all n

$$\frac{d}{dt}(\psi(t),e_n) = \phi_n(t) \text{ exists}$$

and

$$\sum_{n=K+1}^{\infty} |\phi_n(t)|^2 < \infty.$$

In this way $\phi(t) = \Sigma_{n=1}^{\infty} \phi_n(t)e_n$ is the weak differential coefficient of $\psi(t)$.

Proof: The necessity is obvious. For the proof of suffi-

ciency, let, for all $n = 1,2,\cdots$,

$$\phi_n(t) = \frac{d}{dt}(\psi(t),e_n)$$

exist at $t\in(a,b)$ such that $\Sigma_{n=K+1}^{\infty}|\phi_n(t)|^2 < \infty$. Let $y\in\Pi$. Then y can be written in the form

$$y = \sum_{n=1}^{K}\alpha_n e_n(+)\sum_{n=K+1}^{\infty}\alpha_n e_n,$$

where $\Sigma_{n=1}^{\infty}|\alpha_n|^2 < \infty$. On account of Lemma A and the fact that the dual of l_2 is l_2 by virtue of the Schwartz inequality, we conclude that

$$\phi_{(K+1)}(t) = \sum_{n=K+1}^{\infty}\bar{\alpha}_n\frac{d}{dt}(\psi(t),e_n)$$

exists. On the other hand, the existence of $\Sigma_{n=1}^{K}\bar{\alpha}_n(d/dt)(\psi(t),e_n)$ is obvious. Therefore, by linearity of d/dt and Lemma A

$$\phi(t) = \frac{d}{dt}(\psi(t),y)$$

and

$$\phi(t) = \sum_{n=1}^{\infty}\phi_n(t).$$

Lemma C: If $\psi(t)$ is a J-bounded measureable vector-valued function defined over an interval (a,b) into the Pontrjagin space $\Pi = \Pi^+(+)\Pi^-$ referred to in Lemma B, then for all $t\in(a,b)$ outside a set of measure zero

$$\lim_{h\to 0}\int_t^{t+h}(\psi(\tau)-\phi,\psi(\tau)-\phi)\,d\tau = (\psi(t)-\phi,\psi(t)-\phi) \tag{6}$$

for all $\phi\in\Pi$. In particular

$$\int_t^{t+h}(\psi(\tau)-\psi(t),\psi(\tau)-\psi(t))\,d\tau = o(h)$$

almost everywhere. (7)

Proof: Let ϕ be a fixed element of Π and let us write ϕ and $\psi(t)$ in the form

$$\phi = \sum_{n=1}^{\infty}c_n e_n,\quad \psi(t) = \sum_{n=1}^{\infty}\psi_n(t)e_n,$$

where $\Sigma_{n=1}^{\infty}|c_n|^2 < \infty$ and $\Sigma_{n=1}^{\infty}|\psi_n(t)|^2 < \infty$ for each $a < t < b$. Then we have

$$(\psi(t)-\phi,\psi(t)-\phi) = -\sum_{n=1}^{K}|\psi_n(t)-c_n|^2 + \sum_{n=K+1}^{\infty}|\psi_n(t)-c_n|^2.$$

Each term of the right-hand side of the last equality is measurable and bounded and so summable; hence the left-hand side of it is summable. So it is the differential coefficient of its integral almost everywhere, i.e., (6) holds.

Now let ϕ be an arbitrary element of Π and let $\{\theta_n\}$ be a sequence of elements of Π everywhere dense in Π. Then (6) holds with $\phi = \theta_n$ outside a set E_n of measure zero. Hence (6) holds for all n outside a set $E = \cup_{n=1}^{\infty}E_n$ of measure zero. On the other hand, for any subsequence $\{\theta'_n\}$ of $\{\theta_n\}$ tending to ϕ, we have

$$\begin{aligned}
&|(\psi(t)-\phi,\psi(t)-\phi)-(\psi(t)-\theta_{n'},\psi(t)-\theta'_n)|\\
&\quad= |(\psi(t),\theta_{n'}-\phi)+(\theta_{n'}-\phi,\psi(t))+(\phi,\phi-\theta'_n)\\
&\qquad+(\theta'_n,\phi-\theta'_n))|\\
&\quad\leqslant(2\|\psi(t)\|_J+\|\phi\|_J+\|\theta'_n\|_J)\|\phi\\
&\qquad-\theta_{n'}\|_J\leqslant 4M\|\phi-\theta'_n\|_J\to 0
\end{aligned}$$

as $n'\to\infty$, where $M = \max\{\|\psi(t)\|_J, \|\phi\|_J, \|\theta_{N'}\|_J\}$.

$$\begin{aligned}
&\left|\lim_{h\to 0}\frac{1}{h}\int_t^{t+h}(\psi(\tau)-\phi,\psi(\tau)-\phi)d\tau-(\psi(t)-\phi,\psi(t)-\phi)\right|\\
&\quad=\left|\lim_{h\to 0}\frac{1}{h}\int_t^{t+h}(\psi(\tau)-\theta,\psi(\tau)-\phi)d\tau\right.\\
&\quad-\lim_{h\to 0}\frac{1}{h}\int_t^{t+h}(\psi(\tau)-\theta_{n'},\psi(\tau)-\theta_{n'})d\tau\\
&\quad\left.-(\psi(t)-\phi,\psi(t)-\phi)+(\psi(t)-\theta'_n,\psi(t)-\theta'_n)\right|\\
&\quad\leqslant\left|\lim_{h\to 0}\frac{1}{h}\int_t^{t+h}[(\psi(\tau)-\phi,\psi(\tau)-\phi)\right.\\
&\qquad\left.-(\psi(\tau)-\theta'_n,\psi(\tau)-\theta'_n)]d\tau\right|\\
&\quad+\left|-(\psi(t)-\phi,\psi(t)-\phi)+(\psi(t)-\theta_{n'},\psi(t)-\theta'_n)\right|\\
&\quad\leqslant\lim_{h\to 0}\frac{1}{h}\int_t^{t+h}4M\|\phi-\theta_{n'}\|_J d\tau+4M\|\phi-\theta'_n\|_J\\
&\quad=8M\|\phi-\theta'_n\|_J\to 0\quad\text{as } n'\to\infty.
\end{aligned}$$

This gives us that for all $t\in(a,b)$ outside the set E of measure zero (6) holds for all $\phi\in\Pi$. The second part of the lemma follows by putting $\psi(t)$ for ϕ in (6).

Lemma D: If for all $a < t < b$, $\phi(t) = \delta\psi(t)/\delta(t)$ exists and is J-bounded, then ϕ is almost everywhere the strong differential coefficient of $\psi(t)$, i.e.,

$$\lim_{h\to 0}\left\|\frac{\psi(t+h)-\psi(t)}{h}-\phi(t)\right\|_J = 0\quad\text{a.e.}$$

Proof: Let $\Pi = \Pi^+(+)\Pi^-$ and let $\{e_n\}_{n=1}^{\infty}$ be a complete orthonormal system with $\{e_1,\ldots,e_K\}\subset\Pi^-$. Then

$$\begin{aligned}
(\psi(t),e_n) &= (\psi^+(t)+\psi^-(t),e_n^+ + e_n^-)\\
&= (\psi^+(t),e_n^+)+(\psi^-(t),e_n^-)\\
&= 2(\psi^-(t),e_n^-)+(\psi(t),e_n)_J.
\end{aligned}$$

This gives us

$$(\psi(t),e_n)_J = (\psi(t),e_n)-2(\psi^-(t),e_n^-).$$

Thus

$$\frac{d}{dt}(\psi(t),e_n)_J = \frac{d}{dt}(\psi(t),e_n)-\frac{d}{dt}(\psi^-(t),e_n^-).$$

By virtue of the weak differentiability of $\psi(t)$, the differential coefficients on the right-hand side exist by Lemma B and

$$\frac{d}{dt}(\psi(t),e_n)_J = (\phi(t),e_n)-2(\phi^-(t),e_n^-).$$

Thus by Lemma A, $\psi(t)$ is a weakly differentiable vector-valued function with respect to the inner product $[\cdot,\cdot] = (\cdot,\cdot)_J$ and we have

$$\left\| \frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right\|_J^2 = \left(\frac{\psi(t+h)-\psi(t)}{h} - \phi(t), \frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right)_J$$

$$= \left(\left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^+ - \left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^-, \left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^+ + \left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^- \right)$$

$$= \left(\left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^+, \left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^+ \right)$$

$$- \left(\left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^-, \left[\frac{\psi(t+h)-\psi(t)}{h} - \phi(t) \right]^- \right)$$

$$= \sum_{n=K+1}^{\infty} \left| \frac{\psi_n(t+h)-\psi_n(t)}{h} - \phi_n(t) \right|^2 + \sum_{n=1}^{K} \left| \frac{\psi_n(t+h)-\psi_n(t)}{h} - \phi_n(t) \right|^2$$

$$= \sum_{n=K+1}^{\infty} \left| \frac{1}{h}\int_t^{t+h} \phi_n(\tau)\,d\tau - \phi_n(t) \right|^2 + \sum_{n=1}^{K} \left| \frac{1}{h}\int_t^{t+h} \phi_n(\tau)\,d\tau - \phi_n(t) \right|^2$$

$$= \sum_{n=K+1}^{\infty} \left| \frac{1}{h}\int_t^{t+h} (\phi_n(\tau)-\phi_n(t))\,d\tau \right|^2 + \sum_{n=1}^{K} \left| \frac{1}{h}\int_t^{t+h} (\phi_n(\tau)-d\tau-\phi_n(t))\,d\tau \right|^2$$

$$\leqslant \frac{1}{h}\int_t^{t+h} \sum_{n=k+1}^{\infty} |\phi_n(\tau)-\phi_n(t)|^2 d\tau + \frac{1}{h}\int_t^{t+h} \sum_{n=1}^{K} |\phi_n(\tau)-\phi_n(t)|^2\,d\tau$$

$$= \frac{1}{h}\int_t^{t+h} ([\phi(\tau)-\phi(t)]^+, [\phi(\tau)-\phi(t)]^+)\,d\tau - \frac{1}{h}\int_t^{t+h} ([\phi(\tau)-\phi(t)]^-, [\phi(\tau)-\phi(t)]^-)\,d\tau$$

$$= \frac{1}{h}\int_t^{t+h} (\phi(\tau)-\phi(t),\ \phi(\tau)-\phi(t))\,d\tau - 2\frac{1}{h}\int_t^{t+h} ([\phi(\tau)-\phi(t)]^-, [\phi(\tau)-\phi(t)]^-)\,d\tau \to 0$$

as $h \to 0$ almost everywhere (i.e., except possibly on the set E of Lemma C). Therefore, $\phi(t)$ is the differential coefficient of $\psi(t)$ in the strong sense.

Lemma E: Let $\psi_1(t)$ and $\psi_2(t)$ be two differentiable vector-valued functions defined in the interval (a,b) into a Pontrjagin space Π satisfying

$$\frac{\partial\psi_1(t)}{\delta t} = \phi_1(t)$$

and

$$\frac{\delta\psi_2(t)}{\delta t} = \phi_2(t).$$

Then

$$\frac{d}{dt}(\psi_1(t),\psi_2(t)) = (\phi_1(t),\psi_2(t)) + (\psi_1(t),\phi_2(t)). \tag{8}$$

Proof: Consider the difference quotient $(1/h)\{(\psi_1(t+h),\psi_2(t+h)) - (\psi_1(t),\psi_2(t))\}$ equal to

$$(1/h)\{(\psi_1(t+h)-\psi_1(t),\psi_2(t))\}$$

$$+ (1/h)\{(\psi_1(t),\psi_2(t+h)-\psi_2(t))\}$$

$$+ (1/h)\{(\psi_1(t+h)-\psi_1(t),\psi_2(t+h)-\psi_2(t))\}.$$

Since

$$\frac{1}{h}\{(\psi_1(t+h)-\psi_1(t),\psi_2(t+h)-\psi_2(t))\}\Big|$$

$$\leqslant h \left\| \frac{\psi_1(t+h)-\psi_1(t)}{h} \right\|_J \left\| \frac{\psi_2(t+h)-\psi_2(t)}{h} \right\|_J$$

$$= o(h) \tag{9}$$

and the limits of the first two terms in (2) exist as $h \to 0$ by the weak differentiability of $\psi_1(t)$, and $\psi_2(t)$, and are equal to $(\phi_1(t),\psi_2(t))$ and $(\psi_1(t),\phi_2(t))$ so (8) holds.

Lemma F: Let H be a symmetric operator in a Pontrjagin space Π and let $\psi_1(t)$, $\psi_2(t)$ be in the domain of H for all $t \in (a,b)$ and satisfy

$$\frac{\delta\psi_1(t)}{\delta t} = iH\psi_1(t)$$

and

$$\frac{\delta\psi_2(t)}{\delta t} = iH\psi_2(t).$$

Then $(\psi_1(t),\psi_2(t))$ is constant for all t, $a < t < b$.

Proof: Let us calculate $(d/dt)(\psi_1(t),\psi_2(t))$. We have

$$\frac{d}{dt}(\psi_1(t),\psi_2(t)) = (iH\psi_1(t),\psi_2(t)) + (\psi_1(t),iH\psi_2(t))$$

$$= i(\psi_1(t),H\psi_2(t)) - i(\psi_1(t),H\psi_2(t))$$

$$= 0.$$

Thus $(\psi_1(t),\psi_2(t)) = c$.

Proof of Theorem 1: That A is a symmetric operator implies that $(A\chi,y) = (\chi,Ay)$ for any $\chi,y \in D(A)$. This gives [using (8)]

$$(A_{11}\chi_+ + A_{12}\chi_-,y_+) + (A_{21}\chi_+ + A_{22}\chi_-,y_-)$$

$$= (\chi_+,A_{11}y_+ + A_{12}y_-) + (\chi_-,A_{21}y_+ + A_{22}y_-).$$

Putting $\chi_- = y_- = 0$, we get $(A_{11}\chi_+,y_+) = (\chi_+,A_{11}y_+)$ so A_{11} is a symmetric operator. If we put $\chi_+ = y_+ = 0$, it follows that A_{22} is a symmetric operator too. If we put $\chi_+ = y_1 = 0$, we get $(A_{12}\chi_-,y_+) = (\chi_-,A_{21}y_+)$ this will give $A_{21} \subset A_{12}^*$. Similarly, if we put $\chi_- = y_+ = 0$, we get $A_{12} \subset A_{21}^*$.

If A_{12} is bounded, then A_{12}^* is bounded. This fact together with the relation $A_{21} \subset A_{12}^*$ implies that A_{21} is bounded. Now let $\Pi_+ \subseteq D(A)$. Then $D(A_{11}) = \Pi_+$, and since an everywhere defined symmetric operator is self-adjoint and bounded, A_{11} is self-adjoint and bounded. In this case A_{21} is also an everywhere defined operator, and, since $A_{21} = A - A_{11}$ on Π_+, then by virtue of the continuity of A_{11} and closedness of A we get that A_{21} is closed. Hence it is continuous (bounded) by the closed graph theorem.

Proof of Theorem 3: Since A is bounded a solution of (9) is

$$\psi(t)=\sum_{n=0}^{\infty}\frac{(itA)^n}{n!}\phi.$$

For the proof of uniqueness let us put $\psi(t)=\psi_+(t)(+)\psi_-(t)$, where $\psi_+(t)\in H_+$, $\psi_-(t)\in H_-$. Then (9) can be written in the form

$$\frac{d\psi_+}{dt}=i(A_{11}\psi_+ + A_{12}\psi_-),$$

$$\frac{d\psi_-}{dt}=i(A_{21}\psi_+ + A_{22}\psi_-),$$

and we have

$$\begin{aligned}\frac{d}{dt}\|\psi_+(t)\|_J^2 &= \frac{d}{dt}(\psi_+(t),\psi_+(t))\\ &=\left(\frac{d}{dt}\psi_+(t),\psi_+(t),\right)+\left(\psi_+(t),\frac{d}{dt}\psi_+(t)\right)\\ &=(i[A_{11}\psi_+(t)+A_{12}\psi_-(t)],\psi_+(t))\\ &\quad+(\psi_+(t),i[A_{11}\psi_+(t)+A_{12}\psi_-(t)])\\ &=i\{([A_{11}\psi_+(t)+A_{12}\psi_-(t)],\psi_+(t))\\ &\quad-(\psi_+(t),[A_{11}\psi_+(t)+A_{12}\psi_-(t)])\}\\ &=-2\mathrm{Im}(A_{12}\psi_-(t),\psi_+(t)),\end{aligned}$$

whence

$$\left|\frac{d}{dt}\|\psi_+(t)\|_J^2\right|\leqslant 2\|A_{12}\|_J\|\psi_+(t)\|_J\|\psi_-(t)\|_J$$

or

$$2\|\psi_+(t)\|_J\left|\frac{d}{dt}\|\psi_+(t)\|_J\right|\leqslant 2\|A_{12}\|_J\|\psi_+(t)\|_J\|\psi_-(t)\|_J$$

so

$$\frac{d}{dt}\|\psi_+(t)\|_J\leqslant\|A_{12}\|_J\|\psi_-(t)\|_J.$$

Similarly, we can obtain

$$\frac{d}{dt}\|\psi_-(t)\|_J\leqslant\|A_{21}\|_J\|\psi_+(t)\|_J=\|A_{12}\|_J\|\psi_+(t)\|_J.$$

Let us put $\|A_{12}\|_J=\|A_{21}\|_J=M$. Then the above inequalities can be written in the form

$$\frac{d}{dt}\|\psi_+(t)\|_J\leqslant M\|\psi_-(t)\|_J,$$

$$\frac{d}{dt}\|\psi_-(t)\|_J\leqslant M\|\psi_+(t)\|_J.$$

Next let $\chi(t)$, $y(t)$ be a pair of real-valued functions satisfying the following relations:

$$\frac{d}{dt}\chi(t)=My(t),\quad \chi(0)=\|\psi_+(0)\|_J,$$

$$\frac{d}{dt}y(t)=M\chi(t),\quad y(0)=\|\psi_-(0)\|_J.$$

Then we shall prove that $\|\psi_+(t)\|_J\leqslant\chi(t)$, $\|\psi_-(t)\|_J\leqslant y(t)$ for all $t>0$. For this purpose we put $\|\psi_+(t)\|_J=u(t)$, $\|\psi_-(t)\|_J=v(t)$. Clearly $u(t)\geqslant 0$, $v(t)\geqslant 0$. For simplicity we may put $M=1$. Thus we are going to prove that the relations

$$\left.\begin{aligned}&\frac{du}{dt}\leqslant v(t)\\ &\frac{dv}{dt}\leqslant u(t)\\ &\frac{d\chi}{dt}=y(t)\\ &\frac{dy}{dt}=\chi(t)\\ &x(0)=u(0)\\ &y(0)=v(0)\end{aligned}\right\}\quad\text{imply}\quad\begin{cases}u(t)\leqslant\chi(t),\\ v(t)\leqslant y(t),\end{cases}\quad\text{for all } t>0.$$

Let us put $U(t)=u(t)-\chi(t)$ and $V(t)=v(t)-y(t)$. Then we can easily see

$$\frac{dU}{dt}\leqslant V(t),$$

$$\frac{dV}{dt}\leqslant U(t),$$

$$U(0)=V(0)=0.$$

Thus

$$\begin{aligned}U(t)&\leqslant\int_0^t V(\tau)\,d\tau\leqslant\int_0^t d\tau\int_0^t U(\eta)\,d\eta\\ &=\int_0^t(t-\eta)U(\eta)\,d\eta.\end{aligned}\tag{7'}$$

We have to show that $U(t)\leqslant 0$ for $t>0$. It is enough to show that $U(t)\leqslant 0$ for $0<t\leqslant a<1$. Let a be a fixed number <1 and let us put

$$\sup\{U(t):0\leqslant t\leqslant a\}=A.$$

Then by (7') we have

$$U(t)\leqslant A\int_0^t(t-\eta)\,d\eta=\tfrac{1}{2}At^2\leqslant\tfrac{1}{2}A\quad\text{for all } t,0<t\leqslant a.\tag{10}$$

This gives $A\leqslant\frac{1}{2}A$ or $A\leqslant 0$. This means that $U(t)\leqslant 0$. Consequently, $u(t)\leqslant\chi(t)$ for all $t\geqslant 0$.

In the same way we can show that $v(t)\leqslant y(t)$ for all $t\geqslant 0$.

Next consider the system of differential equations:

$$\left.\begin{aligned}&\frac{dx}{dt}=My(t)\\ &\frac{dy}{dt}=M\chi(t)\end{aligned}\right\}\quad\chi(0)=y(0)=0.$$

The general solutions of this system are

$$\chi=Ae^{Mt}+Be^{-Mt},$$

$$y=Ae^{Mt}-Be^{-Mt},$$

and the initial conditions, $\chi(0)=y(0)=0$, gives that $A=B=0$. So $\chi(t)=0$, $y(t)=0$ for $t\geqslant 0$. Consequently, by what we have proved $u(t)=0$, $v(t)=0$ for all $t\geqslant 0$.

Now let ψ^1, ψ^2 both satisfy the equation $d\psi/dt=iA\psi(t)$ and let $\psi^1(0)=\psi^2(0)=\phi$. Then $\psi(t)=\psi^1(t)-\psi^2(t)$ also satisfies (10).

But $\psi(t)$ can be written as follows:

$$\psi(t) = \psi_+(t)(+)\psi_-(t)$$
$$\equiv \psi^1_+(t)(+)\psi^1_-(t) - [\psi^2_+(t)(+)\psi^2_-(t)]$$
$$\equiv \psi^1_+(t) - \psi^2_+(t)(+)[\psi^1_-(t) - \psi^2_-(t)]$$

so $\psi_+(t) \equiv \psi^1_+(t) - \psi^2_+(t)$ and $\psi_-(t) = \psi^1_-(t) - \psi^2_-(t)$ and $\psi^1_+(0) = \psi^2_+(0)$, $\psi^1_-(0) = \psi^2_-(0)$. Therefore, for all $t \geqslant 0$,

$$\psi^1_+(t) = \psi^2_+(t), \quad \psi^1_-(t) = \psi^2_-(t).$$

Hence the solution of (10) is unique.

Next we are going to prove the last part of the theorem. Clearly the functions $\chi(t)$ and $y(t)$, defined by

$$\chi(t) = \chi_0 \cosh Mt + y_0 \sinh Mt,$$
$$y(t) = y_0 \cosh Mt + \chi_0 \sinh Mt,$$

are the solutions of the system

$$\begin{cases} \dfrac{d}{dt}\chi(t) = My(t), \\ \dfrac{d}{dt}y(t) = M\chi(t), \end{cases}$$

corresponding to the initial conditions $\chi(0) = \chi_0$, $y(0) = y_0$. Therefore, if we put $\|\psi_+(0)\|_J = \chi_0$, $\|\psi_-(0)\|_J = y_0$, then by what we have proved we get

$$\|\psi_+(t)\|_J \leqslant \chi(t), \quad \|\psi_-(t)\|_J \leqslant y(t)$$

for all $t \geqslant 0$.

Consequently,

$$\|\psi(t)\|_J = \|\psi_+(t)(+)\psi_-(t)\|_J$$
$$\leqslant \|\psi_+(t)\|_J + \|\psi_-(t)\|_J$$
$$\leqslant \chi(t) + y(t)$$
$$= (\|\psi_+(0)\|_J + \|\psi_-(0)\|_J)e^{MT}.$$

In particular if $M = \|A_{12}\|_J$ in the above system, then we will have

$$\|\psi(t)\|_J \leqslant \exp\|A_{12}\|_J^T(\|\phi_+\|_J + \|\phi_-\|_J)$$
$$= K_T(\|A_{12}\|_J)(\|\phi_+\|_J + \|\phi_-\|_J),$$

which completes the proof.

Proof of Theorem 4: Let $\{e_\alpha\}$, $\{f_\beta\}$ be two complete ortho-normal systems in Π^+, Π^-, respectively, such that $e_{\alpha_0} = \phi_+/\|\phi_+\|_J$ if $\phi_+ \neq 0$ and $f_{\beta_0} = \phi_-/\|\phi - \|_J$ if $\phi_- \neq 0$ and such that the span of all points $\{e_\alpha; A_{11}e_\alpha\}, \{f_\beta; A_{22}f_\beta\}$ is dense in the graphs of A_{11}, A_{22}, respectively. Let S_+, S_- denote projectors on the spans of finite subsets of $\{e_\alpha\}$, $\{f_B\}$, respectively, such that the chosen sets contain e_{α_0}, f_{B_0}. Let us define an operator $B_{S,S}$ on Π, for any choice of sets S_+, S_- as follows:

$$B_{S,S} = \begin{bmatrix} S_+A_{11}S_+ & A_{12} \\ A_{21} & S_-A_{22}S_- \end{bmatrix}.$$

Then for any $\chi, y \in \Pi$ we have

$$B_{S,S}\ \chi = S_+A_{11}S_+\chi_+ + A_{12}\chi_+ + S_-A_{22}S_-\chi_- + A_{21}\chi_+,$$

$$(B_{S,S}\ \chi, y) = (S_+A_{11}S_-\chi_+ + A_{12}\chi_-, y_+) + (A_{21}\chi_+ + S_-A_{22}S_-\chi_-, y_-)$$
$$= (\chi_+, S_+A_{11}s_+y_+ + A_{12}y_-) + (\chi_-, S_-A_{22}S_-y_- + A_{21}y_+)$$
$$= (\chi, B_{S,S}\ y),$$

as we have seen in the proof of Theorem 1.

Thus $B_{S_+S_-}$ is a bounded self-adjoint operator, so on account of Theorem 3 the solutions of equation

$$\frac{d\xi}{dt} = iB_{S,S}\ \xi, \quad \xi(0) = \phi, \tag{11}$$

are bounded over any interval $[-T,T]$ by a bound which does not depend on S_+, S_-. In particular this is true for the solutions $\psi_{S,S}\ (t)$ of (5) with $\psi_{S,S}\ (0) = \phi = \xi(0)$ and of solutions $B_{S,S}\ \psi_{S,S}\ (t)$ of (5) with $B_{S,S}\ \phi = \xi(0)$.

Further we have

$$\left|\frac{d}{dt}(B_{S,S}\ \psi_{S,S}\ (t), e_\alpha)\right|$$
$$= |(B^2{}_{S,S}\ \psi_{S,S}\ (t), e_\alpha)|$$
$$= |(B_{S,S}\ \psi_{S,S}\ (t), B_{S,S}\ e_\alpha)|$$
$$\leqslant \|B_{S_+S_-}\psi_{S,S}\ (t)\|_J \|B_{S_+S_-}e_\alpha\|_J$$
$$\leqslant K_T(\|A_{12}\|_J)(\|\phi\|_J).$$

Similarly, we have

$$\left|\frac{d}{dt}(B_{S,S}\ \psi_{S,S}\ (t), f_\beta)\right| \leqslant K_T(\|A_{12}\|_J)\cdot(\|\phi\|_J).$$

Thus the functions $(B_{S,S}\ \psi_{S,S}\ (t), e_\alpha)$ and $(B_{S,S}\ \psi_{S,S}\ (t), f_\beta)$ are uniformly bounded with uniformly bounded derivatives; therefore, they (i. e., each of them) form a family of equicontinuous functions over any compact interval $[-T,T]$.

Since a set of uniformly continuous uniformly bounded functions is relatively compact in the topology of uniform convergence over any compact interval and since the function

$$(B_{S,S}\ \psi_{S,S}\ (t), e_\alpha) \quad [\text{resp.}(B_{S,S}\ \psi_{S,S}\ (t), f_\beta)]$$

is an element of the topological product space of function spaces, each of them corresponding to a e_α [resp. f_β], it follows from Tychenov's theorem that the family $(B_{S+S}\ \psi_{S,S}\ (t), e_\alpha)$ [resp. $(B_{S,S}\ \psi_{S_+S}\ (t), f_\beta)$] has a limiting point $p_\alpha(t)$ [resp. $q_\beta(t)$], for any compact interval $[-T, T]$. That is, for any given $\epsilon > 0$, and any α [resp. β] and any T we can find a set S_{+1} of α's, indeed a set of projectors [resp. $\mathbf{S}_2$ of β's] such that

$$|P_\alpha(t) - (B_{S,S}\ \psi_{S,S}\ (t), e_\alpha)| < \epsilon,$$
$$|q_\beta(t) - (B_{S,S}\ \psi_{S,S}\ (t), f_\beta)| < \epsilon$$

for all $t \in [-T, T]$ and all $\alpha' \in S_{+1}$ [resp. $\beta' \in S_{-1}$].

From the inequalities

$$\text{(I)} \quad \sum_\alpha |(B_{S,S}\ \psi_{S,S}\ (t), e_\alpha)|^2 = \|P_+B_{S,S}\ \psi_{S,S}\ (t)\|_J^2 \leqslant \mathbf{K}_T(\|\mathbf{A}_{12}\|_J)\cdot\|\phi\|_J,$$

$$\text{(II)} \quad \sum_\beta |B_{S,S}\ \psi_{S,S}\ (t), f_\beta|^2 = \|P_-B_{S,S}\ \psi_{S,S}\ (t)\|_J^2 \leqslant K_T(\|A_{12}\|_J)\cdot\|\phi\|_J,$$

$$\sum_\alpha |P_\alpha(t)|^2 \quad \text{and} \quad \sum_\beta |q_\beta(t)|^2$$

are bounded and have the same bounds as (I), (II). Hence the elements $\theta_+(t)$, $\theta_-(t)$, $\psi_+(t)$, and $\psi_-(t)$ defined by

$$\theta_+(t) = \sum_\alpha p_\alpha(t) e_\alpha, \quad \theta_-(t) = \sum_\beta q_\beta(t) f_\beta,$$

$$\psi_+(t) = \phi_+ + \int_0^t \theta_+(u)\, du, \quad \psi_-(t) = \psi_- + \int_0^t \theta_-(v)\, dv$$

are in Π and satisfy the equations

$$\frac{\delta\psi_+(t)}{\delta t} = \theta_+(t), \quad \frac{\delta\psi_-(t)}{\delta t} = \theta_-(t), \quad \text{respectively.}$$

Next let us put

$$\psi(t) = \psi_+(t) + \psi_-(t),$$

and

$$\theta(t) = \theta_+(t) + \theta_-(t).$$

Then for any e_α we have

$$\begin{aligned}\frac{d}{dt}(\psi(t), e_\alpha) &= (\theta(t), e_\alpha) \\ &= \lim[i(B_{S,S}\,\psi_{S,S}(t), e_\alpha)] \\ &= \lim[i(\psi_{S,S}(t), B_{S,S}\, e_\alpha)] \\ &= i(\psi(t), Ae_\alpha),\end{aligned}$$

where the limit is along some filter F in the sets S_+, S_-.

This shows that

$$(\theta(t), e_\alpha) - i(\psi(t), Ae_\alpha) = 0$$

or

$$\left(\frac{\delta}{\delta t}\psi(t), e_\alpha\right) - i(\psi(t), Ae_\alpha) = 0,$$

from which, by the definition of the scalar product in $\Pi \times \Pi$ [i. e., $(\{a, b\}, \{c, d\}) = (a, c) + (b, d)$], we obtain that

$$\left\{\frac{\delta\psi(t)}{dt}, -i\psi(t)\right\}$$

is orthogonal to the set of all elements $\{e_\alpha, Ae_\alpha\}$. In the same way it is orthogonal to the set of all elements $\{f_\beta, Af_\beta\}$.

Next let us recall that a necessary and sufficient condition for $y = A^*\chi$ for any given χ, $y \in \Pi$ is that the element $\{-y, \chi\} \in \Pi \times \Pi$ be orthogonal to all elements $\{p, Ap\}$ in the graph of A. Now let $p \in D(A)$. Then

$$\begin{aligned}&(\{-y, \chi\}, \{p, Ap\}) \\ &\quad = (-y, p) + (\chi, Ap) \\ &\quad = (-y_+ - y_-, \\ &p_+ + p_-) + (\chi_+ + \chi_-, A_{11}p_+ + A_{12}p_- + A_{21}p_+ + A_{22}p_-) \\ &\quad = (y_+, p_+) - (y_-, p_-) + (\chi_+, A_{11}p_+ + A_{12}p_-) \\ &\quad + (\chi_-, A_{21}p_+ + A_{22}p_-),\end{aligned}$$

and, since for any given $\epsilon > 0$ and any $p \in D(A)$, we can choose a linear combination l_+ of e_α and a linear combination l_- of f_β such that

$$\|p_+ - l_+\|_J < \epsilon, \quad \|p_- - l_-\|_J < \epsilon,$$
$$\|A_{11}p_+ - A_{11}l_+\|_J < \epsilon, \quad \text{and} \quad \|A_{22}p_- - A_{22}l_-\|_J < \epsilon,$$

and, since by the hypothesis the linear combinations of $\{e_\alpha, A_{11}e_\alpha\}$ and of $\{f_\beta, A_{22}f_\beta\}$ are dense in the graphs of A_{11}, A_{22} and since by virtue of the boundedness of A_{12} and A_{21} the values of $\|A_{12}p_- - A_{12}l_-\|_J$, $\|A_{21}p_+ - A_{21}l_+\|_J$ are less than $\epsilon\|A_{12}\|_J$, it follows that if $\{-y, \chi\}$ is orthogonal to all $\{e_\alpha, Ae_A\}$ and to all $\{f_\beta, Af_\beta\}$, then it is orthogonal to all $\{p, Ap\}$ and so $y = A^*\chi$. Consequently, $\psi(t)$ is in the domain of A^* and $\frac{\delta\psi}{\delta t} = iA^*(t)$ and the proof is complete:

$$\left(\left\{\frac{\delta\psi}{\delta t}, -i\psi(t)\right\} \perp \{p, Ap\}\right),$$

i. e.,

$$\left(\left\{\frac{\delta\psi}{\delta t}, -i\psi(t)\right\}, \{p, Ap\}\right) = 0$$

or

$$\frac{\delta\psi}{\delta t} = iA^*\psi(t).$$

Proof of Theorem 5: Since A is a self-adjoint operator by Theorem 4, there is no doubt about the existence of solution. The proof of boundedness and uniqueness is as the proof of boundedness and uniqueness in Theorem 3.

The proof of the last part follows from

$$\begin{aligned}&\frac{d}{dt}(\psi(t), \psi(t)) \\ &\quad = (iA\psi(t), \psi(t)) + (\psi(t), iA\psi(t)) = 0\end{aligned}$$

so $(\psi(t), \psi(t))$ is a constant function of t, i. e., $(\psi(0), \psi(0)) = (\phi, \phi)$.

Proof of Theorem 6: By virtue of Theorem 1, H_{22} is a symmetric operator in Π^-. H_{12}, H_{21} are bounded since H is a cross bounded operator. Therefore, by the self-adjointedness of H_{11} in Π^+ and the cross-boundedness of H in Π, we have

$$\begin{aligned}&\overline{D(H_{11})} = D(H_{21}) = \Pi^+, \\ &\overline{D(H_{22})} = \overline{D(H) \cap \Pi^-} = D(H_{12}) = \Pi^-. \qquad (12)\end{aligned}$$

This shows that H_{22} is densely defined in Π^-.

Closedness: The linear operator H_{22} in Π^- can be written in the form $H_{22} = H - H_{12}$, where H_{12} is bounded in Π^- (consequently, it is closed) and H is closed in Π. Hence H_{22} is closed in Π^-.

Now let H be self-adjoint in Π. Then, in particular $D(H_{12})$ is dense in Π^-. This fact with the symmetry of H and the boundedness of H_{12} implies that

$$(H_{21}\chi^+, y^-) = (\chi^+, \overline{H_{12}y^-}) \quad (\chi^+ \in \Pi^+, y^- \in \Pi^-). \quad (13)$$

Now suppose that for a certain $y^- \in \Pi^-$ there exists a $y_*^- \in \Pi$ such that

$$(H_{22}\chi^-, y^-) = (\chi^-, y_*^-), \quad \text{for all } \chi^- D(H_{22}). \qquad (14)$$

Then for all $\chi \in D(H)$ we have

$$\begin{aligned}(H\chi, y^-) &= (H_{21}\chi^+, y^-) + (H_{22}\chi^-, y^-) \\ &= (\chi^+, \overline{H_{12}y^-}) + (\chi^-, y_*^-) = (\chi, \overline{H_{12}y^-} + y_*^-).\end{aligned}$$

This with H being self-adjoint gives that $y^- \in D(H)$. Consequently we have $y^- \in D(H_{22})$ which, by (14), means that H_{22} is self-adjoint in Π^-.

Let us, conversely, assume that H_{22} is self-adjoint in Π^-. Suppose that for a fixed $y\in\Pi$, there exists an element $y_*\in\Pi$ such that, for all $\chi\in D(H)$, $(H\chi, y) = (\chi, y_*)$.

Let $y = y^+(\dot{+})y^-$, $\chi = \chi^+(\dot{+})\chi^-$. Then

$$\begin{aligned}(H_{22}\chi^-, y^-) &= (H_{22}\chi^-, y) = ((H - H_{12})\chi^-, y)\\ &= (H\chi^-, y) - (H_{12}\chi^-, y)\\ &= (\chi^-, y_* - H_{21}y^+) \quad [\chi\in D(H_{22})].\end{aligned}$$

This gives that $y^-\in D(H_{22})$. Consequently, $y\in D(H)$, i. e., H is self-adjoint in Π.

[1]J. B. L. Cooper, "Symmetric Operators in Hilbert Space," London Math. Soc. Ser. 2, Vol. 50.

A new method for the asymptotic evaluation of a class of path integrals [a)]

J. M. Luttinger
Department of Physics, Columbia University, New York, New York 10027

(Received 17 November 1981; accepted for publication 12 February 1982)

A general method of calculating the asymptotic behavior of a class of Wiener path integrals is given. These integrals are averages of functionals of the "local time." The technique is essentially a variation on the well-known "replica" method now widely used in condensed matter physics, combined with the Laplace method for evaluating integrals containing a large parameter. The leading term is given, and from the construction one sees that the error is typically of order $1/t$.

PACS numbers: 02.50. + s, 02.70. + d

I. INTRODUCTION

In this paper we will discuss the asymptotic behavior (for large time) of a certain class of Wiener integrals. These are the following. Consider a "path" $\mathbf{r}(t)$ in d-dimensions, and define the "local time"

$$L_t(\mathbf{r}) = \frac{1}{t}\int_0^t \delta(\mathbf{r} - \mathbf{r}(\tau))d\tau, \tag{1.1}$$

where $\delta(\mathbf{r})$ is the Dirac delta function in d-dimensions. Let $F_t[L_t(\cdot)]$ be a functional of L_t. {For example,

$$F_t[L_t(\cdot)] = \int L_t(\mathbf{r})v(\mathbf{r})d\mathbf{r} = \frac{1}{t}\int_0^t d\tau\, v(\mathbf{r}(\tau)), \tag{1.2}$$

where $d\mathbf{r}$ is the volume element in d-dimensional Euclidean space and the integral goes over all space.} Then the class of integrals we consider is

$$W_t(\mathbf{r}, \mathbf{r}') \equiv \left\langle e^{-tF_t[L_t(\cdot)]} \,\middle|\, \begin{matrix}\mathbf{r}(t) = \mathbf{r}' \\ \mathbf{r}(0) = \mathbf{r}\end{matrix} \right\rangle, \tag{1.3}$$

where the angular brackets represent the conditional average over all Brownian paths starting at $\mathbf{r}(\tau) = \mathbf{r}$ for $\tau = 0$ and ending at $\mathbf{r}(\tau) = \mathbf{r}'$ for $\tau = t$. That is, we are studying the conditional Wiener average of the functional $\exp\{-tF_t\}$.

The asymptotic behavior of such an integral for large t was first studied in a particular case by Friedberg and myself[1] in connection with an investigation of disordered systems. The method was general but quite heuristic (as will be the treatment in this paper). Independently, Donsker and Varadhan[2] found the general result and strong mathematical proofs. Their result may be stated as follows. Under reasonable assumptions about F_t and the existence of F_∞,

$$\lim_{t\to\infty}\left\{-\frac{1}{t}\ln W_t(\mathbf{r},\mathbf{r}')\right\} = \operatorname{Min}_{(\psi,\psi)=1}\{A(\psi)\}, \tag{1.4}$$

where

$$A(\psi) \equiv \left(\psi, \frac{p^2}{2}\psi\right) + F_\infty[\psi^2(\cdot)], \tag{1.5}$$

$$-p^2 \equiv \nabla^2 \text{ (the Laplacian in } d\text{-dimensions)}, \tag{1.6}$$

$$(\psi, \phi) \equiv \int d\mathbf{r}\, \psi(\mathbf{r})\phi(\mathbf{r}). \tag{1.7}$$

Now while (1.4) is satisfactory for many purposes it clearly gives only the leading term in the exponential behavior of W_t and not the prefactors, nor any systematic way of improving the result. In fact, we shall show (still using typical physicists heuristics) that W_t can be systematically expanded for large t by a method which gives (1.4) as an almost trivial first step. Although this method is very straightforward and elementary in principal it can be fairly tedious to carry through.

The paper is organized as follows. In Sec. II the basic formula relating the path integral to a certain limit of a "functional integral" is established. In Sec. III it is shown how such functional integrals may be evaluated in the limit of large t, and the limiting process carried out.

II. BASIC FORMULA

We begin with the Feynman–Kac[3] representation of the Green's function $G_t(\mathbf{r}, \mathbf{r}')$ of the diffusion equation with absorbers. That is $G_t(\mathbf{r}, \mathbf{r}')$, which is defined by $(t > 0)$

$$\frac{\partial G_t(\mathbf{r}, \mathbf{r}')}{\partial t} + HG_t(\mathbf{r}, \mathbf{r}') = 0, \tag{2.1}$$

$$\lim_{t\to 0^+} G_t(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \tag{2.2}$$

with

$$H = p^2/2 + V(\mathbf{r}) = -\tfrac{1}{2}\nabla_{\mathbf{r}}^2 + V(\mathbf{r}), \tag{2.3}$$

has the representation

$$G_t(\mathbf{r}, \mathbf{r}') = G_t^{(0)}(\mathbf{r}, \mathbf{r}')\left\langle \exp\left[-\int_0^t V(\mathbf{r}(\tau))d\tau\right] \,\middle|\, \begin{matrix}\mathbf{r}(t) = \mathbf{r}' \\ \mathbf{r}(0) = \mathbf{r}\end{matrix} \right\rangle, \tag{2.4}$$

where

$$G_t^{(0)}(\mathbf{r}, \mathbf{r}') = \frac{1}{(2\pi t)^{d/2}}\exp\left\{-\frac{(\mathbf{r} - \mathbf{r}')^2}{2t}\right\}. \tag{2.5}$$

We may also write (2.4) as

$$G_t(\mathbf{r}, \mathbf{r}') = G_t^{(0)}(\mathbf{r}, \mathbf{r}')\left\langle e^{-t\int L_t(\mathbf{r}'')V(\mathbf{r}'')d\mathbf{r}''} \,\middle|\, \begin{matrix}\mathbf{r}(t) = \mathbf{r}' \\ \mathbf{r}(0) = \mathbf{r}\end{matrix} \right\rangle. \tag{2.6}$$

If H has a complete set of orthonormal eigenfunctions

$$H\psi_\mu = E_\mu\psi_\mu, \tag{2.7}$$

the Green's function may also be written[4]

$$G_t(\mathbf{r}, \mathbf{r}') = \sum_\mu \psi_\mu(\mathbf{r})\psi_\mu(\mathbf{r}')e^{-tE_\mu}. \tag{2.8}$$

[a)]This research was supported in part by the National Science Foundation.

 0022-2488/82/061011-06$02.50

We shall now show that $G_t(\mathbf{r}, \mathbf{r}')$ has still another representation. Consider the functional integral[5]

$$\tilde{G}_t(\mathbf{r}, \mathbf{r}') \equiv 2 \lim_{n\to 0} \int \prod_{\sigma=1}^{n} \mathscr{D}\phi_\sigma \phi_1(\mathbf{r})\phi_1(\mathbf{r}')\delta([\phi, \phi] - 1) \times \exp\{-t[\phi, H\phi]\}, \tag{2.9}$$

where $\phi_\sigma(\mathbf{r})$, $\sigma = 1, 2, \ldots, n$ is an n-component real field,

$$[\phi, \phi] = \sum_{\sigma=1}^{n} \int d\mathbf{r}\, \phi_\sigma^2(\mathbf{r}), \tag{2.10}$$

$$[\phi, H\phi] = \sum_{\sigma=1}^{n} \int \phi_\sigma(\mathbf{r}) H \phi_\sigma(\mathbf{r}) d\mathbf{r}. \tag{2.11}$$

$\mathscr{D}\phi_\sigma$ is the "flat" volume element in the space of functions ϕ_σ. This may be taken to be $\lim \Pi_s d\phi_\sigma(s)$, where the product is taken over a discrete set of points s and the lim means making these points dense in R^d.[6] It may also be chosen as $\lim \Pi_{\lambda=1}^{N} da_\lambda^\sigma$, where $\phi_\sigma(\mathbf{r}) = \Sigma_{\lambda=1}^{N} a_\lambda^\sigma \phi_\lambda(\mathbf{r})$, $-\infty < a_\lambda^\sigma < \infty$, ϕ_λ being a complete orthonormal set of functions and the lim means $N \to \infty$ in a suitable sense.

Now put

$$\phi_\sigma(\mathbf{r}) = \sum_\mu a_\mu^\sigma \psi_\mu(\mathbf{r}), \qquad \mathscr{D}\phi_\sigma = \prod_\mu da_\mu^\sigma. \tag{2.12}$$

Then (2.9) becomes

$$\tilde{G}_t(\mathbf{r}, \mathbf{r}') = 2 \lim_{n\to 0} \int_{-\infty}^{\infty} \prod_{\sigma,\mu} da_\mu^\sigma \sum_{\mu_1\mu_2} a_{\mu_1}^{(1)} a_{\mu_2}^{(1)} \times \exp\Big[-t \sum_{\sigma,\mu} (a_\mu^\sigma)^2 E_\mu\Big] \delta\Big(\sum_{\sigma\mu} (a_\mu^\sigma)^2 - 1\Big) \psi_{\mu_1}(\mathbf{r})\psi_{\mu_2}(\mathbf{r}). \tag{2.13}$$

We shall assume that H is bounded from below and therfore without any loss of generality that $E_\mu > 0$ for all μ, so that the a_μ^σ integrals are well defined (though this is actually not necessary). Using the usual representation for the δ-function,

$$\delta\Big(\sum_{\mu\sigma} (a_\mu^\sigma)^2 - 1\Big) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp\Big[-i\omega\Big(\sum_{\mu\sigma} (a_\mu^\sigma)^2 - 1\Big)\Big], \tag{2.14}$$

(2.13) becomes (noticing that μ_1 must equal $\mu_2 \equiv \mu'$ for non-vanishing terms, and carrying the trivial Gaussian integrals)

$$\begin{aligned}\tilde{G}_t(\mathbf{r}, \mathbf{r}') &= 2 \lim_{n\to 0} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{i\omega} \sum_{\mu'} \psi_{\mu'}(\mathbf{r})\psi_{\mu'}(\mathbf{r}') \\ &\quad \times \prod_{\mu\neq\mu'} \Big(\frac{\pi}{tE_\mu + i\omega}\Big)^{n/2} \frac{\frac{1}{2}\sqrt{\pi}}{(tE_\mu + i\omega)^{3/2}} \Big(\frac{\pi}{tE_{\mu'} + i\omega}\Big)^{(n-1)/2} \\ &= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{i\omega} \sum_{\mu'} \psi_{\mu'}(\mathbf{r})\psi_{\mu'}(\mathbf{r}') \frac{1}{tE_{\mu'} + i\omega} \\ &= \sum_{\mu'} \psi_{\mu'}(\mathbf{r})\psi_{\mu'}(\mathbf{r}') e^{-tE_{\mu'}} \\ &= G_t(\mathbf{r}, \mathbf{r}').\end{aligned} \tag{2.15}$$

Using (2.6), (2.8), (2.9), and (2.15) we now have

$$\begin{aligned}&\Big\langle e^{-t\int L_t(\mathbf{r}'')V(\mathbf{r}'')d\mathbf{r}''} \Big| \begin{matrix}\mathbf{r}(t) = \mathbf{r}' \\ \mathbf{r}(0) = \mathbf{r}\end{matrix} \Big\rangle \\ &= [G_t^{(0)}(\mathbf{r}, \mathbf{r}')]^{-1} 2 \lim_{n\to 0} \int \prod_{\sigma=1}^{n} \mathscr{D}\phi_\sigma \phi_1(\mathbf{r})\phi_1(\mathbf{r}')\delta([\phi, \phi] - 1) \\ &\quad \times \exp\{-t[\phi, H\phi]\}.\end{aligned} \tag{2.16}$$

Since

$$[\phi, H\phi] = \sum_\sigma \Big(\phi_\sigma, \frac{p^2}{2}\phi_\sigma\Big) + \int d\mathbf{r}'' \Big(\sum_\sigma \phi_\sigma^2(\mathbf{r}'')\Big) V(\mathbf{r}''),$$

the right-hand side of (2.16) becomes

$$\begin{aligned}&[G_t^{(0)}(\mathbf{r}, \mathbf{r}')]^{-1} 2 \lim_{n\to 0} \int \prod_{\sigma=1}^{n} \mathscr{D}\phi_\sigma \phi_1(\mathbf{r})\phi_1(\mathbf{r}')\delta([\phi, \phi] - 1) \\ &\quad \times \exp\Big\{-t \sum_\sigma \Big(\phi_\sigma, \frac{p^2}{2}\phi_\sigma\Big)\Big\} \\ &\quad \times \exp\Big\{-t \int \sum_\sigma \phi_\sigma^2(\mathbf{r}'')V(\mathbf{r}'')d\mathbf{r}''\Big\}.\end{aligned} \tag{2.17}$$

Now suppose we multiply both sides of (2.16) by an "arbitrary" functional of $V(\mathbf{r})$, $\Gamma_t[V(\cdot)]$, and integrate over V. The left-hand side of (2.16) becomes

$$\Big\langle e^{-tF_t[L_t(\cdot)]} \Big| \begin{matrix}\mathbf{r}(t) = \mathbf{r}' \\ \mathbf{r}(0) = \mathbf{r}\end{matrix} \Big\rangle, \tag{2.18}$$

where F_t is defined by

$$e^{-tF_t[L_t(\cdot)]} = \int \mathscr{D}V \Gamma_t[V(\cdot)] e^{-t\int L_t(\mathbf{r}'')V(\mathbf{r}'')d\mathbf{r}''}. \tag{2.19}$$

Using (2.17), we see that exactly the same integral over V occurs on the right-hand side of (2.16), with $L_t(\mathbf{r}'')$ replaced by $\Sigma_\sigma \phi_\sigma^2(\mathbf{r}'')$. Therefore, we finally obtain

$$\begin{aligned}&\Big\langle e^{-tF_t[L_t(\cdot)]} \Big| \begin{matrix}\mathbf{r}(t) = \mathbf{r}' \\ \mathbf{r}(0) = \mathbf{r}\end{matrix} \Big\rangle \\ &= [G_t^{(0)}(\mathbf{r}, \mathbf{r}')]^{-1} 2 \lim_{n\to 0} \int \prod_\sigma \mathscr{D}\phi_\sigma \phi_1(\mathbf{r})\phi_1(\mathbf{r}')\delta([\phi, \phi] - 1) \\ &\quad \times \exp\{-tS\},\end{aligned} \tag{2.20}$$

$$S = \sum_\sigma \Big(\phi_\sigma, \frac{p^2}{2}\phi_\sigma\Big) + F_t\Big[\sum_\sigma \phi_\sigma^2(\cdot)\Big]. \tag{2.21}$$

This is our main result. Although there are some broad restrictions from the derivation on the possible functionals F_t, we believe (2.20) holds for essentially any F_t for which both sides exist. One can think of (2.19) with $V(\mathbf{r}'')$ replaced by $i\alpha(\mathbf{r}'')$, so that it becomes the functional Fourier transform of Γ_t, and almost any functional may be represented in this way. We have not tried to justify more rigorously this type of functional analytic continuation, but have checked it in a few simple cases.

Equation (2.20) represents a Wiener path integral in terms of a functional integral. There is a great deal of confusion in the physics literature about nomenclature: Wiener integrals are often called functional integrals and functional integrals are often called path integrals. We shall consistently call the Wiener type path integrals, and those of the form (2.9) functional integrals.

We also note the following formula which, while it has nothing to do with the main subject of this paper, may be of some interest in its own right. We notice that if we put $\mathbf{r} = \mathbf{r}'$

in (2.8) and integrate over $\mathbf{r}$ that

$$\int d\mathbf{r}\, G_t(\mathbf{r},\mathbf{r}) = \sum_{\mu} e^{-tE_\mu} = Z(t), \tag{2.22}$$

where $Z(t)$ is the "partition function" corresponding to H. On the other hand, from (2.9) we have

$$Z(t) = 2\lim_{n\to 0}\int \prod_{\sigma=1}^{n} \mathscr{D}\phi_\sigma \left(\int \phi_1(\mathbf{r})\phi_1(\mathbf{r})d\mathbf{r}\right)\delta([\phi,\phi]-1) \times \exp(-t[\phi, H\phi]). \tag{2.23}$$

Since all terms in (2.23), with the exception of

$$\int \phi_1(\mathbf{r})\phi_1(\mathbf{r})d\mathbf{r},$$

are symmetric in $\phi_1, \ldots, \phi_n$ we may replace this integral by

$$\frac{1}{n}\int \sum_{\sigma=1}^{n} \phi_\sigma(\mathbf{r})\phi_\sigma(\mathbf{r})d\mathbf{r} = \frac{1}{n}[\phi,\phi] = \frac{1}{n}$$

on making use of the δ-function in (2.23). Therefore, we find

$$Z(t) = 2\lim_{n\to 0}\frac{1}{n}\int \prod_{\sigma} \mathscr{D}\phi_\sigma \delta([\phi,\phi]-1) \times \exp(-t[\phi, H\phi]). \tag{2.24}$$

This represents the partition function as a kind of "spherical model" for an n-component field, in the limit where the number of components goes to zero.

III. ASYMPTOTIC EXPANSION FOR LARGE t

We now consider the right-hand side of (2.20) for large t. Let

$$F_t = F_\infty + f_t, \tag{3.1}$$

where f_t approaches zero as t approaches infinity. Then, because of the proportionality of the exponent in (2.20) to t, we would expect the leading exponential term to be given by the functional integral analog of the method of Laplace for ordinary integrals with a large parameter. That is,

$$\lim_{t\to\infty}\left\{-\frac{1}{t}\left\langle e^{-tF_t}\middle|\begin{matrix}\mathbf{r}(t)=\mathbf{r}'\\ \mathbf{r}(0)=\mathbf{r}\end{matrix}\right\rangle\right\} = S_0, \tag{3.2}$$

where S_0 is the minimum of S_∞ subject to the constraint $[\phi,\phi]=1$, in the limit $n=0$, and

$$S_\infty = \sum_\sigma \left(\phi_\sigma, \frac{p^2}{2}\phi_\sigma\right) + F_\infty[\phi^2(\cdot)], \tag{3.3}$$

$$\phi^2(\mathbf{r}) = \sum_\sigma \phi_\sigma^2(\mathbf{r}). \tag{3.4}$$

Calling the minimizing function $\phi_{0\sigma}$, we have at once

$$\frac{p^2}{2}\phi_{0\sigma}(\mathbf{r}) + \frac{\delta F_\infty[\phi_0^2(\cdot)]}{\delta\phi_0^2(\mathbf{r})}\phi_{0\sigma}(\mathbf{r}) = \epsilon\phi_{0\sigma}(\mathbf{r}), \tag{3.5}$$

where ϵ is the Lagrange multiplier corresponding to the constraint $[\phi,\phi]=1$, and $\delta F_\infty/\delta\phi_0^2(\mathbf{r})$ is the functional (Frechet) derivative of F_∞. Writing

$$\phi_{0\sigma}(\mathbf{r}) = \phi_0(\mathbf{r})n_\sigma, \tag{3.6}$$

where n_σ is a (not necessarily constant) unit vector[7]

$$\sum_\sigma n_\sigma^2 = 1. \tag{3.7}$$

We now prove that the minimum of S_∞ is achieved when n_σ is an arbitrary constant unit vector. Substituting (3.6) in (3.3) we have

$$S_\infty = -\frac{1}{2}\int d\mathbf{r}\sum_\sigma \phi_0(\mathbf{r})n_\sigma[n_\sigma\nabla^2\phi_0(\mathbf{r}) + 2(\nabla_{\mathbf{r}}n_\sigma) \cdot(\nabla_{\mathbf{r}}\phi_0(\mathbf{r}) + \phi_0(\mathbf{r})\nabla^2 n_\sigma] + F_\infty[\phi_0^2(\cdot)]. \tag{3.8}$$

Applying the gradient operator to (3.7) we obtain

$$\sum_\sigma n_\sigma\nabla_{\mathbf{r}}n_\sigma = 0, \tag{3.9}$$

$$\sum_\sigma [(\nabla_{\mathbf{r}}n_\sigma)\cdot(\nabla_{\mathbf{r}}n_\sigma) + n_\sigma\nabla^2 n_\sigma] = 0. \tag{3.10}$$

Therefore, (3.8) becomes

$$S_0 = -\frac{1}{2}\int d\mathbf{r}\,\phi_0(\mathbf{r})\nabla^2\phi_0(\mathbf{r}) + F_\infty[\phi_0^2(\cdot)] + \frac{1}{2}\int d\mathbf{r}\,\phi_0^2(\mathbf{r})\sum_\sigma(\nabla_{\mathbf{r}}n_\sigma)^2. \tag{3.11}$$

Clearly, the minimum of S_∞ is obtained by making the last term on the right-hand side of (3.11) as small as possible. Since this term is non-negative its minimum is zero, which is attained by making n_σ an arbitrary constant unit vector. From (3.5) and (3.6) for constant n_σ,

$$\frac{p^2}{2}\phi_0(\mathbf{r}) + \left(\frac{\delta F_\infty[\phi_0^2(\cdot)]}{\delta\phi_0^2(\mathbf{r})}\right)\phi_0(\mathbf{r}) = \epsilon\phi_0(\mathbf{r}), \tag{3.12}$$

so that $\phi_0(\mathbf{r})$ is independent of the choice of n_σ. Therefore, combining (3.11) and (3.12) we see that S_0 is given by exactly the prescription (1.4) of Donsker and Varadhan.

To go beyond the Donsker–Varadhan result is fairly straightforward by the method of Laplace, just as it is for ordinary integrals. We write

$$\phi_\sigma(\mathbf{r}) = \phi_0(\mathbf{r})n_\sigma + \chi_\sigma(\mathbf{r}) \tag{3.13}$$

and expand S and the integrand in (2.20) in powers of χ_σ. There are many separate cases to consider, depending on the form of F_t and any invariance properties it may possess. For simplicity we consider only the case of $F_t = F_\infty$, independent of t. (If this is not the case, the additional terms are easily obtained by expanding F_t for large t, and then expanding in these corrections.) More important is the fact that since F_t is a functional of $\Sigma_\sigma\phi_\sigma^2(\mathbf{r})$ alone, we have seen that the minimizing solution is not unique, but contains an arbitrary constant vector n_σ in the internal space.[7] This gives rise to complications which we are now going to discuss. We also mention that if F_t is translationally invariant (which is a common case in physical applications) the minimizing solution has the form $\phi_0(\mathbf{r}-\mathbf{r}_0)n_\sigma$, where $\mathbf{r}_0$ is an arbitrary fixed point in R^d and n_σ a constant vector in the internal space. This extra nonuniqueness also requires special treatment, which we shall give below.

To carry out the method of Laplace directly, it is necessary to expand the exponent S around the minimizing function according to (3.13). Now this works as long as there are no directions in function space away from the point $\phi_0(\mathbf{r})n_\sigma$, for which S does not increase. Then the main contribution comes from small $\chi_\sigma(\sim 1/\sqrt{t})$, and the entire expansion process makes sense. However, for our problem, because of

the rotational invariance of S in the internal space under the same rotation at each point $\mathbf{r}$, changes in the minimizing solution which just change the direction of n_σ have the same S_0. Therefore, there are always directions in function space away from $\phi_0(\mathbf{r})n_\sigma$ (given n_σ) for which S does not change and the components of χ_σ in these directions are not forced to be small. To see how to handle this technical difficulty, it is easiest to proceed indirectly as follows.

Let $\phi_\lambda(\mathbf{r})$ be a complete orthonormal set of functions of of which (ϕ_0) is the minimizing solution in (3.13), and write

$$\phi_\sigma(\mathbf{r}) = \sum_\lambda \alpha_{\lambda\sigma}\phi_\lambda(\mathbf{r}). \tag{3.14}$$

Since this is an orthogonal transformation we have

$$\prod_\sigma \mathscr{D}\,\phi_\sigma = \prod_\sigma \prod_\lambda d\alpha_{\lambda\sigma}. \tag{3.15}$$

Now separate out the ϕ_0 term,

$$\phi_\sigma(\mathbf{r}) = \alpha_{0\sigma}\phi_0(\mathbf{r}) + \sum_{\lambda\neq 0} \alpha_{\lambda\sigma}\phi_\lambda(\mathbf{r}), \tag{3.16}$$

and put

$$\alpha_{0\sigma} = \alpha_0 n_\sigma(\theta_1, \ldots, \theta_{n-1}) \equiv \alpha_0 n_\sigma(\theta), \tag{3.17}$$

$$\alpha_{\lambda\sigma} = \sum_{k=1}^{n} e_\sigma^{(k)}\beta_{\lambda k} \quad (\lambda \neq 0). \tag{3.18}$$

Here $\theta_1 \cdots \theta_{n-1}$ are the $n-1$ angles necessary to specify the direction of the vector n_σ in the internal space, and $e_\sigma^{(k)}$ $(k = 1, \ldots, n)$ are a complete orthonormal set of vectors in the internal space. Then (3.15) becomes

$$\prod_\sigma \mathscr{D}\phi_\sigma = \alpha_0^{n-1}\, d\alpha_0 d\Omega_n(\theta) \prod_{\lambda\neq 0}\prod_{k=1}^{n} d\beta_{\lambda k}, \tag{3.19}$$

where $d\Omega_n(\theta)$ is an element of surface area on a unit sphere in n-dimensional space. If in (2.20) we hold the θ_i integrations till last we can make the further transformation

$$\sum_{k=1}^{n} e_\sigma^{(k)}\beta_{\lambda k} = \gamma_{\lambda 0} n_\sigma(\theta) + \sum_{j=1}^{n-1} \gamma_{\lambda j} t_\sigma^{(j)}(\theta), \tag{3.20}$$

where $n_\sigma(\theta)$, $t_\sigma^{(1)}(\theta)$, ..., $t_\sigma^{(n-1)}(\theta)$ are a complete set of orthonormal vectors in the internal space. Since this is also an orthogonal transformation for fixed θ_i, (3.19) and (3.16) become

$$\prod_\sigma \mathscr{D}\phi_\sigma = \alpha_0^{n-1}\, d\alpha_0 d\Omega_n(\theta) \prod_{\lambda\neq 0}\left(\prod_{j=0}^{n-1} d\gamma_{\lambda j}\right), \tag{3.21}$$

$$\phi_\sigma(\mathbf{r}) = \left[\alpha_0\phi_0(\mathbf{r}) + \sum_{\lambda\neq 0}\gamma_{\lambda 0}\phi_\lambda(\mathbf{r})\right] n_\sigma(\theta) + \sum_{\lambda\neq 0}\sum_{j=1}^{n-1}\gamma_{\lambda j}\phi_\lambda(\mathbf{r}) t_\sigma^{(j)}(\theta). \tag{3.22}$$

The constraint $[\phi, \phi] = 1$ becomes

$$[\phi, \phi] = \alpha_0^2 + \sum_{\lambda\neq 0}\sum_{j=0}^{n-1}\gamma_{\lambda j}^2 = 1. \tag{3.23}$$

We may write (3.22) in the form (3.13), where

$$\chi_0(\mathbf{r}) = \left((\alpha_0 - 1)\phi_0(\mathbf{r}) + \sum_{\lambda\neq 0}\gamma_{\lambda 0}\phi_\lambda(\mathbf{r})\right) n_\sigma(\theta) + \sum_{\lambda\neq 0}\sum_{j=1}^{n-1}\gamma_{\lambda j}\phi_\lambda(\mathbf{r}) t_\sigma^{(j)}(\theta). \tag{3.24}$$

All the γ's will be small because in the integrand of (2.20) they take us in directions orthogonal to the set of functions $\phi_0(\mathbf{r})n_\sigma(\theta)$ (for all θ). On the other hand, from (3.23) $(\alpha_0 > 0)$,

$$\alpha_0 = +\left(1 - \sum_{\substack{\lambda\neq 0 \\ j=0,\ldots,n-1}} \gamma_{\lambda j}^2\right)^{1/2} = 1 - \tfrac{1}{2}\sum_{\substack{\lambda\neq 0 \\ j=0,\ldots,n-1}} \gamma_{\lambda j}^2 + \cdots,$$

$$1 - \alpha_0 = \tfrac{1}{2}\sum \gamma_{\lambda j}^2 + \cdots. \tag{3.25}$$

Thus, we have for the leading term of χ_σ

$$\chi_\sigma(\mathbf{r}) = \left[\left(-\frac{1}{2}\sum_{\substack{\lambda\neq 0 \\ j=0,\ldots,n-1}} \gamma_{\lambda j}^2\right)\phi_0(\mathbf{r}) + \sum_{\lambda\neq 0}\gamma_{\lambda 0}\phi_\lambda(\mathbf{r})\right] n_\sigma(\theta) + \sum_{\substack{\lambda\neq 1,\ldots,n-1 \\ j=1,\ldots,n-1}} \gamma_{\lambda j}\phi_\lambda(r) t_\sigma^{(j)}(\theta). \tag{3.26}$$

Now suppose we put (3.13) in S_∞ and expand in χ_σ to the second order. Making use of (3.12) and the constraint $[\phi, \phi] = 1$, a simple calculation gives

$$S_\infty = S_0 + S_2, \tag{3.27}$$

where

$$S_0 = \left(\phi_0, \frac{p^2}{2}\phi_0\right) + F_\infty(\phi_0^2(\cdot)) \tag{3.28}$$

and

$$S_2 = \sum_{\sigma,\sigma'}\int d\mathbf{r}d\mathbf{r}'\chi_\sigma(\mathbf{r})\chi_{\sigma'}(\mathbf{r}')\langle \mathbf{r}\sigma|M|\mathbf{r}'\sigma'\rangle, \tag{3.29}$$

where

$$\langle \mathbf{r}\sigma|M|\mathbf{r}'\sigma'\rangle = \langle \mathbf{r}|M_L|\mathbf{r}'\rangle n_\sigma n_{\sigma'} + \langle \mathbf{r}|M_T|\mathbf{r}'\rangle(\delta_{\sigma\sigma'} - n_\sigma n_{\sigma'}), \tag{3.30}$$

$$\langle \mathbf{r}|M_T|\mathbf{r}'\rangle \equiv \left\langle \mathbf{r}\left|\frac{p^2}{2}\right|\mathbf{r}'\right\rangle + \left[\left(\frac{\delta F_\infty(\phi^2(\cdot))}{\delta\phi^2(\mathbf{r})}\right)_{\phi^2=\phi_0^2} - \epsilon\right]\delta(\mathbf{r}-\mathbf{r}'), \tag{3.31}$$

$$\langle \mathbf{r}|M_L|\mathbf{r}'\rangle \equiv \langle \mathbf{r}|M_T|\mathbf{r}'\rangle + 2\left(\frac{\delta^2 F_\infty(\phi^2(\cdot))}{\delta\phi^2(\mathbf{r})\delta\phi^2(\mathbf{r}')}\right)_{\phi^2=\phi_0^2}\phi_0(\mathbf{r})\phi_0(\mathbf{r}'), \tag{3.32}$$

$$\left\langle \mathbf{r}\left|\frac{p^2}{2}\right|\mathbf{r}'\right\rangle \equiv -\tfrac{1}{2}\nabla_r^2\,\delta(\mathbf{r}-\mathbf{r}').$$

Since we are only trying to obtain the leading term of the Laplace method, we can drop the $\phi_0 n_\sigma$ component (3.26) (it gives a contribution one order higher in the γ's) and write

$$\chi_\sigma(\mathbf{r}) = \left(\sum_{\lambda\neq 0}\gamma_{\lambda 0}\phi_\lambda(\mathbf{r})\right) n_\sigma + \sum_{\substack{\lambda\neq 0 \\ j=1,\ldots,n-1}} \gamma_{\lambda j}\phi_\lambda(\mathbf{r}) t_\sigma^{(j)}. \tag{3.33}$$

Now choose the ϕ_λ to be the normalized eigenfunctions of M_T, i.e.,

$$\int d\mathbf{r}'\langle \mathbf{r}|M_T|\mathbf{r}'\rangle \phi_\lambda(\mathbf{r}') = E_\lambda^T \phi_\lambda(\mathbf{r}). \tag{3.34}$$

Comparing with (3.12) we see that $E_0^T = 0$, $E_\lambda^T > 0$ ($\lambda \neq 0$) since ϕ_0 minimizes S_0. [ϵ is determined by the nonlinear eigenvalue equation (3.12) and the requirement that ϕ_0 be normalizable.] Substituting (3.33) in (3.29) we obtain

$$S_2 = \sum_{\lambda\neq 0}\sum_{j=1}^{n-1} \gamma_{\lambda j}^2 E_\lambda^T + \sum_{\lambda,\lambda'\neq 0} \gamma_{\lambda 0}\gamma_{\lambda' 0}(\phi_\lambda, M_L \phi_{\lambda'}). \tag{3.35}$$

Substituting all these results in the right-hand side of (2.20) we obtain for the leading term of the Laplace method

$$\left\langle e^{-tF_\infty\{L_i(\cdot)\}} \middle| \begin{matrix}\mathbf{r}(t)=\mathbf{r}'\\ \mathbf{r}(0)=\mathbf{r}\end{matrix}\right\rangle = [G_t^{(0)}(\mathbf{r},\mathbf{r}')]^{-1} I_t e^{-tS_0}\phi_0(\mathbf{r})\phi_0(\mathbf{r}'), \tag{3.36}$$

$$\begin{aligned} I_t &= 2\lim_{n\to 0}\int \alpha_0^{n-1} d\alpha_0 d\Omega_n \prod_{\lambda\neq 0} d\gamma_{\lambda_0} \prod_{j=1}^{n-1} d\gamma_{\lambda j}\delta(\alpha_0^2 - 1) \\ &\quad \times n_1(\theta) n_1(\theta) e^{-tS_2} \\ &= \lim_{n\to 0}\left(\frac{\int d\Omega_n}{n}\right)\int \prod_{\lambda\neq 0} d\gamma_{\lambda_0} \prod_{j=1}^{n-1} d\gamma_{\lambda j} e^{-tS_2}, \end{aligned} \tag{3.37}$$

since by symmetry we may replace $n_1^2(\theta)$ by $(1/n)\,\Sigma_\sigma n_\sigma^2 = 1/n$ in the integrand. Using

$$\frac{1}{n}\int d\Omega_n = \frac{1}{n}\frac{2\pi^{n/2}}{\Gamma(n/2)} \underset{n\to 0}{\to} 1 \tag{3.38}$$

and doing the integrals over $\gamma_{\lambda j}$ we have

$$\begin{aligned} I_t &= \lim_{n\neq 0}\left(\prod_{\lambda\neq 0}(\pi/tE_\lambda^T)^{1/2}\right)^{n-1}\int \prod_{\lambda\neq 0} d\gamma_{\lambda_0} \\ &\quad \times \exp\left[-t\sum_{\lambda,\lambda'\neq 0}\gamma_{\lambda 0}\gamma_{\lambda' 0}(\phi_\lambda, M_L\phi_{\lambda'})\right] \\ &= \left(\prod_{\lambda\neq 0}(\pi/tE_\lambda^T)^{1/2}\right)^{-1}\prod_\rho (\pi/tE_\rho^L)^{1/2}, \end{aligned} \tag{3.39}$$

where E_ρ^L are the eigenvalues of the matrix $(\phi_\lambda, M_L\phi_{\lambda'})\lambda,\lambda'\neq 0$. We may also write (3.39) as

$$I_t = [\mathrm{Det}(M_T')/\mathrm{Det}(M_L')]^{1/2}, \tag{3.40}$$

where M_T and M_L are the matrices of M_T and M_L in the subspace of Hilbert space orthogonal to the state ϕ_0, i.e., the matrices taken with respect to the functions ϕ_λ with $\lambda = 0$ excluded. Using a well-known theorem about determinants we have

$$\frac{\mathrm{Det}(M_L')}{\mathrm{Det}(M_T')} = \mathrm{Det}[(M_T')^{-1}M_L']. \tag{3.41}$$

Therefore, (3.31), (3.32) and (3.41), (3.40) may be put in the form

$$I_t = [\mathrm{Det}(\delta_{\lambda\lambda'} + K_{\lambda\lambda'})]^{-1/2} \quad (\lambda,\lambda'\neq 0), \tag{3.42}$$

where

$$\begin{aligned} K_{\lambda\lambda'} &= \frac{2}{E_\lambda^T}\int d\mathbf{r}\, d\mathbf{r}'\left(\frac{\delta^2 F_\infty(\phi^2(\cdot))}{\delta\phi^2(\mathbf{r})\delta\phi^2(\mathbf{r}')}\right)_{\phi^2=\phi_0^2}\phi_0(\mathbf{r})\phi_0(\mathbf{r}') \\ &\quad \times \phi_{\lambda'}(\mathbf{r})\phi_{\lambda'}(\mathbf{r}'). \end{aligned} \tag{3.43}$$

For small $K_{\lambda\lambda'}$ we may use the Fredholm expansion of the determinant in (3.42).

This completes the calculation of the leading term of the method of Laplace for an F_∞ with no further symmetries. By retaining the higher powers in the γ's which have been dropped, we obtain corrections of order $1/t$ or a systematic expansion in powers of $1/t$. The results get quite complicated, and we shall not give them here. We mention, however, that in the higher orders there are some complications in changing integration variables (see, for example, Gervais and Jevicki[5]).

Finally, we show how to treat the difficulties which arise in the case where F_t is translationally invariant. Again we limit ourselves to F_∞, and discuss only the leading term of the Laplace method. By translational invariance we mean that S_∞ defined by (3.3) has the property that if $\phi_\sigma(\mathbf{r})$ is replaced by $\phi_\sigma(\mathbf{r}-\mathbf{r}_0)$ ($\mathbf{r}_0$ being a fixed vector in R^d) the value of S_∞ does not change. This requires that F_∞ have the same property, since in (3.3) the first term on the right-hand side does. This obviously implies

$$\frac{\delta F_\infty[\phi^2(\cdot)]}{\delta\phi^2(\mathbf{r}-\mathbf{r}_0)} = \left(\frac{\delta F_\infty[\phi^2(\cdot)]}{\delta\phi^2(\mathbf{r})}\right)_{\phi^2(\mathbf{r})\to\phi^2(\mathbf{r}-\mathbf{r}_0)}. \tag{3.44}$$

Therefore, if (3.5) has a solution $\phi_0(\mathbf{r})n_\sigma$ then it also has the solution $\phi_0(\mathbf{r}-\mathbf{r}_0)n_\sigma$ with the same value of S_0. This means that there are again directions in function space (obtained by changing $\mathbf{r}_0$) away from the point $\phi_0(\mathbf{r})n_\sigma$, for which S_∞ does not change. To see the necessary modifications of our procedure we shall first show that M_L now has d zero eigenvalues with the eigenfunctions $\partial\phi_0(\mathbf{r})/\partial x_\alpha, \alpha = 1,2,\ldots,d, \mathbf{r} = (x_1,x_2,\ldots,x_d)$. Differentiating Eq. (3.12) for $\phi_0(\mathbf{r})$ with respect to x_α we find

$$\begin{aligned} &\left\{\frac{p^2}{2} + \left[\left(\frac{\delta F_\infty}{\delta\phi^2(\mathbf{r})}\right)_{\phi^2=\phi_0^2} - \epsilon\right]\right\}\frac{\partial\phi_0(\mathbf{r})}{\partial x_\alpha} \\ &\quad + \left\{\frac{\partial}{\partial x_\alpha}\frac{\delta F_\infty}{\delta\phi^2(\mathbf{r})}\right\}_0 \phi_0(\mathbf{r}) = 0. \end{aligned} \tag{3.45}$$

But, for translationally invariant F_∞,

$$\begin{aligned} \frac{\partial}{\partial x_\alpha}\left(\frac{\delta F_\infty}{\delta\phi^2(\mathbf{r})}\right)_0 &= 2\int d\mathbf{r}'\left(\frac{\delta^2 F_\infty}{\delta\phi^2(\mathbf{r}')\delta\phi^2(\mathbf{r})}\right)_0\phi_0(\mathbf{r}') \\ &\quad \times \frac{\partial\phi_0(\mathbf{r}')}{\partial x_\alpha'}, \end{aligned} \tag{3.46}$$

as can be seen at once using (3.44) and taking $\mathbf{r}_0$ to be an infinitesimal vector in the α direction. Therefore, (3.45) becomes

$$\begin{aligned} &\left\{\frac{p^2}{2} + \left[\left(\frac{\delta F_\infty}{\delta\phi^2(\mathbf{r})}\right)_{\phi^2=\phi_0^2} - \epsilon\right]\right\}\frac{\partial\phi_0(\mathbf{r})}{\partial x_\alpha} \\ &\quad + 2\int d\mathbf{r}'\left(\frac{\delta^2 F_\infty}{\delta\phi^2(\mathbf{r})\delta\phi^2(\mathbf{r}')}\right)_{\phi^2=\phi_0^2}\phi_0(\mathbf{r})\phi_0(\mathbf{r}')\frac{\partial\phi_0(\mathbf{r}')}{\partial x_\alpha'}. \end{aligned} \tag{3.47}$$

However, comparing with (3.32) this shows the $[\partial\phi_0(\mathbf{r})/\partial x_\alpha](\alpha = 1,\ldots,d)$ are eigenfunctions of M_L with eigenvalues zero. Therefore, the combinations of $\gamma_{\lambda 0}$'s which correspond to moving χ_σ [see (3.24)] in these directions in function space are not limited to small values, and must be eliminated from χ_σ for the expansion to make sense. To see how to do this, instead of (3.22) we write

$$\phi_\sigma(\mathbf{r}) = \left[\alpha_0\phi_0(\mathbf{r}-\mathbf{r}_0) + \sum_{\lambda\neq 0}\gamma_{\lambda 0}\phi_\lambda(\mathbf{r}-\mathbf{r}_0)\right]n_\sigma(\theta) + \sum_{\lambda\neq 0}\gamma_{\lambda j}\phi_\lambda(\mathbf{r}-\mathbf{r}_0)t_\sigma^{(j)}(\theta), \tag{3.48}$$

where $\mathbf{r}_0$ is a fixed vector in R^d. This is just as satisfactory as (3.22), since $\phi_\lambda(\mathbf{r}-\mathbf{r}_0)$ is just as good a complete orthonormal set as $\phi_\lambda(\mathbf{r})$ is. The volume element is still given by (3.21). Now consider the functions $\psi_\mu(\mathbf{r})$ ($\mu = 1,...,d$), chosen as linear combinations of the $\partial\phi_0(\mathbf{r})/\partial x_\alpha$ such that

$$(\psi_{\mu'},\psi_\mu) = \delta_{\mu\mu'}.$$

[If $\phi_0(\mathbf{r})$ is a function of $|\mathbf{r}|$ alone, then the $\partial\phi_0/\partial x_\alpha$ are orthogonal for different α, and all we need do is normalize them.] Since ϕ_0 is orthogonal to the $\partial\phi_0/\partial x_\alpha$, it is equally orthogonal to the ψ_μ. Thus, we can choose as our orthonormal basis $\{\phi_\lambda\} = \phi_0,\phi_1,\phi_2,...,\psi_1,...,\psi_d = \{\phi_{\bar\lambda}\}, \psi_1,\psi_2,...,\psi_d$, where $\bar\lambda = 0,1,2,...$ ranges over the (incomplete) set of functions $\phi_{\bar\lambda}$. We now want to eliminate the $\gamma_{\mu 0}$ from (3.48). This is done by the following (usual) trick. We note that, putting $\Sigma_\sigma n_\sigma\phi_\sigma(\mathbf{r}) = n\cdot\phi(\mathbf{r})$,

$$\begin{aligned}\int d\mathbf{r}_0 \prod_{\mu=1}^{d}\delta\left(\int d\mathbf{r}\, n\cdot\phi(\mathbf{r})\psi_\mu(\mathbf{r}-\mathbf{r}_0)\right) &= \int d\mathbf{r}_0 \prod_{\mu=1}^{d}\delta\left(\sum_{\beta=+}^{d} A_{\mu\beta}(x_{0\beta}-R_\beta)\right)\\ &= |\mathrm{Det}(A_{\mu\beta})|^{-1}\end{aligned} \tag{3.49}$$

by a well-known result. In (3.49)

$$A_{\mu\beta} \equiv \int d\mathbf{r}\, n\cdot\phi(\mathbf{r})\frac{\partial\psi_\mu(\mathbf{r}-\mathbf{R})}{\partial x_\beta} \tag{3.50}$$

and $\mathbf{R}$ is the solution[8] of

$$\int n\cdot\phi(\mathbf{r})\psi_\mu(\mathbf{r}-\mathbf{R})d\mathbf{r} = 0, \quad \mu = 1,...,d. \tag{3.51}$$

Therefore, we can place the left-hand side of (3.49) multiplied by $|\mathrm{Det}(A_{\mu\beta})|$ in the functional integral (2.20) without changing its value. Using the expansion (3.48) in (2.20) we obtain, because of the translational invariance of S, the same value as we would get dropping the $\mathbf{r}_0$, i.e., our previous results for S_0 and S_2. Equation (3.48) must also be inserted the normalization condition $[\phi,\phi] = 1$ (which it leaves unchanged), and in the inserted factors. The δ function in (3.49) becomes $\Pi_{\mu=1}^{d}\delta(\gamma_{\mu 0})$, which means that we can simply set the $\gamma_\mu = 0$ instead of integrating over them. For the determinant of $A_{\mu\beta}$ we need

$$A^0_{\mu\beta} = \int d\mathbf{r}\,\phi_0(\mathbf{r})\frac{\partial\psi_\mu}{\partial x_\beta} = -\int d\mathbf{r}\frac{\partial\phi_0}{\partial x_\beta}\psi_\mu(\mathbf{r}). \tag{3.52}$$

Now

$$\psi_\mu = \sum_\alpha C_{\mu\alpha}\frac{\partial\phi_0(\mathbf{r})}{\partial x_\alpha}, \tag{3.53}$$

where the $C_{\mu\alpha}$ are so chosen that $(\psi_\mu,\psi_{\mu'}) = \delta_{\mu\mu'}$. That is,

$$\delta_{\mu,\mu'} = \sum_{\alpha,\beta=1}^{d} C_{\mu'\beta}C_{\mu\alpha}\left(\frac{\partial\phi_0}{\partial x_\beta},\frac{\partial\phi_0}{\partial x_\alpha}\right) = \sum_{\alpha,\beta=1}^{d} C_{\mu'\beta}C_{\mu\alpha}D_{\beta\alpha}, \tag{3.54}$$

$$D_{\beta\alpha} \equiv \left(\frac{\partial\phi_0}{\partial x_\beta},\frac{\partial\phi_0}{\partial x_\alpha}\right). \tag{3.55}$$

Taking the determinant of both sides of (3.54), we find at once

$$C^2D = 1, \tag{3.56}$$

where C is the determinant of $C_{\mu\alpha}$ and D the determinant of $D_{\beta\alpha}$. Substituting (3.53) in (3.52), we have

$$A^0_{\mu\beta} = -\sum_\alpha C_{\mu\alpha}D_{\alpha\beta}. \tag{3.57}$$

Again take determinants of both sides of (3.57),

$$|\mathrm{Det}(A^0_{\mu\beta})| = |C|\,|D| = |D|^{1/2}, \tag{3.58}$$

by (3.56).

The rest of the calculation is exactly as in the previous case without translational invariance, and (3.36) is simply replaced by

$$\left\langle e^{-tF_\infty[L_A(\cdot)]}\middle|\begin{matrix}\mathbf{r}'\\ \mathbf{r}\end{matrix}\right\rangle = [G_t^{(0)}(\mathbf{r},\mathbf{r}')]^{-1}\left(\int d\mathbf{r}_0\phi_0(\mathbf{r}-\mathbf{r}_0)\phi_0(\mathbf{r}'-\mathbf{r}_0)\right)e^{-tS_\bullet}\tilde I_t, \tag{3.59}$$

$$\tilde I_t = |D|^{1/2}\left(\frac{\pi}{t}\right)^{-d/2}[\mathrm{Det}(M'_T)/\mathrm{Det}(M''_L)]^{1/2}, \tag{3.60}$$

where M''_L is the matrix of M_L in the subspace of Hilbert space orthogonal to the states $\phi_0(\mathbf{r}),\partial\phi_0(\mathbf{r})/\partial x_\alpha$.

[1] R. Friedberg and J. M. Luttinger, Phys. Rev. B **12**, 4460 (1975).

[2] M. D. Donsker and S. R. S. Varadhan, Commun. Pure Appl. Math. **28**, 1 (1975); **28**, 279 (1975); **28**, 525 (1975). See also *Proc. Int. Conf. on Function Space Integration* (Oxford U.P., New York, 1974), for a simple preliminary report of the work by Donsker and Varadhan.

[3] M. Kac, *Probability and Related Topics in Physical Science* (Interscience, New York, 1959), Vol. 1, p. 161 ff.

[4] We assume $V(\mathbf{r})$ goes to infinity as $|r|\to\infty$ sufficiently rapidly so that the spectrum of H is discrete. This is not an essential limitation, but only simplifies the presentation.

[5] Functional integrals of this sort have been widely used by theoretical physicists since the early fifties without, however, mathematically rigorous discussions of the needed limiting processes. Two clear discussion of how one operates with them are J. L. Gervais and A. Jevecki, Nucl. Phys. B **110**, 93–112 (1976); L. D. Faddeev and A. A. Slavnov, *Gauge Fields* (Benjamin/Cummings, New York, 1980), pp. 59–69. We shall operate in the same formal manner as these authors, though it is not absolutely certain that this *always* leads to correct results.

[6] It is necessary, of course, to also discretize operators such as H in some manner if we are to use this definition. We shall not discuss this here, as it is not necessary for our purposes.

[7] We mean vector in the n-dimensional vector space at each point $\mathbf{r}$ of R^d. We shall call this the "internal space" at point $\mathbf{r}$. A constant vector n_σ in this space means one which is independent of $\mathbf{r}$.

[8] This solution exists and is unique in the following sense. Since we are using the method of Laplace, $n\cdot\phi(\mathbf{r})$ can be taken as $\phi_0(\mathbf{r})$ in the leading term. Since $\psi_\mu(\mathbf{r})$ is orthogonal to $\phi_0(\mathbf{r})$, (3.51) will vanish for $\mathbf{R} = 0$. We have not found it possible to find the conditions such that this is the only root, but have found it to be so for all simple examples considered. We shall assume this to be the case; if it is not, it is easy to modify the formulas to take this fact into account.

Stochastic optimal control and quantum mechanics

Lech Papiez
Department of Applications of Probability Theory, Institute of Mathematics, Silesian University, Katowice, Poland

(Received 28 October 1980; accepted for publication 17 April 1981)

Classical mechanics is formulated as a kind of deterministic optimal control. The simplest, from the optimal control point of view, stochastic generalization of classical mechanics is submitted. The connections between stochastic mechanics and quantum mechanics are shown, thanks to the stochastic optimal control method.

PACS numbers: 02.50.Ey, 02.30.Wd

1. INTRODUCTION

The aim of this paper is to lay stress on close connections between quantum dynamics (Schrödinger equation) and the probabilistic variational principle. First of all, we try to show here that, from the optimal control point of view, quantum dynamics is the simplest and the most natural generalization of classical mechanics laws of motion. In order to gain this destination, we begin with deterministic optimal formulation of classical dynamics and apply it to the derivation of the Hamilton–Jacobi equation. This allow us to give the formulation of the analogous optimal control problem for the stochastic system. Then, applying to this problem results of the stochastic control theory we formally obtain the Hamilton–Jacobi equation for this system, which is equivalent to the Schrödinger equation.

Contrary to previous stochastic variational derivations of the Schrödinger equation[1,2] we use not a positive but an imaginal diffusion constant. For our considerations this constant is physically required. Unfortunately, in this case, it is difficult to keep the mathematical method rigorous.[3] Anyway, a justification of the present method may be given by understanding it as a Euclidean method.

In the paper we introduce only those notions and results of the optimal control theory[4] which are useful for our purposes.

2. DETERMINISTIC OPTIMAL CONTROL AND CLASSICAL MECHANICS

Let us be given a dynamical system

$$\dot{x} = f(t,x,u), \tag{2.1}$$

where $t \in [t_0,t_1] \subset R$ is a time, $x \in R^n$ is a position of a system, $u \in U$ (U—a subset of R^m) is some parameter, and f is a sufficiently regular function $f: R^1 \times R^n \times R^m \to R^n$. Let us assume also that a set $\mathcal{U}$ of continuous functions $u: [t_0,t_1] \to U$ is given (though a wider classes than $\mathcal{U}$ may also be considered[4]). For a fixed function $u(t)$ from a set $\mathcal{U}$ the expression (2.1) becomes a well defined differential equation and its solution $x_u(t)$ is called a trajectory of a system relevant to a control function $u(t)$. In general, any trajectory $x(t)$ should also fulfill some initial and final conditions.

Moreover, let us connect with a system (2.1) some functional $J(t_0,x_0,u) = \phi(t_0,x_0,t_1,x_u(t_1))$, depending on a control function u and initial conditions t_0, x_0 of a trajectory. Now, the problem of finding such a control function $u(t)$ that the relevant trajectory $x_u(t)$ given by (2.1) fulfills boundary conditions and gives a minimum value for $J(t_0,x_0,u)$ is called a deterministic optimal control.

We have two different ways to solve the optimal control problem. The Pontriagin maximum principle[4,5] is one of them and the Bellman principle of dynamical programming[4,6,7] is the second. More interesting for our purposes is the second method. Thus we describe, in a few words, only this one. The crucial notion for the method of dynamical programming is the function $S(t,x)$, defined as[4]

$$S(t,x) = \inf_{u \in \mathcal{U}_{t,x}} \phi(t,x,t_1,x_u(t_1)), \tag{2.2}$$

where $\mathcal{U}_{t,x} \subset \mathcal{U}$ denotes a set of all possible $u(t)$ functions for which trajectories $x(t)$ equal x for the initial moment t. The following equality may be proved[4] for this function:

$$\frac{\partial S(t,x)}{\partial t} + \frac{\partial S(t,x)}{\partial x} f(t,x,u^*(t)) = 0. \tag{2.3}$$

Here $u^*(t)$ is an optimal control function. We neglect boundary conditions for (2.3), as they are not important for our considerations.

Let us concentrate now on the special case of a dynamical system (2.1). We assume that

(i) no restrictions exist for values of a control function $u(t)$, i.e., $U = R^n$,

(ii) Eq. (2.1) is of the simple form

$$\dot{x} = u, \tag{2.4}$$

(iii) the functional which constitutes the criterion is given as

$$J(t,x,u) = \int_t^{t_1} L(t',x(t'),u(t'))\,dt', \tag{2.5}$$

where $L(t,x,u)$ is a Lagrange function,

(iv) t_1 and $x(t_1) = x_1$ are fixed values.

For the optimal control problem defined above Eq. (2.3) becomes

$$\frac{\partial S(t,x)}{\partial t} + \frac{\partial S(t,x)}{\partial x} \dot{x}^*(t) + L(t,x,\dot{x}^*(t)) = 0. \tag{2.6}$$

Taking into account the standard form of the Lagrange function,

$$L(t,x,u) = (m/2)u^2 - V(x), \tag{2.7}$$

where $V(x)$ is a potential, we may easily obtain the equality

 0022-2488/82/061017-03$02.50

$$\frac{\partial S(t,x)}{\partial x} = -mu^{*T}(t) = -m\dot{x}^{*T}(t). \tag{2.8}$$

Now, applying (2.7) and (2.8) we have, from (2.6),

$$\frac{\partial S(t,x)}{\partial t} - \frac{1}{2m}\left[\frac{\partial S(t,x)}{\partial x}\right]^2 - V(x) = 0. \tag{2.9}$$

It is not difficult to notice that any dynamical system of classical Lagrangian mechanics is of the type considered above. Moreover, we realize that the action of classical mechanics,[8]

$$S^i_{cm}(t,x) = \int_t^{t_1} L(t',x^*(t'),\dot{x}^*(t'))\,dt', \tag{2.10}$$

treated as a function of the trajectories initial conditions t,x,[9] and the $S(t,x)$ function, defined for the system considered above, are identical. This helps us to see that $S^i_{cm}(t,x)$ fulfills Eq. (2.9). Equation (2.9) for $S^i_{cm}(t,x)$ is, of course, well known in classical mechanics and is called the Hamilton–Jacobi equation [for "initial conditions" action function $S^i_{cm}(t,x)$].

Thus, we may really regard classical Lagrangian mechanics as a section of deterministic optimal control.

3. STOCHASTIC OPTIMAL CONTROL AND QUANTUM MECHANICS

Let us now consider a situation when a dynamical system undergoes deterministic as well as stochastic excitations. The purely stochastic excitation may be represented, in the best way, by a white noise. Thus, the most natural generalization of the equation of motion (2.1) is[4]

$$\dot{x} = f(t,x,u) + \sigma(t,x,u)\gamma, \tag{3.1}$$

where $f, t, x, u, x(t)$, and $u(t)$ are variables identical as in (2.1), γ is the n-dimensional white noise, i.e., the generalized derivative of the Wiener process,[10] and $\sigma(t,x,u)$ is a function interpreted as a diffusion coefficient. We restrict ourselves to such a stochastic system of type (3.1) which is the simplest probabilistic generalization of the classical mechanical one. As the dynamical equation (2.4) in the optimal control formulation of classical mechanics is $\dot{x} = u$ so its simplest random equivalent, in accordance with (3.1), should have a form

$$\dot{x} = u + a\gamma, \tag{3.2}$$

where a is some constant matrix. For a purpose which will be clear later we take $a = (-i\hbar/m)^{1/2}I$, where I is the unity matrix. The natural generalization of the criterion (2.5) is

$$J(t,x,u) = E_{tx}\left\{\int_t^{t_1} L(t',x(t'),u(t'))\,dt'\right\}, \tag{3.3}$$

where $E_{tx}\{\ \}$ denotes a kind of a mean value[4,11] and $u(t)$ is a so-called feedback control function[4] $u(t) = u(t,x(t))$. The mean value operation E_{tx} is uniquely determined by a stochastic process $x(t)$ fulfilling (3.2) and the initial condition $x(t) = x$ (x is considered here as a common, nonrandom variable). We may define a nonrandom action function $\tilde{S}(t,x)$ for any stochastic system the same way as in (2.2). We have[12]

$$\tilde{S}(t,x) = \inf_u J(t,x,u). \tag{3.4}$$

It may be proved that the function $\tilde{S}(t,x)$ fulfills, for a general system (3.1) and a criterion (3.3), the equation[4]

$$\frac{\partial S(t,x)}{\partial t} + A^{u^*}(t)\tilde{S}(t,x) + L(t,x,u^*(t)) = 0, \tag{3.5}$$

where

$$A^{u^*}(t) = \tfrac{1}{2}\sum_{i,j=1}^{n} a_{ij}(t,x,u^*(t))\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} f_i(t,x,u^*(t))\frac{\partial}{\partial x_i}, \tag{3.6}$$

$a_{ij}(t,x,u)$ is the i,j element of the $\sigma(t,x,u)\sigma^T(t,x,u)$ matrix, $f_i(t,x,u)$ is the ith element of a vector function $f(t,x,u)$, $i,j = 1,2,\ldots,n$, $u^*(t) = u^*(t,x^*(t)) = u^*(t,x)$ is a feedback optimal control function. If we consider a simple stochastic generalization (3.2) of a classical mechanical system we shall get the following simplified version of Eq. (3.5):

$$\frac{\partial \tilde{S}(t,x)}{\partial t} - \frac{i\hbar}{2m}\sum_{i=1}^{n}\frac{\partial^2 S(t,x)}{\partial x_i \partial x_i} + \sum_{i=1}^{n} u_i^*(t)\frac{\partial \tilde{S}(t,x)}{\partial x_i} + L(t,x,u^*(t)) = 0. \tag{3.7}$$

Let us now consider the following equation, which is equivalent to (3.7)[4]:

$$\frac{\partial \tilde{S}(t,x)}{\partial t} + \min_{u\in R_n}\left\{-\frac{i\hbar}{2m}\sum_{i=1}^{n}\frac{\partial^2 \tilde{S}(t,x)}{\partial x_i \partial x_i} + u\frac{\partial \tilde{S}(t,x)}{\partial x} + L(t,x,u)\right\} = 0. \tag{3.8}$$

For the standard form (2.7) of the Lagrange function, (3.8) transforms into

$$\frac{\partial \tilde{S}(t,x)}{\partial t} + \min_{u\in R^n}\left\{-\frac{i\hbar}{2m}\sum_{i=1}^{n}\frac{\partial^2 \tilde{S}(t,x)}{\partial x_i \partial x_i} + u\frac{\partial \tilde{S}(t,x)}{\partial x} + \frac{m}{2}u^2 - V(x)\right\} = 0. \tag{3.9}$$

We easily get that the minimum value for the expression in the bracket above is gained for

$$u^T = -\frac{1}{m}\frac{\partial \tilde{S}(t,x)}{\partial x}, \tag{3.10}$$

where u^T is a transposition of a column vector u. Thus it is clear from (3.7), (3.8), and (3.10) that

$$u^{*T}(t) = u^{*T}(t,x) = -\frac{1}{m}\frac{\partial \tilde{S}(t,x)}{\partial x}. \tag{3.11}$$

Applying (3.11) into (3.7) we get finally

$$\frac{\partial \tilde{S}(t,x)}{\partial t} - \frac{i\hbar}{2m}\sum_{i=1}^{n}\frac{\partial^2 \tilde{S}(t,x)}{\partial x_i \partial x_i} - \frac{1}{2m}\times\sum_{i=1}^{n}\left[\frac{\partial \tilde{S}(t,x)}{\partial x_i}\right]^2 - V(x) = 0. \tag{3.12}$$

The crucial fact of this paper is that Eq. (3.12) is a kind of Schrödinger equation. Namely, if to a common form of Schrödinger equation,

$$i\hbar\frac{\partial\psi(t,x)}{\partial t} = -\frac{\hbar^2}{2m}\sum_{i=1}^{n}\frac{\partial^2\psi(t,x)}{\partial x_i\partial x_i} + V(x)\psi(t,x), \tag{3.13}$$

we apply a substitution

$$\psi(t,x) = \exp\{-(i/\hbar)S_{qm}(t,x)\}, \tag{3.14}$$

we shall obtain

$$\frac{\partial S_{qm}(t,x)}{\partial t} + \frac{i\hbar}{2m}\sum_{i=1}^{n}\frac{\partial^2 S_{qm}(t,x)}{\partial x_i\partial x_i} - \frac{1}{2m} \times\sum_{i=1}^{n}\left[\frac{\partial S_{qm}(t,x)}{\partial x_i}\right]^2 - V(x) = 0. \tag{3.15}$$

Thus we see now that the last expression is identical to (3.12).

It is worth noticing at last that for $\hbar\to 0$ Eqs. (3.12) and (3.15) tend to a classical limit given in (2.7).

4. DISCUSSION

Let us notice that $\tilde{S}(t,x)$ is a complex valued function. Thus it seems convincing to interpret the function $\psi(t,x) = \exp\{(i/\hbar)\tilde{S}(t,x)\}$ identically as in common quantum mechanics, i.e., that $|\psi(t,x)|^2$ is a probability density function for a system to be in position x at time t. Briefly speaking, we accept the standard interpretation of quantum mechanics and the problem of hidden variables does not exist here. This makes us, however, understand complex valued stochastic processes $\dot{x}(t)$ and $x(t)$ to be only theoretically useful variables connected to the system under consideration and not give them a physical interpretation of velocity and position. This interpretation explains the fact that processes $\dot{x}(t)$ and $x(t)$ do not determine the state of the system.

It is worth noticing, at last, that we shall avoid bothering our head about the complex space–time if complex valued trajectories are understood as starting from a real position x.[13] In this case $\tilde{S}(t,x)$ is a complex valued function given on the real space–time.

[1] A. Blaquiere, J. Optimization Theor. Appl. **27**, 71 (1979). *Dynamical Systems and Microphysics*, edited by A. Blaquiere *et al.* (Springer, Berlin, 1980).

[2] K. Yasue, Lett. Math. Phys. **4**, 357 (1980); J. Math. Phys. **22**, 1010 (1981); J. Funct. Anal. (1981, to be published).

[3] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. II* (Academic, New York, 1975).

[4] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control* (Springer, Berlin, 1975).

[5] L. C. Pontriagin, V. G. Boltanski, R. V. Gamkrelidze, and E. F. Mistschenko, *Matematitsheskaia Teoria Optimalnykh Prcessov* (Nauka, Moscow, 1976).

[6] J. N. Roitenberg, *Avtomatitscheskie Upravlenie* (Nauka, Moscow, 1976).

[7] V. I. Zubov, Teoria Uravneni Upravlaemogo Dvizhenia, Izd. Len. Univ. Leningrad, 1980.

[8] L. Landau and E. Lipshitz, *Mekhanika* (Moscow, 1958).

[9] Usually, authors of handbooks in classical mechanics used functions analogous to $S^i_{cm}(t,x)$ but depending on final variables. See, for example, L. Landau and E. Lipshitz, Ref. 8.

[10] I. I. Gikhman, A. V. Skoronkhod, and M. I. Iadrienko, *Teoria Veroiatnostei i Matematitsheskaia Statistika* (Vistsha Shkola, Kiev, 1979).

[11] A. D. Ventzel, *Kurs Teorii Slutshainykh Processov* (Nauka, Moscow, 1975).

[12] To avoid difficulties let us take here the mass m to be purely imaginary (as in the definition of Feynman's formula by analytic continuation).

[13] The similar problem is analyzed in V. P. Maslov's book *Kompleksnyi Metod VKB v Nelineinykh Uravneniakh* (Nauka, Moscow, 1977).

On the dynamics of diffusions and the related general electromagnetic potentials

L. M. Morato

Istituto di Elettrotecnica ed Elettronica dell'Università di Padova and Laboratorio per ricerche di Dinamica dei Sistemi e di Bioingegneria del CNR, Corso Stati Uniti, Camin, Padova, Italy

(Received 21 July 1981; accepted for publication 6 October 1981)

In this paper the Nelson's stochastic mechanics is extended to general diffusion motions. A representation theorem is proved which gives a one-to-one correspondence between solutions of certain Schrödinger equations and diffusion processes satisfying appropriate regularity conditions. Exploiting results of stochastic mechanics on Riemannian manifolds it is shown that the real part of the Schrödinger equations corresponding to the considered diffusions can be interpreted as Newton's second law where the force is produced by generalized electromagnetic potentials.

PACS numbers: 02.50.Ey, 02.40.Ky

I. INTRODUCTION

Suppose that the motion of a particle is modeled by a diffusion taking values in $\mathbb{R}^3$ with diffusion coefficient $(\hbar/m)I$, h being the Planck's constant, m the mass of the particle, and I the 3×3 identity matrix. Then, defining in a convenient fashion the stochastic kinematical variables, we know from Nelson's work on stochastic mechanics[1,2] that the continuity equation of the process and Newton's second law $ma = F$, where a is a certain mean acceleration and F the classical force, can be transformed by means of a change of variables into the corresponding Schrödinger equation. The theory has been developed for the cases when F is proportional to the gradient of some scalar potential V or comes from an electromagnetic potential $(\mathscr{A},\phi)$.

The present work is mainly concerned with systems with a finite number of degrees of freedom such that the time evolution of the configuration variables, for given initial conditions, can be well described by a diffusion process. This process satisfies an Itô stochastic differential equation with general diffusion matrix. Our main result is that any general n-dimensional diffusion, provided some regularity conditions are satisfied, has a mean acceleration which comes from a gauge invariant class on $n+1$-vector potentials.

In Nelson's work the fact that the mean acceleration a is proportional to the classical force plays the role of a constraint and determines the drift of the diffusion process. In this paper the reverse line of reasoning is followed. First the kinematics proposed by Nelson is generalized to diffusions having general diffusion matrix. Then a theorem is proved which provides, for any given diffusion, a correspondence between diffusion processes and solutions of Schrödinger equations where the Hamiltonians come in general from $n+1$-vector potentials. Under some regularity conditions the correspondence is seen to be one-to-one, provided a stochastic process is considered to be defined by all its joint finite-dimensional distributions. In our setting the process is determined by the Markov property and by all its transition probabilities.

The case of a strictly stationary process leads to real-time-independent Schrödinger equations, the eigenvalues of which are the expected values of the kinetic terms and of the potential energy. If the process represents the time evolution of the values of the configuration variables of some physical system one can also seek a physical interpretation. The imaginary part of the Schrödinger equation represents the continuity equation of the given process. The real part is expected to represent Newton's second law, where F is the "Lorentz force" produced by the $n+1$-vector potential. It turns out that if the diffusion matrix is constant with respect to the values taken by the process (nonlocal diffusion) one finds directly a Lorentz-force type expression for F.[3,4]

For a general local diffusion one needs to exploit methods of stochastic mechanics on Riemannian manifolds.[5–7] The idea is that of associating in a canonical way a Riemannian manifold $\mathbb{M}$ to the diffusion matrix of the process. Then the given diffusion, which takes values in the Euclidean space, can be considered as a coordinate representation of a nonlocal diffusion taking values on $\mathbb{M}$.

In this setting it becomes possible to prove that the mean acceleration of such a diffusion is in fact proportional to a Lorentz force.

II. BASIC THEOREMS FOR DIFFUSION PROCESSES TAKING VALUES IN n-DIMENSONAL EUCLIDEAN SPACES

Consider the diffusion process x_t described by Itô stochastic differential equation

$$dx_t = b_+(x_t,t)dt + G_+(x_t,t)dw_t, \tag{2.1}$$

with probability density $\rho(x,t,)$, which is assumed to be positive on $X\times[0,\infty)$, $X\subseteq\mathbb{R}^n$. w_t is the standard m-dimensional Brownian motion, independent of x_0, and b_+ and G_+ are, respectively, n-vector- and $n\times m$ matrix-valued functions of x and t, satisfying the usual conditions that insure the existence and uniqueness of the solutions of (2.1).[8] When $G(x,t)$ is not constant with respect to x we shall call x_t a "local diffusion."

We define the mean forward and backward derivatives of a function $f(x,t)$ of x and t by

$$\begin{aligned} D_+f(x_t,t) &= \lim_{\Delta\downarrow 0}\mathscr{E}^{x_t}\frac{1}{\Delta}\left[f(x_{t+\Delta},t+\Delta)-f(x_t,t)\right], \\ D_-f(x_t,t) &= \lim_{\Delta\downarrow 0}\mathscr{E}^{x_t}\frac{1}{\Delta}\left[f(x_t,t)-f(x_{t-\Delta},t-\Delta)\right], \end{aligned} \tag{2.2}$$

 0022-2488/82/061020-05$02.50

where $\mathscr{E}^{x_t}$ denotes the conditional expectation given x_t.

(h_1) Assume that x_t admits a backward representation through the reversed Itô equation

$$dx_t = b_-(x_t,t)dt + G_-(x_t,t)dw_t^*, \tag{2.3}$$

where w_t^* has the same properties as w_t except that the increments $w_t^* - w_s^*$ are independent of x_t for $s \leqslant t$. From (2.1) and h_1 one immediately has $D_+x_t = b_+$ and $D_-x_t = b_-$. The probability density of x_t is then solution of the two Fokker–Planck equations

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}(\rho b_+) + \tfrac{1}{2}\partial_i\partial_j(G_+G_+^T)_{ij}\rho, \tag{2.4}$$

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}(\rho b_-) - \tfrac{1}{2}\partial_i\partial_j(G_-G_-^T)_{ij}\rho. \tag{2.5}$$

The problem of constructing the backward representation (2.3) has been solved for nonlocal diffusions taking values in Euclidean spaces with diffusion matrix proportional to the identity[1,2] and also in n-dimensional Riemannian manifolds.[5–7] The method is here extended to general diffusions taking values in Euclidean spaces.

We make use of the following theorem[2]:

Theorem 1: Let D_+f and D_-f exist as a limit in L^1 for any $f \in C^\infty$. Then if f and g are real-valued functions having compact support in $X \times [0,\infty)$ one has

$$\frac{d}{dt}\mathscr{E}\{f(x_t,t)g(x_t,t)\} = \mathscr{E}\{(D_+f(x_t,t)g(x_t,t) + f(x_t,t)(D_-g(x_t,t))\}. \tag{2.6}$$

■

Suppose f and g can be expanded in Taylor series. One has from (2.1) and (2.3)

$$D_+f = \left(\frac{\partial}{\partial t} + b_{+i}\partial_i + \tfrac{1}{2}(G_+G_+^T)_{ij}\partial_i\partial_j\right)f, \tag{2.7}$$

$$D_-f = \left(\frac{\partial}{\partial t} + b_{-i}\partial_i - \tfrac{1}{2}(G_-G_-^T)_{ij}\partial_i\partial_j\right)f, \tag{2.8}$$

and denoting by $D_+^\dagger$, the formal adjoint of D_+, as defined by (2.7), from Theorem 1 the following equality holds:

$$\int_X \int_{-\infty}^{\infty} fD_+^\dagger(g\rho)\,dx\,dt = -\int_X (fD_-g)\rho\,dx\,dt \tag{2.9}$$

or, formally, for all $x \in X$ and $t \in [0,\infty)$

$$D_- = -\rho^{-1}D_+^\dagger\rho. \tag{2.10}$$

Standard calculations involving (2.4) yield

$$D_-g = \left(\frac{\partial}{\partial t} + b_{+i}\partial_i - \frac{1}{\rho}\partial_j(G_+G_+^T)_{ij}\rho\partial_i - \tfrac{1}{2}(G_+G_+^T)_{ij}\partial_i\partial_j\right)g \tag{2.11}$$

so that from (2.8)

$$G_-G_-^T = G_+G_+^T, \tag{2.12}$$

$$b_{-i} = b_{+i} - (1/\rho)\partial_j[(G_+G_+^T)_{ij}\rho]. \tag{2.13}$$

Denoting $G_+G_+^T$ simply by GG^T, the "current velocity" $v(x_t,t) \equiv [(D_+ + D_-)/2]x_t$ and the "osmotic velocity" $u(x_t,t) \equiv [(D_+ - D_-)/2]x_t$ have components

$$v_i = \frac{b_{+i} - b_{-i}}{2} = b_{+i} - \frac{1}{2}\frac{1}{\rho}\partial_j[(GG^T)_{ij}\rho], \tag{2.14}$$

$$u_i = \frac{b_{+i} - b_{-i}}{2} = \frac{1}{2}\frac{1}{\rho}\partial_j[(GG^T)_{ij}\rho]. \tag{2.15}$$

(h_2) Assume D_+b_- and D_-b_+ exist as limits in L^1 and b_{+i} and b_{-i} are sufficiently smooth for D_+ and D_- to act on them as the differential operators defined by (2.7) and (2.8). In this case the "mean acceleration" $a(x_t,t) \equiv \tfrac{1}{2}(D_+D_- + D_-D_+)x_t$ can be written in terms of v and u by

$$a_i = \frac{D_+b_{-i} + D_-b_{+i}}{2} = \frac{\partial}{\partial t}v_i + v_j\partial_jv_i - \{\tfrac{1}{2}(GG^T)_{kj}\partial_k\partial_ju_i + u_j\partial_ju_i\}. \tag{2.16}$$

We can now state the main fact concerning processes in Euclidean spaces in the form of the following representation theorem.

Theorem 2: Let x_t be a diffusion process described by (2.1) having probability density $\rho(x,t)$ which does not vanish on $X \times [0,\infty)$. Let assumption h_1 hold and assume (h_3) that the $m \times m$ matrix-valued function GG^T is strictly positive for all $x \in X$ and $t \in [0,\infty)$. Then denoting by S the set of all real-valued differentiable functions $S(x,t)$ of x and t, there exists a class of $n+1$-vector potentials $\{(A_S,\phi_S)\}_{S\in S}$ that is invariant under the gauge transformation

$$A_S \to A_{S'} = A_S + \operatorname{grad}(S' - S),$$
$$\phi_S \to \phi'_S = \phi_S - \frac{\partial(S' - S)}{\partial t}, \tag{2.17}$$

such that the operator

$$\mathscr{H} = \tfrac{1}{2}(i\partial_i + A_i^S)(GG^T)_{ij}(i\partial_j + A_j^S) + \phi_S + \alpha(t), \tag{2.18}$$

where $\alpha(t)$ is an arbitrary real function of the time, determines the evolution of the wave function $\psi \equiv \rho^{1/2}e^{iS}$ through the Schrödinger equation

$$i\frac{\partial \psi}{\partial t} = \mathscr{H}\psi. \tag{2.19}$$

Conversely, given any solution $\psi' = e^{R' + i(S' + S)}$ of (2.19) there exists a unique diffusion process x'_t satisfying the Itô equation

$$dx'_t = b'_+(x'_t,t)dt + G'(x,t)dw_t \tag{2.20}$$

that is uniquely defined up to an orthogonal transformation of G' which satisfies the conditions

$$G'G'^T = GG^T, \tag{2.21a}$$

$$\rho' = |\psi'|^2, \tag{2.21b}$$

$$v' = GG^T[\nabla(S' + S) - A_S]. \tag{2.21c}$$

Proof: By h_1 the current velocity v of the process x_t is well defined for all $(x,t) \in X \times [0,\infty)$ and h_3 allows one to define the vector field

$$A_S \equiv -(GG^T)^{-1}v + \nabla S. \tag{2.22}$$

Substituting (2.22) in (2.18) and writing (2.19) for $\psi = \rho^{1/2}e^{iS}$

one gets, separating imaginary and real parts,

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}(v\rho), \tag{2.23}$$

$$\phi_S = \tfrac{1}{2}\{\partial_i[(GG^T)_{ij}\partial_j R] + (GG^T)_{ij}\partial_i R\partial_j R - (GG^T)^{-1}_{ij} v_i v_j\} - \frac{\partial S}{\partial t} - \alpha(t), \tag{2.24}$$

where $R \equiv \frac{1}{2}\ln\rho$.

Equation (2.23) represents the correct continuity equation of x_t, as follows from (2.4) and (2.5). The sought gauge-invariant class of $n+1$-vector potentials is defined by (2.22) and (2.24). Conversely, given any solution $\psi' \equiv e^{R' + i(S'+S)}$ of the Schrödinger equation (2.19) and (2.18), one has that the diffusion process x'_t described by the Itô stochastic differential equation

$$dx'_t = b'_+(x'_t, t)dt + GTdw_t, \tag{2.25}$$

where T is any $m \times m$ orthogonal matrix and b'_+ is the vector field with components

$$(b'_+)_i = (GG^T)_{ij}(\partial_j(S'+S) - A^S_j) + \frac{1}{2}\frac{1}{\rho}\partial_j[(GG^T)_{ij}\rho], \tag{2.26}$$

satisfies (2.21).

The uniqueness of the diffusion process easily follows from the fact that if any other diffusion x''_t with density $\rho'' = \rho'$ and matrix diffusion $G'' = GT$ satisfies (2.21c) then one has $v'' = v'$ and thus by (2.15) $b''_+ = v'' + u''$ $= v' + u' = b'_+$. Then, since the drift b'_+ and covariance matrix GG^T uniquely determine the evolution of transition probability densities through the Fokker–Planck equation, and since (2.21b) also defines the initial probability density, by the Markov property all joint finite dimensional distributions are uniquely determined.

■

In Nelson's stochastic mechanics the analogous Schrödinger equation for $n = 3$, $GG^T = (\hbar/m)I$, where I is the 3×3 identity matrix, $\hbar$ Plank's constant, and m the mass of a particle subjected to the electromagnetic field $(\mathscr{A}, \phi)$, is derived by constraining the mean acceleration a to satisfy Newton's second law $ma = F$, where F is the Lorentz force produced by $(\mathscr{A}, \phi)$. In that case the real part of the Schrödinger equation is simply an integrated version of Newton's second law.

It can be shown[3,4] that for nonlocal diffusions satisfying h_1, h_3, and h_2 one has

$$a_p = (GG^T)_{pk}\left\{-\partial_k\phi_S - \frac{\partial}{\partial t}A^S_k + v_i[\partial_k A^S_i - \partial_i A^S_k]\right\} - (GG^T)_{pk}\left\{\left[\frac{\partial}{\partial t}(GG^T)^{-1}\right]v\right\}_k, \tag{2.27}$$

that, for $n = 3$ and $GG^T = I$ (I being the identity matrix) is the familiar Lorentz's force due to the electromagnetic potential (A_S, ϕ_S). In fact, by taking the gradient of both sides of (2.24) and denoting $\partial_i R$ by r_i and $(GG^T)^{-1}_{ij} v_j$ by z_i one gets

$$\partial_k\phi_S = \tfrac{1}{2}\partial_k\partial_i[(GG^T)_{ij}r_j] + \tfrac{1}{2}\partial_k[(GG^T)_{ij}r_i r_j] - \tfrac{1}{2}\partial_k[(GG^T)_{ij}z_i z_j] - \frac{\partial}{\partial t}(\partial_k S). \tag{2.28}$$

Recalling that for any $n \times n$ symmetric matrix Q and $y \in \mathbb{R}^n$ one has

$$\tfrac{1}{2}\partial_k(Q_{ij}y_i y_j) = (Q_{ij}\, y_i\partial_j)y_k + Q_{ij}y_k(\partial_k\, y_i - \partial_i\, y_k) \tag{2.29}$$

and that r is a gradient and $z = \nabla S - A^S$, (2.28) becomes

$$\partial_k\phi_S = \tfrac{1}{2}\partial_k[\partial_i(GG^T)_{ij}r_j] + [(GG^T)_{ij}r_i\partial_j]r_k - [(GG^T)_{ij}z_i\partial_j]z_k + (GG^T)_{ij}z_j[\partial_k A^S_i - \partial_i A^S_k] - \frac{\partial}{\partial t}(\partial_k S). \tag{2.30}$$

Thus the multiplication by $(GG^T)_{pk}$ and summation over k's, together with the substitutions $GG^T r \equiv u$ and $GG^T z \equiv v$, give the desired result.

For general diffusions taking values in Euclidean spaces the problem of the connection between the real part of the Schrödinger equation (2.19) and Newton's second law is not simple. It will be solved in Sec. IV, exploiting some results concerning stochastic mechanics on Riemannian manifolds.[5-7]

III. STRICTLY STATIONARY CASE

Consider a diffusion x_t with time-independent probability density $\rho(x,t) \equiv \rho(x)$, $\forall t \in [0, \infty)$, that is assumed not to vanish on $X \subseteq \mathbb{R}^n$, and assume it is the solution of the time-independent Itô stochastic differential equation

$$dx_t = b_+(x_t)dt + G(x_t)dw_t \tag{3.1}$$

with respect to the initial condition $x(0) = x_0$, where x_0 is a random variable having probability density equal to $\rho(x)$. Let h_1 and h_3 hold. Then applying Theorem 2 for $S_k \equiv kt$ and separating variables one finds that $\rho^{1/2}$ satisfies the time-independent Schrödinger equation

$$\{\tfrac{1}{2}(i\nabla + A_{S_k})(GG^T)_{ij}(i\nabla + A_{S_k}) + \phi_{S_k} + k - E\}\rho^{1/2} = 0, \tag{3.2}$$

where (A_{S_k}, ϕ_{S_k}) does not depend on t and E is a real constant. Recalling (2.24) one can see that $\phi_0 \equiv \phi_S + \partial_S/\partial_t$ is gauge independent. (3.2) can be rewritten as

$$\{-\tfrac{1}{2}\partial_i(GG^T)_{ij}\partial_j + \tfrac{1}{2}v_i(GG^T)^{-1}_{ij}v_j + \phi_0\}\rho^{1/2} = E\rho^{1/2}. \tag{3.3}$$

The constant E is fixed by averaging both sides of (3.3). Integrating after multiplication by $\rho^{1/2}$ and recalling (2.15) one gets

$$E = \int_X \{\tfrac{1}{2}u_i(GG^T)^{-1}_{ij}u_j + \tfrac{1}{2}v_i(GG^T)^{-1}_{ij}v_j + \phi_0\}\rho\, dx. \tag{3.4}$$

The integrand on the right side of (3.4) plays the role of "energy function". We want now briefly to discuss the case when GG^T is not strictly positive for all $(x,t) \in X \times [0, \infty)$. Actually assumption h_3 is crucial in giving the law (2.26) that constructs the drift from a given wave function. On the other side it would be meaningless if at the deterministic limit the energy function does go to infinity. In fact by the definition (2.15) of u the osmotic kinetic term remains finite also if singular values of the matrix-valued function $G(x)G(x)^T$ are allowed. Furthermore by the definition (2.24) of ϕ_S, which is

true if h_1 and h_3 hold, one has, in the strictly stationary case,

$$\tfrac{1}{2}v_i(GG^T)^{-1}_{ij}v_j + \phi_0 - E = \tfrac{1}{2}\partial_i[(GG^T)_{ij}\partial_j R] + (GG^T)_{ij}\partial_i R\partial_j R. \quad (3.5)$$

Thus if one considers the limit $(GG^T)\to(G_0G_0^T)$, where $G_0G_0^T$ is not positive definite for all $(x,t)\in X\times[0,\infty)$ and the limit process is again a diffusion satisfying h_1, then the sum of the terms on the left-hand side remains finite for all $(x,t)\in X_0\times[0,\infty)$, X_0 being the set where the limit probability density does not vanish. This implies, in these assumptions, that the energy function is well defined up to an additive constant and that (3.3) remains correct.

Finally, one can note that the form of the energy function suggests that the natural metric in considering the dynamics of general diffusions is that defined by the metric tensor $g_{ij}(x) = [(G(x)G^T(x))^{-1}]_{ij}$.

IV. EXTENSION OF THE REPRESENTATION THEOREM AND PHYSICAL INTERPRETATION FOR NONLOCAL DIFFUSIONS ON RIEMANNIAN MANIFOLDS

Let $\mathbb{M}$ be an n-dimensional Riemannian manifold with metric tensor $g_{ij}(x)$. Let q_t be a diffusion process on $\mathbb{M}$, satisfying the stochastic equation in the Itô sense.[9]

$$dq_t = \hat{b}_+(q_t,t)dt + d\hat{w}_t, \quad (4.1)$$

where $\hat{w}_t$ is the Brownian motion on $\mathbb{M}$, which satisfies the Itô stochastic differential equation

$$d\hat{w}_t^i = m^i(\hat{w}_t)dt + \sigma_k^i(\hat{w}_t)dw_{tk}, \quad (4.2)$$

where w_t is the standard n-dimensional Brownian motion and m^i and σ_k^i are related to the tensor g_{ij} by the equalities

$$m^i = -\tfrac{1}{2}g^{jk}\Gamma^i_{jk}, \quad (4.3)$$

$$\sigma_k^i\sigma_k^j = g^{ij}, \quad (4.4)$$

Γ^i_{jk} being the Christoffel symbols and $g^{ij}g_{ij} = \delta_{ij}$.

(h_1') Assume that the process q_t admits the backward representation

$$dq_t = \hat{b}_-(q_t,t)dt + d\hat{w}_t^*, \quad (4.5)$$

where $\hat{w}_t^*$ is the backward representation of $\hat{w}_t$. Defining the invariant probability density of q_t by the equality

$$\mathscr{E}[F(q(t))] = \int_{\mathbb{M}}\hat{\rho}(q,t)F(q)(|g|)^{1/2}dq, \quad (4.6)$$

where $F(q)$ is any function of the process, it can be shown by standard techniques that it satisfies the two Fokker–Planck equations

$$\frac{d\hat{\rho}}{\partial t} = -\operatorname{div}_{\mathbb{M}}(\hat{\rho}\hat{b}_+) + \tfrac{1}{2}\Delta_{\mathbb{M}}\hat{\rho}, \quad (4.7)$$

$$\frac{\partial\hat{\rho}}{\partial t} = -\operatorname{div}_{\mathbb{M}}(\hat{\rho}\hat{b}_-) - \tfrac{1}{2}\Delta_{\mathbb{M}}\hat{\rho}, \quad (4.8)$$

where $\operatorname{div}_{\mathbb{M}} y \equiv \nabla_i^{\mathbb{M}}y^i \equiv |g|^{-1/2}\partial_i(|g|)^{1/2}y^i$ and $\Delta_{\mathbb{M}} \equiv \nabla_{\mathbb{M}}^i\nabla_i^{\mathbb{M}} \equiv (|g|)^{1/2}\partial_i(|g|)^{1/2}g^{ij}\partial_j$, $\nabla_{\mathbb{M}}$ denoting covariant derivative.

The main difficulty in extending stochastic mechanics to processes taking values on manifolds is connected with the definition of the mean forward and backward derivatives. In fact the generalization of (2.2), when f is a vector-valued function, implies that one needs some notion of transport of $f(q_{t+\Delta},t+\Delta)$ from $q(t+\Delta)$ to $q(t)$ and of $f(q(t),t)$ from q_t to $q_{t+\Delta}$.

The problem has been solved in Refs. 4 and 5 by introducing the notion of "stochastic geodesic displacement". We refer directly to Refs. 6 or 7 for the definition. Let $T^G_{q_1,q_2}F(q_1)$ denote the stochastic geodesic displacement of $F(q_1)$ from q_1 to q_2. We define the mean forward and backward derivatives by

$$\hat{D}_+F(q_t,t) = \lim_{\Delta\downarrow 0}\mathscr{E}^{q_t}(1/\Delta)\times[T^G_{q_{t+\Delta},q_t}F(q_{t+\Delta},t+\Delta) - F(q_t,t)], \quad (4.9)$$

$$\hat{D}_-F(q_t,t) = \lim_{\Delta\downarrow 0}\mathscr{E}^{q_t}(1/\Delta)\times[T^G_{q_t,q_{t-\Delta}}F(q_t,t) - F(q_{t-\Delta},t-\Delta)]$$

so that, with suitable assumptions on F, one has[4,5]

$$(\hat{D}_+F)^i = \left[\left(\frac{\partial}{\partial t} + \hat{b}_+\cdot\nabla_{\mathbb{M}} + \tfrac{1}{2}\Delta_D\right)F\right]^i, \quad (4.10)$$

$$(\hat{D}_-F)^i = \left[\left(\frac{\partial}{\partial t} + \hat{b}_-\nabla_{\mathbb{M}} - \tfrac{1}{2}\Delta_D\right)F\right]^i, \quad (4.11)$$

where Δ_D is the Laplace–de Rham operator, given by

$$(\Delta_DF)^i = \Delta_{\mathbb{M}}F^i + R^i_jF^j, \quad (4.12)$$

R_{ij} being the curvature tensor associated with g. One has $\hat{D}_+q_t = \hat{b}_+$, $\hat{D}_-q_t = \hat{b}_-$ and, following the same line as in Sec. II, $\hat{b}_+$ and $\hat{b}_-$ are related by the equality

$$\hat{b}^i{}_- = \hat{b}^i{}_+ - (1/\hat{\rho})\nabla^i_{\mathbb{M}}\hat{\rho} \quad (4.13)$$

for all $q\in Q$ where $\hat{\rho}$ does not vanish and $t\in[0,\infty)$.

(h_2') Assume $\hat{D}_+\hat{b}_-$ and $\hat{D}_-\hat{b}_+$ exist as limits in L^1 and that $\hat{b}_+$ and $\hat{b}_-$ are sufficiently smooth for $\hat{D}_+$ and $\hat{D}_-$ to act on them as the operators defined by (4.10) and (4.11). Then the mean acceleration $\hat{a} = \tfrac{1}{2}(\hat{D}_+\hat{D}_- + \hat{D}_-\hat{D}_+)q_t$ can be written in the form

$$\hat{a}^i = \frac{\partial\hat{v}^i}{\partial t} + (\hat{v}^j\nabla_j^{\mathbb{M}})\hat{v}^i - \tfrac{1}{2}\{[(\Delta_D\hat{u})^i + (\hat{u}^j\nabla_j^{\mathbb{M}})\hat{u}^i\}, \quad (4.14)$$

where $\hat{v}^i = \hat{b}^i{}_+ + \hat{b}^i{}_-/2$ and $\hat{u}^i = \hat{b}^i{}_+ - \hat{b}^i{}_-/2$.

In the spirit of Theorem 2 one can state the following.

Theorem 3: Let q_t be a diffusion process on $\mathbb{M}$ described by (4.1) and let its invariant density $\hat{\rho}$ not vanish on $Q\subseteq\mathbb{M}$ for any $t\in[0,\infty)$. Let assumption (h_1') hold. Then there exists a class of $n+1$-vector potentials $\{(A_S,\phi_S)\}_{S\in\mathbb{S}}$ where $\mathbb{S}$ is the set of all real-valued differentiable functions of q and t which are invariant under the gauge transformations (2.17) such that the operator

$$\hat{\mathscr{H}} = \tfrac{1}{2}g^{ij}(i\nabla_i^{\mathbb{M}} + A_i^S)(i\nabla_j^{\mathbb{M}} + A_j^S) + \phi_S + \alpha(t) \quad (4.15)$$

determines the evolution of the wave function $\hat{\psi} = \hat{\rho}^{1/2}e^{iS}$ through the Schrödinger equation

$$i\frac{\partial\hat{\psi}}{\partial t} = \hat{\mathscr{H}}\hat{\psi}. \quad (4.16)$$

Conversely, given any solution $\hat{\psi}' = (\hat{\rho}')^{1/2}e^{i(S'+S)}$ of (4.16) there exists an unique diffusion process q_t' satisfying the Itô equation on $\mathbb{M}$,

$$dq_t' = b'(q_t',t)dt + d\hat{w}_t, \quad (4.17)$$

where $\hat{w}_t$ satisfies (4.2) and is defined up to an orthogonal transformation of the matrix with elements σ^i_k, such that

$$\hat{\rho}(q',t) = |\hat{\psi}(q',t)|^2,$$
$$\hat{v}'_i = \nabla^{\mathbf{M}}_i(S' + S) - A_S. \tag{4.18}$$

Proof: The proof is analogous to that of Theorem 2: Separating the real and imaginary parts of (4.16), making the substitution $\hat{\psi} = \hat{\rho}^{1/2}e^{iS}$, and with

$$A^{\ i}_S = \nabla^i_{\mathbf{M}} S - \hat{v}^i, \tag{4.19}$$

where $\hat{v}$ is well defined by h'_1 for all $q \in Q$ and $t \in [0, \infty)$ one gets

$$\frac{\partial \hat{\rho}}{\partial t} = -\operatorname{div}_{\mathbf{M}}(\hat{v}\cdot\hat{\rho}), \tag{4.20}$$

$$\phi_S = \tfrac{1}{2}\{\Delta_{\mathbf{M}}\hat{R} + \nabla^i_{\mathbf{M}}\nabla^{\mathbf{M}}_j\hat{R}\} - \tfrac{1}{2}\hat{v}^i\hat{v}_j - \frac{\partial S}{\partial t} - \alpha(t). \tag{4.21}$$

Again (4.20) is the correct continuity equation for q_t, as follows from (4.7), (4.8), and the definition of $\hat{v}$. The sought gauge invariant class is defined by (4.19) and (4.21). The converse part of the theorem runs exactly as in Theorem 2 and will be omitted.

■

The nonlocality on $\mathbb{M}$ of the diffusion q_t allows one to interpret (4.21) as the integrated version of Newton's second law. With the same line followed to get (2.27) one finds[3]

$$\hat{a}_k = \left[-\partial_k\phi - \frac{\partial A_k}{\partial t} + \hat{v}^i(\nabla^{\mathbf{M}}_k A_i - \nabla^{\mathbf{M}}_i A_k)\right]. \tag{4.22}$$

One can see,[5] using Hamilton equations, that the right-hand side of (4.22) is the intrinsic time differentiation of $g_{kj}\dot{q}^j$, provided $\dot{q}^j$ is identified with $\hat{v}^j$.

V. NEWTON'S SECOND LAW FOR LOCAL DIFFUSIONS ON EUCLIDEAN SPACES

We consider in this section a local diffusion x_t described by (1.1) with matrix diffusion G independent of t (the time-dependent case could also be treated), and let h_1 and h_3 hold. Then by Theorem 2 the wave function $\psi = \rho^{1/2}e^{iS}$ satisfies the Schrödinger equation

$$i\frac{\partial \psi}{\partial t} = \{(i\partial_i + A_i^{\ S})(GG^T)_{ij}(i\partial_j + A_j^{\ S}) + \phi_S\}\psi, \tag{5.1}$$

where GG^T is independent of t.

Consider now an n-dimensional Riemannian manifold with metric tensor g defined by the equality

$$g_{ij}(x) = [(G(x)G^T(x))^{-1}]_{ij}. \tag{5.2}$$

Then one can easily prove that the wavefunction $\hat{\psi} = \hat{\rho}^{1/2}e^{iS}$, where $\hat{\rho} = |g|^{-1/2}\rho$, satisfies the Schrödinger equation on $\mathbb{M}$,

$$i\frac{\partial \hat{\psi}}{\partial t} = \{(i\nabla^i_{\mathbf{M}} + A^{\ i}_S)(i\nabla^{\mathbf{M}}_j + A^{\ S}_j) + \phi_S + \mu_g\}\hat{\psi}, \tag{5.3}$$

where μ_g is a correction term depending only on the metric tensor g. In fact consider the diffusion process q_t represented by $\hat{\psi}$, which has invariant probability density $\hat{\rho}$ and current velocity $\hat{v}^i = \nabla^i_{\mathbf{M}}S - A^{\ i}_S$. Its continuity equation is

$$\frac{\partial \hat{\rho}}{\partial t} = -\operatorname{div}_{\mathbf{M}}(\hat{\rho}\hat{v}) = -|g|^{-1/2}\partial_i(|g|)^{1/2}\hat{\rho}\hat{v}^i). \tag{5.4}$$

By the definition of $\hat{\rho}$, and since by (4.18) and (2.21) $g^{ij}\hat{v}_j = v_i$, (5.4) becomes

$$\frac{\partial \rho}{\partial t} = -\partial_i(\rho\cdot\hat{v}^i) = -\operatorname{div}(\rho\cdot v) \tag{5.5}$$

and then the imaginary part of (5.1) implies the imaginary part of (5.3) and vice versa.

Finally, replacing R with $\hat{R} + \frac{1}{2}\ln|g|^{1/2}$ in the real part of (5.1) one finds

$$-\frac{\partial S}{\partial t} = -\tfrac{1}{2}\{|g|^{-1/2}\partial_i[(|g|)^{1/2}g^{ij}\partial_j\hat{R}] + g^{ij}\partial_i\hat{R}\partial_j\hat{R} - g_{ij}v_iv_j\} + \phi_S + \mu_g \tag{5.6}$$

where μ_g is defined by

$$\mu_g = \tfrac{1}{2}\{\partial_i[g^{ij}\partial_j(\tfrac{1}{2}\ln|g|^{1/2})] + g^{ij}\partial_i(\tfrac{1}{2}\ln|g|^{1/2})\partial_j(\tfrac{1}{2}\ln|g|^{1/2})\}. \tag{5.7}$$

We are now in a position to give the physical interpretation of the real part of Eq. (5.1): if in place of the Euclidean metric we endow $\mathbb{R}^n$ with the metric $g(x) = (G(x)G(x)^T)^{-1}$ or, which is the same, we consider $\mathbb{R}^n$ as the coordinate space of an n-dimensional Riemannian manifold with metric tensor $(GG^T)^{-1}$, then the local diffusion x_t on $\mathbb{R}^n$, which is represented by the solution $\psi = \rho^{1/2}e^{iS}$ of the Schrödinger equation (2.19)–(2.18) related to the vector potential (A_S, ϕ_S), plays the role of the coordinate representation of the nonlocal diffusion q_t on $\mathbb{M}$, described by Eqs. (4.1)–(4.4) and represented by $\hat{\psi} = \hat{\rho}^{1/2}e^{iS}$. By (4.22) the mean acceleration of q_t satisfies Newton's second law $\hat{a} \propto \hat{F}$, where $\hat{F}$ is the classical force generated by the vector potential $(A_S, \phi_S + \mu_g)$.

ACKNOWLEDGMENT

I wish to thank Professor F. Guerra for encouragement and helpful discussions.

[1] E. Nelson, Phys. Rev. **150**, 1079 (1966).
[2] E. Nelson, *Dynamical theories of Brownian motion* (Princeton U. P., Princeton, NJ, 1967).
[3] L. Morato, "Notes on Stochastic Mechanics", Rapporto interno LADSEB–CNR 78/06 (1978) (unpublished).
[4] L. Morato, "Kinematics and Dynamics of Itô processes," *Proceedings of the Udine International Conference on Dynamical Systems and Microphysics*. CISM N. 261 (Springer, New York, 1980).
[5] T. G. Dankel, Jr., Arch. Rational Mech. Anal. **37**, 192 (1970).
[6] D. Dohrn and F. Guerra, *Geodesic correction to stochastic parallel displacement of tensors*, Springer Lecture Notes in Physics **93** (Springer, New York, 1979).
[7] D. Dohrn and F. Guerra, Lett. Nuovo Cimento **22**, 121 (1978).
[8] I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes* (Saunders, Philadelphia, 1969).
[9] K. Itô, "The Brownian Motion and Tensor Fields on Riemannian Manifolds," in *Proceedings of the International Congress of Mathematics* (Stockholm, 1962), pp. 536–39.

Properties of three-dimensional Cartesian tensors. I. Some properties of irreducible tensors

Muchere C. Russ and C. A. Hollingsworth
Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 11 November 1980; accepted for publication 3 April 1981)

It is shown that the decomposition of three-dimensional Cartesian tensors into their parts which are irreducible under the rotation group is made conceptually simple by explicitly expressing these parts in an embedded form.

PACS numbers: 02.90. + p, 03.40. − t, 02.20. + b, 46.20. + e

I. INTRODUCTION

Since any irreducible tensor of weight w transforms under rotation as a tensor of rank w, it is often advantageous to be able to express any given tensor explicitly as the sum of its irreducible parts. A general Cartesian tensor of rank r ($r \geqslant 2$) can be reduced into irreducible tensors of weights r and less. Thus, a general third-rank tensor reduces into irreducible parts as follows: one of weight 0, three of weight 1, two of weight 2, and one of weight 3. A general fourth-rank tensor can be reduced into the following irreducible parts: three of weight 0, six of weight 1, six of weight 2, three of weight 3, and one of weight 4. The one, or more, irreducible tensors which make up the weight-w irreducible part of a reducible tensor of rank r is called the isotypic[1] subspace of weight w.

In the case of rank-2 tensors the reduction is unique and well known. The third-rank tensor and special fourth-rank tensors have been reduced in more than one way.[1–5] Russ[6] and Jerphagnon, Chemla, and Bonneville[7] have given explicit reductions for the general third- and fourth-rank tensors. Different reduction schemes in general lead to different reductions of those isotypic spaces that are composed of more than one tensor. The results of the various reductions are linearly related to one another, but the relationships are usually not obvious.

There is a way of visualizing the reduction which is conceptually very simple. We shall use this, somewhat heuristic, approach as a vehicle for presenting the terminology and symbolism needed for our further considerations.

The components of a general reducible Cartesian tensor of rank r will be represented by the symbol $T_{i_1,i_2,\ldots,i_r}$, where the subscripts take the values x, y, and z.

We shall have occasion to use three different symbols for the components of an irreducible tensor of weight w.

One symbol is $t^{r,w}_{i_1,i_2,\ldots,i_w}(n)$, where the superscript r indicates that these components are from a reducible tensor of rank r and are, therefore, expressible as linear homogeneous functions of the components of $T_{i_1,i_2,\ldots,i_r}$. The (n) is to distinguish between the various irreducible tensors making up the isotypic space of weight w and takes on numerical values 1, 2, ···. This notation gives the irreducible tensor in its *natural form*, in which the tensor is completely symmetric and traceless in its indices $i_1,i_2,\ldots,i_w$.

Only $2w + 1$ of the 3^w components $t^{r,w}_{i_1 i_2 \cdots i_w}(n)$ are independent. It is often convenient to use only one subscript to distinguish between the components of an independent set. We shall use the notation $t^{r,w}_q(n)$ $(q = 1, 2, \ldots, 2w + 1)$, where we have chosen and numbered the components as shown in Table I. The remaining of the 3^w components are expressible in terms of these by the conditions of the complete symmetry and tracelessness.

We sometimes have need to express an irreducible tensor of rank r and weight w as a rank-r tensor formed by embedding a weight-w irreducible tensor in a rank-r tensor space. For such a structure we use the symbol $\tau^{r,w}_{i_1 i_2 \cdots i_r}(n)$.

For the whole isotypic space of weight w (in embedded form) we use $\tilde{T}^{r,w}_{i_1 i_2 \cdots i_r}$. Thus, we have

$$\tilde{T}^{r,w}_{i_1 i_2 \cdots i_r} = \sum_n \tau^{r,w}_{j_1 j_2 \cdots j_r}(n), \tag{1}$$

where the subscripts $j_1, j_2, \ldots, j_r$ are the $i_1, i_2, \ldots, i_r$ in some order, which varies with n. The reducible tensor is given by

$$T_{i_1 i_2 \cdots i_r} = \sum_w \tilde{T}^{r,w}_{i_1 i_2 \cdots i_r}. \tag{2}$$

Embedding to give the $\tau^{r,w}(n)$ is obtained by an appropriate combination of the $t^{r,w}_{i_1 i_2 \cdots i_w}(n)$ with unit tensors δ and the alternating tensor ϵ. For example, in the case of the second rank tensor the results are the simple ones given below. The label (n) is not needed here because there is only one irreducible tensor of each weight. We use the summation convention, in which a repeated index implies summation over that index.

$$T_{ij}{}^{2,0} = t^{2,0}\delta_{ij},$$

$$t_{ij}{}^{2,1} = t_s{}^{2,1}\epsilon_{sij},$$

$$T_{ij}{}^{2,2} = t_{ij}{}^{2,2}.$$

Equation (2) becomes

TABLE I. Definition of the independent components $t^{r,w}_q = t^{r,w}_{i_1 \cdots i_w}$.

$t^{r,w}_q$	$t^{r,1}_i$	$t^{r,2}_{ij}$	$t^{r,3}_{ijk}$	$t^{r,4}_{ijkl}$
1	x	xx	xxy	$xxxy$
2	y	xy	xxz	$xxxz$
3	z	xz	xyy	$xxyy$
4		yy	xyz	$xxzz$
5		yz	xzz	$xyyy$
6			yyz	$xzzz$
7			yzz	$yyyz$
8				$yyzz$
9				$yzzz$

 0022-2488/82/061025-03$02.50

$$T_{ij} = t^{2,0}\delta_{ij} + t_s^{2,1}\epsilon_{sij} + t_{ij}^{22}. \tag{3}$$

For the reducible tensor of rank r, Eq. (2) represents 3^r equations in 3^r unknowns. The reduction process corresponds to the solving of these equations for the unknowns $t^{r,w}_{i_1 i_2 \cdots i_w}(n)$ in terms of the $T_{i_1 i_2 \cdots i_r}$. This has been done, and the results have been given elsewhere[6] in complete detail for tensors through ranks 4.

Of course, if the reverse process is desired, that is, if the irreducible components are known and one wishes to obtain the reducible rank-r tensor corresponding to these, then substitution into Eq. (2) gives the required result.

II. SOME PROPERTIES OF ISOTYPIC SUBSPACES

In this section we consider briefly some general aspect of isotypic subspaces with special attention to third- and fourth-rank tensors. First, however, we make note of the following points:

For all ranks r, the isotypic subspace of weight 0 is composed of the rotation tensors of rank r, that is, the tensors of rank r that are invariant under all proper rotations.[8] The rotation tensors through rank four are well known and often used.[9] Smith[8] has given a method for obtaining the independent rotation tensors of any rank.

For cases in which improper rotations must be considered, and, therefore, the distinction between polar (also called true) and axial (also called pseudo-) tensors becomes important, the following rule applies[1,7]: If the rank of the tensor and the weight of its irreducible part differ by an even (odd) number, then the tensor and its irreducible part have the same (opposite) character.

Gel'fand *et al.*[1] have obtained an equation whose solutions are the tensors $\tau^{r,w}_{i_1 i_2 \cdots i_r}(n)$. We can use that equation to check that the tensors we obtain by our embedding process are actually irreducible tensors of the desired rank and weight.

A. Third-rank tensors

For rank 3 the general equation given by Gel'fand *et al.* becomes

$$\delta_{ij}\tau^{3,2}_{ssk} + \delta_{ij}\tau^{3,w}_{iss} + \delta_{ki}\tau^{3,w}_{sjs} - \tau^{3,w}_{jik} - \tau^{3,w}_{ikj} - \tau^{3,w}_{kji} + [\tfrac{1}{2}w(w+1) - 3]\tau^{3,w}_{ijk} = 0 \tag{4}$$

From Eq. (4) follow various properties of the irreducible third-rank tensors of different weights in embedded form. Among these properties a general one is the following: Except for $w = 1$, these tensors are traceless. A particular result is that for $w = 0$ the tensor is skew-symmetric, but this follows directly from other considerations also [See Eq. (5), and note that in general the minimum required symmetry of any particular $\tau^{r,w}_{i_1 i_2 \cdots i_r}$ is given explicitly by the embedded form].

Weight 0: All rotation tensors of rank[9] 3 are multiples of the alternating tensor ϵ_{ijk}. Therefore, the irreducible tensor of weight zero is given by

$$\tilde{T}^{3,0}_{ijk} = t^{3,0}\epsilon_{ijk}, \tag{5}$$

where $t^{3,0}$ is a scalar.

Weight 1: The isotypic space of weight 1 can be expressed as follows:

$$\tilde{T}^{3,1}_{ijk} = t_i^{3,1}(1)\delta_{jk} + t_j^{3,1}(2)\delta_{ik} + t_k^{3,1}(3)\delta_{ij}, \tag{6}$$

where $t_i^{3,1}(n)$ $(n = 1, 2, 3)$ are vectors. All vectors are irreducible rank-1, weight-1 tensors. Since the tensors on the right in Eq. (6) are formed by the embedding of vectors, and since they satisfy Eq. (4), they must be irreducible tensors of rank 3 and weight 1. Therefore, it follows that Eq. (6) represents a resolution of the weight-1 isotypic space.

Weight 2: There are three obvious ways that irreducible rank-2 and weight-2 tensors can be embedded in third-rank space, but they are independent only in pairs. The three possibilities are $t_{is}^{3,2}\epsilon_{sjk}$, $t_{js}^{3,2}\epsilon_{sik}$, and $t_{ks}^{3,2}\epsilon_{sij}$, where to be irreducible the tensor $t_{ij}^{3,2}$ must be symmetric and traceless. Among the three possible resolutions that can be obtained by taking these tensors in pairs, one is

$$\tilde{T}^{3,2}_{ijk} = t_{is}^{3,2}(1)\epsilon_{sjk} + t_{js}^{3,2}(2)\epsilon_{sik}. \tag{7}$$

All of these resolutions divide the isotypic subspace into two parts, each of which belongs to a tensor that is skew-symmetric in a pair of indices. Many other resolutions are possible. For example, this subspace can be resolved into the sum of a tensor that is symmetric in a pair of indices and one that is skew-symmetric in the same pair of indices. Gel'fand *et al.*[1] resolve the subspace by assuming cyclic symmetry. All the various resolutions are, of course, linear combinations of one another.

Weight 3: This is simply the unique rank-3 and weight-3 irreducible tensor $t_{ijk}^{3,3}$ in the natural form and is symmetric and traceless with respect to each pair of indices.

B. Fourth-rank tensors

The equation given by Gel'fand *et al.*[1] becomes in this case

$$\delta_{ij}\tau^{4,w}_{sskl} + \delta_{ik}\tau^{4,w}_{sjsl} + \delta_{il}\tau^{4,w}_{sjks} + \delta_{jk}\tau^{4,w}_{issl} + \delta_{jl}\tau^{4,w}_{isks} + \delta_{kl}\tau^{4,w}_{ijss} - \tau^{4,w}_{jikl} - \tau^{4,w}_{kjil} - \tau^{4,w}_{ljki} - \tau^{4,w}_{ikjl} - \tau^{4,w}_{ilkj} - \tau^{4,w}_{ijlk} + [\tfrac{1}{2}w(w+1) - 4]\tau^{4,w}_{ijkl} = 0. \tag{8}$$

Some general results that we have derived from this equation are the following:

(i) Except when $w = 0$, all double traces are zero.

(ii) All single traces are zero when $w = 3$ or 4.

(iii) Single traces are zero when $w = 0$ and the other two indices are different.

(iv) Single traces are skew-symmetric in the other two indices for $w = 1$.

(v) Single traces are symmetric in the other two indices when $w = 2$.

Since the general procedure is the same as that for the resolution of the third-rank tensor, we give the following results with little comment.

Weight 0: There are three independent rotation tensors of rank 4, and they give the following resolution of the rank-4, weight-0, subspace:

$$\tilde{T}^{4,0}_{ijkl} = t^{4,0}(1)\delta_{ij}\delta_{kl} + t^{4,0}(2)\delta_{ik}\delta_{jl} + t^{4,0}(3)\delta_{il}\delta_{jk}. \tag{9}$$

Weight 1: Vectors can be embedded in a fourth-rank space six ways, giving the following resolution:

$$\tilde{T}^{4,1}_{ijkl} = t^{4,1}_{s}(1)\epsilon_{sij}\delta_{kl} + t^{4,1}_{s}(2)\epsilon_{sik}\delta_{jl} + t^{4,1}_{s}(3)\epsilon_{sjk}\delta_{il} + t^{4,1}_{s}(4)\epsilon_{sil}\delta_{jk} + t^{4,1}_{s}(5)\epsilon_{sjl}\delta_{ik} + t^{4,1}_{s}(6)\epsilon_{skl}\delta_{ij}. \tag{10}$$

Weight 2: Irreducible rank-2, weight-2 tensors can be embedded in rank-4 space in six ways, giving the following resolution:

$$\tilde{T}^{4,2}_{ijkl} = t^{4,2}_{ij}(1)\delta_{kl} + t^{4,2}_{ik}(2)\delta_{jl} + t^{4,2}_{jk}(3)\delta_{il} + t^{4,2}_{il}(4)\delta_{jk} + t^{4,2}_{jl}(5)\delta_{ik} + t^{4,2}_{kl}(6)\delta_{ij}. \tag{11}$$

Weight 3: Irreducible tensors of rank and weight three can be embedded in rank-4 space to give fourth-rank irreducible tensors in several (not all independent) ways. This is done by using the alternating tensor ϵ_{ijk}. It happens that to be independent these tensors must have a common index in ϵ_{ijk}. For example, in the following case j is the common index:

$$\tilde{T}^{4,3}_{ijkl} = t^{4,3}_{kls}(1)\epsilon_{sij} + t^{4,3}_{lis}(2)\epsilon_{sjk} + t^{4,3}_{iks}(3)\epsilon_{sjl}. \tag{12}$$

Weight 4: This is the unique rank-4 and weight-4 tensor $t^{4,4}_{ijkl}$ in its natural form.

[1] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Macmillan, New York, 1963).
[2] J. A. R. Coope, R. F. Snider, and F. R. McCourt, J. Chem. Phys. **43**, 2269 (1965).
[3] J. A. R. Coope and R. F. Snider, J. Math. Phys. **11**, 1003 (1970).
[4] H. J. Juretschke, *Crystal Physics: Macroscopic Physics of Anisotropic Solids* (Benjamin, Reading, Mass., 1974).
[5] P. R. Morris, J. Appl. Phys. **40**, 447 (1969).
[6] M. C. Russ, PhD. thesis, University of Pittsburgh, 1975.
[7] J. Jerphagnon, D. Chemla, and R. Bonneville, Adv. in Phys. **27**, 609 (1978). In addition to giving explicit expressions for the reduction of tensors of ranks 3 and 4, this paper contains a good discussion of the application of irreducible tensors and the relationship between their Cartesian and spherical forms.
[8] G. F. Smith, Tensor **19**, 79 (1968).
[9] For example, see G. Temple, *Cartesian Tensors* (Methuen, London, 1960); also, L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Addison-Wesley, New York, 1960), p. 381.

Properties of three-dimensional Cartesian tensors. II. Arbitrarily oriented tensors expressed in generalized spherical functions. Ranks through 4

Muchere C. Russ and C. A. Hollingsworth
Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 11 November 1980; accepted for publication 3 April 1981)

Explicit expressions for the rotation of Cartesian tensors of ranks through 4 are obtained. These are given in terms of the irreducible components and the generalized spherical functions. They can be obtained in terms of the reducible components by the use of results which were obtained previously. As a simple illustration of an application of the results, an expression is obtained for the energy of an octupole as a function of the orientation of the octupole in an inhomogeneous field.

PACS numbers: 02.90. + p, 03.40. − t,02.20. + b,46.20. + e

I. INTRODUCTION

In application of tensors to physical problems it is often necessary to know the components of a given tensor in a coordinate system that differs by an arbitrary rotation from the one for which the components were given. For example, the components may be known for a coordinate system fixed in some "body" and the expressions for the components may be needed in a "laboratory" system. These expressions are given by the equations for the transformation of tensors. Thus, if we designate the components of a tensor in the body system by $B_{i_1 i_2 \cdots}$ and those in the laboratory system by $L_{j_1 j_2 \cdots}$, then we have

$$L_{j_1 j_2 \cdots} = R_{j_1 i_1} R_{j_2 i_2} \cdots B_{i_1 i_2 \cdots}. \tag{1}$$

R is the orthogonal matrix representing the rotation which takes the body system into the laboratory system, and the repeated index summation convention has been used. The result expressed by Eq. (1) can take a number of different forms, depending upon how the rotation R is expressed. All of these results are rather unwieldy for tensors of ranks greater than 2. One form which can be convenient for many purposes is that in which the components $L_{j_1 j_2 \cdots}$ are expressed in terms of the generalized spherical functions[1,2] $D^{w}_{m'm}(\alpha\beta\gamma)$ of the Euler angles $(\alpha\beta\gamma)$ which describe the rotation R. For a tensor of rank r we would have

$$L_{j_1 j_2 \cdots} = \sum_{w=0}^{r} \sum_{m'=-w}^{w} \sum_{m=-w}^{w} A^{r,w}_{j_1 j_2 \cdots; m'm} D^{w}_{m'm}(\alpha\beta\gamma), \tag{2}$$

where the coefficients $A^{r,w}_{j_1 j_2 \cdots; m'm}$ are functions of the components $B_{i_1 i_2 \cdots}$, and they are always the same functions for any given weight. It is the purpose of this paper to give these functions for tensors of ranks through 4.

II. METHOD OF CALCULATION

There are several different methods that could, in principle, be used to obtain the coefficients in Eq. (2). However, for ranks greater than 2 most of these methods would involve much tedious algebra. The method to be described in this section was relatively easy for ranks through 4. Much of the advantage provided by this method resulted from the use of the tensors in explicitly reduced form, and these forms are given in Paper I of this series [3] for tensors through rank 4.

In its general aspects the calculation goes as follows: The irreducible Cartesian tensor of rank w and weight w is transformed to the so-called spherical basis.[1] The tensor is rotated in the spherical basis, where the effect of rotation is given explicitly in terms of the generalized functions $D^{w}_{m'm}(\alpha\beta\gamma)$. The rotated tensor is then returned to the Cartesian system. This gives the rotated irreducible Cartesian tensor in terms of the functions $D^{w}_{m'm}(\alpha\beta\gamma)$. Since any Cartesian tensor of rank r $(r \leqslant 4)$ has already been expressed explicitly[3] in terms of its irreducible tensors of weight w $(w \leqslant r)$, the desired result has been obtained.

A few further comments about the procedure outlined above seems warranted. The transformation from the Cartesian basis to the spherical basis can be effected by the use of the following unitary transformation:

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -i & 0 \\ 0 & 0 & \sqrt{2} \\ 1 & -i & 0 \end{bmatrix}. \tag{3}$$

Upon transformation to the spherical basis each of the components (3^{w} in number) of a rank-w, weight-w tensor becomes some scalar times one of the spherical harmonics Y^{w}_{m} $(m = -w,\ldots,0,\ldots,w)$. In the spherical basis, therefore, each component transforms under proper rotation according to the law[1]

TABLE I. Values of β^{w}_{q}. The value zero is given for $w = 1$, $q = 3$, because the only nonzero value of $\sigma^{1}_{3|m|}$ is for $m = 0$, and whenever $m = 0$ we take $\beta^{w}_{q} = 0$.

q \ w	1	2	3	4
1	−1	+1	+1	−1
2	+1	−1	+1	−1
3	0	−1	−1	+1
4		+1	−1	+1
5		+1	−1	−1
6			+1	−1
7			+1	+1
8				+1
9				+1

 0022-2488/82/061028-03$02.50

TABLE II. Values of $\gamma^w_{m'}$.

m' \ w	1	2	3	4
0	$-1/\sqrt{2}$	$-(1/2)\sqrt{3/2}$	$\sqrt{5/4}$	$\sqrt{70}/16$
1	1/2	$-1/2$	$-\sqrt{15}/8$	$\sqrt{14}/8$
2		1/4	$\sqrt{6}/8$	$-\sqrt{7}/8$
3			$-1/8$	$\sqrt{2}/8$
4				$-1/16$

$$R Y^w_m = \sum_{m'} D^w_{m'm}(\alpha\beta\gamma) Y^w_{m'}. \tag{4}$$

The return to a Cartesian basis is effected by the use of the unitary transformation U^{-1}, which is the conjugate of the transpose of U.

We can designate the $2w+1$ independent components of the irreducible rank-r, weight-w, Cartesian tensors (original and rotated, respectively) by $b^{r,w}_p(n)$ and $l^{r,w}_q(n)$, where the indices p and q take the integral values from 1 to $2w+1$. The index (n) is to distinguish between the different irreducible tensors of rank r and weight w, since for $w<r$ and $r>2$ (except for $r=3$ and $w=0$) there are more than one of these irreducible tensors. With this notation the results of the present calculation can be expressed as follows:

$$l^{r,w}_q(n) = \sum_{m'}\sum_{m} a^w_{qm'm} D^w_{m'm}(\alpha\beta\gamma), \tag{5}$$

where the coefficients $a^w_{qm'm}$ are the same functions of the components $b^{r,w}_p(n)$ for all irreducible tensors of weight w. The effects of r and n come in implicitly through the $b^{r,w}_p(n)$, which are functions of the rank-r components $B_{i_1 i_2 \cdots i_r}$, and different values of (n) correspond to different functions. The functions $A^{r,w}_{j_1 j_2 \cdots ;m'm}$ in Eq. (2) can be obtained from the functions $a^w_{qm'm}$ by embedding the irreducible components in the rank-r space.[3] The functions $a^w_{qm'm}$ are given in the next section, and the functions $b^{r,w}_p(n)$ are already available.[3]

III. RESULTS

We define the independent components $b^{4,w}_p(n)$ and $l^{4,w}_q(n)$ according to the scheme that was used previously.[3] It turns out that Eqs. (5) can be expressed in the following way:

$$l^{r,w}_q = \sum_{m'|m|} \gamma^w_{m'} \sigma^w_{q|m|} N^{r,w}_{m'} \left[D^w_{m'm}(\alpha\beta\gamma) + \beta^w_q D^w_{m'(-m)}(\alpha\beta\gamma) \right], \tag{6}$$

where $N^{r,2}_{m'}$ are functions of the components $b^{r,w}_p(n)$ and satisfy

$$N^{r,w}_{-m'} = N^{r,w*}_{m'} \tag{7}$$

where the asterisk indicates complex conjugation. $\gamma^w_{m'}$, $\sigma_{q|m|}$, and β^w_q are numerical coefficients, and they satisfy

$$\gamma^0_0 \sigma^0_{00} = 1, \tag{8}$$

$$\gamma^w_{m'} = (-1)^{m'} \gamma^w_{-m'}. \tag{9}$$

Values of these numerical coefficients are given in Tables I, II, and III, and the functions $N^{r,w}_{m'}$ are in Eqs. (10)–(14); these functions hold for all $r \geqslant w$. The composite coefficients $\gamma^w_{m'} \sigma^w_{q|m|} N^{r,w}_{m'}$ can be factored into coefficients $\gamma^w_{m'}$, $\sigma^w_{q|m|}$, and $N^{r,w}_{m'}$ in many different ways. We have chosen a factoring that seems to be convenient:

$$N^{r,0}_0 = b^{r,0}; \tag{10}$$

$$N^{r,1}_0 = b^{r,1}_3,$$
$$N^{r,1}_1 = b^{r,1}_1 + i b^{r,1}_2; \tag{11}$$

TABLE III. Values of $\sigma^w_{q|m|}$.

$\lvert m\rvert$ \ q	1	2	3	4	5	6	7	8	9
				$w=1$					
0	0	0	$-\sqrt{2}$						
1	1	$-i$	0						
				$w=2$					
0	$-\sqrt{2/3}$	0	0	$-\sqrt{2/3}$	0				
1	0	0	-1	0	i				
2	1	$-i$	0	-1	0				
				$w=3$					
0	0	$-2/\sqrt{5}$	0	0	0	$-2/\sqrt{5}$	0		
1	$-i/\sqrt{15}$	0	$1/\sqrt{15}$	0	$-4/\sqrt{15}$	0	$4i/\sqrt{15}$		
2	0	$\sqrt{2/3}$	0	$-i\sqrt{2/3}$	0	$-\sqrt{2/3}$	0		
3	i	0	1	0	0	0	0		
				$w=4$					
0	0	0	$-2/\sqrt{70}$	$8/\sqrt{70}$	0	0	0	$8/\sqrt{70}$	0
1	0	$-3/\sqrt{14}$	0	0	0	$4/\sqrt{14}$	$3i/\sqrt{14}$	0	$-4i/\sqrt{14}$
2	$-i/\sqrt{7}$	0	0	$-2/\sqrt{7}$	$-i/\sqrt{7}$	0	0	$2/\sqrt{7}$	0
3	0	$1/\sqrt{2}$	0	0	0	0	$i\sqrt{/2}$	0	0
4	i	0	1	0	$-i$	0	0	0	0

$$
\begin{aligned}
N_0^{r,2} &= b_1^{r,2} + b_4^{r,2},\\
N_1^{r,2} &= b_3^{r,2} + ib_5^{r,2},\\
N_2^{r,2} &= b_1^{r,2} + 2ib_2^{r,2} - b_4^{r,2};
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
N_0^{r,3} &= -(b_2^{r,3} + b_6^{r,3}),\\
N_1^{r,3} &= b_5^{r,3} + ib_7^{r,3},\\
N_2^{r,3} &= b_2^{r,3} + 2ib_4^{r,3} - b_6^{r,3},\\
N_3^{r,3} &= 4ib_1^{r,3} - 4b_3^{r,3} - b_5^{r,3} + ib_7^{r,3};
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
N_0^{r,4} &= b_4^{r,4} + b_8^{r,4},\\
N_1^{r,4} &= b_6^{r,4} + ib_9^{r,4},\\
N_2^{r,4} &= -2ib_1^{r,4} + b_4^{r,4} - 2ib_5^{r,4} - b_8^{r,4},\\
N_3^{r,4} &= 4b_2^{r,4} + 3b_6^{r,4} - 4ib_7^{r,4} - 3ib_9^{r,4},\\
N_4^{r,4} &= 4ib_1^{r,4} - 8b_3^{r,4} - b_4^{r,4} - 4ib_5^{r,4} - b_8^{r,4}.
\end{aligned}
\tag{14}
$$

The coefficients given in Tables II and III correspond to the Euler angles as defined by Rose,[1] and we have taken α, β, and γ to be the angles required to bring the body coordinates into coincidence with the laboratory coordinates. Therefore, β and α are the ordinary spherical angles (polar and azimuthal, respectively) giving the laboratory z axis in the body coordinates. It will sometimes be more convenient to use the angles representing the orientation of the body coordinates in the laboratory coordinates, i.e., to use the rotation required to bring the laboratory coordinates into coincidence with the body coordinates. We designate these latter Euler angles, ϕ, θ, and χ. It follows that ϕ, θ, χ are, respectively, $-\gamma$, $-\beta$, and $-\alpha$. θ and ϕ are the polar and azimuthal angles (respectively) of the body z axis in the laboratory coordinates. The results given in Tables II and III then apply if $D^{w*}_{mm'}(\phi\theta\chi)$ replaces $D^{w}_{m'm}(\alpha\beta\gamma)$ in Eqs. (6).

IV. APPLICATION TO THE ENERGY OF AN OCTUPOLE

We illustrate a simple application of the preceding results by considering an equation for the energy of an octupole as function of the orientation of the octupole in a field whose potential has nonzero third-order derivatives.

The octupole is a completely symmetric third-rank tensor O_{ijk}. It interacts with the third-order gradient Φ_{ijk} of the potential Φ. These gradient terms are also the components of a completely symmetric tensor. The energy of interaction of the octupole and the field can be expressed as the following scalar product[4]:

$$
U = O_{ijk}\Phi_{ijk} = \tilde{O}^{3,1}_{ijk}\tilde{\Phi}^{3,1}_{ijk} + \tilde{O}^{3,3}_{ijk}\tilde{\Phi}^{3,3}_{ijk},
\tag{15}
$$

where the symbols $\tilde{O}^{3,1}_{ijk}$,etc., designate the isotypic spaces. If we use the notation $o^{3,w}_q$ and $\phi^{3,w}_q$ to represent the irreducible components, we have

$$
\tilde{O}^{3,1}_{ijk}\tilde{\Phi}^{3,1}_{ijk} = 9(o_1^{3,1}\phi_1^{3,1} + o_2^{3,1}\phi_2^{3,1} + o_3^{3,1}\phi_3^{3,1}),
\tag{16}
$$

$$
\begin{aligned}
\tilde{O}^{3,3}_{ijk}\tilde{O}^{3,3}_{ijk} = {} & 4o_1^{3,3}\phi_1^{3,3} + 4o_2^{3,3}\phi_2^{3,3} + 4o_3^{3,3}\phi_3^{3,3} + 6o_4^{3,3}\phi_4^{3,3}\\
& + 4o_5^{3,3}\phi_5^{3,3} + 4o_6^{3,3}\phi_6^{3,3} + 4o_7^{3,3}\phi_7^{3,3}\\
& + o_3^{3,3}\phi_5^{3,3} + o_5^{3,3}\phi_3^{3,3} + o_2^{3,3}\phi_6^{3,3} + o_6^{3,3}\phi_2^{3,3}\\
& + o_1^{3,3}\phi_7^{3,3} + o_7^{3,3}\phi_1^{3,3}.
\end{aligned}
\tag{17}
$$

Suppose that these tensors are expressed in a coordinate system which has been defined for the field tensor Φ_{ijk} and that the octupole components, in a coordinate system defined for the octupole, are B_{ijk} (reducible) and $b^{3,w}_q$ (irreducible). If the Euler angles α, β, γ are the angles through which the octupole coordinates must be rotated to obtain coincidence with the field coordinates, then the components $o^{3,w}_q$ that go into Eqs. (16) and (17) are given by Eqs. (6), (11), and (13). Substitution of Eqs. (16) and (17) into Eq. (15) then gives U as a function of the orientation of the field coordinates in the octupole system (that is, α and β are the spherical angles of the field z axis in the octupole coordinates). Substitution of $D^{w*}_{mm'}(\phi\theta\chi)$ in place of $D^{w}_{m'm}(\alpha\beta\gamma)$ gives an equation expressing U in terms of the orientation of the octupole coordinates in the field coordinates (that is, ϕ and θ are the spherical angles of the octupole z axis in the field coordinates).

[1] M. E. Rose, *Elementay Theory of Angular Momentum* (Wiley, New York, 1957).

[2] D. M. Brink and G. R. Satchler, *Angular Momentum*, 2nd ed. (Oxford, London, 1968).

[3] Muchere C. Russ and C. A. Hollingsworth, J. Math. Phys. **22**, 1025 (1981), Part I.

[4] We should note that mutipole tensors are often defined so that they are traceless and are, therefore, irreducible. For an octupole tensor so defined only the weight-3 part would occur in Eq. (15). If Φ obeys the Laplace equation, then $\tilde{\Phi}^{3,1}_{ijk}$ is zero, and $\tilde{O}^{3,1}_{ijk}$ has no effect anyway.

Properties of three-dimensional Cartesian tensors. III. Concerning tensor averages

C. A. Hollingsworth and Muchere C. Russ
Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 11 November 1980; accepted for publication 3 April 1981)

Results from the two earlier papers in this series are used to investigate some average tensorial properties of systems in terms of the tensorial properties of their subsystems. Special attention is given to the Boltzmann distribution in which the orientation energy of a subsystem is assumed to be expressible as a power series in the components of a single vector which is fixed in the laboratory system. The orientation probability density is obtained in general form for subsystems with no symmetry (point group C_1), and they are given in more explicit form for a few cases of higher symmetry. General expressions for tensor averages are given, and they are applied to a special case as an example.

PACS numbers: 02.90. + p, 05.90. + m, 03.40. − t

I. INTRODUCTION

In this paper we shall use some results given in earlier publications[1,2] to calculate some average tensorial properties of systems in terms of the tensorial properties of their subsystems. For convenience we restrict our consideration to systems with only one type of subsystem. Let $L_{i\ldots}$ and $B_{i\ldots}$ be the tensor components of a subsystem in the laboratory and subsystem (body) coordinates, respectively. Let $l_q^{r,w}$ and $b_q^{r,w}$ be the corresponding irreducible components.[1] Suppose that the angular orientations of the subsystems relative to the laboratory basis are given in terms of the probability density $W(\alpha\beta\gamma)$, or $W(\phi\theta\chi)$, where α,β,γ and ϕ,θ,χ are the Euler angles defined previously.[2] Further, suppose that W is expanded in terms of the generalized spherical functions as follows:

$$W(\phi\theta\chi) = \sum_{wm'm} g^w_{m'm} D^w_{m'm}(\alpha\beta\gamma) = \sum_{wm'm} g^w_{m'm} D^{w*}_{mm'}(\phi\theta\chi). \quad (1)$$

Also, we have

$$l_q^{r,w} = \sum_{m'm} a^w_{qm'm} D^w_{m'm}(\alpha\beta\gamma) = \sum_{m'm} a^w_{qm'm} D^{w*}_{mm'}(\phi\theta\chi). \quad (2)$$

Averaging over all the subsystems gives

$$\langle l_q^{r,w}\rangle = \int l_q^{r,w} W\, d\Omega = \frac{8\pi^2}{2w+1}\sum_{m'm} a^w_{qm'm} g^{w*}_{m'm} = \frac{8\pi^2}{2w+1}\sum_{m'm} a^{w*}_{qm'm} g^w_{m'm}. \quad (3)$$

The averages $\langle L_{i\ldots}\rangle$ can be obtained from the expressions which give the tensor components in terms of their irreducible parts.[1]

Any indicial symmetry possessed by the tensors, and any geometrical symmetry possessed by the subsystems, will provide simplification via the coefficients $a^w_{qm'm}$, while any geometrical symmetry possessed by W will affect the coefficients $g^w_{m'm}$. We note the following important special cases of the effects of geometrical symmetry: If the subsystems have a C_n axis, then all the $a^w_{qm'm}$ with $w < n$ are zero unless $m' = 0$. If W has a C_n axis, then all $g^w_{m'm}$ with $w < n$ are zero unless $m = 0$. If the subsystems have a center of inversion, then only the $a^w_{qm'm}$ with even w are nonzero. If W has a center of inversion, then only the $g^w_{m'm}$ with even w are nonzero. Note that

$$g^0_{00} = 1/8\pi^2 \quad (4)$$

and that this first term in Eq. (1) corresponds to random orientation.

Often the quantities of interest are composed of scalar products of the property tensors $L_{i\ldots}$ of the subsystems and some physical tensors $F_{i\ldots}$, $G_{j\ldots}$, etc., fixed in the laboratory coordinates. The averages of such scalar products are the scalar products of the averages $\langle L_{i\ldots}\rangle$ with the tensors $F_{i\ldots}$, $G_{i\ldots}$, etc. Some examples of this type are considered in the following sections. In order to keep the algebra from becoming too cumbersome, we shall consider only Boltzmann distributions in which the orientation energy is assumed to be a function of a single vector which is fixed in the laboratory coordinates. Extension to more complicated problems would, in many cases, be straightforward in principle. Our considerations are limited to ranks 4 and below, since explicit expressions for the reducible tensors in terms of the irreducible components are not available for higher ranks.

II. THE PROBABILITY DENSITY FUNCTION

We assume that the orientation energy u can be expressed as a power series in the components of some vector F, which is fixed in the laboratory system. For convenience we take the laboratory z axis to be in the direction of F, so that the series can be expressed in terms of $F_z (= F)$ as follows:

$$u = -(L_z F + L_{zz}F^2 + L_{zzz}F^3 + L_{zzzz}F^4 + \cdots), \quad (5)$$

where the coefficients $L_{z\ldots}$ are the components of some property tensor of the subsystems.

We have, taking $\phi = 0$ because of the cylindrical symmetry,

 0022-2488/82/061031-06$02.50

$$W(\phi\theta\chi) = (1/q)\exp(-u/kT)$$
$$= (1/q)\sum_{wm'} h^{w}_{m'0} D^{w*}_{0m'}(\phi\theta\chi), \tag{6}$$

where q is a partition function given by

$$q = \int \exp[-u/kT]\, d\Omega = 8\pi^2 h^{0}_{00}, \tag{7}$$

and the coefficients in Eq. (1) are given by

$$g^{w}_{m'0} = h^{w}_{m'0}/q. \tag{8}$$

We also use the power series

$$W(\phi\theta\chi) = (1/q)\sum_{n} C_n F^n, \tag{9}$$

where by the Maclaurin expansion we have

$$\begin{aligned}
C_0 &= 1,\\
C_1 &= L_z/kT,\\
C_2 &= \tfrac{1}{2}(L_z/kT)^2 + L_{zz}/kT,\\
C_3 &= \tfrac{1}{6}(L_z/kT)^3 + L_z L_{zz}/(kT)^2 + L_{zzz}/kT,\\
C_4 &= \tfrac{1}{24}(L_z/kT)^4 + \tfrac{1}{2}(L_z^2 L_{zz})/(kT)^3\\
&\quad + \tfrac{1}{2}(2L_z L_{zzz} + L_{zz}^2)/(kT)^2 + L_{zzzz}/kT.
\end{aligned} \tag{10}$$

Since Eq. (5) is also a Maclaurin series, the tensors $L_{z\ldots}$ must be completely symmetric in their indices. Therefore,

$$\begin{aligned}
L_z &= l^{1,1}_3,\\
L_{zz} &= l^{2,0} - l^{2,2}_1 - l^{2,2}_4,\\
L_{zzz} &= 3l^{3,1}_3 - l^{3,3}_2 - l^{3,3}_6,\\
L_{zzzz} &= 3l^{4,0} - 6(l^{4,2}_1 + l^{4,2}_4) - l^{4,4}_4 - l^{4,4}_8.
\end{aligned} \tag{11}$$

The irreducible components $l^{r,w}_q$ are known[2] in terms of the subsystem irreducible components $b^{r,w}_q$ and the functions $D^{w}_{m'm}$. In turn, the components $b^{r,w}_q$ are known[1] in terms of the reducible subsystem components $B_{i\ldots}$. Expressions for q and the coefficients of $h^{w}_{m'0}$ can now be obtained. The most general results are those for subsystems with no geometric symmetry (point group C_1). These results are given in the next subsection. All the results for subsystems with symmetry can be obtained by direct substitution into these general results. This is illustrated subsequently by considering examples.

A. Subsystems with no geometric symmetry (point group C_1)

It is found that

$$h^{w}_{-m'0} = (-1)^{m'} h^{w*}_{m'0}, \tag{12}$$

and we use the following notation:

$$h^{w}_{m'0} = \sum_{s=0}^{4} \frac{H^{w}_{m's}}{(kT)^s}. \tag{13}$$

The expressions for the nonzero $H^{w}_{m's}$ and for q are listed below in terms of the $N^{r,w}_{m'}$ given previously[2]:

$$\begin{aligned}
H^0_{00} &= 1,\\
H^0_{01} &= N^{2,0}_0 F^2 + 3N^{4,0}_0 F^4,\\
H^0_{02} &= [N^{1,1}_1 N^{1,1*}_1 + (N^{1,1}_0)^2]F^2/6 + [30N^{3,1}_0 N^{1,1}_0 + 15N^{3,1*}_1 N^{1,1}_1\\
&\quad + 15N^{3,1}_3 N^{1,1*}_1 + N^{2,2*}_2 N^{2,2}_2 + 3(N^{2,2}_0)^2 + 15(N^{2,0}_0)^2]F^4/30,\\
H^0_{03} &= [N^{2,2*}_2 (N^{1,1}_1)^2 + N^{2,2}_2 (N^{1,1*}_1)^2 + 10(N^{1,1}_0)^2 N^{2,0}_0\\
&\quad + 2N^{2,2}_0 N^{1,1*}_1 N^{1,1}_1 - 4N^{2,2}_0 (N^{1,1}_0)^2 + 10N^{1,1*}_1 N^{1,1}_1 N^{2,0}_0]F^4/60,\\
H^0_{04} &= [N^{1,1*}_1)^2 (N^{1,1}_1)^2 + 2N^{1,1}_1 N^{1,1*}_1 (N^{1,1}_0)^2 + (N^{1,1}_0)^4]F^4/120,\\
H^1_{11} &= (-N^{1,1}_1 F - 3N^{3,1}_1 F^3)/\sqrt{2},\\
H^1_{12} &= (-N^{2,2}_2 N^{1,1*}_1 - 2N^{2,2}_1 N^{1,1}_0 - N^{2,2}_0 N^{1,1}_1 - 5N^{1,1}_1 N^{2,0}_0)F^3/5\sqrt{2},\\
H^1_{13} &= [-(N^{1,1}_1)^2 N^{1,1*}_1 - N^{1,1}_1 (N^{1,1}_0)^2]F^3/10\sqrt{2},\\
H^1_{01} &= N^{1,1}_0 F + 3N^{3,1}_0 F^3,\\
H^1_{02} &= (N^{2,2*}_1 N^{1,1}_1 + N^{2,2}_1 N^{1,1*}_1 - 2N^{2,2}_0 N^{1,1}_0 + 5N^{1,1}_0 N^{2,0}_0)F^3/5,\\
H^1_{03} &= [N^{1,1}_1 N^{1,1*}_1 N^{1,1}_0 + (N^{1,1}_0)^3]F^3/10,\\
H^2_{21} &= N^{2,2}_2 F^2/\sqrt{6} + (\sqrt{6})N^{4,2}_2 F^4,\\
H^2_{22} &= (N^{1,1}_1)^2 F^2/2\sqrt{6} + [3N^{3,3}_3 N^{1,1*}_1 + 6N^{3,3}_2 N^{1,1}_0 - 3N^{3,3}_1 N^{1,1}_1 + 42N^{3,1}_1 N^{1,1}_1\\
&\quad + 4(N^{2,2}_1)^2 + 14N^{2,2}_2 N^{2,0}_0 + 4N^{2,2}_2 N^{2,2}_0](\sqrt{6})F^4/84,\\
H^2_{23} &= [3N^{2,2}_2 N^{1,1*}_1 N^{1,1}_1 + N^{2,2}_2 (N^{1,1}_0)^2 + 4N^{2,2}_1 N^{1,1}_1 N^{1,1}_0 + 2N^{2,2}_0 (N^{1,1}_1)^2\\
&\quad + 7(N^{1,1}_1)^2 N^{2,0}_0](\sqrt{6})F^4/84,\\
H^2_{24} &= [(N^{1,1}_1)^3 N^{1,1*}_1 + 10(N^{1,1}_1)^2 (N^{1,1}_0)^2](\sqrt{6})F^4/1680,
\end{aligned}$$

$$H^2_{11} = -(\sqrt{6})N^{2,2}_1F^2/3 - 2(\sqrt{6})N^{4,2}_1F^4,$$

$$H^2_{12} = -N^{1,1}_1N^{1,1}_0F^2/\sqrt{6} + (-3N^{3,3}_2N^{1,1*}_1 - 6N^{3,3}_1N^{1,1}_0 + 3N^{3,3}_0N^{11}_1 - 21N^{3,1}_1N^{1,1}_0$$
$$- 21N^{3,1}_0N^{1,1}_1 - 2N^{2,2}_2N^{2,2*}_1 - 2N^{2,2}_1N^{2,2}_0 + 14N^{2,2}_1N^{2,0}_0)(\sqrt{2})F^4/42,$$

$$H^2_{13} = [N^{2,2}_2N^{1,1*}_1N^{1,1}_0 + 2N^{2,2}_1N^{1,1*}_1N^{1,1}_1 - N^{2,2*}_1(N^{1,1}_1)^2 - 3N^{2,2}_1(N^{1,1}_0)^2$$
$$+ N^{2,2}_0N^{1,1}_1N^{1,1}_0 - 7N^{1,1}_1N^{1,1}_0N^{2,0}_0](\sqrt{6})F^4/42,$$

$$H^2_{14} = [-(N^{1,1}_1)^2N^{1,1*}_1N^{1,1}_0 - N^{1,1}_1(N^{1,1}_0)^3](\sqrt{6})F^4/84,$$

$$H^2_{01} = -N^{2,2}_0F^2 - 6N^{4,2}_0F^4,$$

$$H^2_{02} = [-N^{1,1*}_1N^{1,1}_1 + 2(N^{1,1}_0)^2]F^2/6 + [9N^{3,3*}_1N^{1,1}_1 + 9N^{3,3}_1N^{1,1*}_1 + 18N^{3,3}_0N^{1,1}_0$$
$$+ 84N^{3,1}_0N^{1,1}_0 - 21N^{3,1*}_1N^{1,1}_1 - 21N^{3,1}_1N^{1,1*}_1 - 2N^{2,2*}_2N^{2,2}_2 + 4N^{2,2*}_1N^{2,2}_1$$
$$- 42N^{2,2}_0N^{2,0}_0 + 6(N^{2,2}_0)^2]F^4/42,$$

$$H^2_{03} = [-N^{2,2*}_2(N^{1,1}_1)^2 - N^{2,2}_2(N^{1,1*}_1)^2 + 2N^{2,2*}_1N^{1,1}_1N^{1,1}_0 + 2N^{2,2}_1N^{1,1*}_1N^{1,1}_0$$
$$- 5N^{2,2}_0N^{1,1*}_1N^{1,1}_1 - 11N^{2,2}_0(N^{1,1}_0)^2 - 7N^{1,1*}_1N^{1,1}_1N^{2,0}_0$$
$$+ 14(N^{1,1}_0)^2N^{2,0}_0]F^4/42,$$

$$H^2_{04} = [-(N^{1,1*}_1)^2(N^{1,1}_1)^2 + N^{1,1}_1N^{1,1*}_1(N^{1,1}_0)^2 + 2(N^{1,1}_0)^4]F^4/84,$$

$$H^3_{31} = N^{3,3}_3F^3/2\sqrt{5},$$

$$H^3_{32} = -N^{2,2}_2N^{1,1}_1F^3/2\sqrt{5},$$

$$H^3_{33} = -(N^{1,1}_1)^3F^3/12\sqrt{5},$$

$$H^3_{21} = 3N^{3,3}_2F^3/\sqrt{30},$$

$$H^3_{22} = (N^{2,2}_2N^{1,1}_0 + 2N^{2,2}_1N^{1,1}_1)F^3/\sqrt{30},$$

$$H^3_{23} = (N^{1,1}_1)^2N^{1,1}_0F^3/2\sqrt{30},$$

$$H^3_{11} = -(\sqrt{3})N^{3,3}_1F^3/2,$$

$$H^3_{12} = (N^{2,2}_2N^{1,1*}_1 - 8N^{2,2}_1N^{1,1}_0 + 6N^{2,2}_0N^{1,1}_1)F^3/10\sqrt{3},$$

$$H^3_{13} = [(N^{1,1}_1)^2N^{1,1*}_1 - 4N^{1,1}_1(N^{1,1}_0)^2]F^3/20\sqrt{3},$$

$$H^3_{01} = N^{3,3}_0F^3,$$

$$H^3_{02} = [-N^{2,2*}_1N^{1,1}_1 - N^{2,2}_1N^{1,1*}_1 - 3N^{2,2}_0N^{1,1}_0]F^3/5,$$

$$H^3_{03} = [-3N^{1,1}_1N^{1,1*}_1N^{1,1}_0 + 2(N^{1,1}_0)^3]F^3/30,$$

$$H^4_{41} = N^{4,4}_4F^4/\sqrt{70},$$

$$H^4_{42} = [2N^{3,3}_3N^{1,1}_1 + (N^{2,2}_2)^2]F^4/2\sqrt{70},$$

$$H^4_{43} = N^{2,2}_2(N^{1,1}_1)^2F^4/2\sqrt{70},$$

$$H^4_{44} = (N^{1,1}_1)^4F^4/24\sqrt{70},$$

$$H^4_{31} = 2N^{4,4}_3F^4/\sqrt{35},$$

$$H^4_{32} = (-N^{3,3}_3N^{1,1}_0 - 3N^{3,3}_2N^{1,1}_1 - 2N^{2,2}_2N^{2,2}_1)F^4/2\sqrt{35},$$

$$H^4_{33} = [-N^{2,2}_1(N^{1,1}_1)^2 - N^{2,2}_2N^{1,1}_1N^{1,1}_0]F^4/2\sqrt{35},$$

$$H^4_{34} = -(N^{1,1}_1)^3N^{1,1}_0F^4/12\sqrt{35},$$

$$H^4_{21} = (\sqrt{10})N^{4,4}_2F^4/5,$$

$$H^4_{22} = [-N^{3,3}_3N^{1,1*}_1 + 12N^{3,3}_2N^{1,1}_0 + 15N^{3,3*}_1N^{1,1}_1 + 8(N^{2,2}_1)^2$$
$$- 6N^{2,2}_2N^{2,2}_0]F^4/14\sqrt{10},$$

$$H^4_{23} = [-N^{2,2}_2N^{1,1}_1N^{1,1*}_1 + 8N^{2,2}_1N^{1,1}_1N^{1,1}_0 + 2N^{2,2}_2(N^{1,1}_0)^2$$
$$- 3N^{2,2}_0(N^{1,1}_1)^2]F^4/14\sqrt{10},$$

$$H^4_{24} = [-(N^{1,1}_1)^3N^{1,1*}_1 + 6(N^{1,1}_1)^2(N^{1,1}_0)^2]F^4/86\sqrt{10},$$

$$H^4_{11} = -2N^{4,4}_1F^4/\sqrt{5},$$

$$H^4_{12} = [3N^{3,3}_2 N^{1,1*}_1 - 15N^{3,3}_1 N^{1,1}_0 - 10N^{3,3}_0 N^{1,1}_1 + 2N^{2,2}_2 N^{2,2*}_1]F^4/14\sqrt{5},$$
$$H^4_{13} = [N^{2,2}_2 N^{1,1*}_0 N^{1,1}_0 + 2N^{2,2}_1 N^{1,1*}_1 N^{1,1}_1 - 4N^{2,2}_1 (N^{1,1}_0)^2 + N^{2,2*}_1 (N^{1,1}_1)^2 + 6N^{2,2}_0 N^{1,1}_1 N^{1,1}_0]F^4/14\sqrt{5},$$
$$H^4_{14} = [6(N^{1,1}_1)^2 N^{1,1*}_1 N^{1,1}_0 - 8N^{1,1}_1 (N^{1,1}_0)^3]F^4/172\sqrt{5},$$
$$H^4_{01} = -N^{4,4}_0 F^4,$$
$$H^4_{02} = [-15N^{3,3*}_1 N^{1,1}_1 - 15N^{3,3}_1 N^{1,1*}_1 + 40N^{3,3}_0 N^{1,1}_0 + N^{2,2*}_2 N^{2,2}_2 - 16N^{2,2*}_1 N^{2,2}_1 + 18(N^{2,2}_0)]F^4/70,$$
$$H^4_{03} = [N^{2,2*}_2 (N^{1,1}_1)^2 + N^{2,2}_2 (N^{1,1*}_1)^2 - 16N^{2,2*}_1 N^{1,1}_1 N^{1,1}_0 - 16N^{2,2}_1 N^{1,1*}_1 N^{1,1}_0 + 12N^{2,2}_0 N^{1,1*}_1 N^{1,1}_1 - 24N^{2,2}_0 (N^{1,1}_0)^2]F^4/140,$$
$$H^4_{04} = [3(N^{1,1}_1)^2 (N^{1,1*}_1)^2 - 24N^{1,1}_1 N^{1,1*}_1 (N^{1,1}_0)^2 + 8(N^{1,1}_0)^4]F^4/840.$$

The partition function is

$$q = 8\pi^2\{1 + N^{2,0}_0 F^2/kT + [F^2/6(kT)^2][N^{1,1}_1 N^{1,1*}_1 + (N^{1,1}_0)^2] + 3N^{4,0}_0 F^4/kT + [F^4/10(kT)^2][15N^{3,1}_0 N^{1,1}_0 + 5N^{3,1*}_1 N^{1,1}_1 + 5N^{3,1}_3 N^{1,1*}_1 + N^{2,2*}_2 N^{2,2}_2/3 + (N^{2,2}_0)^2 + 5(N^{2,0}_0)^2] + [F^4/60(kT)^3][N^{2,2*}_2 (N^{1,1}_1)^2 + N^{2,2}_2 (N^{1,1*}_1)^2 + 2N^{2,2}_0 N^{1,1*}_1 N^{1,1}_1 - 4N^{2,2}_0 (N^{1,1}_0)^2 + 10N^{1,1*}_1 N^{1,1}_1 N^{2,0}_0 + 10(N^{1,1}_0)^2 N^{2,0}_0] + [F^4/120(kT)^4][(N^{1,1*}_1)^2 (N^{1,1}_1)^2 + 2N^{1,1}_1 N^{1,1*}_1 (N^{1,1}_0)^2 + (N^{1,1}_0)^4]\}.$$

B. Subsystems with point symmetries T, O, and O_h

Since the tensors $B_{i\cdots}$ are symmetric in their indices, the independent components are related as follows[3]:

$$B_{xx} = B_{yy} = B_{zz},$$
$$B_{xxxx} = B_{yyyy} = B_{zzzz}, \tag{14}$$
$$B_{xxyy} = B_{xxzz} = B_{yyzz}.$$

The nonzero irreducible components $b^{r,w}_q$ are given by the following relations:

$$b^{2,0} = B_{xx},$$
$$b^{4,0}(n) = (2B_{xxyy} + B_{xxxx})/5, \quad n = 1,2,3, \tag{15}$$
$$b^{4,4}_3 = b^{4,4}_4 = b^{4,4}_8 = (3B_{xxyy} - B_{xxxx})/5.$$

The following results for q and the nonzero coefficients $h^{w}_{m'0}$ are obtained by use of the results given in Sec. IIA:

$$q = 8\pi^2\{1 + (b^{2,0}/2kT)^2 + \tfrac{1}{2}[(3b^{4,0}/2kT)] + (b^{2,0}/2kT)^2]F^4 + \cdots\}, \tag{16}$$
$$h^4_{00} = -b^{4,4}_3 F^4/(2kT) + \cdots,$$
$$h^4_{40} = h^4_{-40} = -(\sqrt{70})b^{4,4}_3 F^4/(28kT) + \cdots. \tag{17}$$

C. Subsystems with point symmetries D_{nh} with $n \geq 5$

In this case we have the independent components B_{zz}, B_{zzzz}, and the following[4]:

$$B_{xx} = B_{yy},$$
$$B_{xxxx} = B_{yyyy} = 3B_{xxyy}, \tag{18}$$
$$B_{xxzz} = B_{yyzz}.$$

Also, we have

$$b^{2,0} = (2B_{xx} + B_{zz})/3,$$
$$b^{2,2}_1 = b^{2,2}_4 = (B_{xx} - B_{zz})/3,$$
$$b^{4,0}(n) = (8B_{xxyy} + 4B_{xxzz} + B_{zzzz})/15, \quad n = 1,2,3, \tag{19}$$
$$b^{4,2}_1(n) = b^{4,2}_4(n) = (4B_{xxyy} - B_{xxzz} - B_{zzzz})/21, \quad n = 1,\ldots 6,$$
$$b^{4,4}_3 = -\tfrac{1}{4}b^{4,4}_4 = -\tfrac{1}{4}b^{4,4}_8 = (3B_{xxyy} - 6B_{xxzz} + B_{zzzz})/35,$$
$$q = 8\pi^2\left\{1 + \frac{b^{2,0}}{2kT}F^2 + \left[\frac{5(b^{2,0})^2 + 4(b^{2,2}_1)^2}{40(kT)^2} + \frac{3b^{4,0}}{4kT}\right]F^4 + \cdots\right\}. \tag{20}$$

The nonzero coefficients $h^{w}_{m'0}$ are

$$h^2_{00} = -\frac{b^{2,2}_1 F^2}{kT} + \left[\frac{(b^{2,2}_1)^2}{7(kT)^2} - \frac{b^{2,0}b^{2,2}_1}{2(kT)^2} - \frac{3b^{4,2}_1}{kT}\right]F^4 + \cdots, \tag{21}$$
$$h^4_{00} = \left[\frac{9(b^{2,2}_1)^2}{35(kT)^2} + \frac{2b^{4,4}_3}{kT}\right]F^4 + \cdots.$$

III. TENSOR AVERAGES

Suppose that $\mathscr{L}_{i\cdots}$ are tensor properties of the subsystems expressed in the laboratory coordinates. They will in general depend upon the vector F and so can be expressed (through terms of rank 4) as follows:

$$\mathscr{L}_i = \hat{L}_i + \hat{L}_{i;z}F + \hat{L}_{i;zz}F^2 + \hat{L}_{i;zzz}F^3 + \cdots,$$
$$\mathscr{L}_{ij} = \hat{L}_{ij} + \hat{L}_{ij;z}F + \hat{L}_{ij;zz}F^2 + \cdots,$$
$$\mathscr{L}_{ijk} = \hat{L}_{ijk} + \hat{L}_{ijk;z}F + \cdots, \tag{22}$$
$$\mathscr{L}_{ijkl} = \hat{L}_{ijkl} + \cdots,$$

where the tensors $\hat{L}_{i\cdots}$ have the circumflex accent to distinguish them from the tensors in Eq. (5). The tensors $\hat{L}_{i\cdots}$ are completely symmetric in the indices which are to the right of the semicolon. The type of symmetry which applies to indices on the left of the semicolon is that of the $\mathscr{L}_{i\cdots}$.

The following averages can be nonzero when $W(\phi\theta\chi)$ has the cylindrical symmetry possessed by the functions in Sec. II:

$$\langle \hat{L}_z \rangle, \quad \langle \hat{L}_{z;z} \rangle, \quad \langle \hat{L}_{z;zz} \rangle, \quad \langle \hat{L}_{z;zzz} \rangle;$$
$$\langle \hat{L}_{xx} \rangle = \langle \hat{L}_{yy} \rangle, \quad \langle \hat{L}_{zz} \rangle, \quad \langle \hat{L}_{xx;z} \rangle = \langle \hat{L}_{yy;z} \rangle, \quad \langle \hat{L}_{zz;z} \rangle,$$
$$\langle \hat{L}_{xx;zz} \rangle = \langle \hat{L}_{yy;zz} \rangle, \quad \langle \hat{L}_{zz;zz} \rangle;$$
$$\langle \hat{L}_{xxz} \rangle = \langle \hat{L}_{yyz} \rangle, \quad \langle \hat{L}_{zzz} \rangle,$$
$$\langle \hat{L}_{xxz;z} \rangle = \langle \hat{L}_{yyz;z} \rangle, \quad \langle \hat{L}_{zzz;z} \rangle;$$
$$\langle \hat{L}_{xxzz} \rangle = \langle \hat{L}_{yyzz} \rangle, \quad \langle \hat{L}_{zzzz} \rangle,$$
$$\langle \hat{L}_{xxxx} \rangle = \langle \hat{L}_{yyyy} \rangle = 3\langle \hat{L}_{xxyy} \rangle.$$

In what follows we use $\hat{l}^{r,w}_{(R)q}(n)$ to represent the irreducible components of the reducible tensor $\hat{L}_{i_1\cdots i_R;i_{R+1}\cdots i_r}$, where R is the rank of $\mathscr{L}_{i_1\cdots i_R}$ and is needed as an additional label to identify the $\hat{L}_{i\cdots}$ to which the irreducible component belongs. We use $\hat{b}^{r,w}_{(R)q}(n)$ to represent the irreducible components in the subsystem coordinates. The reducible components of $\hat{L}_{i_1\cdots i_R;i_{R+1}\cdots i_r}$ in the subsystem coordinates are designated $\hat{B}_{i_1\cdots i_R;i_{R+1}\cdots i_r}$.

The following constitutes an outline of a convenient procedure for calculating the averages $\langle \mathscr{L}_{i\cdots} \rangle$:

(i) The indicial symmetry of the tensor $\mathscr{L}_{i\cdots}$ and the geometric symmetry of the subsystem are used to obtain the relationships between the nonzero components $\hat{B}_{i_1\cdots i_R;\cdots}$.

(ii) The nonzero components $b^{r,w}_{(R)q}(n)$ are obtained in terms of the $\hat{B}_{i_1\cdots i_R;\cdots}$.

(iii) The $\hat{l}^{r,w}_{(R)q}(n)$ are expressed in terms of the $\hat{b}^{r,w}_{(R)q}(n)$ and the $D^{w}_{m'm}$.

(iv) The $\hat{L}_{i_1\cdots i_R;\cdots}$ are expresed in terms of the $\hat{l}^{r,w}_{(R)q}(n)$.

(v) Expressions

$$\mathscr{L}_{i_1\cdots i_R} = \sum_{rwm'm} A^{r,w}_{m'm} D^{w*}_{mm'}(\phi\theta\chi) \tag{23}$$

are then obtained, where the coefficients are linear combinations of the coefficients $a^{w}_{qm'm}$ occurring in Eq. (2). (Actually the coefficients $A^{r,w}_{m'm}$ should bear the indices $i_1\cdots i_R$ also, but we have suppressed these.) Note that Eq. (23) may be useful *per se*, since it gives the tensor $\mathscr{L}_{i\cdots}$ as a function of the orientation of the subsystem relative to the vector F.

(vi) The tensor average is given by

$$\langle \mathscr{L}_{i_1\cdots i_R} \rangle = 8\pi^2 \sum_{rwm'm} \left(\frac{1}{2w+1}\right) A^{r,w*}_{m'm} g^{w}_{m'm}. \tag{24}$$

The averages can all be expressed as the sum of a zero-weight part that corresponds to random orientation of the subsystems and does not, therefore, show an explicit temperature dependence and of a nonrandom part that shows an explicit temperature dependence. That is, we can write

$$\langle \mathscr{L}_{i_1\cdots i_R} \rangle = \langle L_{i_1\cdots i_R} \rangle^0 + \langle L_{i_1\cdots i_R} \rangle^T, \tag{25}$$

where

$$\langle \mathscr{L}_{i_1\cdots i_R} \rangle^0 = \sum_r A^{r,0}_{00} \tag{26}$$

and

$$\langle \mathscr{L}_{i_1\cdots i_R} \rangle^T = 8\pi^2 \sum_{r,w\neq 0,m'm} \left(\frac{1}{2w+1}\right) A^{r,w*}_{m'm} g^{w}_{m'm}. \tag{27}$$

It should be noted that the explicit temperature dependence indicated by the notation in Eq. (25) does not necessarily express the total effect of the temperature, since in general the tensor components $\hat{B}_{i\cdots}$ will have a temperature dependence, which we are not considering in our present treatment.

Example

As an illustration we given the results for the case of completely symmetric $\mathscr{L}_{i\cdots}$ and subsystem symmetries O and O_h. (These two symmetries give the same results for completely symmetric $\mathscr{L}_{i\cdots}$.)

The results are as follows[5]:

$$\langle \mathscr{L}_z \rangle^0 = \hat{b}^{2,0}_{(1)} + 3\hat{b}^{4,0}_{(1)} F^3 + \cdots,$$
$$\langle \mathscr{L}_z \rangle^T = 4\hat{b}^{4,4}_{(1)3} b^{4,4}_{3} F^7/(21kT) + \cdots; \tag{28}$$
$$\langle \mathscr{L}_{xx} \rangle^0 = \langle \mathscr{L}_{yy} \rangle^0 = \hat{b}^{2,0}_{(2)} + \hat{b}^{4,0}_{(2)}(1) F^2 + \cdots,$$
$$\langle \mathscr{L}_{zz} \rangle^0 = \hat{b}^{2,0}_{(2)} + [\hat{b}^{4,0}_{(2)}(1) + 2\hat{b}^{4,0}_{(2)}(2)] F^2 + \cdots,$$
$$\langle \mathscr{L}_{zz} \rangle^T = -2\langle \mathscr{L}_{xx} \rangle^T = -2\langle \mathscr{L}_{yy} \rangle^T$$
$$= 4\hat{b}^{4,4}_{(2)3} b^{4,4}_{3} F^6/(21kT)\cdots, \tag{29}$$
$$\langle \mathscr{L}_{zzz} \rangle^0 = 3\langle \mathscr{L}_{xxz} \rangle^0 = 3\langle \mathscr{L}_{yyz} \rangle^0 = 3\hat{b}^{4,0}_{(3)} F + \cdots,$$
$$\langle \mathscr{L}_{zzz} \rangle^T = -2\langle \mathscr{L}_{xxz} \rangle^T = -2\langle \mathscr{L}_{yyz} \rangle^T$$
$$= 4\hat{b}^{4,4}_{(3)3} b^{4,4}_{3} F^5/(21kT) + \cdots; \tag{30}$$
$$\langle \mathscr{L}_{zzzz} \rangle^0 = \langle \mathscr{L}_{xxxx} \rangle^0 = \langle \mathscr{L}_{yyyy} \rangle^0$$
$$= 3\langle \mathscr{L}_{xxyy} \rangle^0 = 3\langle \mathscr{L}_{xxzz} \rangle^0 = 3\langle \mathscr{L}_{yyzz} \rangle^0$$
$$= 3\hat{b}^{4,0}_{(4)} + \cdots,$$
$$\langle \mathscr{L}_{zzzz} \rangle^T = \tfrac{8}{3}\langle \mathscr{L}_{xxxx} \rangle^T = \tfrac{8}{3}\langle \mathscr{L}_{yyyy} \rangle^T = 8\langle \mathscr{L}_{xxyy} \rangle^T$$
$$= -2\langle \mathscr{L}_{xxzz} \rangle^T = -2\langle \mathscr{L}_{yyzz} \rangle^T$$
$$= 4\hat{b}^{4,4}_{(4)3} b^{4,4}_{3} F^4/(21kT) + \cdots. \tag{31}$$

IV. A DESCRIPTION OF POSSIBLE APPLICATIONS

In order to give the foregoing results more concrete significance, we shall briefly discuss some possible applications and explain the physical meanings of the various tensors that occur.

Perhaps the simplest examples are those in which the subsystem is a molecule (or small particle) in a strong, homogeneous, static electric field. In this case F_z is the static field, and Eq. (5) gives the energy u of the molecule in the field as a function of the magnitude of the field and the orientation of the molecule relative to the field. L_z, L_{zz}, L_{zzz}, and L_{zzzz} are the dipole moment, the polarizability, and the first and second hyperpolarizabilities,[6] respectively, in the absence of the field and expressed in the laboratory coordinates. They are completely symmetric tensors. $l^{r,w}_{q}$ are the irreducible components of these tensors. The corresponding tensors in the molecular coordinates are $B_{i\cdots}$ and $b^{r,w}_{q}$.

The tensors $\mathscr{L}_{i\cdots}$ are some properties of the molecule in the presence of the static field. We shall consider two examples of these.

First, suppose that $\mathscr{L}_i$, $\mathscr{L}_{ij}$, $\mathscr{L}_{ijk}$, and $\mathscr{L}_{ijkl}$ are the dipole and higher multipole moments of the molecule in the static field. Equations (22) give these moments as functions of field.[6] The tensors $\hat{L}_{i\cdots}$ (no semicolon) are the multipole moments in the absence of the field. The tensors $\hat{L}_{i\cdots}$ (with semicolon) give the effect of the field on the moments. The irreducible tensors corresponding to $\hat{L}_{i\cdots}$ with and without

the semicolon, respectively, are $\hat{l}^{R,w}_{(R)q}(r=R)$ and $\hat{l}^{r,w}_{(R)q}(n)$ $(r\neq R)$. The (n) must be indicated in the last of the preceding tensors because those tensors are not in general symmetric in indices that occur on opposite sides of the semicolon. The analogous notations for the tensors in the molecular coordinates are $\hat{B}_{i\cdots}$ (with and without semicolon), $\hat{b}^{R,w}_{(R)q}$ and $b^{r,w}_{(R)q}(n)$. Once the $\hat{L}_{i\cdots}$ in Eqs. (22) have been expressed in terms of the $b^{r,w}_{q}$ and the $\hat{b}^{r,w}_{(R)q}(n)$, this equation gives the multipoles as functions of the static field and the angular orientations of the molecule relative to that field's direction. The molecular multipoles are usually (but not always) defined as irreducible tensors.[6,7] Usually only those with weight equal to the rank are considered. They are, therefore, usually the tensors $\hat{b}^{R,R}_{(R)q}$. If one is interested in the interaction of the molecule with fields that satisfies Laplace's equation, the tensors $b^{R,w}_{(R)q}$ with $w\neq R$ have no effect. The tensors $\hat{b}^{r,w}_{(R)q}(n)$ $(r\neq R)$ give the effect of the static field on the multipole of rank R, and they are symmetric in the sets of R indices and r-R indices separately, but not with respect to interchange of indices between these sets.

The tensor averages $\langle \mathscr{L}_{i\cdots} \rangle$ would be the time averages of the multipoles of an individual molecule, or the ensemble averages of a system of independent molecules, with a Boltzmann distribution of angular orientations in the field.

As a second illustration of the possible meanings of the tensors $\mathscr{L}_{i\cdots}$, suppose that they are frequency-dependent polarizabilities and hyperpolarizabilities. That is, suppose they represent the response of the molecule to a periodic field, which may be in an arbitrary direction relative to the static field F_z. Equations (22) then express the effect of the static field upon these frequency dependent tensors. $\hat{L}_{i\cdots}$ (no semicolon) are these tensor in the absence of the static field; $\hat{b}^{R,w}_{(R)q}$ are the corresponding irreducible tensors in the molecular coordinates. Once the $\hat{L}_{i\cdots}$ (with and without the semicolon) have been expressed in terms of the molecular tensors $b^{r,w}_{q}$, $\hat{b}^{R,w}_{(R)q}$, and $\hat{b}^{r,w}_{(R)q}(n)$, then Eqs. (22) give the frequency dependent polarizabilites and hyperpolarizabilities $\mathscr{L}_{i\cdots}$ as functions of the magnitude of the static field and the orientation of the molecule relative to that field. The anisotropy produced by the static field leads to the Kerr effect. If the periodic field is of sufficiently high frequency, even though it may be a strong field, its effect on the Boltzmann distribution produced by the static field can be neglected. $\langle \mathscr{L}_{i\cdots} \rangle$ are the averages of the frequency-dependent polarizabilities and hyperpolarizabilities.

Finally, we mention that the partition function q holds for any system of independent subsystems whose energies depend upon a vector in accordance with Eq. (5), provided these subsystems are free to attain thermal equilibrium with respect to the energy u.

[1] M. C. Russ and C. A. Hollingsworth, J. Math. Phys. **22**, 1025 (1981), Part I.

[2] M. C. Russ and C. A. Hollingsworth, J. Math. Phys. **22**, 1028 (1981), Part II.

[3] A convenient table giving the relationships between the nonzero components of tensors through rank 4 and for each of the 32 point groups is given by R. R. Bires, *Symmetry and Magnetism* (North-Holland, Amsterdam, 1964).

[4] Since D_{nh} is the same as $D_{\infty h}$ for tensors of ranks less than n, one can use the point group D_{6h}. See A. A. Reznik and C. A. Hollingsworth, J. Chem. Phys. **59**, 2054 (1973).

[5] Because of high symmetry of these subsystems it would be necessary to consider ranks through 8 to obtain a zero-weight term with a power of F equal to that of the lowest power in the nonrandom terms.

[6] Actually the hyperpolarizabilities are defined as $2L_{zzz}$ and $6L_{zzzz}$. See, for example, A. D. Buckingham, Quart. Rev. **13**, 183 (1959); also in Vol. XII of *Advances in Chemical Physics* (Wiley, Interscience, New York, 1967).

[7] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).

Construction of new integrable Hamiltonians in two degrees of freedom

Craig R. Holt[a)]
Department of Physics, University of Colorado and Joint Institute for Laboratory Astrophysics, University of Colorado and National Bureau of Standards, Boulder, Colorado 80309

(Received 29 October 1981; accepted for publication 12 February 1982)

A new procedure for deriving integrable Hamiltonians and their constants of the motion is introduced. We term this procedure the truncation program. Integrable Hamiltonians occurring in the truncation program possess constants of the motion which are polynomials in a perturbation parameter ϵ. The relationship between this program and the Whittaker program in two degrees of freedom is discussed. Integrable Hamiltonians occurring in the Whittaker program (a generalization of Whittaker's work) possess constants of the motion which are polynomials in the momentum coordinates. Many previously known integrable Hamiltonians are derived. A new family of integrable double resonance Hamiltonians and a new family of integrable Hamiltonians of the form $(p_1^2 + p_2^2)/2 + V(q_1, q_2)$ are derived.

PACS numbers: 03.20. + i

I. INTRODUCTION: INTEGRABLE HAMILTONIANS IN TWO DEGREES OF FREEDOM

In spite of the scarcity of integrable Hamiltonians relative to nonintegrable Hamiltonians,[1] several have been discovered either by construction[2–4] or by accident.[5,6] What follows is a partial list of these Hamiltonians and some references to other systems in two and more degrees of freedom.

The first systematic construction procedure was introduced by Whittaker.[2] He found all integrable Hamiltonians of the form

$$\mathscr{H} = \frac{p_1^2 + p_2^2}{2} + \epsilon V(q_1, q_2),$$

such that the second constant of the motion is linear or quadratic in momentum. Drach[3] considered Hamiltonians of the form

$$\mathscr{H} = p_2 p_2 \lambda(q_1, q_2) + \epsilon V(q_1, q_2),$$

and found all those which possessed constants of the motion which also were quadratic in momentum. In addition, he found a few with the same form such that the second constant of the motion is cubic in momentum. Havas[4] (and references contained therein) listed all Hamiltonians of the form

$$\mathscr{H} = g_{11}p_1^2 + g_{12}p_1p_2 + g_{22}p_2^2 + \epsilon V(q_1, q_2),$$

where the g_{ij} are functions of q_1 and q_2, such that the Hamilton–Jacobi equation separates after a coordinate transformation in the q plane. This is a generalization of Whittaker's work because all of Havas' constants of the motion are also quadratic in the momentum.

In addition to the above integrable systems, others were obtained less systematically. The Toda lattice,[5]

$$\mathscr{H} = \frac{p_1^2 + p_2^2}{2} + \epsilon[e^{q_2 + \sqrt{3}q_1} + e^{q_2 - \sqrt{3}q_1} + e^{-2q_2}],$$

for example, is an integrable (yet nonseparable in the sense noted above) Hamiltonian with a second constant of the motion which is cubic in the momentum. Additional, possibly nonseparable, integrable Hamiltonians have since been discovered.[6] The property shared by the systems in Ref. 6 is that the second constant of the motion is a polynomial in p_1 and p_2. Consequently, the class of integrable Hamiltonians, of the form $g_{11}p_1^2 + g_{12}p_1p_2 + g_{22}p_2^2 + \epsilon V(q_1, q_2)$, possessing this property contains all previously known integrable Hamiltonians of this type.

Single resonance Hamiltonians[7] comprise the last family of known integrable Hamiltonians:

$$\mathscr{H} = H(I_1, I_2) + \sum_{M=-\infty}^{\infty} V_M(I_1, I_2)e^{iMn\cdot\phi},$$

where $n\cdot\phi = n_1\phi_1 + n_2\phi_2$. The second constant of the motion, in this case, is given by $K = n_2I_1 - n_1I_2$.

In this paper, a systematic method is presented which is used not only to construct many of the above systems, but in addition, it is used to construct whole new families of integrable Hamiltonians. At the foundation of the method is the perturbation theory of McNamara and Whiteman[8] as applied to Hamiltonians of the form $\mathscr{H} = H(I_1, I_2) + \epsilon V(I_1, I_2, \phi_1, \phi_2)$. In their papers, a perturbation expansion in ϵ for the second constant of the motion is developed. In the present paper, the condition that the second constant of the motion be a polynomial in ϵ of finite order is imposed (i.e., the perturbation expansion truncates). As a consequence, H, V, and the zero-order invariant (which is perturbed into the second constant of the motion) are required to satisfy special partial differential equations. In some cases, these equations can be solved to obtain families of integrable Hamiltonians.

This method motivates the introduction of what we term the truncation program. The goal of the truncation program of order L is to derive all integrable Hamiltonians of the form $H + \epsilon V$ (where H is integrable) such that the constants of the motion are polynomials in ϵ of order $<L$.

Another method for deriving integrable Hamiltonians, we term the Whittaker program (a generalization of Whittaker's work[2]). The goal of the Whittaker program of order M is to find all integrable Hamiltonians of the form $(p_1^2 + \cdots + p_N^2)/2 + V(q_1,\ldots,q_n)$ such that the N constants of the motion are polynomials in the momentum coordinates of order $<M$. In this paper, the relationship between the trun-

[a)]Present address: Fermi National Accelerator Lab, Box 500, Batavia, Illinois 60510.

 0022-2488/82/061037-10$02.50

cation program and the Whittaker program is discussed in the case of two degrees of freedom.

The organization of the paper is as follows. In Sec. II, Poisson coordinates are introduced. These coordinates are used in Sec. III to construct an operator which maps invariants of H into those for $H + \epsilon V$. It is shown that this operator possesses an undesirable property: It maps discontinuous functions into continuous functions. A new operator lacking this property is introduced in Sec. IV. It is shown under certain conditions that this operator also maps invariants of H into those for $H + \epsilon V$. In Sec. V, it is shown how this operator can be used to obtain integrable Hamiltonians and their second constants of the motion. The constants of the motion which result are shown to be polynomials in ϵ of finite order. In Sec. VI, integrable Hamiltonians of the form $(p_1^2 + p_2^2)/2 + \epsilon V(q_1, q_2)$ are considered. It is shown that if the second constant of the motion is a polynomial in ϵ of finite order, that under one further assumption it also is a polynomial in momentum of finite order. In Sec. VII, sample potentials V satisfying these properties are derived.

II. POISSON COORDINATES FROM ACTION ANGLE COORDINATES

An integrable Hamiltonian H in N degrees of freedom possesses N constants of the motion $\{I_n, n = 1,...,N\}$. For such systems, there is an action-angle coordinate system such that

$$\{\phi_n, I_m\} = \delta_{nm} \tag{1}$$

and

$$H = H(I_1,...,I_N) \tag{2}$$

(where $\{,\}$ are the Poisson brackets). The ϕ_n coordinates are angle coordinates on the invariant torus generated by $\{I_n = I_n^0, n = 1,...,N\}$. The time evolution on a torus is obtained by integrating

$$\frac{d\phi_N}{dt} = \omega_n(I), \tag{3}$$

where $\omega_n(I) \equiv \partial H/\partial I_n$.

In addition to the I_n, there are N-1 more time-independent invariants given by

$$\psi_n = \frac{\phi_n}{\omega_n} - \frac{\phi_{n+1}}{\omega_{n+1}}, \quad n = 1,...,N-1. \tag{4}$$

The above quantities are invariants in the sense that $d\psi_n/dt = 0$ [as can readily be verified using (3)]. To be a constant of the motion, however, it is necessary that the invariant be differentiable everywhere on the phase space manifold.[9] Clearly the ψ_n are not differentiable everywhere for they fail to be continuous at $\phi_n = 0$. Nevertheless, the invariants ψ_n can in special cases be used to form constants of the motion. For example, assume that $H = H(m_1 I_1 + m_2 I_2, I_3,...,I_N)$ (where m_1, m_2 are integers), then an extra (i.e., in addition to $I_1,...,I_N$) constant of the motion is given by

$$K = (I_1)^{m_2/2}(I_2)^{m_1/2} \sin(m_2\phi_1 - m_1\phi_2). \tag{5}$$

A new canonical coordinate system can be introduced which uses as coordinates the $2N$-1 invariants and an additional coordinate termed a time coordinate.[10] A time coordinate T is a function on phase space defined by $dT/dt = 1$, i.e., it is conjugate to H when used as a coordinate:

$$\{T, H\} = \frac{dT}{dt} = 1. \tag{6}$$

For example, one choice for the time coordinate associated with (2) is given by

$$T = \frac{\phi_1}{\omega_1}. \tag{7}$$

In general, it can be shown[10] that a new canonical coordinate system $(\theta_n, F_n, n = 1,...,N$, where $\{\theta_n, F_m\} = \delta_{nm})$, which we term a Poisson coordinate system, can be formed from the invariants and T such that

$$F_1 = H, \tag{8}$$

$$\theta_1 = T, \tag{9}$$

with the remaining coordinates $(\theta_n, F_n, n = 2,...,N)$ being invariants:

$$\{\theta_n, H\} = 0, \tag{10}$$

$$\{F_n, H\} = 0. \tag{11}$$

The relationship (needed in a later section) between ϕ_n, θ_n, and ϕ_n is obtained using two generating functions of the second kind,[11]

$$S_1 = \sum_{l=1}^{N} F_l(I)\theta_l, \tag{12}$$

$$S_2 = \sum_{l=1}^{N} I_l(F)\phi_l, \tag{13}$$

where the F_l are chosen to be arbitrary differentiable functions of the I_l with the restriction that they be functionally independent and $F_1 = H$. Since the F_l are N independent functions of the I_l, the I_l are likewise N independent functions of the F_l. Consequently, the coordinates conjugate to the F_l are obtained from

$$\theta_n = \frac{\partial S_2}{\partial F_n}, \tag{14}$$

$$\theta_n = \sum_{l=1}^{N} \phi_l \frac{\partial I_l}{\partial F_n}. \tag{15}$$

Alternatively, the angle coordinates ϕ_n can be expressed in terms of the θ_n coordinates $(\phi_n = \partial S_1/\partial I_n)$:

$$\phi_n = \sum_{l=1}^{N} \theta_l \frac{\partial F_l}{\partial I_n}. \tag{16}$$

The invariants ψ_n are simply linear combinations of the θ_n $(n = 2,...,N)$.

III. INVARIANTS OF $H + \epsilon V$

Given the invariants of H (namely $F_1,...,F_N, \theta_2,...,\theta_N$), an expression can be obtained for the invariants of $H + \epsilon V$ (where V is a differentiable function on phase space and ϵ is a tracking parameter).[8] An invariant denoted by $\mathcal{K}$, of $H + \epsilon V$ must satisfy

$$\{\mathcal{K}, H + \epsilon V\} = 0. \tag{17}$$

In Poisson coordinates, (17) becomes

$$\frac{\partial \mathscr{K}}{\partial T} + \epsilon\{\mathscr{K}, V\} = 0. \tag{18}$$

Integration of (18) with respect to T, holding the remaining Poisson coordinates fixed, yields

$$\int_{T_0}^{T} dT \frac{\partial \mathscr{K}}{\partial T} + \epsilon \int_{T_0}^{T} dT \{\mathscr{K}, V\} = 0, \tag{19}$$

where T_0 is a constant. Application of the fundamental theorem of calculus to (19) yields

$$\mathscr{K}|_T - \mathscr{K}|_{T_0} + \epsilon \int_{T_0}^{T} dT \{\mathscr{K}, V\} = 0. \tag{20}$$

Equation (20) can be rewritten as

$$R(\mathscr{K}) = \mathscr{K}|_{T_0}, \tag{21}$$

where $R(\ )$ is an operator defined by

$$R(\mathscr{K}) = \mathscr{K} - \epsilon D(\mathscr{K}), \tag{22}$$

with

$$D(\mathscr{K}) = \int_{T_0}^{T} dT \{V, \mathscr{K}\}. \tag{23}$$

Equation (21) shows that R maps invariants of $H + \epsilon V$ into invariants of H. This follows because $\mathscr{K}|_{T_0}$ depends only on the remaining $2N$-1 coordinates, which are also invariants of H: i.e.,

$$R(\mathscr{K}) = K(F_1,...,F_N, \theta_2,...,\theta_N), \tag{24}$$

where $K(,,...)$ is an arbitrary function of its arguments. In addition, the solution to (18) exists and is unique (assuming the boundary condition $\mathscr{K}|_{T_0} = K$).[12] Consequently, Eq. (24) can be inverted to obtain

$$\mathscr{K} = R^{-1}(K). \tag{25}$$

Equation (25) shows that R^{-1} maps invariants of H into invariants of $H + \epsilon V$ [the perturbative expansion of (25) appears in Ref. 8].

Since constants of the motion must be continuous functions, it is desirable that all operators, such as R and D, map continuous functions into continuous functions. The main disadvantage of R, however, is that it maps continuous functions into discontinuous functions. Consider, for example, an $N = 2$ system where $V = \sin(\phi_1 + \phi_2)$. Assuming $T_0 = 0$, one obtains using (21)

$$R(H + \epsilon V) = H + \epsilon V|_{T_0 = 0}, \tag{26}$$

where

$$V|_{T_0=0} = \sin\left[\frac{\dfrac{\partial F_2}{\partial I_1} + \dfrac{\partial F_2}{\partial I_2}}{\dfrac{\partial F_2}{\partial I_1}\omega_2 - \dfrac{\partial F_2}{\partial I_2}\omega_1}\right](\omega_2\phi_1 - \omega_1\phi_2). \tag{27}$$

The Hamiltonian $H + \epsilon V$ is continuous, yet a necessary condition for $V|_{T_0}$ to be continuous is that ω_2, ω_1 be commensurable. This is not the case if H is nonlinear in the action coordinates. Consequently, R maps continuous functions into discontinuous functions for most H. Using R^{-1}, therefore, the constants of the motion for $H + \epsilon V$ can only be generated from discontinuous invariants.

This is the motivation for seeking an operator which maps continuous functions into continuous functions. Such as operator is constructed in the next section.

IV. DOMAIN AND EXISTENCE OF $\hat{R}^{-1}$

A new operation $\hat{D}$ (in terms of which $\hat{R} = 1 - \epsilon\hat{D}$), analogous to D, is defined here by showing explicitly how it operates on C^∞ functions. Let $A(\phi, I)$ be such a function, then

$$A(\phi, I) = \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_N=-\infty}^{\infty} A_{n_1 \cdots n_N} e^{i(n_1\phi_1 + \cdots + n_N\phi_N)}, \tag{28}$$

where $A_{n_1 \cdots n_N}$ is a C^∞ function of the I_n. For simplicity of notation, (28) can be rewritten as

$$A = \sum_n A_n e^{in\cdot\phi}, \tag{29}$$

where $n = (n_1, \cdots, n_N)$ and $n\cdot\phi = n_1\phi_1 + \cdots + n_N\phi_N$. Likewise, the Fourier expansion for V is given by

$$V = \sum_n V_n e^{in\cdot\phi}, \tag{30}$$

with $V_0 = 0$. (For the purposes of this paper, it is no loss of generality to assume $V_0 = 0$, because the V_0 component can be absorbed into H, i.e., $H \to H + \epsilon V_0$.)

The result of operating $D(\ )$ on A is obtained by calculating $\{V, A\}$, converting it to Poisson coordinates using (16), and integrating with respect to T (equivalently θ_1) holding the remaining Poisson coordinates fixed. The resulting expression is given by

$$D(A) = \sum_{n\neq 0} \frac{Z_n}{n\cdot\omega} e^{i\beta_n\cdot\theta}[e^{in\cdot\omega T} - e^{in\cdot\omega T_0}] + Z_0(T - T_0), \tag{31}$$

where

$$n\cdot\omega = n_1\frac{\partial H}{\partial I_1} + \cdots + n_N\frac{\partial H}{\partial I_N}, \tag{32}$$

$$Z_n = \sum_l [V_l \nabla_l \cdot A_{l-n} - A_{l-n} \nabla_{l-n} \cdot V_l], \tag{33}$$

$$\beta_n \cdot \theta \equiv \sum_{\mu=2}^{N} \theta_\mu \sum_{\nu=1}^{N} \frac{\partial F_\mu}{\partial I_\nu} n_\nu, \tag{34}$$

$$\nabla_l = l_1\frac{\partial}{\partial I_1} + \cdots + l_N\frac{\partial}{\partial I_N}, \tag{35}$$

and H is assumed to be nondegenerate.[13] Since the lower limit of integration in (23) causes the mapping of continuous functions into discontinuous functions, one is motivated to consider the following operator (analogous to D):

$$\hat{D}(A) \equiv \sum_{n\neq 0} \frac{Z_n}{n\cdot\omega} e^{in\cdot\phi}, \tag{36}$$

where the Z_0T term is deleted from (31) because the function T is also discontinuous. Assuming (36) converges, $\hat{D}$ maps continuous functions into continuous functions.

A new operator $\hat{R}$ (analogous to R), defined by

$$\hat{R} \equiv 1 - \epsilon\hat{D}, \tag{37}$$

possesses an essential property similar to R, i.e., it maps constants of the motion for $H + \epsilon V$ into constants of the motion for H. More explicitly: If A is a constant of the motion for $H + \epsilon V$, one finds that

$$\hat{R}(A) = A_0. \tag{38}$$

This follows from

$$0 = \{A, H + \epsilon V\} \tag{39}$$

$$= \sum_n (n\cdot\omega A_n - \epsilon Z_n) e^{in\cdot\phi}, \tag{40}$$

which implies

$$n\cdot\omega A_n - \epsilon Z_n = 0. \tag{41}$$

Substitution of (41) into (36) and using (37) yields (38). Since A_0 is a constant of the motion for H, $\hat{R}$ maps constants of the motion of $H + \epsilon V$ into constants of the motion of H. This still holds when H is partially degenerate ($\nabla_m \cdot H = 0$ for fixed m and all $I_1, \cdots, I_N$): i.e, one can show that

$$\hat{R}(A) = \sum_{L=-\infty}^{\infty} A_{Lm} e^{iLm\cdot\phi}. \tag{42}$$

Owing to $\nabla_m \cdot H = 0$, the rhs of (42) is a constant of the motion for H.

Of greater interest is to obtain constants of the motion for $H + \epsilon V$ from those for H. For this, the inverse (if it exists) of $\hat{R}$ must be obtained. The operator $\hat{R}^{-1}$ exists on a family of domains defined by

$$\Omega_L \equiv \{A \mid \hat{D}^L(A) = 0\}, \tag{43}$$

as can be seen from the following considerations. Consider the following expression:

$$\hat{R}_L^{-1} \equiv \sum_{l=1}^{L-1} \epsilon^l \hat{D}^l. \tag{44}$$

The operator $\hat{R}_L^{-1}$ is equal to $\hat{R}^{-1}$ on Ω_L because

$$\hat{R}_L^{-1}\hat{R} = \hat{R}\hat{R}_L^{-1} = \sum_{l=0}^{L-1} \epsilon^l \hat{D}^l - \sum_{l=1}^{L} \epsilon^l \hat{D}^l \tag{45}$$

$$= 1 - \epsilon^L \hat{D}^L, \tag{46}$$

and the rhs of (46) is equal to the identity on Ω_L.

The domains Ω_L are not empty. One can show, for example, that the function

$$f(H) = \sum_{l=0}^{L-1} C_l H^l, \tag{47}$$

is in Ω_L where the C_l are constants. As a consequence, one obtains the unsurprising result that

$$\hat{R}_2^{-1}(H) = H + \epsilon V, \tag{48}$$

i.e, $\hat{R}_2^{-1}$ maps the old Hamiltonian H into the new Hamiltonian, $H + \epsilon V$. Are there other $K \in \Omega_L$ such that $\{K, H\} = 0$, and such that $K \neq f(H)$? If so, then an additional constant of the motion for $H + \epsilon V$ is given by

$$\mathcal{K} = \hat{R}_L^{-1}(K). \tag{49}$$

In the next section of this paper, conditions on H and V are obtained (in special cases) such that Ω_L contains such functions. Families of Hamiltonians are then derived which satisfy these conditions.

V. CONSTRUCTION OF INTEGRABLE HAMILTONIANS: $V(\phi_1, \phi_2, I_1, I_2)$

In two degrees of freedom, only one additional constant of the motion, besides the Hamiltonian, is required for the system to be integrable. Such Hamiltonians can be constructed by demanding that H, V, and K satisfy the following conditions:

$$\hat{D}^L(K) = 0, \tag{50}$$

$$\{K, H\} = 0, \tag{51}$$

where $K \neq f(H)$. Hidden in (50) and (51) are partial differential equations for H, V, and K. When these equations can be solved, new families of integrable Hamiltonians result. This procedure is termed the truncation program of order L because the second constant of the motion (49) is a polynomial in ϵ of order $\leqslant L$. In this section, the $L = 1$ and $L = 2$ cases are considered.

Consider (50) and (51) for $L = 1$ and H nondegenerate. The general solution to (51) is $K = K(I_1, I_2)$. Applying $\hat{D}$ to $K(I_1, I_2)$ and substituting the result into (50) yields

$$0 = \sum_n V_n \frac{\nabla_n \cdot K}{\nabla_n \cdot H} e^{in\cdot\phi}. \tag{52}$$

Setting each Fourier coefficient to zero yields

$$0 = V_n \frac{\nabla_n \cdot K}{\nabla_n \cdot H} \tag{53}$$

for each n. The trivial solution to (53) is $\nabla_n \cdot K = 0$ and $\nabla_m \cdot K = 0$, where n and m are not parallel. In this case, K is the constant function which is of no interest because the constant function is a constant of the motion for all Hamiltonians. The only nontrivial solution is $\nabla_m \cdot K = 0$ for a fixed direction m, with $V_n = 0$ for n nonparallel to m. The resulting family of integrable Hamiltonians is the well-known single resonance Hamiltonians[7] given by

$$\mathcal{K} = H + \epsilon \sum_{N=-\infty}^{\infty} V_N e^{iNm\cdot\phi}, \tag{54}$$

with H and V_N being arbitrary functions of I_1 and I_2, and the second constant of the motion is given by

$$K = K(J_m), \tag{55}$$

where

$$J_m = m_2 I_1 - m_1 I_2. \tag{56}$$

There are no other nontrivial solutions to (52), and therefore the Hamiltonians (54) are the only $L = 1$ integrable Hamiltonians for H nondegenerate.

There exists additional $L = 1$ integrable Hamiltonians, however, for H degenerate $[H = H(J_m)]$. One such family is given by

$$\mathcal{K} = H(J_m) + \epsilon \sum_n V_n(J_m) e^{in\cdot\phi + i\nabla_n\cdot\psi}, \tag{57}$$

with the second invariant given by

$$K = \sin(m\cdot\phi + \nabla_m \cdot \psi), \tag{58}$$

where $V_n(\,)$ is arbitrary and $\psi(I_1, I_2)$ is arbitrary. This family is new yet it has not been derived here. For this reason it does not necessarily exhaust all $L = 1$ integrable Hamiltonians for degenerate H. It is exhibited here only to demonstrate that there are additional classes of $L = 1$ integrable Hamiltonians beyond those of the form of (54).

Next consider Eqs. (50) and (51) for $L = 2$, with H nondegenerate. The equations for resulting from (50) and (51) for $L = 2$, assuming all Fourier terms in (30) are nonzero, are intractable. Nevertheless, the equations which result when V contains only two Fourier components,

$$V = V_n \sin(n \cdot \phi + \delta_n) + V_m \sin(m \cdot \phi + \delta_m), \quad (59)$$

can be solved to obtain the dependences of V_n, V_m, δ_n, and δ_m on I_1 and I_2. Under these assumptions, one obtains

$$\hat{D}(K) = B_n \sin(n \cdot \phi + \delta_n) + B_m \sin(m \cdot \phi + \delta_m), \quad (60)$$

where

$$B_n = V_n \frac{\nabla_n \cdot K}{\nabla_n \cdot H}, \quad (61)$$

$$B_m = V_m \frac{\nabla_m \cdot K}{\nabla_m \cdot H}. \quad (62)$$

Equating Fourier coefficients of the $\hat{D}^2 = 0$ equation equal to zero results in the following partial differential equations (p.d. eqs.):

$$V_n \nabla_n \cdot B_n - B_n \nabla_n \cdot V_n = 0, \quad (63)$$

$$V_m \nabla_m \cdot B_m - B_m \nabla_m \cdot V_m = 0, \quad (64)$$

$$V_n \nabla_n \cdot B_m - B_n \nabla_n \cdot V_m = 0, \quad (65)$$

$$V_m \nabla_m \cdot B_n - B_m \nabla_m \cdot V_n = 0, \quad (66)$$

$$(V_n B_m - V_m B_n)(\nabla_n \cdot \delta_m - \nabla_m \cdot \delta_n) = 0. \quad (67)$$

The solution to (63) and (64) is obtained by rewriting them as

$$V_n B_n \nabla_n \log\left(\frac{B_n}{V_n}\right) = 0, \quad (68)$$

$$V_m B_m \nabla_m \log\left(\frac{B_m}{V_m}\right) = 0. \quad (69)$$

The solution to (68) and (69) is

$$B_n = f_n(J_n) V_n, \quad (70)$$

$$B_m = f_m(J_m) V_m, \quad (71)$$

where $f_n(\,)$ and $f_m(\,)$ are arbitrary functions, J_m is given by (56), and

$$J_n = n_2 I_1 - n_1 I_2. \quad (72)$$

Substitution of (70) and (71) into (65) and (66) and solving for V_n and V_m yields

$$V_n = \frac{h_m(J_m)}{f_n - f_m}, \quad (73)$$

$$V_m = \frac{h_n(J_n)}{f_m - f_n}, \quad (74)$$

where $h_m(\,)$ and $h_n(\,)$ are arbitrary functions. Equation (67) is solved for δ_n and δ_m by observing that it is similar to the divergence equal zero equation in electromagnetism. Thus,

$$\delta_n = \nabla_n \cdot \psi, \quad (75)$$

$$\delta_m = \nabla_m \cdot \psi, \quad (76)$$

where $\psi(I_1, I_2)$ is arbitrary. By substituting (70) and (71) into (61) and (62), the following equations result:

$$f_n \nabla_n \cdot H = \nabla_n \cdot K, \quad (77)$$

$$f_m \nabla_m \cdot H = \nabla_m \cdot K. \quad (78)$$

These equations are solved by rewriting them as

$$\nabla_n \cdot (f_n H - K) = 0, \quad (79)$$

$$\nabla_m \cdot (f_m H - K) = 0. \quad (80)$$

The solution is

$$f_n H - K = g_n(J_n), \quad (81)$$

$$f_m H - K = g_m(J_m), \quad (82)$$

where $g_n(\,)$ and $g_m(\,)$ are arbitrary. Solving (81) and (82) for H and K yields

$$H = \frac{g_n - g_m}{f_n - f_m}, \quad (83)$$

$$K = \frac{f_n g_m - f_m g_n}{f_m - f_n}. \quad (84)$$

The final form for this family of integrable Hamiltonians is given by

$$\mathscr{H} = H + \epsilon[V_n \sin(n \cdot \phi + \delta_n) + V_m \sin(m \cdot \phi + \delta_m)], \quad (85)$$

where H, V_n, V_m, δ_n, and δ_m are given by (83), (73), (74), (75), and (76), respectively, and the second constant of the motion is given by

$$\mathscr{K} = K + \epsilon[B_n \sin(n \cdot \phi + \delta_n) + B_m \sin(m \cdot \phi + \delta_m)], \quad (86)$$

where K, B_n, and B_m are given by (84), (70), and (71), respectively.

The family of Hamiltonians given by (85) does not necessarily exhaust the $L = 2$ integrable Hamiltonians even for H nondegenerate. This family, however, does represent all the $L = 2$ integrable Hamiltonians when V contains only two Fourier components (H nondegenerate). Furthermore, it represents a new family of integrable Hamiltonians which is not previously found in the literature.

VI. POTENTIAL SYSTEMS: $V(q_1, q_2)$

In this section, all integrable Hamiltonians of the form

$$\mathscr{H} = \frac{p_1^2 + p_2^2}{2} + \epsilon V(q_1, q_2), \quad (87)$$

where $H = (p_1^2 + p_2^2)/2$, are considered such that the second constant of the motion is a polynomial in ϵ of finite order[14]:

$$\mathscr{K} = \sum_{l=0}^{L} \epsilon^l K_l, \quad (88)$$

where V and the K_l are assumed to be independent of ϵ (i.e., the perturbation expansion in ϵ truncates). An expression for the K_l is obtained in this section, but the form for some of the potentials is derived in the following section within the context of the Whittaker program of order M. In the Whittaker program (which we generalized from Whittaker's work[2]), integrable Hamiltonians are sought such that the second constant of the motion is a polynomial in p_1 and p_2 of order $\leqslant M$. The purpose of this section is to show, under one further assumption, that (88) is also a polynomial in p_1 and p_2 of order $\leqslant 3L$, and thus occurs in the Whittaker program of order $M = 3L$.

In order that (88) be a constant of the motion for (87), it must satisfy $\{\mathscr{K}, \mathscr{H}\} = 0$, i.e.,

$$0 = \left\{H + \epsilon V, \sum_{l=0}^{L} \epsilon^l K_l\right\}, \quad (89)$$

$$= \sum_{l=0}^{L} \epsilon^l \{H,K_l\} + \sum_{l=1}^{L+1} \epsilon^l \{V,K_{l-1}\}. \tag{90}$$

where V and K_l are assumed to be independent of ϵ, Eq. (90) implies

$$0 = \{H,K_0\}, \tag{91}$$

$$0 = \{V,K_L\}, \tag{92}$$

$$0 = \{H,K_{l+1}\} + \{V,K_l\}, \tag{93}$$

for $0 \leqslant l \leqslant L-1$.

The general expression for function K_l satisfying (93) is obtained by induction on the general solution to (92). Equation (92) is a p.d. eq. for K_L:

$$\frac{\partial V}{\partial q_1}\frac{\partial K_L}{\partial p_1} + \frac{\partial V}{\partial q_2}\frac{\partial K_L}{\partial p_2} = 0, \tag{94}$$

with a general solution given by

$$K_L = B_L(q_1,q_2,\xi), \tag{95}$$

where $B_L(\,,\,)$ is arbitrary, and

$$\xi = p_1 V_2 - p_2 V_1, \tag{96}$$

with $V_1 \equiv \partial V/\partial q_1$, $V_2 \equiv \partial V/\partial q_2$, $V_{12} = \partial^2 V/\partial q_1 \partial q_2$, etc. By induction, Eq. (93) is a p.d. eq. for K_l, given K_{l+1}:

$$V_1 \frac{\partial K_l}{\partial p_1} + V_2 \frac{\partial K_l}{\partial p_2} = \{K_{l+1},H\}. \tag{97}$$

By changing variables from p_1 and p_2 to ξ and η, where

$$\eta = a_1 p_1 + a_2 p_2 \tag{98}$$

(a_1 and a_2 are constants), Eq. (97) becomes

$$\frac{\partial K_l}{\partial \eta} = \frac{\{K_{l+1},H\}}{\lambda}, \tag{99}$$

where $\lambda = a_1 V_1 + a_2 V_2$. Equation (99) is then integrated to obtain the general expression for K_l:

$$K_l = \int^{\eta} d\eta' \left[\frac{\{K_{l+1},H\}}{\lambda}\right]_{q_1,q_2,\xi \sim \text{constant}} + B_l(q_1,q_2,\xi), \tag{100}$$

where $B_l(\,,\,,\,)$ is arbitrary and $0 \leqslant l \leqslant L-1$.

The general expression for functions K_l satisfying (92) and (93) is given by (100) and depends on the $L+1$ arbitrary functions $B_l(\,,\,,\,)$ and the potential V. The final conditions on the B_l and V are obtained by substituting Eq. (100), for $l=0$, into (91).

For arbitrary L, these conditions are complicated and are not considered here except in the $L=0$ case. In this case, the conditions are simple, and the corresponding potentials can be obtained. Consider (91) for $L=0$:

$$\frac{\partial K_0}{\partial q_1} p_1 + \frac{\partial K_0}{\partial q_2} p_2 = 0. \tag{101}$$

Equation (101) is solved by the method of Frobenius by expanding $K_0(q_1,q_2,\xi)$ in a power series:

$$K_0 = \sum_{l=0}^{\infty} b_l \xi^l, \tag{102}$$

where the b_l depend only on q_1,q_2. Substitution of (102) into (101) yields

$$\begin{aligned} 0 = p_1 &\left[\frac{\partial b_0}{\partial q_1} + p_1\left(\frac{\partial b_1}{\partial q_1} V_2 + b_1 V_{12}\right)\right. \\ &\left. + p_2\left(-\frac{\partial b_1}{\partial q_1} V_1 - b_1 V_{11}\right) + O(p^2)\right] \\ + p_2 &\left[\frac{\partial b_0}{\partial q_2} + p_1\left(\frac{\partial b_1}{\partial q_2} V_2 + b_1 V_{22}\right)\right. \\ &\left. + p_2\left(-\frac{\partial b_1}{\partial q_2} V_1 - b_1 V_{12}\right) + O(p^2)\right]. \end{aligned} \tag{103}$$

Equating coefficients of $p_1^l\, p_2^m$ to zero yields

$$\frac{\partial b_0}{\partial q_1} = \frac{\partial b_0}{\partial q_2} = 0, \tag{104}$$

$$\frac{\partial b_1}{\partial q_1} V_2 + b_1 V_{12} = 0, \tag{105}$$

$$\frac{\partial b_1}{\partial q_2} V_1 + b_1 V_{12} = 0, \tag{106}$$

$$\frac{\partial b_1}{\partial q_2} V_2 + b_1 V_{22} - \frac{\partial b_1}{\partial q_1} V_1 - b_1 V_{11} = 0. \tag{107}$$

There are higher-order terms, but they are not displayed because it will be seen that (104)–(107) completely determine V (up to an arbitrary function). Besides, the higher-order equations are satisfied if $b_l = 0$ for $l \geqslant 2$.

Equations (105) and (106) are rewritten as

$$\frac{\partial}{\partial q_1}(b_1 V_2) = 0, \tag{108}$$

$$\frac{\partial}{\partial q_2}(b_1 V_1) = 0, \tag{109}$$

and integrated to obtain

$$b_1 V_2 = f_2(q_2), \tag{110}$$

$$b_1 V_1 = f_1(q_1), \tag{111}$$

where $f_2(\,)$ and $f_1(\,)$ are arbitrary functions. After rewriting (107) as

$$V_2 \frac{\partial}{\partial q_2} \ln(b_1 V_2) - V_1 \frac{\partial}{\partial q_1} \ln(b_1 V_1) = 0, \tag{112}$$

and substituting (110) and (111) into (112), one obtains

$$\frac{V_2}{f_2}\frac{df_2}{dq_2} - \frac{V_1}{f_1}\frac{df_1}{dq_1} = 0. \tag{113}$$

By dividing (110) by (111), a second equation for V is obtained:

$$f_1 V_2 - V_1 f_2 = 0. \tag{114}$$

Equations (113) and (114) are two homogeneous equations for V_1 and V_2. The only nontrivial solution is when the determinant is zero, i.e.,

$$\frac{df_2}{dq_2} - \frac{df_1}{dq_1} = 0. \tag{115}$$

Separation of variables yields the solution

$$f_1 = 2\alpha q_1 + C_1, \tag{116}$$

$$f_2 = 2\alpha q_2 + C_2, \tag{117}$$

where α, C_1, and C_2 are constants. Substitution of (116) and (117) into (114) yields a partial differential equation for V with the general solution given by

$$V = F\left[\alpha(q_1^2 + q_2^2) + C_1 q_1 + C_2 q_2\right], \tag{118}$$

where $F[\,]$ is arbitrary. The functional form for b_1 is obtained from (110):

$$b_1 = \left(\frac{dF}{d\rho}\right)^{-1}, \tag{119}$$

where $\rho = \alpha(q_1^2 + q_2^2) + C_1 q_1 + C_2 q_2$. Thus, the second constant of the motion is given by

$$= b_1 \xi, \tag{120}$$

$$= C_2 p_1 - C_1 p_2 - 2\alpha(q_1 p_2 - q_2 p_1). \tag{121}$$

This family of integrable Hamiltonians ($\alpha \neq 0$) all have cylindrical symmetry about a point in the q plane and represents all the $L = 0$ integrable Hamiltonians.

The $L = 1$ integrable Hamiltonians are tedious to obtain using the above techniques. Consequently, the remainder of this section is concerned not with deriving potentials, but with showing, under a special assumption in the $L \geqslant 1$ case, that the second constant of the motion is a polynomial in p_1 and p_2 (or equivalently in ξ and η) of order $\leqslant 3L$. Not until Sec. VII will the corresponding potentials be derived within the Whittaker program.

It is first shown that the second constant of the motion, given by (88), is a polynomial in η of order $\leqslant 3L$. This is shown using mathematical induction that K_l, given by (100), is a polynomial in η of order $\leqslant 3(L - l)$. The function K_L, given by (95), is independent of η and therefore is a polynomial of order zero in η. The next step in the proof is to assume that the function K_{l+1} [generated inductively from K_L using (100)] is a polynomial in η of order $\leqslant 3(L - l - 1)$:

$$K_{l+1} = \sum_{m=0}^{3(L-l-1)} \alpha_m \eta^m, \tag{122}$$

where $\alpha_m = \alpha_m(q_1, q_2, \xi)$; Eq. (100) becomes

$$K_l = \int^{\eta} d\eta' \frac{1}{\lambda} \sum_{m=0}^{3(L-l-1)} \eta^m \{\alpha_m, H\} + B_l(q_1, q_2, \xi). \tag{123}$$

Computation of $\{\alpha_m, H\}$ in the ξ, η coordinate yields

$$\{\alpha_m, H\} = C_2^m \eta^2 + C_1^m \eta + C_0^m, \tag{124}$$

where

$$C_2^m = \frac{\partial \alpha_m}{\partial \xi} \frac{1}{\lambda_2}\left[(V_1^2 - V_2^2)V_{12} + V_1 V_2 (V_{22} - V_{11})\right], \tag{125}$$

$$C_1^m = \frac{V_1}{\lambda}\frac{\partial \alpha_m}{\partial q_1} + \frac{V_2}{\lambda}\frac{\partial \alpha_m}{\partial q_2} + \frac{\xi}{\lambda^2}\frac{\partial \alpha_m}{\partial \xi} \times \left[2(V_1 a_2 + a_1 V_2)V_{12} + (a_2 V_2 - a_1 V_1)(V_{22} - V_{11})\right], \tag{126}$$

$$C_0^m = \frac{\xi}{\lambda}\left(a_2 \frac{\partial \alpha_m}{\partial q_1} - a_1 \frac{\partial \alpha_m}{\partial q_2}\right) + \frac{\xi^2}{\lambda^2}\frac{\partial \alpha_m}{\partial \xi} \times \left[(a_2^2 - a_1^2)V_{12} - a_1 a_2 (V_{22} - V_{11})\right]. \tag{127}$$

Since the C's are independent of η, Eq. (123) becomes

$$K_l = \frac{1}{\lambda} \sum_{m=0}^{3(L-l-1)} \frac{C_2^m \eta^{m+3}}{m+3} + \frac{C_1^m \eta^{m+2}}{m+2} + \frac{C_0^m \eta^{m+1}}{m+1} + B_l(q_1, q_2, \xi). \tag{128}$$

Equation (128) is a polynomial in η of order $\leqslant 3(L - l)$. As a consequence, the second constant of the motion given by (88) is also a polynomial in η of order $\leqslant 3L$, as was to be shown.

Under one further assumption, it is now shown by induction that (88) is a polynomial in both ξ and η (and therefore p_1 and p_2) of order $\leqslant 3L$. Condition (91) implies that

$$K_0 = K(p_1, p_2, \theta), \tag{129}$$

where $\theta = q_1 p_2 - q_2 p_1$ (= angular momentum) and $K(\,,\,)$ is arbitrary. If one assumes that $K(\,,\,)$ is a polynomial in its arguments (or equivalently in η and ξ) or order $\leqslant 3L$, then K_0 must also be a polynomial in η and ξ or order $\leqslant 3L$. The next step in the proof is to assume that K_l is a polynomial in η and ξ of order $\leqslant 3L - 2l$. It remains to be shown that K_{l+1} is a polynomial in ξ and η of order $\leqslant 3L - 2(l+1) \leqslant 3L$ (for $l > 0$). Inspection of (125)–(128) reveals that C_i^m is a polynomial in ξ or order $\leqslant 3L - m - i - 2l - 1$, and thus α_m is a polynomial in ξ of order $\leqslant 3L - m - 2l$. This and (122) imply that K_{l+1} is a polynomial in η and ξ of order $\leqslant 3L - 2(l+1)$. As a consequence, the second constant of the motion given by (88) is a polynomial in p_1 and p_2 of order $\leqslant 3L$, as was to be shown.

If the assumption that K in (129) is a polynomial in momentum of order $\leqslant 3L$ is relaxed, then it can be shown (in the Appendix) that the potential V must satisfy one or more partial differential equations. In addition to these, V must also satisfy other partial differential equations, as derived in the next section. Should too many restrictions be imposed on V, the only solution is the trivial $V =$ constant solution. To avoid this possibility, the above assumption on K is made.

The purpose of the above analysis is to show the relationship between the integrable Hamiltonians derived within the truncation program and those obtained within the Whittaker program: i.e., under the assumption that K_0 is a polynomial in momentum, both programs generate constants of the motion which are polynomials in the momentum coordinates. This represents a large class of integrable Hamiltonians, and it contains all previously known integrable Hamiltonians (in two degrees of freedom) of the form (87). In the next section, some $L = 1$ integrable Hamiltonians are derived not within the truncation program but within the Whittaker program.

VII. THE WHITTAKER PROGRAM: M = 2,3

In this section, p.d. eqs. are derived which the potential V must satisfy in order that there exists a second constant of the motion which is a polynomial in momentum of order $\leqslant 3$.[15] In addition, some solutions to these equations are exhibited.

Consider first the case of $M = 2$ (second constant of the motion is quadratic in the momentum). The second constant of the motion is assumed to take the form

$$\mathcal{K} = ap_1^2 + bp_2^2 + cp_1 p_2 + ep_1 + fp_2 + g, \tag{130}$$

where a, b, c, e, f, and g depend only on q_1 and q_2. When the coefficients of $p_1^n p_2^m$ in the equation $\{\mathcal{K}, \mathcal{H}\} = 0$ are set equal to zero, the following partial differential equations result:

$$\frac{\partial a}{\partial q_1} = 0, \tag{131}$$

$$\frac{\partial b}{\partial q_2} = 0, \tag{132}$$

$$\frac{\partial b}{\partial q_1} + \frac{\partial c}{\partial q_2} = 0, \tag{133}$$

$$\frac{\partial c}{\partial q_1} + \frac{\partial a}{\partial q_2} = 0, \tag{134}$$

$$\frac{\partial e}{\partial q_1} = 0, \tag{135}$$

$$\frac{\partial f}{\partial q_2} = 0, \tag{136}$$

$$\frac{\partial f}{\partial q_1} + \frac{\partial e}{\partial q_2} = 0, \tag{137}$$

$$\frac{1}{\epsilon}\frac{\partial g}{\partial q_1} - 2aV_1 - cV_2 = 0, \tag{138}$$

$$\frac{1}{\epsilon}\frac{\partial g}{\partial q_2} - 2bV_2 - cV_1 = 0, \tag{139}$$

$$eV_1 + fV_2 = 0. \tag{140}$$

Assume $e = f = 0$; otherwise the potential V resulting from solving (135)–(137), (140) is identical to (118). Equations (131)–(134) have the general solution

$$a = a_0 + a_1 q_2 + a_2 q_2^2, \tag{141}$$

$$b = b_0 + b_1 q_1 + a_2 q_1^2, \tag{142}$$

$$c = c_0 - b_1 q_2 - a_1 q_1 - 2a_2 q_1 q_2, \tag{143}$$

where a_0, a_1, a_2, b_0, b_1, and c_0 are constants. Application of integrability conditions to (138) and (139) yield the following p.d. eq. for V:

$$\frac{\partial}{\partial q_2}(2aV_1 + cV_2) + \frac{\partial}{\partial q_1}(2bV_2 + cV_1). \tag{144}$$

When the partial derivatives in (144) are carried out, the following second order p.d. eq. for V results:

$$c\left(\frac{\partial^2 V}{\partial q_1^2} - \frac{\partial^2 V}{\partial q_2^2}\right) + 2(b-a)\frac{\partial^2 V}{\partial q_1 \partial q_2} + \left(\frac{\partial c}{\partial q_1} - 2\frac{\partial a}{\partial q_2}\right)\frac{\partial V}{\partial q_1} + \left(2\frac{\partial b}{\partial q_1} - \frac{\partial c}{\partial q_2}\right)\frac{\partial V}{\partial q_2} = 0. \tag{145}$$

The function g is obtained by solving (145) for V, substituting it into (138) and (139), and integrating. The second constant of the motion is then given by (130).

Whittaker[2] obtained the general solution to (145) and the second constant of the motion. More useful, however, is that all separable potentials in two dimensions satisfy (145). This follows[4] from the fact that if the Hamilton–Jacobi equation separates (separates after a coordinate transformation in the q plane) a constant of the motion exists which is second order in p_1 and p_2. A necessary condition then, for a potential to be separable, is if the constants in (141)–(143) can be chosen so that V satisfies (145). Equation (145) provides a check of the separability of potentials, derived in the Whittaker program of order 3.

The p.d. eqs. satisfied by the $M = 3$ potentials is derived similarly to (148). The second constant of the motion is assumed to take the form

$$\mathcal{K} = Ap_1^3 + Bp_1^2 p_2 + Cp_1 p_2^2 + Ep_2^3 + Fp_1 + Gp_2, \tag{146}$$

where A, B, C, E, F, and G depend only on q_1 and q_2. The even powers of p_1 and p_2 are absent from (146) because the equations resulting from these are identical to (131)–(140), and solving them would result in (145) independent of the cubic terms in (146). Equating the coefficients of $p_1^n p_2^m$ in the $\{\mathcal{K}, \mathcal{H}\} = 0$ equation to zero results in the following p.d. eqs:

$$\frac{\partial A}{\partial q_1} = 0, \tag{147}$$

$$\frac{\partial E}{\partial q_2} = 0, \tag{148}$$

$$\frac{\partial B}{\partial q_1} + \frac{\partial A}{\partial q_2} = 0, \tag{149}$$

$$\frac{\partial C}{\partial q_1} + \frac{\partial B}{\partial q_2} = 0, \tag{150}$$

$$\frac{\partial E}{\partial q_1} + \frac{\partial C}{\partial q_2} = 0, \tag{151}$$

$$\frac{1}{\epsilon}\frac{\partial F}{\partial q_1} - 3AV_1 - BV_2 = 0, \tag{152}$$

$$\frac{1}{\epsilon}\frac{\partial G}{\partial q_2} - 3EV_2 - CV_1 = 0, \tag{153}$$

$$\frac{1}{\epsilon}\left(\frac{\partial G}{\partial q_1} + \frac{\partial F}{\partial q_2}\right) - 2BV_1 - 2CV_2 = 0, \tag{154}$$

$$FV_1 + GV_2 = 0. \tag{155}$$

Equations (147)–(151) have a polynomial solution:

$$A = A_0 + A_1 q_2 + A_2 q_2^2 + A_3 q_2^3, \tag{156}$$

$$B = B_0 + B_1 q_2 + E_2 q_2^2 - A_1 q_1 - 2A_2 q_1 q_2 - 3A_3 q_2^2 q_1, \tag{157}$$

$$C = C_0 - B_1 q_1 + A_2 q_1^2 - E_1 q_2 - 2E_2 q_1 q_2 + 3A_3 q_2 q_1^2, \tag{158}$$

$$E = E_0 + E_1 q_1 + E_2 q_1^2 - A_3 q_1^3, \tag{159}$$

where A_0, A_1, A_2, E_0, E_1, E_2, B_0, C_0, B_1, and A_3 are constants. Equation (155) is solved by defining a function Z:

$$F = \epsilon Z V_2, \tag{160}$$

$$G = -\epsilon Z V_1. \tag{161}$$

Substitution of (160) and (161) into (152)–(154) yields

$$ZV_{12} + \frac{\partial Z}{\partial q_1}V_2 - 3AV_1 - BV_2 = 0, \tag{162}$$

$$-ZV_{12} - \frac{\partial Z}{\partial q_2}V_1 - 3EV_2 - CV_1 = 0, \tag{163}$$

$$Z(V_{22} - V_{11}) + \frac{\partial Z}{\partial q_2}V_2 - \frac{\partial Z}{\partial q_1}V_1 - 2BV_1 - 2CV_2 = 0. \tag{164}$$

Adding (162) and (157) yields

$$V_1\left(3A + C + \frac{\partial Z}{\partial q_2}\right) + V_2\left(3E + B - \frac{\partial Z}{\partial q_1}\right) = 0. \tag{165}$$

This equation can be solved for Z by first introducing a new function Y such that

$$\frac{\partial Y}{\partial q_2} = -3A - C, \tag{166}$$

$$\frac{\partial Y}{\partial q_1} = 3E + B. \tag{167}$$

Such a function exists by virtue of (147)–(151), and integration of (166) and (167) yields

$$\begin{aligned} Y = Z_0 &+ (3E_0 + B_0)q_1 + \frac{(3E_1 - A_1)}{2} q_1^2 + E_2 q_1^3 \\ &- (3A_0 + C_0)q_2 - \frac{(3A_1 - E_1)}{2} q_2^2 - A_2 q_2^3 \\ &- A_2 q_1^2 q_2 + E_2 q_1 q_2^2 + B_1 q_1 q_2 - \frac{3A_3}{4}(q_1^2 + q_2^2)^2, \end{aligned} \tag{168}$$

where Z_0 is a constant. In terms of Y and V, the general solution to (165) is

$$Z = Y + \Phi(V), \tag{169}$$

where $\Phi(\,)$ is arbitrary. Substitution of (169) into (162) and (164) yields *two* nonlinear p.d. eqs. which V must satisfy:

$$YV_{21} + 3EV_2 - 3AV_1 = -\frac{\partial^2 \psi(V)}{\partial q_1 \partial q_2}, \tag{170}$$

$$\begin{aligned} Y(V_{22} - V_{11}) - 3(A + C)V_2 - 3(E + B)V_1 \\ = \left(\frac{\partial^2}{\partial q_1^2} - \frac{\partial^2}{\partial q_2^2}\right)\psi(V), \end{aligned} \tag{171}$$

where $d\psi/dV \equiv \Phi(V)$.

For each choice of $\psi(V)$, Eqs. (170) and (171) are the p.d. eqs. which V must satisfy in order that there exists a second constant of the motion which is cubic in the momentum. Unlike the $M = 2$ case, there is no general solution available. Nevertheless, a few particular solutions can be exhibited. The Toda Lattice,[5]

$$V = \alpha_+ E^{q_2 + \sqrt{3}q_1} + \alpha_- e^{q_2 - \sqrt{3}q_1} + \beta e^{-2q_2}, \tag{172}$$

is a solution when all the constants in (156)–(159), (168), and ψ are set to zero except $A_0 = 1$, $C_0 = -3$, $Z_0 = 3$. This potential is nonseparable, as can be verified by showing that it does not satisfy (145).

A new nonseparable potential satisfying (170) and (171) is

$$V = (q_1^2 + \delta)q_2^{-2/3} + \tfrac{3}{4}q_2^{4/3}, \tag{173}$$

when again $\psi = 0$ and all constants are zero except $A_0 = 1$, $C_0 = \frac{3}{2}$. This potential is nonseparable, as can be checked by showing that it does not satisfy (145). The second constant of the motion is given by

$$\begin{aligned} \mathcal{K} = 2p_1^3 &+ 3p_1 p_2^2 + 3\epsilon p_1[2(q_1^2 + \delta)q_2^{-2/3} \\ &- 3q_2^{4/3}] + 18\epsilon p_2 q_1 q_2^{1/3}. \end{aligned} \tag{174}$$

Not all potentials which satisfy (170) and (171) are nonseparable. In the case of all cosntants and $\psi = 0$ except $B_0 = 1$, the potential is given by

$$V = q_1^2 + 4q_2^2 + \delta q_1^{-2}, \tag{175}$$

and the second constant of the motion is given by

$$\mathcal{K} = p_1^2 p_2 + 2\epsilon[4q_1 q_2 p_1 - p_2 q_1^2 + p_2 \delta q_1^{-2}]. \tag{176}$$

The Hamiltonian constructed from this potential has a constant of the motion which is cubic in the momentum, and it must also possess a constant of the motion which is quadradic in the momentum ($p_2^2 + 8\epsilon q_2^2$, for example) due to the Hamilton–Jacobi equation separating in a Cartesian coordinate system.

The Hamiltonians occurring in the Whittaker program of order 2 and 3 are identical to the $L = 1$ Hamiltonians considered in Sec. VI. This follows because all the second constants of motion occurring in the Whittaker program of order $M = 2, 3$ [(130) and (140)] are first order in ϵ [see (138), (139), (160), and (161)], and all the second constants of the motion in the $L = 1$ case in Sec. VI are polynomials in p_1 and p_2 of order $\leqslant 3$.

Potentials (172), (173), and (175) by no means exhaust the $L = 1$ potentials. Solving (170) and (171) for other choices of the constants in (156)–(159) is a method for obtaining novel integrable potentials. Potential (173) is, in fact, such a potential.

ACKNOWLEDGMENTS

I wish to thank William P. Reinhardt for helpful conversations, and the University of Colorado for generous fellowships. This work was supported in part by National Science Foundation grants CHE80-11442 and PHY79-04928, to the University of Colorado.

APPENDIX

In this appendix, it is shown that if K in (129), for $L = 1$, is a polynomial in its arguments of order $\geqslant 4$, then V must satisfy a subsidiary (i.e., in addition to those in the text) partial differential equation(s). Recall the form for $K_0 = K$,

$$K = B_0 + b_1\eta + b_2\eta^2 + b_2\eta^3, \tag{A1}$$

where B_0, b_1, b_2, b_3 depend only on ξ, q_1, and q_2. Since $K(p_1, p_2, q_1 p_2 - q_2 p_1)$ has no terms in η of order > 3, one obtains

$$\left(\frac{\partial}{\partial \eta}\right)^4_{q_1, q_2, \xi \sim \text{ const}} K = 0. \tag{A2}$$

By changing from the coordinates η and ξ to p_1 and p_2, Eq. (A2) becomes

$$\begin{aligned} \Bigg[V_1 \left|\frac{\partial}{\partial p_1}\right|_{p_2, \theta \sim \text{ const}} &+ V_2 \left|\frac{\partial}{\partial p_2}\right|_{p_1, \theta \sim \text{ const}} \\ &+ (q_1 V_2 - q_2 V_1) \left|\frac{\partial}{\partial \theta}\right|_{p_1, p_2 \sim \text{ const}} \Bigg]^4 \cdot K = 0, \end{aligned} \tag{A3}$$

where $\theta = q_1 p_2 - q_2 p_1$.

When (A3) is evaluated at $p_1 = p_2 = 0$, a p.d. eq. for V results. Take, for example, the case of K independent of θ. Then (A3) becomes

$$AV_1^4 + 4BV_1^3 V_2 + 6CV_1^2 V_2^2 + 4EV_1 V_2^3 + GV_2^4 = 0, \tag{A4}$$

where $A = \partial^4 K/\partial p_1^4|_{p_1 = p_2 = 0}$, $B = \partial^4 K/\partial p_1^3 \partial p_1|_{p_1 = p_2 = 0}$, etc. Equation (A4) is a p.d. eq. for V with a general solution given by

$$V = f(\alpha q_1 + \beta q_2), \tag{A5}$$

where $f(\,)$ is arbitrary and α, β satisfy

$$0 = A\alpha^4 + 4\beta\alpha^3\beta + 6C\alpha^2B^2 + 4E\alpha\beta^3 + G\beta^4. \quad \text{(A6)}$$

For the more general case, $K(p_1, p_2, \theta)$, Eq. (A3) is a p.d. eq. in V of the form

$$F(V_1, V_2, q_1 V_2 - q_2 V_1) = 0, \quad \text{(A7)}$$

where $F(\ ,\ ,\ )$ is a fourth-order polynomial in its arguments.

If K is assumed to be a polynomial in its arguments of order $\geqslant 5$, then the condition

$$\left(\frac{\partial}{\partial\eta}\right)^5 K = 0 \quad \text{(A8)}$$

results in a second p.d. eq. for V of the form (A7). By induction, one concludes that if K is a polynomial in its arguments of order $N \geqslant 4$, then V must satisfy $N-3$ subsidiary p.d. eqs. Furthermore, these arguments can be generalized for the $L > 1$ case of (129), i.e., if K is a polynomial in its arguments of order $N \geqslant 3L+1$, then V must satisfy $N-3L$ subsidiary p.d. eqs.

[1]See, for example, J. Moser, *Stable and Random Motions in Dynamical Systems* (Princeton U. P., Princeton, N.J., 1973).

[2]E. T. Whittaker, *Analytical Dynamics* (Cambridge U. P., London, 1927), pp. 332–335.

[3]M. J. Drach, Compt. Rend. **200**, 22 (1935).

[4]P. Havas, J. Math. Phys. **16**, 1461 (1975); **16**, 2476 (1975).

[5]M. Toda, J. Phys. Soc. Jpn. **23**, 501 (1967); J. Ford, S. P. Stoddard, and J. S. Turner, Prog. Theor. Phys. **50**, 1547 (1973).

[6]F. Calogero, J. Math. Phys. **12**, 419 (1971); **15**, 1425 (1975); J. Moser, Adv. Math. **16**, 197 (1975); M. Adler, Commun. Math. Phys. **55**, 195 (1977).

[7]This is the case of cyclic coordinates: H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950), p. 220.

[8]B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2024 (1967); **9**, 1365 (1968); J. Lacina, Ann. Phys. (N. Y.) **51**, 381 (1969).

[9]See, for example, W. Thirring, *Classical Dynamical Systems* (Springer, New York, 1978), pp. 92–108.

[10]N. Mukunda, Phys. Rev. **155**, 1383 (1967).

[11]Reference 7, pp. 237–247.

[12]See, for example, V. I. Smirnov, *A Course in Higher Mathematics* (Pergamon, Oxford, 1964), Vol. IV.

[13]In this paper, a Hamiltonian is termed nondegenerate if $\nabla_m \cdot H \neq 0$ almost everywhere in the I plane for any choice of the integer valued vector m. A partially degenerate Hamiltonian satisfies $\nabla_m \cdot H = 0$ everywhere in the I plane for a fixed integer valued vector m.

[14]A treatment of this problem using the D operator is found in C. R. Holt, "Construction of Integrable Hamiltonians & Multidimensional Quantum Tunneling" (Ph.D. Dissertation, University of Colorado, Boulder, 1981).

[15]L. S. Hall (Lawrence Livermore Laboratory internal document UCID-18980/REV-I) has generalized the Whittaker program of order M and has obtained an expression for the constant of the motion (when it exists). His results are more general than those in this section because (1) he takes into account the variational dependence of p_1 and p_2 due to conservation of energy, and (2) he includes the effect of a magnetic field. (The author learned of Hall's work after the present work was completed.)

Dyadic Green functions for the time-dependent wave equation

Egon Marx and Daniel Maystre[a)]
National Bureau of Standards, Washington, D. C. 20234

(Received 18 September 1981; accepted for publication 18 December 1981)

The theory of dyadic Green functions for a transient electromagnetic field, which obeys the vector wave equation, is presented within the framework of the theory of distributions. First, the elementary solution of the scalar wave equation is derived, and then it is used to find the general solution of that equation. After establishing the equivalence between Maxwell's equations and the time-dependent vector wave equation, the dyadic elementary solution is derived and applied to solve the equation. Further properties of dyadic Green functions for the wave equation are derived within the heuristic approach to the theory of Green's functions. The paper includes a collection of formulas from the theory of distributions intended to help readers who are not familiar with the subject.

PACS numbers: 03.50.De, 03.40.Kf, 02.30. + g

I. INTRODUCTION

Transient electromagnetic fields obey a set of partial differential equations, Maxwell's equations, that can be reduced to the vector wave equation.

Green functions are often used in the solution of linear partial differential equations such as the wave equation. These functions correspond to solutions of differential equations with impulsive sources; sometimes they satisfy homogeneous boundary conditions and have causality or the radiation condition built into them. The source (a Dirac delta function) and some Green functions are not functions in the usual sense of the word; they are generalized functions or distributions. Delta functions were used in physics before they received a firm mathematical foundation,[1,2] and they continue to be used in a heuristic manner for most applications. Such a situation can easily lead to ambiguous or incorrect results, and a mathematically well-defined theory such as the theory of distributions should be used whenever practicable; unfortunately, many scientists and engineers are unfamiliar with the power and elegance of the theory of distributions. An excellent example of its effective use can be found in the theory of gratings.[3]

One of us[4] used the heuristic approach to derive integral equations for transient electromagnetic fields. The shortcomings of such a method were particularly evident in the determination of the free-space dyadic Green function. One of the aims of this paper is to present a proper derivation of this Green function.

Although Green functions for the scalar wave equation can be used to write integral equations for the electric and magnetic vector fields, dyadic Green functions provide a more natural way to relate a vector field to its vector sources and offer greater flexibility in solving a vector wave equation. Dyadic Green functions for the vector Helmholtz equations are developed in detail by Tai[5]; the free-space Green function for this equation is expressed in terms of the elementary solution of the scalar Helmholtz equation, which is a locally integrable function, and its derivatives. The corresponding treatment of the time-dependent equation involves derivatives of distributions. Generalized functions are also used to define the derivatives of the Green function for the scalar Helmholtz equation.[6]

We propose to introduce the basic concepts used in the computation of Green functions via the theory of distributions by presenting the analysis of the scalar wave equation in Sec. II. In Sec. III we briefly discuss the equivalence of Maxwell's equations and the vector wave equation, and then introduce dyadic Green functions. We find the elementary solution of the vector wave equation, which is essentially the free-space dyadic Green function, and show how it is applied to solve that equation.

To assist the reader who is not familiar with the theory of distributions, we collect in Appendix A some of the most important definitions and properties of distributions that we use in this paper.

There are some aspects of the heuristic approach to Green functions, such as homogeneous boundary conditions, that are not easily expressed in the language of distributions. Nevertheless, when a Green function that obeys homogeneous boundary conditions can be found, the solution of the differential equations is reduced to integrals over known functions. These Green functions and their symmetry properties are discussed in Appendix B.

We hope this paper will encourage other scientists and engineers to make wider use of a powerful tool: generalized functions or distributions.

II. THE SCALAR WAVE EQUATION

There are many kinds of phenomena in acoustics, electromagnetism, elasticity, and other branches of physics that are described by a function of space and time ψ that obeys the partial differential equation

$$\Box\psi(x) = \alpha(x), \tag{1}$$

where x stands for the four-vector $(ct,\mathbf{x})$, $\alpha(x)$ is the appropriate source term, and the d'Alembertian is defined by

$$\Box = \partial^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2. \tag{2}$$

In this paper we focus our attention on the electromagnetic

[a)]On leave from Laboratoire d'Optique Electromagnetique, Faculté des Sciences et Techniques de St. Jérôme, Rue H. Poincaré, 13397 Marseille, France.

field; applications to other fields require only minor changes.

The solutions of the wave equation can be found with the help of retarded Green functions $G_R(x,x')$. They satisfy

$$\Box\, G_R(x,x') = \delta^{(4)}(x-x'), \tag{3}$$

$$G_R(x,x') = 0, \quad t<t', \tag{4}$$

where the source in Eq. (3) is the four-dimensional Dirac delta distribution and Eq. (4) expresses the physical condition of causality. Since the source of the Green function is a distribution, we should use the theory of distributions to find such a function, which actually may itself be a distribution.

Closely related to the free-space Green function is the elementary solution of Eq. (1), which satisfies, in the sense of distributions as defined in Appendix A,

$$\Box\, \mathscr{G}(x) = \delta^{(4)}(x), \tag{5}$$

$$\mathscr{G}(x) = 0, \quad t<0, \tag{6}$$

and we can find the free-space Green function from

$$G_R^{(0)}(x,x') = \mathscr{G}(x-x'). \tag{7}$$

To find $\mathscr{G}$, we assume that it is a tempered distribution and find its Fourier transform $\mathscr{K}$. Equations (A43) and (A44) allow us to reduce Eq. (5) to

$$(-\omega^2/c^2 + \mathbf{k}^2)\mathscr{K}(\mathbf{k},\omega) = (2\pi)^{-2}. \tag{8}$$

To solve for $\mathscr{G}$, we have to specify which reciprocal of $\mathbf{k}^2 - \omega^2/c^2$ we must choose. The causality condition (6), when compared to Eq. (A51), indicates that we should add a small positive imaginary part to ω, and set

$$\mathscr{K}(\mathbf{k},\omega) = (2\pi)^{-2}\,[\mathbf{k}^2 - (\omega+i\epsilon)^2/c^2]^{-1}, \tag{9}$$

if we use $\mathscr{F}_-$ for the inverse Fourier transform of the time variable; it is understood that ϵ will tend to zero after the appropriate integrations are performed, a limit that is well defined in the theory of distributions. In the complex ω plane, the poles at $\omega = \pm|\mathbf{k}|c$ are moved slightly below the real axis and, since we are using $e^{-i\omega t}$ in the inverse Fourier transform, the contour can be closed around the upper half-plane for $t<0$ without changing the integral, which vanishes because no singularities are enclosed in the contour.

Actually, we know that $\mathscr{G}(x)$ is not a function, and, when the inverse Fourier transform is calculated by integrations, we have to use integrals that are mathematically ill defined to obtain delta distributions. For this reason, and because we can use the result in the computation of the dyadic elementary solution, we first derive the inverse Fourier transform of

$$K(\mathbf{k},\omega) = (\omega+i\epsilon)^{-2}[k^2 - (\omega+i\epsilon)^2/c^2]^{-1}$$

$$= k^{-2}(\omega+i\epsilon)^{-2} - k^{-2}[(\omega+i\epsilon)^2 - k^2c^2]^{-1}, \tag{10}$$

where $k = |\mathbf{k}|$; the transform of K is a function. Then $\mathscr{G}(x)$ can be obtained by differentiation, and the derivatives always exist for distributions. By convention we change the sign of the space part in the exponential, and compute the inverse Fourier transform

$$F(\mathbf{x},t) = (2\pi)^{-2}\int d^3k\, d\omega\, \exp[-i(\omega t - \mathbf{k}\cdot\mathbf{x})]\, K(\mathbf{k},\omega). \tag{11}$$

Equation (A52) gives

$$\int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}(\omega+i\epsilon)^{-2} = -2\pi t\theta(t), \tag{12}$$

where $\theta(t)$ is the unit step function. To find the inverse transform of k^{-2}, we use Eq. (A55) to derive

$$\int_{-\infty}^{\infty} d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}}\, k^{-2} = 4\pi r^{-1}\int_0^{\infty} dk\, \sin(kr)k^{-1}, \tag{13}$$

where $r = |\mathbf{x}|$. The integrand of the convergent integral on the right-hand side of Eq. (13) is an even function of k, so that we do not change the value of the integral if we extend the range of integration to $-\infty$ and divide by 2. The integrand is not singular at the origin, and so we can deform the contour or add a small negative imaginary part to k; then Eq. (A51) gives

$$\int_{-\infty}^{\infty} dk\, e^{ikr}(k-i\epsilon)^{-1} = 2\pi i\theta(r) \tag{14}$$

and Eq. (13) becomes

$$\int_{-\infty}^{\infty} d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}}\, k^{-2} = 2\pi^2 r^{-1}[\theta(r) - \theta(-r)] = 2\pi^2 r^{-1}. \tag{15}$$

Similarly, we compute

$$\int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}\,[(\omega+i\epsilon)^2 - k^2c^2]^{-1}$$

$$= -2\pi\sin(kct)(kc)^{-1}\theta(t), \tag{16}$$

$$\int_{-\infty}^{\infty} d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}}\sin(kct)k^{-3}$$

$$= 4\pi r^{-1}\int_0^{\infty} dk\, \sin(kct)\sin(kr)k^{-2}, \tag{17}$$

$$\int_{-\infty}^{\infty} dk\, \cos(ku)(k-i\epsilon)^{-2} = -\pi|u|, \tag{18}$$

$$\int_{-\infty}^{\infty} dk\, \sin(kct)\sin(kr)k^{-2} = \tfrac{1}{2}\pi(|ct+r| - |ct-r|), \tag{19}$$

whence, collecting all the necessary terms, we find

$$F(\mathbf{x},t) = \tfrac{1}{2}\pi\,\theta(t)r^{-1}(|t+r/c| - |t-r/c| - 2t), \tag{20}$$

which can be rewritten as

$$F(\mathbf{x},t) = -\pi r^{-1}(t-r/c)\,\theta(t-r/c). \tag{21}$$

By Eqs. (9) and (A43), we can set

$$\mathscr{G} = -(2\pi)^{-2}\frac{\partial^2 F}{\partial t^2}, \tag{22}$$

and we have

$$\frac{\partial F}{\partial t} = -\pi r^{-1}\,\theta\left(t-\frac{r}{c}\right), \tag{23}$$

$$\frac{\partial^2 F}{\partial t^2} = -\pi r^{-1}\delta\left(t-\frac{r}{c}\right), \tag{24}$$

where we have used Eq. (A14). Thus, the elementary solution of the wave equation is

$$\mathscr{G}(x) = (4\pi r)^{-1}\,\delta(t-r/c). \tag{25}$$

We note that $1/r$ corresponds to an integrable singularity, and we can rigorously define this distribution by its value on

a test function ϕ,

$$\langle \mathcal{G},\phi \rangle = \int_0^\infty dt\,(4\pi t)^{-1} \oint_{S(ct)} dS\,\phi(\mathbf{x},t), \tag{26}$$

where we use $S(a)$ to denote the sphere of radius a centered at the origin. The support of this distribution, as defined in the Appendix A, is the future light cone given by

$$c^2t^2 - \mathbf{x}^2 = 0, \quad t \geqslant 0. \tag{27}$$

In Ref. 7 we find an elegant direct proof that the distribution $\mathcal{G}$ defined by Eq. (26) actually satisfies Eq. (5). Since the integration is only over positive values of t, $\langle \mathcal{G},\phi \rangle$ vanishes whenever the support of ϕ lies in the region $t<0$, so that the distribution $\mathcal{G}$ vanishes for $t<0$. Also in Ref. 7 we find the solution of the Cauchy initial value problem for the homogeneous wave equation.

If the source $\alpha(x)$ is given for all space and time, we can find the solution of Eq. (1) by a convolution product as defined in Eq. (A32). We write

$$\psi = \mathcal{G} * \check{\alpha}, \tag{28}$$

where we use the inverted caret to indicate that the source is a distribution that can include singular terms. We use Eqs. (A39), (5), and (A36) to verify that ψ satisfies the wave equation,

$$\Box\psi = (\Box\,\mathcal{G}) * \check{\alpha} = \delta * \check{\alpha} = \check{\alpha}. \tag{29}$$

The convolution product is well defined when the support of α is bounded below, but this is not a necessary condition; we assume that α is a locally integrable function that vanishes sufficiently rapidly when $t \to -\infty$ to define the integrals in what follows. We use the definitions (26) and (A32) to write

$$\langle \mathcal{G} * \alpha,\phi \rangle = \langle \mathcal{G}(\xi), \langle \alpha(\xi'),\phi(\xi+\xi') \rangle \rangle, \tag{30}$$

$$\langle \mathcal{G} * \alpha,\phi \rangle = \int_0^\infty \frac{d\tau}{4\pi\tau} \oint_{S(c\tau)} d\sigma \int_{-\infty}^\infty d^4\xi'\,\alpha(\xi')\,\phi(\xi+\xi'), \tag{31}$$

where $\xi = (c\tau,\boldsymbol{\xi})$, $\xi' = (c\tau',\boldsymbol{\xi}')$, and $d\sigma$ is the surface element in $\boldsymbol{\xi}$-space. We change the variable ξ' to $x-\xi$ and keep ξ to obtain

$$\langle \mathcal{G} * \alpha,\phi \rangle = \int_0^\infty \frac{d\tau}{4\pi\tau} \oint_{S(c\tau)} d\sigma \int_{-\infty}^\infty d^4x\,\alpha(x-\xi)\,\phi(x) \tag{32}$$

We now change the order of integration and replace ξ by $x-x'$, keeping x, and find

$$\langle \mathcal{G} * \alpha,\phi \rangle = \int_{-\infty}^\infty d^4x \left[\int_{-\infty}^t \frac{dt'}{4\pi(t-t')} \oint_{S\{\mathbf{x},c(t-t')\}} dS'\,\alpha(x') \right] \phi(x), \tag{33}$$

where $S(\mathbf{x},a)$ represents the sphere of radius a centered at $\mathbf{x}$. Consequently we can write Eq. (28) as

$$\psi(x) = \int_{-\infty}^t \frac{dt'}{4\pi(t-t')} \oint_{S\{\mathbf{x},c(t-t')\}} dS'\,\alpha(x'). \tag{34}$$

When the unknown field ψ is defined only in a region of space V bounded by a surface S, and, when the sources are given starting at a time t_0, we also have to specify the initial values of the field and its time derivative, and either the field or the normal derivative on S. In this case, we can extend the function ψ by assuming that it is zero outside V and for times before t_0. Then the given values become jumps in the function and its derivatives, and we use Eqs. (A14) and (A21) to write, if the normal $\hat{n}$ points out of V,

$$\begin{aligned}\Box\psi = \{\Box\psi\} &+ c^{-2}\,\psi(\mathbf{x},t_0)\,\delta'(t-t_0) \\ &+ \frac{1}{c^2}\frac{\partial\psi(\mathbf{x},t_0)}{\partial t}\,\delta(t-t_0) \\ &+ \nabla\cdot[\hat{n}\psi\delta(S)] + \frac{\partial\psi}{\partial n}\,\delta(S) = \check{\alpha},\end{aligned} \tag{35}$$

where the curly brackets indicate derivatives in the sense of functions. From Eq. (28) we can derive

$$\begin{aligned}\psi = \mathcal{G} * \alpha &+ \frac{1}{c^2}\frac{\partial\mathcal{G}}{\partial t} * \psi(\mathbf{x},t_0)\,\delta(t-t_0) \\ &+ \frac{1}{c^2}\mathcal{G} * \frac{\partial\psi(\mathbf{x},t_0)}{\partial t}\,\delta(t-t_0) \\ &+ \frac{\partial\mathcal{G}}{\partial n} * \psi\delta(S) + \mathcal{G} * \frac{\partial\psi}{\partial n}\,\delta(S).\end{aligned} \tag{36}$$

which is equivalent to the well-known formula

$$\begin{aligned}\psi(x) = &\int_{t_0}^t dt' \int_V dV'\,\alpha(x')\,G_R^{(0)}(x,x') \\ &- \frac{1}{c^2}\int_V dV' \left[\psi(x') \frac{\partial G_R^{(0)}(x,x')}{\partial t'} - \frac{\partial\psi(x')}{\partial t'}\,G_R^{(0)}(x,x') \right]_{t'=t_0} \\ &- \int_{t_0}^t dt' \oint_S dS' \left[\psi(x') \frac{\partial G_R^{(0)}(x,x')}{\partial n'} - \frac{\partial\psi(x')}{\partial n'}\,G_R^{(0)}(x,x') \right].\end{aligned} \tag{37}$$

Equation (36) or (37) can be used to compute ψ by integrations only if ψ and $\partial\psi/\partial n$ are both known on S. Since only one of these boundary values is required to determine ψ, Eq. (37) reduces to an integral equation for the function on S that is not given when we let $\mathbf{x}$ approach the surface from the outside, where ψ vanishes. If two arbitrary functions are given on S for ψ and $\partial\psi/\partial n$, and Eq. (37) is used to compute $\psi(x)$, this function does not vanish outside V unless the given functions satisfy the above-mentioned integral equation; the jumps in ψ and $\partial\psi/\partial n$ will be as given, but the boundary conditions will not be satisfied.

III. MAXWELL'S EQUATIONS AND DYADIC GREEN FUNCTIONS

The free-space Maxwell equations for the electromagnetic fields **E** and **B** are

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\epsilon_0}, \tag{38}$$

$$\nabla\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}, \tag{39}$$

$$\nabla\cdot\mathbf{B} = 0, \tag{40}$$

$$\nabla\times\mathbf{B} = \mu_0\mathbf{j} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}, \tag{41}$$

where the sources are the current density **j** and the charge density ρ, and ϵ_0 and μ_0 are the permittivity and permeability of the vacuum which are related to the speed of light by $\epsilon_0\mu_0 = c^{-2}$. These equations are consistent only if the equation that expresses charge conservation,

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{j} = 0, \tag{42}$$

is satisfied. Consequently, the charge density has to be given only at the initial time t_0, and we can find it at other times from

$$\rho(\mathbf{x},t) = \rho(\mathbf{x},t_0) - \int_{t_0}^{t} dt'\,\nabla\cdot\mathbf{j}(\mathbf{x},t'). \tag{43}$$

It should also be remembered that only Eqs. (39) and (41) are equations of motion, while (38) and (40) are constraints that the initial values of **E** and **B** have to satisfy; for later times, the equations of motion and charge conservation ensure that these constraints are satisfied.

If we restrict the fields to a region V bounded by a surface S, we need to know boundary values of the fields. The normal component of one field can be obtained from the tangential component of the other field by means of the relations

$$\frac{\partial}{\partial t}(\hat{n}\cdot\mathbf{E}) = -c^2\,\nabla_s\cdot(\hat{n}\times\mathbf{B}) - \frac{\mathbf{j}}{\epsilon_0}, \tag{44}$$

$$\frac{\partial}{\partial t}(\hat{n}\cdot\mathbf{B}) = \nabla_s\cdot(\hat{n}\times\mathbf{E}), \tag{45}$$

where $\hat{n}$ is the normal to S and ∇_s is the surface gradient operator. Furthermore, the energy balance equation,

$$\oint_S dS\,\hat{n}\cdot\mathbf{E}\times\mathbf{B} = -\frac{1}{2}\frac{d}{dt}\int_V dV\left(\mathbf{B}^2 + \frac{\mathbf{E}^2}{c^2}\right) - \mu_0\int_V dV\,\mathbf{j}\cdot\mathbf{E}, \tag{46}$$

when applied to the difference of two solutions of Maxwell's equations with the same sources, initial values, and boundary values, implies that the fields are uniquely determined by the current density, the initial values of **E** and **B** subject to the constraints, and the tangential component of either **E** or **B**.

We can express the fields in terms of the sources, initial values, and boundary values by means of the equations[4]

$$\mathbf{E}(x) = \int_{t_0}^{t} dt'\int_V dV'\left[\mu_0\,\mathbf{j}(x')\frac{\partial G_R(x,x')}{\partial t'} + \frac{\rho(x')}{\epsilon_0}\nabla' G_R(x,x')\right] - \int_V dV'\left[\frac{1}{c^2}\mathbf{E}(x')\frac{\partial G_R(x,x')}{\partial t'} - \mathbf{B}(x')\times\nabla' G_R(x,x')\right]_{t'=t_0}$$
$$+ \int_{t_0}^{t} dt'\left[\oint_S d\mathbf{S}'\mathbf{E}(x')\cdot\nabla' G_R(x,x') - \oint_S d\mathbf{S}'\cdot\mathbf{E}(x')\,\nabla' G_R(x,x') - \oint_S d\mathbf{S}'\cdot\nabla' G_R(x,x')\mathbf{E}(x') - \oint_S d\mathbf{S}'\times\mathbf{B}(x')\frac{\partial G_R(x,x')}{\partial t'}\right], \tag{47}$$

$$\mathbf{B}(x) = \mu_0\int_{t_0}^{t} dt'\int_V dV'\,\mathbf{j}(x')\times\nabla' G_R(x,x') - \frac{1}{c^2}\int_V dV'\left[\mathbf{E}(x')\times\nabla' G_R(x,x') + \mathbf{B}(x')\frac{\partial G_R(x,x')}{\partial t'}\right]_{t'=t_0}$$
$$+ \int_{t_0}^{t} dt'\left[\frac{1}{c^2}\oint_S d\mathbf{S}'\times\mathbf{E}(x')\frac{\partial G_R(x,x')}{\partial t'} - \oint_S d\mathbf{S}'\cdot\nabla' G_R(x,x')\mathbf{B}(x') + \oint_S d\mathbf{S}'\mathbf{B}(x')\cdot\nabla' G_R(x,x') - \oint_S d\mathbf{S}'\cdot\mathbf{B}(x')\,\nabla' G_R(x,x')\right], \tag{48}$$

where $G_R(x,x')$ is a Green function for the scalar wave equation.

An alternative approach to the solution of Maxwell's equations is their reduction to a single vector wave equation. We now show how this equation can be solved with the help of dyadic Green functions.

We eliminate **B** from Eqs. (39) and (41) to obtain

$$\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} + \nabla\times(\nabla\times\mathbf{E}) = -\mu_0\frac{\partial\mathbf{j}}{\partial t}. \tag{49}$$

The initial value of the time derivative of **E** is obtained from **B** through Eq. (41). The boundary conditions on the field **E** are given either as the tangential component of **E** or the tangential component of $\nabla\times\mathbf{E}$, which can be obtained from the tangential component of **B** through Eq. (39). Once the field **E** is determined, the field **B** can be found from Eq. (39) and its initial value; these fields satisfy Maxwell's equations.

The elementary solution of the vector wave equation is a dyadic Q that satisfies

$$\frac{1}{c^2}\frac{\partial^2 Q(x)}{\partial t^2} + \nabla\times[\nabla\times Q(x)] = \delta^{(4)}(x)\mathsf{I}, \tag{50}$$

where I is the unit 3×3 dyadic, and the causality condition

$$Q(x) = 0, \quad t<0. \tag{51}$$

We can find the elementary solution following the procedure in Sec. II. If U is the Fourier transform of Q, it satisfies

$$-(\omega^2/c^2)\mathsf{U} - \mathbf{k}\times(\mathbf{k}\times\mathsf{U}) = (2\pi)^{-2}\mathsf{I}. \tag{52}$$

Scalar multiplication by k on the left allows us to find $\mathbf{k}\cdot\mathsf{U}$, which is then substituted into the expansion of the triple vector product. We again add a small positive imaginary part to ω to satisfy the causality condition (51), and solve for U to find

$$\mathsf{U}(\mathbf{k},\omega) = \frac{1}{4\pi^2}\frac{(\omega+i\epsilon)^2\mathsf{I} - c^2\mathbf{k}\,\mathbf{k}}{(\omega+i\epsilon)^2[\mathbf{k}^2 - (\omega+i\epsilon)^2/c^2]}. \tag{53}$$

The denominator, apart from the numerical factor, is precisely the function $K(\mathbf{k},\omega)$ in Eq. (10). Equation (A43) then implies that

$$Q(\mathbf{x},t) = \frac{1}{4\pi^2}\left[-\mathsf{I}\frac{\partial^2 F(\mathbf{x},t)}{\partial t^2} + c^2\nabla\nabla F(\mathbf{x},t)\right]. \tag{54}$$

The time derivative is given by Eq. (24), and we have to compute the gradient of the gradient of F. To handle the singularity of the function at the origin, we consider F the limit as $\epsilon\to 0$ of a function F_ϵ that vanishes inside a sphere of radius ϵ centered at the origin; we write

$$F_\epsilon(\mathbf{x},t) = F(\mathbf{x},t)\,\theta(r-\epsilon). \tag{55}$$

For a sequence of distributions T_j that tend to a limit T, it has been shown[1] that the derivatives T_j' tend to T' (which is not necessarily the case for a sequence of functions). We thus consider a sequence of positive values ϵ_j that tend to zero, compute first ∇F_ϵ, and obtain ∇F by letting ϵ tend to zero. The function F_ϵ is discontinuous on the spherical surface $S(\epsilon)$, and we use Eq. (A18) to find

$$\nabla F_\epsilon = \pi t r^{-3}\mathbf{x}\,\theta(t-r/c)\,\theta(r-\epsilon) - \pi\epsilon^{-1}(t-\epsilon/c)\hat{r}\delta[S(\epsilon)], \tag{56}$$

where $\hat{r}$ is the unit normal to $S(\epsilon)$. When $\epsilon\to 0$, the first term tends to a locally integrable function, and we only have to examine the surface integral of the coefficient of $\delta[S(\epsilon)]$. If $f(\mathbf{x})$ is a locally integrable function, we can write

$$\oint_{S(\epsilon)} dS f(\mathbf{x})\phi(\mathbf{x}) \approx \phi(0)\oint_{S(\epsilon)} dS f(\mathbf{x}), \tag{57}$$

which is true to lowest order in ϵ because the continuous test function ϕ satisfies

$$|\phi(\mathbf{x}) - \phi(0)| < a\epsilon. \tag{58}$$

for a constant a when x is on the sphere $S(\epsilon)$. The surface integral reduces to an integral over the solid angle and the limit is

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(t-\frac{\epsilon}{c}\right)\oint_{S(1)}\epsilon^2\hat{r}\,d\Omega = 0. \tag{59}$$

Thus, the gradient of F is

$$\nabla F = \pi t r^{-3}\,\mathbf{x}\,\theta(t-r/c). \tag{60}$$

We use the same procedure to compute the second-order derivatives. If we multiply a component of ∇F by $\theta(r-\epsilon)$, where we choose ϵ such that $\epsilon<ct$, we obtain a function that is discontinuous on the spherical surfaces $r=\epsilon$ and $r=ct$. The second derivatives of the function F for $\epsilon<r<ct$ are given by

$$\left\{\frac{\partial^2 F}{\partial x_i\partial x_j}\right\} = \pi t\frac{\partial}{\partial x_i}\left(\frac{x_j}{r^3}\right) = \pi t\left(\frac{\delta_{ij}}{r^3} - \frac{3x_ix_j}{r^5}\right), \tag{61}$$

and, adding the singular parts, we have

$$\nabla(\nabla F)_\epsilon = \pi t(r^{-3}\mathsf{I} - 3r^{-5}\,\mathbf{x}\,\mathbf{x})\,\theta(r-\epsilon)\,\theta(t-r/c) + \pi t r^{-4}\,\mathbf{x}\,\mathbf{x}\delta[S(\epsilon)] - \pi t r^{-4}\,\mathbf{x}\,\mathbf{x}\,\delta[S(ct)]. \tag{62}$$

We note that the function in Eq. (61) is not Lebesgue-integrable at the origin, but the limit of the distribution must exist since those of the other terms in Eq. (62) all exist. For the coefficient of $\delta[S(\epsilon)]$ we have

$$\lim_{\epsilon\to 0+}\oint_{S(\epsilon)} dS\frac{x_ix_j}{r^4}\phi(\mathbf{x}) = \frac{4\pi}{3}\delta_{ij}\,\phi(0), \tag{63}$$

and we write

$$\nabla\nabla F = \pi t r^{-3}(\mathsf{I} - 3r^{-2}\,\mathbf{x}\,\mathbf{x})\,\theta(t-r/c) + \tfrac{4}{3}\pi t\delta^{(3)}(\mathbf{x})\mathsf{I} - \pi c^{-1}r^{-3}\mathbf{x}\,\mathbf{x}\,\delta[S(ct)], \tag{64}$$

where the first term on the right-hand side is a kind of principal value. We can use symmetry arguments to prove that the integral of $r^{-3}\delta_{ij} - 3r^{-5}x_ix_j$ over a sphere vanishes; for $i\neq j$ the integrand is antisymmetric in the coordinates, and for $i=j$ we obtain three equal integrals whose sum vanishes. Finally, Eq. (54) becomes

$$Q(\mathbf{x},t) = (4\pi r)^{-1}\delta(t-r/c)(\mathsf{I} - r^{-2}\,\mathbf{x}\,\mathbf{x}) + (4\pi r^3)^{-1}c^2 t\,\theta(t-r/c)(\mathsf{I} - 3r^{-2}\,\mathbf{x}\,\mathbf{x}) + \tfrac{1}{3}c^2 t\,\theta(t)\,\delta^{(3)}(\mathbf{x})\mathsf{I}. \tag{65}$$

This distribution can be defined by an equation similar to (26) for the scalar elementary solution. For a component Q_{ij} of Q, we have

$$\langle Q_{ij},\phi\rangle = \int_0^\infty\frac{dt}{4\pi t}\oint_{S(ct)} dS(\delta_{ij} - \alpha_i\alpha_j)\,\phi(\mathbf{x},t) + \lim_{\epsilon\to 0+}\int_{\epsilon/c}^\infty\frac{c^2 t\,dt}{4\pi}\int_\epsilon^{ct}\frac{dr}{r^3}\oint_{S(r)} dS(\delta_{ij} - 3\alpha_i\alpha_j)\,\phi(\mathbf{x},t) + \int_0^\infty\frac{c^2 t\,dt}{3}\delta_{ij}\,\phi(0,t), \tag{66}$$

where $\alpha_i = x_i/r$. The limit of the second integral must exist, as discussed after Eq. (62). Note that Q is symmetric.

To solve the inhomogeneous wave equation (49) when the sources are known for all space and for all past times, we use

$$\mathbf{E} = -\mu_0 Q^* \cdot \frac{\partial\mathbf{j}}{\partial t}, \tag{67}$$

where we have both a convolution product of distributions and a dot product of the dyadic and the vector. We can rewrite this equation in the form

$$\mathbf{E}(x) = -\int_{-\infty}^{t} \frac{\mu_0 dt'}{4\pi(t-t')} \oint_{S[\mathbf{x},c(t-t')]} dS' \left[\frac{\partial \mathbf{j}(x')}{\partial t'} - \frac{\mathbf{R}}{R^2}\mathbf{R}\cdot\frac{\partial \mathbf{j}(x')}{\partial t'}\right]$$
$$-\int_{-\infty}^{t} \frac{(t-t')dt'}{4\pi\epsilon_0} \int_0^{c(t-t')} \frac{dr'}{R^3} \oint_{S(\mathbf{x},r')} dS' \left[\frac{\partial \mathbf{j}(x')}{\partial t'} - 3\frac{\mathbf{R}}{R^2}\mathbf{R}\cdot\frac{\partial \mathbf{j}(x')}{\partial t'}\right] - \int_{-\infty}^{t} \frac{(t-t')dt'}{3\epsilon_0}\frac{\partial \mathbf{j}(\mathbf{x},t')}{\partial t'}, \tag{68}$$

where $\mathbf{R} = \mathbf{x} - \mathbf{x}'$ and $R = |\mathbf{R}|$. This result has to agree, of course, with the corresponding expression obtained from Eq. (47), that is,

$$\mathbf{E}(x) = \int_{-\infty}^{t} dt' \int dV' \left[\mu_0 \mathbf{j}(x')\frac{\partial G_R^{(0)}(x,x')}{\partial t'} + \frac{1}{\epsilon_0}\rho(x')\nabla' G_R^{(0)}(x,x')\right], \tag{69}$$

which is also a convolution product of distributions. To show the equivalence of these expressions, we first use Eq. (A39) to obtain

$$\mathbf{E}(x) = -\int_{-\infty}^{t} dt' \int dV' \left[\mu_0 \frac{\partial \mathbf{j}(x')}{\partial t'} + \frac{1}{\epsilon_0}\nabla'\rho(x')\right] G_R^{(0)}(x,x'), \tag{70}$$

and, taking into account Eqs. (7), (22), and (42), we show that

$$\nabla\rho * \frac{\partial^2 F}{\partial t^2} = \nabla\frac{\partial\rho}{\partial t} * \frac{\partial F}{\partial t}$$
$$= -\nabla\nabla\cdot\mathbf{j} * \frac{\partial F}{\partial t} = -\nabla\nabla\cdot\frac{\partial \mathbf{j}}{\partial t} * F$$
$$= -\nabla\cdot\frac{\partial \mathbf{j}}{\partial t} * \nabla F = -\frac{\partial \mathbf{j}}{\partial t} *\cdot \nabla\nabla F, \tag{71}$$

and Eq. (68) becomes

$$\mathbf{E} = -\mu_0 \frac{\partial \mathbf{j}}{\partial t} *\cdot \frac{-|\partial^2 F/\partial t^2 + c^2\nabla\nabla F}{4\pi^2}, \tag{72}$$

which is the same as Eq. (67).

As done in Sec. II for the scalar wave equation, initial conditions and boundary conditions can be built into the sources of the vector wave equation. We use Eq. (A20) twice to obtain

$$\nabla\times(\nabla\times\mathbf{E})$$
$$= \{\nabla\times(\nabla\times\mathbf{E})\} + \hat{n}\times\Delta\,(\nabla\times\mathbf{E})\,\delta(S) + \nabla\times[\hat{n}\times\Delta\,\mathbf{E}\,\delta(S)], \tag{73}$$

where the quantity in the square brackets is a singular distribution. We assume that the fields vanish for $t < t_0$ and outside V, and find

$$\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} + \nabla\times(\nabla\times\mathbf{E})$$
$$= -\mu_0\frac{\partial \mathbf{j}}{\partial t} + \mathbf{E}(\mathbf{x},t_0)\delta'(t-t_0) + \frac{\partial\mathbf{E}(\mathbf{x},t_0)}{\partial t}\delta(t-t_0)$$
$$+ \Delta\,[\hat{n}\times(\nabla\times\mathbf{E})]\delta(S) + \nabla\times[\Delta\,(\hat{n}\times\mathbf{E})\,\delta(S)]. \tag{74}$$

If the initial and boundary conditions are given for $\mathbf{B}$, we use Eqs. (39) and (41) to change Eq. (74) into

$$\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} + \nabla\times(\nabla\times\mathbf{E})$$
$$= -\mu_0\frac{\partial \mathbf{j}}{\partial t} + \mathbf{E}(\mathbf{x},t_0)\,\delta'(t-t_0)$$
$$+ \left[c^2\nabla\times\mathbf{B}(\mathbf{x},t_0) - \frac{\mathbf{j}(\mathbf{x},t_0)}{\epsilon_0}\right]\delta(t-t_0)$$
$$-\frac{\partial}{\partial t}\Delta\,(\hat{n}\times\mathbf{B})\,\delta(S) + \nabla\times[\Delta\,(\hat{n}\times\mathbf{E})\,\delta(S)], \tag{75}$$

Convolution with the dyadic elementary solution provides a formal solution of the wave equation that also satisfies the initial and the boundary conditons; we obtain

$$\mathbf{E} = -\mu_0\frac{\partial \mathsf{Q}}{\partial t} *\cdot \mathbf{j} + \frac{\partial \mathsf{Q}}{\partial t} *\cdot \mathbf{E}(\mathbf{x},t_0)\delta(t-t_0)$$
$$+ \mathsf{Q} *\cdot \left[c^2\,\nabla\times\mathbf{B}(\mathbf{x},t_0) - \frac{\mathbf{j}(\mathbf{x},t_0)}{\epsilon_0}\right]\delta(t-t_0)$$
$$+ \mathsf{Q} *\cdot \hat{n}\times\frac{\partial \mathbf{B}}{\partial t}\delta(S) + \hat{n}\times\mathbf{E}\delta(S) *\cdot \nabla\times\mathsf{Q}, \tag{76}$$

where we assume that the given functions vanish for $t < t_0$ and outside V. This equation is equivalent to Eq. (B22) when the free-space dyadic Green function is used for G_R.

In general we do not know both the tangential components of $\mathbf{E}$ and $\mathbf{B}$ on the boundary S, and Eq. (76) leads to an integral equation for the unknown boundary value of the field. We can use other Green functions that satisfy homogeneous boundary conditions, but such a condition cannot be imposed on a distribution; this possibility is discussed further in Appendix B, where the heuristic approach to Green functions is used.

The dyadic elementary solution could also have been obtained from the elementary solution for the vector Helmholtz equation[5] that satisfies the outgoing-wave condition at infinity. These two distributions are related by a Fourier transform that leads from the time variable to the frequency variable.

IV. CONCLUDING REMARKS

In this paper we have shown how the free-space Maxwell equations can be solved, or reduced to relatively simple integral equations, within the framework of the theory of distributions.

We first found the elementary solution of the scalar wave equation and showed how it is used to find an integral equation for boundary values of the field, which can then be used to compute a solution by integration. These formulations are well known, especially in the heuristic approach to Green functions, but it is useful to have a mathematically well-defined derivation that can be generalized to more difficult problems without ambiguities.

We then followed the same procedure to find the dyadic elementary solution of the vector wave equation that was derived from Maxwell's equations, and found the corresponding expression of the field in terms of the sources, the initial values, and the boundary values. This relationship reduces to an integral equation for either the tangential component of **E** or the tangential component of **B** when the field point tends to the boundary surface. The scalar elementary solution can also be used to solve Maxwell's equations, but the boundary terms include the normal components of the fields.

There are difficulties in the definition of Green functions that obey homogeneous boundary conditions within the framework of the theory of distributions, and we had to use the heuristic approach to obtain some further results, as shown in Appendix B. These Green functions are no longer invariant under translations, and the integrals that give the fields are not convolutions. It would be useful to extend the theory of distributions to cover these subjects.

In addition to presenting new results for dyadic Green functions for the time-dependent vector wave equation, we have demonstrated how the theory of distributions can be used to obtain rigorous results in problems where the heuristic approach is hazardous.

APPENDIX A: REVIEW OF DISTRIBUTIONS

In this appendix we briefly review the main concepts from the theory of distributions,[1,2] and give enough of the results to present the equations used in the paper.

A distribution is a continuous linear functional on a space of test functions. The most general set of distributions is obtained when the test functions belong to the space $\mathscr{D}$ of infinitely differentiable functions of bounded support. A distribution T is defined when a complex number c is associated with each test function ϕ, and we use the notation

$$\langle T,\phi \rangle = c. \tag{A1}$$

For instance, the Dirac delta distribution is defined by

$$\langle \delta,\phi \rangle = \phi(0). \tag{A2}$$

A distribution T_f can be associated to any locally integrable function f by

$$\langle T_f,\phi \rangle = \int_{-\infty}^{\infty} f(x)\phi(x)\,dx. \tag{A3}$$

It is customary to use f both for the function and the distribution, and in most cases relations are valid for both. When T corresponds to a function, it is called a regular distribution; otherwise, it is called singular. Although, in general, it does not make sense to talk about the value of a distribution at a point x, we often write $T(x)$ even when T does not correspond to a function mainly to indicate what variable is involved; such a liberty if often taken with the delta distribution. These concepts can be generalized to distributions in spaces of higher dimensions. For instance,

$$\langle T(\mathbf{x},t),\phi(\mathbf{x},t)\rangle = c \tag{A4}$$

defines a distribution in a four-dimensional space–time.

A linear combination of distributions is defined by

$$\langle \alpha T_1 + \beta T_2,\phi \rangle = \alpha \langle T_1,\phi \rangle + \beta \langle T_2,\phi \rangle, \tag{A5}$$

and the null distribution is given by

$$\langle T_0,\phi \rangle = 0, \tag{A6}$$

for arbitrary ϕ. Consequently, two distributions T_1 and T_2 are equal if and only if, for any test function ϕ,

$$\langle T_1,\phi \rangle = \langle T_2,\phi \rangle. \tag{A7}$$

A distribution T is zero in an open region of space if the value of the functional vanishes for all test functions ϕ whose support is in that region. The complement of the union of all regions in which T is zero is called the support of the distribution. For instance, the support of the delta distribution is the origin.

A distribution can be shifted by an amount a according to

$$\langle T(x-a),\phi(x)\rangle = \langle T(x),\phi(x+a)\rangle; \tag{A8}$$

we define the symmetrically transposed distribution by

$$\langle T(-x),\phi(x)\rangle = \langle T(x),\phi(-x)\rangle, \tag{A9}$$

and we define a change of scale by

$$\langle T(ax),\phi(x)\rangle = |a|^{-1}\langle T(x),\phi(x/a)\rangle. \tag{A10}$$

Any distribution T has a derivative $T' = dT/dx$ defined by

$$\langle T',\phi \rangle = -\langle T,\phi' \rangle, \tag{A11}$$

or, for a distribution in a four-dimensional space–time, a partial derivative is defined by

$$\left\langle \frac{\partial T(\mathbf{x},t)}{\partial t},\phi(\mathbf{x},t)\right\rangle = -\left\langle T(\mathbf{x},t),\frac{\partial \phi(\mathbf{x},t)}{\partial t}\right\rangle. \tag{A12}$$

If θ is the unit step function, Eqs. (A2), (A8), and (A11) imply that

$$\theta' = \delta. \tag{A13}$$

Consider a function $f(x)$ that has a locally integrable derivative defined almost everywhere, except at a set of points x_i where f can have finite jumps. Equations (A3) and (A10) then lead to the relation

$$\frac{df}{dx} = \left\{\frac{df}{dx}\right\} + \sum_i (\Delta_i f)\delta(x-x_i), \tag{A14}$$

where the left-hand side is the derivative of f in the sense of distributions, which always exists, the derivative in curly brackets is taken in the sense of functions, $\Delta_i f$ is the jump

$$\Delta_i f = f(x_i+0) - f(x_i-0), \tag{A15}$$

and $\delta(x - x_i)$ is the shifted delta distribution.

Equation (A14) can be generalized to functions of several variables. If $f(\mathbf{x})$ is differentiable with respect to x_1 except on a surface S,

$$\frac{\partial f}{\partial x_1} = \left\{\frac{\partial f}{\partial x_1}\right\} + n_1 \Delta f \delta(S), \tag{A16}$$

where n_1 is a component of the unit normal, Δf is the jump across S in the direction of the normal (reversing $\hat{n}$ changes the sign of Δf and leaves the product unchanged), and the definition of the singular distribution is

$$\langle u(\mathbf{x})\delta(S),\phi(\mathbf{x})\rangle = \int_S dS\, u(\mathbf{x})\phi(\mathbf{x}), \tag{A17}$$

where $u(\mathbf{x})$ is a function that needs to be defined on S only. In the usual three-dimensional space, Eq. (A16) leads to the relations

$$\nabla f = \{\nabla f\} + \hat{n}\,\Delta f \delta(S), \tag{A18}$$

$$\nabla\cdot\mathbf{f} = \{\nabla\cdot\mathbf{f}\} + \hat{n}\cdot\Delta\,\mathbf{f}\,\delta(S), \tag{A19}$$

$$\nabla\times\mathbf{f} = \{\nabla\times\mathbf{f}\} + \hat{n}\times\Delta\,\mathbf{f}\,\delta(S). \tag{A20}$$

Combining Eqs. (A18) and (A19), we obtain for the Laplacian

$$\nabla^2 f = \{\nabla^2 f\} + \Delta\frac{\partial f}{\partial n}\delta(S) + \nabla\cdot[\hat{n}\,\Delta f \delta(S)], \tag{A21}$$

where $\Delta\,(\partial f/\partial n)$ is the jump in the normal derivative that comes from the term

$$\hat{n}\cdot\Delta\,\{\nabla f\} = \Delta\,[\hat{n}\cdot\{\nabla F\}] = \Delta\frac{\partial f}{\partial n}, \tag{A22}$$

and the derivatives of the singular distribution in the last term of Eq. (A21) have to be taken according to the general definition (A12).

A generalization of the ordinary product of two functions to a product of arbitrary distributions is not possible because, for instance, the product of two locally integrable functions is not necessarily locally integrable. On the other hand, the direct product of two distributions can always be defined by

$$\langle T(x)U(y),\phi(x,y)\rangle = \langle T(x),\psi(x)\rangle, \tag{A23}$$

where

$$\psi(x) = \langle U(y),\phi(x,y)\rangle \tag{A24}$$

is an indefinitely differentiable function of x that satisfies

$$\frac{d\psi}{dx} = \left\langle U(y),\frac{\partial\phi(x,y)}{\partial x}\right\rangle. \tag{A25}$$

This product is commutative and distributive over a sum, that is,

$$T(x)U(y) = U(y)T(x), \tag{A26}$$

$$T(x)[U_1(y) + U_2(y)] = T(x)U_1(y) + T(x)U_2(y), \tag{A27}$$

where Eq. (A26) means

$$\langle T(x),\langle U(y),\phi(x,y)\rangle\rangle = \langle U(y),\langle T(x),\phi(x,y)\rangle\rangle. \tag{A28}$$

The product can be generalized to more than two factors, and it is associative. We have, for instance,

$$\delta(x)\delta(y)\delta(z) = \delta^{(3)}(\mathbf{x}), \tag{A29}$$

where the three-dimensional delta distribution is defined by

$$\langle\delta^{(3)}(\mathbf{x}),\phi(\mathbf{x})\rangle = \phi(0) = \phi(0,0,0). \tag{A30}$$

The direct product of two distributions results in another distribution in a space of higher dimension. It is also often possible to define a convolution product in the same space of the original distributions, indicated by

$$V(x) = T(x)*U(x) \tag{A31}$$

and defined by

$$\langle V,\phi\rangle = \langle T(x)U(y),\phi(x+y)\rangle \tag{A32}$$

in terms of the direct product (A23). It should be remembered that Eq. (A31) does not mean that the value of the convolution product at x is the product of the values of the factors at x, even when these quantities are defined. The convolution product of two arbitrary distributions is not always defined because, even though $\phi(x)$ has a compact support, the support of $\phi(x+y)$ in the plane is essentially the union of unbounded diagonal strips and the right-hand side of Eq. (A32) need not be defined. The convolution product exists when one of the distributions has a compact support or when both distributions have supports bounded on the same side; these are sufficient conditions, and it is not necessary for either to be satisfied. When the convolution product exists, it is commutative and distributive,

$$T*U = U*T, \tag{A33}$$

$$T*(U_1 + U_2) = T*U_1 + T*U_2. \tag{A34}$$

When T and U are functions, Eq. (A32) can be rewritten in the form

$$V(x) = \int_{-\infty}^{\infty} T(x - x')U(x')\,dx'. \tag{A35}$$

Some useful particular relations are

$$\delta*T = T, \tag{A36}$$

$$\delta'*T = T', \tag{A37}$$

$$\delta(x - a)*T(x) = T(x - a), \tag{A38}$$

$$\frac{d}{dx}(T*U) = \frac{dT}{dx}*U = T*\frac{dU}{dx}. \tag{A39}$$

The convolution product can be extended in a straightforward way to more than two factors and to distributions of several variables.

The definition of the Fourier transform of a distribution is

$$\langle\mathcal{F}(T),\phi\rangle = \langle T,\mathcal{F}(\phi)\rangle, \tag{A40}$$

but this definition is not applicable in general because the Fourier transform of a function of compact support does not have a compact support. We restrict ourselves to the space of tempered distributions where the Fourier transform is defined by enlarging the space of test functions to the space $\mathcal{S}$ of indefinitely differentiable functions that decrease rapidly at infinity. By rapidly decreasing we mean that, for arbitrary nonnegative integers m and n, $x^m\phi^{(n)}(x)$ is bounded; in other words, $\phi(x)$ tends to zero faster than any power of $1/|x|$ when

$|x|\to\infty$. The Fourier transform of such a function is also in $\mathscr{S}$, and the Fourier transform of a tempered distribution is well defined by Eq. (A40). In particular, distributions of bounded support are tempered distributions, as are the distributions that correspond to locally integrable functions that increase slowly at infinity (more slowly than some power of $|x|$).

The sign of the exponential in a Fourier transform is arbitrary, and we define

$$\mathscr{F}_{\pm}(\phi) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} \phi(x)\exp(\pm ikx)\,dx; \tag{A41}$$

then, $\mathscr{F}_{-}$ is the inverse transform for $\mathscr{F}_{+}$ and vice versa. Thus, for a function f we have

$$\mathscr{F}_{\pm}(f) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} f(x)\exp(\pm ikx)\,dx, \tag{A42}$$

if the integral exists; the Fourier transform of a distribution T satisfies

$$\mathscr{F}_{\pm}(T') = \mp ik\,\mathscr{F}_{\pm}(T), \tag{A43}$$

$$\mathscr{F}_{\pm}[T(x-a)] = \exp(\pm iak)\,\mathscr{F}_{\pm}[T(x)]; \tag{A44}$$

and, for the convolution product of two distributions of compact support, we have

$$\mathscr{F}_{\pm}(T*U) = \mathscr{F}_{\pm}(T)\,\mathscr{F}_{\pm}(U), \tag{A45}$$

since the Fourier transform of a distribution of compact support is a function.

We now give some examples of distributions and their Fourier transforms. The principal value of $1/x$, which is defined by

$$\left\langle \mathrm{P}\frac{1}{x},\phi(x)\right\rangle = \lim_{\epsilon\to 0+}\left[\int_{-\infty}^{-\epsilon}\frac{\phi(x)}{x}\,dx + \int_{\epsilon}^{\infty}\frac{\phi(x)}{x}\,dx\right], \tag{A46}$$

satisfies

$$\frac{d}{dx}\log|x| = \mathrm{P}\frac{1}{x}, \tag{A47}$$

$$\mathscr{F}_{\pm}\left[\mathrm{P}\frac{1}{x}\right] = \pm i(\tfrac{1}{2}\pi)^{1/2}\,\mathrm{sgn}(k). \tag{A48}$$

We also have

$$\mathscr{F}_{\pm}(\delta) = (2\pi)^{-1/2}, \tag{A49}$$

whence

$$\mathscr{F}_{\pm}[\mathrm{P}(1/x) \pm i\pi\delta] = \pm i(2\pi)^{1/2}\,\theta(k), \tag{A50}$$

where upper or lower signs have to be taken together. It is customary to write the argument of the Fourier transform as $(x\mp i\epsilon)^{-1}$, and we have

$$\mathscr{F}_{\pm}[(x\mp i\epsilon)^{-1}] = \pm i(2\pi)^{1/2}\,\theta(k), \tag{A51}$$

where the limit $\epsilon\to 0+$ is implicit. The Fourier transform of the derivative of this distribution is, by Eq. (A43),

$$\mathscr{F}_{\pm}[(x\mp i\epsilon)^{-2}] = -(2\pi)^{1/2}\,k\theta(k). \tag{A52}$$

We can write

$$(x\mp i\epsilon)^{-2} = \mathrm{F.P.}(1/x^2) \mp i\pi\delta', \tag{A53}$$

where the derivative of $\mathrm{P}(1/x)$ is related to the finite part of Hadamard of a divergent integral,

$$\left\langle \mathrm{F.P.}\frac{1}{x^2},\phi\right\rangle = \mathrm{F.P.}\int_{-\infty}^{\infty}\frac{\phi(x)}{x^2}\,dx = \lim_{\epsilon\to 0+}\left[\int_{-\infty}^{-\epsilon}\frac{\phi(x)}{x^2}\,dx + \int_{\epsilon}^{\infty}\frac{\phi(x)}{x^2}\,dx - \frac{2\phi(0)}{\epsilon}\right]. \tag{A54}$$

A relationship that is useful for functions is

$$\int_{-\infty}^{\infty} d^3k\,f(k)\exp(\pm i\,\mathbf{k}\cdot\mathbf{x}) = 4\pi r^{-1}\int_0^{\infty} kdk\,f(k)\sin(kr), \tag{A55}$$

where $k = |\mathbf{k}|$ and $r = |\mathbf{x}|$.

APPENDIX B: HOMOGENEOUS BOUNDARY CONDITIONS

When boundary conditions are given for the unknown field in a linear partial differential equation, it is often useful to define Green functions that obey homogeneous boundary conditions. For the scalar wave equation, we can define $G_R^{(1)}(x,x')$ which vanishes when $\mathbf{x}$ is on S, and $G_R^{(2)}$, whose normal derivative vanishes on S. When $G_R^{(0)}$ is replaced by $G_R^{(1)}$ or $G_R^{(2)}$ in Eq. (37), one of the surface integrals vanishes, and we obtain a solution by integration over known functions instead of an integral equation.

There are two problems that keep us from using the theory of distributions to handle these Green functions: They are no longer functions of $x-x'$, so that Eq. (36) no longer represents convolution products, and a distribution would have to vanish on a surface, which is not an open region of space. Thus, in this context, we use the heuristic approach to Green functions.

The Green functions $G_R^{(1)}$, and $G_R^{(2)}$ and the corresponding advanced Green functions obey symmetry relations

$$G_R(x,x') = G_A(x',x), \tag{B1}$$

$$G_R(\mathbf{x},t;\mathbf{x}',t') = G_R(\mathbf{x}',-t';\mathbf{x},-t), \tag{B2}$$

$$\frac{\partial G_R(x,x')}{\partial t'} = -\frac{\partial G_R(x,x')}{\partial t}, \tag{B3}$$

but, in general,

$$\nabla G_R(x,x') \neq -\nabla' G_R(x',x). \tag{B4}$$

When V is the half-plane $z\geqslant 0$,

$$G_R^{(1)}(x,x') = \mathscr{G}(x-x') - \mathscr{G}(x-x_I'), \tag{B5}$$

where x_I' is the image point obtained by setting $z_I' = -z'$. Even though the translated distributions are defined through Eq. (A8), giving a precise meaning to $G_R^{(1)}$, we cannot say what is meant by $G_R^{(1)}$ vanishing when $z=0$.

If we substitute either $G_R^{(1)}$ or $G_R^{(2)}$ in Eqs. (46) and (47), unknown fields in the surface integrals are not eliminated; this constitutes some of the motivation to consider dyadic Green functions.

The appropriate form of Green theorem for the vector wave equation is

$$\int_{t_0}^{t_1} dt \int_V dV \left[\left\{ \frac{1}{c^2} \frac{\partial^2 \mathbf{u}}{\partial t^2} + \nabla \times (\nabla \times \mathbf{u}) \right\} \cdot \mathbf{v} - \left\{ \frac{1}{c^2} \frac{\partial^2 \mathbf{v}}{\partial t^2} + \nabla \times (\nabla \times \mathbf{v}) \right\} \cdot \mathbf{u} \right]$$
$$= \frac{1}{c^2} \int_V dV \left[\mathbf{v} \cdot \frac{\partial \mathbf{u}}{\partial t} - \mathbf{u} \cdot \frac{\partial \mathbf{v}}{\partial t} \right]_{t_0}^{t_1} + \int_{t_0}^{t_1} dt \oint_S d\mathbf{S} \cdot [\mathbf{u} \times (\nabla \times \mathbf{v}) - \mathbf{v} \times (\nabla \times \mathbf{u})]. \quad \text{(B6)}$$

Retarded and advanced dyadic Green functions are defined by

$$\frac{1}{c^2} \frac{\partial^2 \mathsf{G}(x,x')}{\partial t^2} + \nabla \times [\nabla \times \mathsf{G}(x,x')] = \delta^{(4)}(x - x')\mathsf{I}, \quad \text{(B7)}$$

$$\mathsf{G}_R(x,x') = 0, \quad t < t', \quad \text{(B8)}$$

$$\mathsf{G}_A(x,x') = 0, \quad t > t'. \quad \text{(B9)}$$

The two simplest types of homogeneous boundary conditions are

$$\hat{n} \cdot \mathsf{G}^{(1)}(x,x')|_{\mathbf{x} \in S} = 0, \quad \text{(B10)}$$

$$\hat{n} \cdot \nabla \times \mathsf{G}^{(2)}(x,x')|_{\mathbf{x} \in S} = 0. \quad \text{(B11)}$$

In either case, we can use x_μ'' as variables of integration in Eq. (B6), let $t_0 \to -\infty$, and substitute

$$\mathbf{u}(x'') = \mathsf{G}_A(x'',x) \cdot \mathbf{a}, \quad \text{(B12)}$$

$$\mathbf{v}(x'') = \mathsf{G}_R(x'',x') \cdot \mathbf{b}, \quad \text{(B13)}$$

where $\mathbf{a}$ and $\mathbf{b}$ are arbitrary constant vectors. We then use Eqs. (B7)–(B9) and the boundary condition (B10) or (B11) to show the symmetry relation

$$\mathsf{G}_R(x,x') = \tilde{\mathsf{G}}_A(x',x), \quad \text{(B14)}$$

where the tilde indicates the transpose of the dyadic. Similarly, if instead of Eq. (B12) we use

$$\mathbf{u}(x'') = \mathsf{G}_R(\mathbf{x}'', -t''; \mathbf{x}', -t') \cdot \mathbf{a}, \quad \text{(B15)}$$

we find

$$\mathsf{G}_R(\mathbf{x},t;\mathbf{x}',t') = \tilde{\mathsf{G}}_R(\mathbf{x}', -t'; \mathbf{x}, -t), \quad \text{(B16)}$$

with the corresponding result for G_A. This equation implies that

$$\frac{1}{c^2} \frac{\partial^2 \tilde{\mathsf{G}}(x,x')}{\partial t'^2} + \nabla' \times [\nabla' \times \tilde{\mathsf{G}}(x,x')] = \delta^{(4)}(x - x')\mathsf{I}, \quad \text{(B17)}$$

$$\hat{n}' \cdot \tilde{\mathsf{G}}^{(1)}(x,x')|_{x' \in S} = 0, \quad \text{(B18)}$$

$$\hat{n}' \cdot \nabla' \times \tilde{\mathsf{G}}^{(2)}(x,x')|_{x' \in S} = 0. \quad \text{(B19)}$$

If we use x_μ' as variables of integration in Eq. (B6) and set

$$\mathbf{u}(x') = \mathbf{E}(x'), \quad \text{(B20)}$$

$$\mathbf{v}(x') = \tilde{\mathsf{G}}_R(x,x') \cdot \mathbf{a}, \quad \text{(B21)}$$

we find

$$\mathbf{E}(x) = -\mu_0 \int_{t_0}^{t} dt' \int_V dV' \, \mathsf{G}_R(x,x') \cdot \frac{\partial \mathbf{j}(x')}{\partial t'} + \frac{1}{c^2} \int_V dV' \left[\mathsf{G}_R(x,x') \cdot \left\{ c^2 \nabla' \times \mathbf{B}(x') - \frac{\mathbf{j}(x')}{\epsilon_0} \right\} - \frac{\partial \mathsf{G}_R(x,x')}{\partial t'} \cdot \mathbf{E}(x') \right]_{t' = t_0}$$
$$- \int_{t_0}^{t} dt' \oint_S dS' \, \hat{n}' \cdot \left[\mathbf{E}(x') \times \{ \nabla' \times \tilde{\mathsf{G}}_R(x,x') \} - \frac{\partial \mathbf{B}(x')}{\partial t'} \times \tilde{\mathsf{G}}_R(x,x') \right]. \quad \text{(B22)}$$

If we know $\hat{n} \times \mathbf{E}$ on S, we can use $\mathsf{G}_R^{(1)}$ to eliminate the second term in the surface integral, and, if we know $\hat{n} \times \mathbf{B}$ on S, $\mathsf{G}_R^{(2)}$ serves to eliminate the first term. In either case, Eq. (B22) gives $\mathbf{E}$ as a sum of integrals over known functions when the right Green function can be found.

In practice, we probably would not be able to find $G_R^{(1)}$ or $G_R^{(2)}$. Alternatively, we can use the free-space Green function $G_R^{(0)}$ in Eq. (B22) and let $\mathbf{E}$ approach S from the outside, where $\mathbf{E}$ vanishes. We obtain an integral equation where either $\hat{n} \times \mathbf{E}$ or $\hat{n} \times \mathbf{B}$ is known and the other is to be determined. We could also use $\hat{n} \times (\nabla \times \mathbf{E})$ instead of $\hat{n} \times \mathbf{B}$ in this formulation. Once we know both tangential fields, Eq. (B22) gives $\mathbf{E}$ everywhere.

[1] L. Schwartz, *Théorie des distributions*, Vol. 1 (Hermann, Paris, 1950) and Vol. 2 (1951).

[2] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1.

[3] R. Petit, Ed., *Electromagnetic Theory of Gratings* (Springer-Verlag, Berlin, 1980).

[4] E. Marx, Natl. Bur. Std. (U.S.) Tech. Note 1157, Washington, D. C., February 1982.

[5] C.-T. Tai, *Dyadic Green's Functions in Electromagnetic Theory* (Intext, Scranton, 1971).

[6] S.-W. Lee, J. Boersma, C.-L. Law, and G. A. Deschamps, IEEE Trans. Antennas Propagation **AP-28**, 311 (1980).

[7] L. Schwartz, *Mathematics for the Physical Sciences* (Addison-Wesley, Reading, Mass., 1966), pp. 281–4.

Single integral equation for wave scattering

Egon Marx
National Bureau of Standards, Washington, D.C. 20234

(Received 18 September 1981; accepted for publication 18 December 1981)

When a wave interacts with an obstacle, the scattered and transmitted fields can be found by solving a system of integral equations for two unknown fields defined on the surface of the body. By choosing a more appropriate unknown function, the system of equations is reduced to a single singular integral equation of the first kind. This reduction is done here for transient and monochromatic waves, for a scalar field that obeys the wave equation, and for electromagnetic fields that obey Maxwell's equations.

PACS numbers: 03.50.De, 03.40.Kf, 02.30.Rz

I. INTRODUCTION

The main physical problem under consideration in this paper is the scattering of an electromagnetic wave at the interface between two homogeneous regions of space. The determination of the scattered and transmitted fields throughout space has previously been reduced[1] to the solution of integral equations for unknown functions defined only on the surface.

If the surface is that of a perfect conductor, it is sufficient to find the surface current density,[1-3] which is a tangential vector field; the scattered fields are then determined by integrations. If the scatterer is a dielectric, it is possible to obtain[1] two integral equations for two tangential vector fields. Especially for numerical calculations, such a doubling in the number of unknown functions can be a problem because it increases computer storage requirements.

Maystre has shown[4] how the number of unknown functions in the problem of diffraction of electromagnetic waves by a grating can remain unchanged when a dielectric replaces a perfect conductor. We extend this method to three-dimensional scattering.

In Sec. II we consider the scattering of a transient scalar wave. The incident field is specified by the initial values of the field and its time derivative at time $t = 0$ in one of the regions. We assume that the field obeys the homogeneous wave equation, that the field is continuous across the surface, and that the normal derivative on one side is a fixed multiple of that on the other side. We define a single unknown function on the surface that obeys a singular integral equation (65) of the first kind; the scattered and transmitted fields can then be found by integrations, as shown in Eqs. (70) and (71). The solution of this problem is of interest, for instance, in acoustics and also is a model for a suitable generalization that can be applied to electromagnetic scattering. In Sec. III we start with an incident electromagnetic wave specified by $\mathbf{E}$ and $\mathbf{B}$ at $t = 0$. These fields obey the homogeneous Maxwell equations, the tangential component of the electric field is continuous across the surface, and the magnetic induction on one side is a constant multiple of that on the other side. We define a tangential vector field on the surface and find the integral equation (136) it obeys; the fields are then given by Eqs. (141)–(144). We use the Green function for the scalar wave equation, but we also sketch how the equivalent problem of the electric field that obeys the vector wave equation can be reduced with the help of a dyadic Green function. The related problems for monochromatic waves are reduced to similar integral equations in Sec. IV. In this form, it is easy to take into account finite conductivity and dispersive media.

Although most scientists and engineers prefer to use a heuristic approach to Dirac delta functions and Green functions, we use the theory of distributions[5-7] to provide a firm mathematical foundation to our derivations. We give more details in the derivations than we would if we were using a more familiar method, but we do not look for the least restrictive conditions or stop to prove all our assertions. In Appdendix A of Ref. 7 we collect a number of formulas that are used in this paper.

The method used in this paper can be applied to still other scattering problems, although details differ in each case.

II. SCALAR WAVES

We first consider the problem of the scattering of a scalar field ψ by an obstacle. We assume that the field is initially located in a homogeneous medium in a region V_1, separated from another homogeneous medium in a region V_2 by a surface S.

We will define auxiliary fields ψ_1 and ψ_2, which depend on a single unknown function η defined on S. This function obeys an integral equation that is an expression of the original boundary conditions. Then the field ψ can be obtained from η and the incident field by integrations.

The real field $\psi(\mathbf{x},t)$ satisfies the wave equations

$$\ddot{\psi}/v_1^2 - \nabla^2\psi = 0, \quad \mathbf{x}\in V_1, \tag{1}$$

$$\ddot{\psi}/v_2^2 - \nabla^2\psi = 0, \quad \mathbf{x}\in V_2, \tag{2}$$

where a dot is used to indicate a derivative with respect to time and v_1 and v_2 are the speeds of propagation of the wave in V_1 and V_2, respectively. In addition, ψ satisfies the boundary conditions

$$\psi_+ = \psi_-, \quad \mathbf{x}\in S, \tag{3}$$

$$\left(\frac{\partial\psi}{\partial n}\right)_+ = \alpha\left(\frac{\partial\psi}{\partial n}\right)_-, \quad \mathbf{x}\in S, \tag{4}$$

where the normal derivative is defined by

$$\frac{\partial\psi}{\partial n} = \hat{n}\cdot\nabla\psi, \tag{5}$$

in terms of the unit normal $\hat{n}$ to S that points from V_1 to V_2, and where the subscripts $+$ and $-$ indicate the limiting values of fields in V_2 and V_1, respectively. The initial values of the field ψ and its time derivative $\dot{\psi}$ are also given, and they vanish in V_2.

We apply the results derived in the theory of distributions to a field χ that obeys the same wave equation in V_1 and V_2, vanishes for times before the initial time $t=0$, and is discontinuous on the surface S. We use the same symbol χ for both the distribution and the function, and indicate derivatives in the sense of functions by curly brackets. Then, when the second derivatives of χ are integrable functions almost everywhere, χ satisfies

$$\{\ddot{\chi}\}/v^2 - \{\nabla^2\chi\} = \beta, \tag{6}$$

where $\beta(\mathbf{x},t)$ represents the sources of the field. The wave equation in the sense of distributions reduces to[5,7]

$$\ddot{\chi}/v^2 - \nabla^2\chi = \beta + v^{-2}\chi(\mathbf{x},0)\delta'(t) + v^{-2}\dot{\chi}(\mathbf{x},0)\delta(t) - \eta\delta(S) - \nabla\cdot[\hat{n}\phi\delta(S)] = \breve{\beta}, \tag{7}$$

where $\breve{\beta}$ is a source distribution that include the source function β, the initial values of χ and $\dot{\chi}$, and the jumps $\phi = \Delta\chi$ of the function and $\eta = \Delta(\partial\chi/\partial n)$ of the normal derivative on S. If $\mathcal{G}$ is the elementary solution of the wave equation, the solution of Eq. (7) is given by the convolution product

$$\chi = \mathcal{G} * \breve{\beta}. \tag{8}$$

When the sources vanish and the surface S is disregarded, Eq. (8) has the form

$$\chi = \mathcal{G} * f(\mathbf{x})\delta'(t) + \mathcal{G} * g(\mathbf{x})\delta(t), \tag{9}$$

while, if the sources and initial values vanish, Eq. (8) becomes

$$\chi = -\mathcal{G} * \eta\delta(S) - \mathcal{G} * \nabla\cdot[\hat{n}\phi\delta(S)]. \tag{10}$$

We now examine in what sense the jumps in χ and $\partial\chi/\partial n$ have the specified values. The elementary solution of the wave equation is given by[5,7]

$$\langle \mathcal{G},\varphi\rangle = \int_0^\infty \frac{dt}{4\pi t}\oint_{S(vt)} dS\,\varphi(\mathbf{x},t), \tag{11}$$

where $S(a)$ is a sphere of radius a centered at the origin and φ is a test function. We rewrite this equation in the form

$$\langle \mathcal{G},\varphi\rangle = \int dV\,\varphi(\mathbf{x},r/v)/4\pi r, \tag{12}$$

where $r = |\mathbf{x}|$. We use Eq. (11) to write

$$\langle \mathcal{G} * g(\mathbf{x})\delta(t),\varphi\rangle = \int_0^\infty \frac{dt}{4\pi t}\oint_{S(vt)} dS \int dV'\, g(\mathbf{x}')\varphi(\mathbf{x}+\mathbf{x}',t), \tag{13}$$

and we change the variables $\mathbf{x}'$ to $\mathbf{x}'' = \mathbf{x} + \mathbf{x}'$ and the order of integration to obtain

$$\langle \mathcal{G} * g(\mathbf{x})\delta(t),\varphi\rangle = \int_0^\infty \frac{dt}{4\pi t}\int dV''\oint_{S(vt)} dS\, g(\mathbf{x}''-\mathbf{x})\varphi(\mathbf{x}'',t). \tag{14}$$

We change dummy variables from $\mathbf{x}$ to $\mathbf{x}'$ and from $\mathbf{x}''$ to $\mathbf{x}$, and find

$$\mathcal{G} * g(\mathbf{x})\delta(t) = \frac{\theta(t)}{4\pi t}\oint_{S(vt)} dS'\, g(\mathbf{x}-\mathbf{x}'), \tag{15}$$

where $\theta(t)$ is the unit step function. Similarly, we find

$$\mathcal{G} * f(\mathbf{x})\delta'(t) = -\frac{\theta(t)}{4\pi t^2}\oint_{S(vt)} dS'\, f(\mathbf{x}-\mathbf{x}'). \tag{16}$$

We note that the surface element dS' in Eqs. (15) and (16) has a factor v^2t^2, so that the expressions do not diverge when $t\to 0$. The integrals in Eqs. (15) and (16) are essentially the averages of the functions f and g over a sphere of radius vt centered at $\mathbf{x}$.

The singular distribution $\eta\delta(S)$ is defined by

$$\langle \eta\delta(S),\varphi\rangle = \int_{-\infty}^{\infty} dt\oint_S dS\,\eta(\mathbf{x},t)\varphi(\mathbf{x},t), \tag{17}$$

and, combining this expression with Eq. (12), we find

$$\langle \mathcal{G} * \eta\delta(S),\varphi\rangle = \int \frac{dV'}{4\pi r'}\int_{-\infty}^{\infty} dt\oint_S dS\,\eta(\mathbf{x},t)\varphi(\mathbf{x}+\mathbf{x}',t+r'/v). \tag{18}$$

Changing the variables of integration $\mathbf{x}'$ and t to $\mathbf{x}'' = \mathbf{x} + \mathbf{x}'$ and $t'' = t + r'/v$, we obtain

$$\langle \mathcal{G} * \eta\delta(S),\varphi\rangle = \int_{-\infty}^{\infty} dt'' \int dV''\left[\oint_S dS\,\frac{\eta(\mathbf{x},t''-|\mathbf{x}''-\mathbf{x}|/v)}{4\pi|\mathbf{x}''-\mathbf{x}|}\right] \times\varphi(\mathbf{x}'',t''), \tag{19}$$

whence

$$\mathcal{G} * \eta\delta(S) = \oint_S dS'\,\eta(\mathbf{x}',t-R/v)/4\pi R, \tag{20}$$

where $\mathbf{R} = \mathbf{x} - \mathbf{x}'$ and $R = |\mathbf{R}|$. Similarly,

$$\langle \nabla\cdot[\hat{n}\phi\delta(S)],\varphi\rangle = -\int_{-\infty}^{\infty} dt\oint_S dS\,\phi(\mathbf{x},t)\hat{n}\cdot\nabla\varphi(\mathbf{x},t), \tag{21}$$

and we have, after the same substitutions,

$$\langle \mathcal{G} * \nabla\cdot[\hat{n}\phi\delta(S)],\varphi\rangle = -\int_{-\infty}^{\infty} dt'' \int dV''\left[\oint_S d\mathbf{S}\,\frac{\phi(\mathbf{x},t''-|\mathbf{x}''-\mathbf{x}|/v)}{4\pi|\mathbf{x}''-\mathbf{x}|}\right] \cdot\nabla''\varphi(\mathbf{x}'',t''), \tag{22}$$

where the right-hand side of Eq. (22) corresponds to the divergence of the vector-valued distribution in the square brackets. To simplify the notation, we define retarded functions by setting

$$\eta_{\rm ret}(\mathbf{x}',t') = \eta(\mathbf{x}',t-R/v), \tag{23}$$

and we use the functional notation

$$G\{\eta\} = -\oint_S dS'\,\eta_{\rm ret}(\mathbf{x}',t')/4\pi R \tag{24}$$

for minus the integral in Eq. (20).

If the field point $\mathbf{x}$ is not on S, the integrand in Eq. (22) does not become singular, and we can write

$$\left\{\nabla\cdot\oint_S dS'\,\frac{\phi_{\rm ret}(\mathbf{x}',t')}{4\pi R}\right\} = -\oint_S d\mathbf{S}'\cdot\left[\frac{\dot{\phi}_{\rm ret}(\mathbf{x}',t')\mathbf{R}}{4\pi vR^2} + \frac{\phi_{\rm ret}(\mathbf{x}',t')\mathbf{R}}{4\pi R^3}\right]. \tag{25}$$

When $\mathbf{x}$ is on S, the second term on the right-hand side of Eq.

(25) is not an integrable function. Since the distribution χ is independent of the values of the function χ on the surface, we can define the function on the surface by choosing the principal values of the integrals. The principal value of a surface integral at the point $\mathbf{x}$ on S is defined as the limit of the integral over the surface excluding a small patch about $\mathbf{x}$ when the size of the patch tends to zero. We now compute the limit of $\chi(\mathbf{x},t)$ and $\mathbf{x}$ approaches a point $\mathbf{x}_0$ on S from either side of the surface along the normal $\hat{n}$; we have

$$\mathbf{x} = \mathbf{x}_0 + h\hat{n}. \tag{26}$$

We separate the surface S into a patch S_1 about $\mathbf{x}_0$ and the remainder $\bar{S}$. When ϕ is well behaved on S_1, we can approximate $\phi_{\rm ret}(\mathbf{x}',t')$ by the constant value $\phi(\mathbf{x}_0,t)$ and show that the difference between the two is of the order of R; the integral of this difference then tends to a quantity of the order of the measure of the patch as $h\to 0$. The integral of $\phi(\mathbf{x}_0,t)$ over the patch is then proportional to

$$\int_{S_1} d\mathbf{S}'\cdot\mathbf{R}/R^3 = \Omega, \tag{27}$$

where Ω is the solid angle subtended by S_1 and $\mathbf{x}$, which also remains finite as $h\to 0$, and so does the integral over $\bar{S}$. Thus, the limits of the function χ as $\mathbf{x}$ approaches $\mathbf{x}_0$ from either side of the surface are finite. We characterize the size of the patch by a parameter a and, if the surface is smooth at $\mathbf{x}_0$, we find

$$\lim_{a\to 0}\lim_{h\to 0\pm}\int_{S_1}\frac{d\mathbf{S}'\cdot\mathbf{R}}{R^3} = \pm 2\pi. \tag{28}$$

Similarly,

$$\lim_{a\to 0}\lim_{h\to 0\pm}\int_{S_1}\frac{d\mathbf{S}'\cdot\mathbf{R}}{R^2} = 0, \tag{29}$$

$$\lim_{a\to 0}\lim_{h\to 0\pm}\int_{S_1}\frac{dS'}{R} = 0. \tag{30}$$

We write the function χ in the form

$$\chi\{\mathbf{x},t\} = G\{\eta\} + N\{\phi\}, \tag{31}$$

where $G\{\eta\}$ is defined in Eq. (24) and $N\{\phi\}$ stands for

$$N\{\phi\} = \oint_S d\mathbf{S}'\cdot\frac{\dot{\phi}_{\rm ret}(\mathbf{x}',t')\mathbf{R}}{4\pi v R^2} + \mathrm{P}\oint_S d\mathbf{S}'\cdot\frac{\phi_{\rm ret}(\mathbf{x}',t')\mathbf{R}}{4\pi R^3}. \tag{32}$$

In the limit $a\to 0$, the integrals over $\bar{S}$ tend to the principal values that define χ on the surface, the integrals over S_1 are evaluated with the help of Eqs. (28)–(30), and we obtain

$$\chi_\pm(\mathbf{x},t) = \pm\tfrac{1}{2}\phi(\mathbf{x},t) + \chi(\mathbf{x},t), \quad \mathbf{x}\in S. \tag{33}$$

The discontinuity of χ across S is equal to ϕ regardless of the definition of χ on S.

We can compute the gradient of χ from Eq. (10), which gives

$$\langle\nabla\chi,\varphi\rangle = \int_{-\infty}^{\infty} dt\int dV\left[\oint_S dS'\,\frac{\eta_{\rm ret}(\mathbf{x}',t')}{4\pi R}\nabla\varphi(\mathbf{x},t) - \oint_S d\mathbf{S}'\,\frac{\phi_{\rm ret}(\mathbf{x}',t')}{4\pi R}\cdot\nabla\nabla\varphi(\mathbf{x},t)\right]. \tag{34}$$

If $\mathbf{x}$ is not on S, we can integrate by parts to find the derivatives in the sense of functions,

$$\{\nabla\chi\} = \frac{1}{4\pi}\oint_S dS'\left(\frac{\dot{\eta}_{\rm ret}\mathbf{R}}{vR^2} + \frac{\eta_{\rm ret}\mathbf{R}}{R^3}\right) + \frac{1}{4\pi}\oint_S d\mathbf{S}'\cdot\left[\frac{\ddot{\phi}_{\rm ret}\mathbf{RR}}{v^2R^3} + \left(\frac{\dot{\phi}_{\rm ret}}{v} + \frac{\phi_{\rm ret}}{R}\right)\frac{3\mathbf{RR}-R^2\mathbf{I}}{R^4}\right], \tag{35}$$

where $\mathbf{I}$ is the unit 3×3 dyadic. In general, we see that the function $\nabla\chi$ diverges as $\mathbf{x}$ approaches S unless $\phi = 0$. To determine in what sense the discontinuity in the normal derivative of χ is actually η, we compute

$$\lim_{a\to 0}\lim_{h\to 0\pm}\int_{S_1} dS'\,\frac{\mathbf{R}}{R^2} = 0, \tag{36}$$

$$\lim_{a\to 0}\lim_{h\to 0\pm}\int_{S_1} dS'\,\frac{\mathbf{R}}{R^3} = \pm 2\pi\hat{n}, \tag{37}$$

$$\lim_{a\to 0}\lim_{h\to 0\pm}\int_{S_1} dS'\cdot\frac{\mathbf{RR}}{R^3} = 0. \tag{38}$$

The other two integrals diverge. If we approximate S_1 by a flat circular disk of radius a centered at $\mathbf{x}_0$, we have

$$\int_{S_1} d\mathbf{S}'\cdot\frac{3\mathbf{RR}-R^2\mathbf{I}}{R^4} = \hat{n}\int_0^a 2\pi\rho\,d\rho\,\frac{2h^2-\rho^2}{(h^2+\rho^2)^2} = \pi\left(\frac{3a^2}{a^2+h^2} - \log_e\frac{h^2+a^2}{h^2}\right)\hat{n}, \tag{39}$$

which diverges as $h\to 0$, and

$$\int_{S_1} d\mathbf{S}'\cdot\frac{3\mathbf{RR}-R^2\mathbf{I}}{R^5} = \hat{n}\int_0^a 2\pi\rho\,d\rho\,\frac{2h^2-\rho^2}{(h^2+\rho^2)^{5/2}} = \frac{2\pi a^2}{(h^2+a^2)^{3/2}}, \tag{40}$$

which implies that, for this special S_1,

$$\lim_{h\to 0\pm}\int_{S_1} d\mathbf{S}'\cdot\frac{3\mathbf{RR}-R^2\mathbf{I}}{R^5} = \frac{2\pi\hat{n}}{a}, \tag{41}$$

although the integral is not absolutely convergent. Nevertheless, these integrals do not depend on the sign of h and, if their contributions cancel when the difference of the limits from either side of S is taken, we can write

$$(\nabla\chi)_+ - (\nabla\chi)_- = \eta\hat{n}. \tag{42}$$

For the special case in which $\phi = 0$, we can again define the value of $\nabla\chi$ on the surface by the principal value integrals, and we multiply by the normal to find

$$\left(\frac{\partial\chi}{\partial n}\right)_\pm = \pm\frac{1}{2}\eta + \frac{\partial\chi}{\partial n}, \tag{43}$$

$$\left(\frac{\partial\chi}{\partial n}\right)_{\phi=0} = N'\{\eta\}, \tag{44}$$

where

$$N'\{\eta\} = \hat{n}\cdot\mathrm{P}\oint_S dS'\,\frac{1}{4\pi}\left(\frac{\dot{\eta}_{\rm ret}\mathbf{R}}{vR^2} + \frac{\eta_{\rm ret}\mathbf{R}}{R^3}\right). \tag{45}$$

We return to the original problem and first decompose the field in V_1 into the incident and scattered fields:

$$\psi = \psi^{\text{in}} + \psi^{\text{sc}}, \quad \mathbf{x}\in V_1. \tag{46}$$

The incident field satisfies the initial conditions and the homogeneous wave equation, and is defined everywhere by assuming that the medium in V_1 fills all space. Then ψ^{in} is given by Eq. (9), which becomes

$$\psi^{\text{in}} = \mathcal{G}_1 * [\psi(\mathbf{x},0)\delta'(t) + \dot{\psi}(\mathbf{x},0)\delta(t)], \tag{47}$$

where we use the subscript 1 on $\mathcal{G}_1$ to indicate that we have to set $v = v_1$ in Eq. (11). Equation (47), with the help of Eqs. (15) and (16), gives ψ^{in} in terms of the initial conditions for ψ, and we can assume that it is a known field.

We first show how a pair of coupled integral equations can be derived for two unknown functions on S. We assume that ψ^{sc} vanishes in V_2 and that its boundary values in V_1 are ϕ_1 and η_1 for ψ^{sc} and $\partial\psi^{\text{sc}}/\partial n$, respectively. Then Eq. (10) gives

$$\psi^{\text{sc}} = \mathcal{G}_1 * \eta_1 \delta(S) + \mathcal{G}_1 * \nabla\cdot[\hat{n}\phi_1\delta(S)]. \tag{48}$$

Similarly, if the transmitted field vanishes in V_1, by Eqs. (3), (4), and (10) we have

$$\psi^{\text{tr}} = -\alpha\mathcal{G}_2 * \left(\frac{\partial\psi^{\text{in}}}{\partial n} + \eta_1\right)\delta(S) - \mathcal{G}_2 * \nabla\cdot[\hat{n}(\psi^{\text{in}} + \phi_1)\delta(S)]. \tag{49}$$

We obtain two integral equations for ϕ_1 and η_1 if we let the field point in Eqs. (48) and (49) tend to S and set

$$\psi^{\text{sc}}_+ = 0, \tag{50}$$

$$\psi^{\text{tr}}_- = 0. \tag{51}$$

Once ϕ_1 and η_1 are known, Eqs. (48) and (49) give ψ^{sc} in V_1 and ψ^{tr} in V_2.

To reduce the number of unknown functions on S to one, we define two auxiliary fields ψ_1 and ψ_2 which satisfy

$$\psi_1(\mathbf{x},t) = \psi^{\text{sc}}(\mathbf{x},t), \quad \mathbf{x}\in V_1, \tag{52}$$

$$\ddot{\psi}_1(\mathbf{x},t)/v_1^2 - \nabla^2\psi_1(\mathbf{x},t) = 0, \quad \mathbf{x}\in V_2, \tag{53}$$

$$\psi_{1+} = \psi_{1-} = \psi^{\text{sc}}_-, \quad \mathbf{x}\in S, \tag{54}$$

$$\Delta\left(\frac{\partial\psi_1}{\partial n}\right) = \eta, \quad \mathbf{x}\in S, \tag{55}$$

$$\psi_2(\mathbf{x},t) = 0, \quad \mathbf{x}\in V_1, \tag{56}$$

$$\psi_2(\mathbf{x},t) = \psi(\mathbf{x},t), \quad \mathbf{x}\in V_2, \tag{57}$$

plus homogeneous initial conditions for both ψ_1 and ψ_2; our unknown function is η, the discontinuity of the normal derivative of ψ_1 across S. Since ψ_1 is continuous across S and satisfies the same wave equation in V_1 and V_2, Eq. (31) gives

$$\psi_1 = G_1\{\eta\}. \tag{58}$$

The field ψ_2 also obeys the same wave equation with $v = v_2$ in V_1 and V_2, and the jumps of ψ_2 and $\partial\psi_2/\partial n$ are given by

$$\Delta\psi_2 = \psi_{2+} = \psi_+ = \psi_- = \psi^{\text{in}} + \psi_{1-}, \tag{59}$$

$$\Delta\left(\frac{\partial\psi_2}{\partial n}\right) = \left(\frac{\partial\psi_2}{\partial n}\right)_+ = \left(\frac{\partial\psi}{\partial n}\right)_+ = \alpha\left(\frac{\partial\psi}{\partial n}\right)_- = \alpha\frac{\partial\psi^{\text{in}}}{\partial n} + \alpha\left(\frac{\partial\psi_1}{\partial n}\right)_-, \tag{60}$$

that is, from Eqs. (58), (43), and (44),

$$\Delta\psi_2 = \psi^{\text{in}} + G_1\{\eta\}, \quad \mathbf{x}\in S, \tag{61}$$

$$\Delta\left(\frac{\partial\psi_2}{\partial n}\right) = \alpha\frac{\partial\psi^{\text{in}}}{\partial n} - \tfrac{1}{2}\alpha\eta + \alpha N_1'\{\eta\}, \quad \mathbf{x}\in S, \tag{62}$$

where the jumps are expressed in terms on η. Equation (31) then gives

$$\psi_2 = \alpha G_2\left\{\frac{\partial\psi^{\text{in}}}{\partial n} - \tfrac{1}{2}\eta + N_1'\{\eta\}\right\} + N_2\{\psi^{\text{in}} + G_1\{\eta\}\}. \tag{63}$$

This function ψ_2 does not vanish in V_1 for arbitrary η, so we obtain the integral equation for η by imposing the condition

$$\psi_{2-} = 0. \tag{64}$$

We use Eq. (33) to express ψ_{2-} in terms of the jump of ψ_2 across S, given by Eq. (61), and the value of ψ_2 on S from Eq. (63). After rearranging the terms, we obtain

$$[\tfrac{1}{2}G_1 - N_2G_1 + \alpha(\tfrac{1}{2}G_2 - G_2N_1')]\{\eta\} + \tfrac{1}{2}\psi^{\text{in}} - N_2\{\psi^{\text{in}}\} - \alpha G_2\left\{\frac{\partial\psi^{\text{in}}}{\partial n}\right\} = 0, \tag{65}$$

where the composite operators are defined as in

$$N_2G_1\{\eta\} = N_2\{G_1\{\eta\}\}, \tag{66}$$

in terms of the operators (24), (32), and (45). Equation (65) is a singular integral equation of the first kind.

To show that the field ψ_2 actually vanishes in V_1, we use the uniqueness theorem that follows from the energy relation

$$\nabla\cdot(\dot{\chi}\nabla\chi) = \frac{1}{2}\frac{\partial}{\partial t}\left[\frac{\dot{\chi}^2}{v^2} + (\nabla\chi)^2\right], \tag{67}$$

a consequence of the homogeneous wave equation. We integrate over the volume V_1 and use Gauss's divergence theorem to derive

$$\oint_S dS\,\dot{\chi}\frac{\partial\chi}{\partial n} = \frac{1}{2}\frac{d}{dt}\int_{V_1} dV\left[\frac{\dot{\chi}^2}{v^2} + (\nabla\chi)^2\right]. \tag{68}$$

If the volume V_1 goes out to infinity, we use causality to show that there is no contribution from that part of the surface integral. If χ vanishes on S, so does $\dot{\chi}$ and Eq. (68) shows that the integral on the right-hand side is a constant. If χ and $\dot{\chi}$ vanish initially in V_1, this integral vanishes and, since the integrand is nonnegative, both $\dot{\chi}$ and $\nabla\chi$ vanish throughout V_1, that is, χ is a constant field in V_1, and this constant is zero because χ vanishes on S. The field ψ_2 satisfies these conditions, showing that it has to vanish in V_1; in particular,

$$\left(\frac{\partial\psi_2}{\partial n}\right)_- = 0. \tag{69}$$

Once η is determined from Eq. (65), the scattered and transmitted fields can be found from Eqs. (58) and (63), that is,

$$\psi(\mathbf{x},t) = \psi^{\text{in}}(\mathbf{x},t) + G_1\{\eta\}, \quad \mathbf{x}\in V_1, \tag{70}$$

$$\psi(\mathbf{x},t) = N_2\{\psi^{\text{in}}\} + \alpha G_2\left\{\frac{\partial\psi^{\text{in}}}{\partial n}\right\} + [N_2G_1 + \alpha(-\tfrac{1}{2}G_2 + G_2N_1')]\{\eta\}, \quad \mathbf{x}\in V_2. \tag{71}$$

It is obvious that this field satisfies Eqs. (1) and (2), as well as the initial conditions. To verify that the boundary conditions

(3) and (4) are satisfied, we first note that, since $\psi = \psi_2$ in V_2 and ψ_2 satisfies the boundary conditions (64) and (69), the boundary values ψ_+ and $(\partial\psi/\partial n)_+$ are equal to the jumps of ψ_2 and $\partial\psi_2/\partial n$ given by Eqs. (61) and (62). Since both terms in Eq. (70) are continuous on S, we have

$$\psi_- = \psi^{\text{in}} + G_1\{\eta\}, \quad \mathbf{x}\in S, \tag{72}$$

and Eq. (3) is satisfied. Furthermore, if we apply Eqs. (43) and (44) to the second term in Eq. (70), we obtain

$$\left(\frac{\partial\psi}{\partial n}\right)_- = \frac{\partial\psi^{\text{in}}}{\partial n} - \frac{1}{2}\eta + N_1'\{\eta\}, \quad \mathbf{x}\in S; \tag{73}$$

comparison with Eq. (62) shows that the boundary condition (4) is also satisfied. Thus, ψ satisfies all the imposed conditions.

III. ELECTROMAGNETIC FIELDS

We represent the electromagnetic fields by four vector-valued distributions that vanish for negative times and correspond to locally integrable functions with derivatives defined in the sense of functions in regions V_1 and V_2, separated by a surface S.

We extend the procedure used in the previous section for the scalar field ψ to apply to the pair of vector fields $\mathbf{E}$ and $\mathbf{B}$. We will use auxiliary fields $\mathbf{E}_1$, $\mathbf{B}_1$, $\mathbf{E}_2$, and $\mathbf{B}_2$ to express the fields $\mathbf{E}$ and $\mathbf{B}$ in terms of the incident fields and a single tangential vector field $\hat{n}\times\boldsymbol{\eta}$ defined on S. This field obeys an integral equation which is a consequence of the original boundary conditions.

Maxwell's equations in the sense of distributions are

$$\nabla\cdot\mathbf{D} = \check{\rho}, \tag{74}$$

$$\nabla\cdot\mathbf{B} = 0, \tag{75}$$

$$\nabla\times\mathbf{E} + \dot{\mathbf{B}} = \mathbf{B}_0(\mathbf{x})\delta(t), \tag{76}$$

$$\nabla\times\mathbf{H} - \dot{\mathbf{D}} = \check{\mathbf{j}} - \mathbf{D}_0(\mathbf{x})\delta(t), \tag{77}$$

where we use the inverted caret on the sources to indicate that they are distributions that may include singular parts, that is,

$$\check{\rho} = \rho_0 + \rho_s\delta(S) + \sum q_i\delta^{(3)}(\mathbf{x} - \mathbf{x}_i), \tag{78}$$

where ρ is the volume charge density, ρ_s is the surface charge density, and the last term represents point charges, which will not be considered further in this paper. The current density includes the surface current density $\mathbf{J}_s$ in

$$\check{\mathbf{j}} = \mathbf{j} + \mathbf{J}_s\delta(S). \tag{79}$$

In Eqs. (76) and (77) the meaning of the source distributions differ from that of $\check{\beta}$ in Eq. (7), since they do not include the initial values and jump terms.

The terms proportional to $\delta(t)$ represent impulses that cause the initial conditions to be satisfied. In this form, Maxwell's equations contain both the equations in the sense of functions and the boundary conditions. The derivatives of the distributions are

$$\nabla\cdot\mathbf{D} = \{\nabla\cdot\mathbf{D}\} + \Delta(\hat{n}\cdot\mathbf{D})\,\delta(S), \tag{80}$$

$$\nabla\cdot\mathbf{B} = \{\nabla\cdot\mathbf{B}\} + \Delta(\hat{n}\cdot\mathbf{B})\,\delta(S), \tag{81}$$

$$\nabla\times\mathbf{E} + \dot{\mathbf{B}} = \{\nabla\times\mathbf{E}\} + \Delta(\hat{n}\times\mathbf{E})\,\delta(S) + \{\dot{\mathbf{B}}\} + \mathbf{B}(\mathbf{x},0)\delta(t), \tag{82}$$

$$\nabla\times\mathbf{H} - \dot{\mathbf{D}} = \{\nabla\times\mathbf{H}\} + \Delta(\hat{n}\times\mathbf{H})\,\delta(S) - \{\dot{\mathbf{D}}\} - \mathbf{D}(\mathbf{x},0)\delta(t), \tag{83}$$

and we can equate the regular and the singular parts in each equation to derive

$$\{\nabla\cdot\mathbf{D}\} = \rho, \tag{84}$$

$$\{\nabla\cdot\mathbf{B}\} = 0, \tag{85}$$

$$\{\nabla\times\mathbf{E}\} + \{\dot{\mathbf{B}}\} = 0, \tag{86}$$

$$\{\nabla\times\mathbf{H}\} - \{\dot{\mathbf{D}}\} = \mathbf{j}, \tag{87}$$

$$\Delta(\hat{n}\cdot\mathbf{D}) = \rho_s, \tag{88}$$

$$\Delta(\hat{n}\cdot\mathbf{B}) = 0, \tag{89}$$

$$\Delta(\hat{n}\times\mathbf{E}) = 0, \tag{90}$$

$$\Delta(\hat{n}\times\mathbf{H}) = \mathbf{J}_s. \tag{91}$$

The constitutive relations

$$\mathbf{D} = \epsilon\mathbf{E}, \tag{92}$$

$$\mathbf{B} = \mu\mathbf{H} \tag{93}$$

are not generally defined in terms of distributions. In this section we assume that ϵ and μ are equal to different constants throughout V_1 and V_2, where the fields correspond to functions.

We now consider an incident electromagnetic field given by the initial values $\mathbf{E}_0(\mathbf{x})$ and $\mathbf{B}_0(\mathbf{x})$ in V_1, and we seek to determine the scattered and transmitted fields. As before, we separate the fields in V_1 into incident and scattered fields,

$$\mathbf{E} = \mathbf{E}^{\text{in}} + \mathbf{E}^{\text{sc}}, \quad \mathbf{x}\in V_1, \tag{94}$$

$$\mathbf{B} = \mathbf{B}^{\text{in}} + \mathbf{B}^{\text{sc}}, \quad \mathbf{x}\in V_1, \tag{95}$$

where the incident fields satisfy the initial conditions. We also have to satisfy the boundary conditions (90) and (91); the surface current density vanishes unless one of the regions corresponds to a perfect conductor. These conditions can be written in the form

$$\hat{n}\times\mathbf{E}_+ = \hat{n}\times\mathbf{E}_-, \tag{96}$$

$$\hat{n}\times\mathbf{B}_+ = \alpha\hat{n}\times\mathbf{B}_-, \tag{97}$$

where

$$\alpha = \mu_2/\mu_1. \tag{98}$$

The incident fields are given by[8]

$$\mathbf{E}^{\text{in}} = v_1^{-2}\dot{\mathscr{G}}_1 * \mathbf{E}_0\delta(t) + \nabla\mathscr{G}_1 * \times\mathbf{B}_0\delta(t), \tag{99}$$

$$\mathbf{B}^{\text{in}} = v_1^{-2}[\dot{\mathscr{G}}_{1+} * \mathbf{B}_0\delta(t) - \nabla\mathscr{G}_1 * \times\mathbf{E}_0\delta(t)], \tag{100}$$

where we have both a convolution product and a vector product in the gradient terms and the speed of propagation is given by

$$v_1^2 = 1/(\epsilon_1\mu_1), \tag{101}$$

and the initial values satisfy the constraints

$$\nabla\cdot\mathbf{E}_0 = 0, \tag{102}$$

$$\nabla\cdot\mathbf{B}_0 = 0. \tag{103}$$

We verify that these fields obey the appropriate Maxwell equations; we have

$$\nabla\cdot\mathbf{E}^{\text{in}} = v_1^{-2}\dot{\mathscr{G}}_1 * \nabla\cdot\mathbf{E}_0\delta(t) + \mathscr{G}_1 * \nabla\cdot\nabla\times\mathbf{B}_0\delta(t) = 0, \tag{104}$$

$$\nabla\cdot\mathbf{B}^{\text{in}} = v_1^{-2}[\dot{\mathscr{G}}_1 * \nabla\cdot\mathbf{B}_0\delta(t) - \mathscr{G}_1 * \nabla\cdot\nabla\times\mathbf{E}_0\delta(t)] = 0, \tag{105}$$

$$\nabla\times\mathbf{E}^{\rm in}+\dot{\mathbf{B}}^{\rm in}$$
$$=v_1^{-2}\nabla\dot{\mathscr{G}}_1*\times\mathbf{E}_0\delta(t)+\mathbf{B}_0\delta(t)*\cdot\nabla\nabla\mathscr{G}_1$$
$$-\nabla^2\mathscr{G}_1*\mathbf{B}_0\delta(t)+v_1^{-2}\ddot{\mathscr{G}}_1*\mathbf{B}_0\delta(t)$$
$$-v_1^{-2}\nabla\dot{\mathscr{G}}_1*\times\mathbf{E}_0\delta(t)$$
$$=\nabla\cdot\mathbf{B}_0\delta(t)*\nabla\mathscr{G}_1+\delta*\mathbf{B}_0\delta(t)=\mathbf{B}_0\delta(t),\tag{106}$$
$$v_1^2\nabla\times\mathbf{B}^{\rm in}-\dot{\mathbf{E}}^{\rm in}$$
$$=\nabla\dot{\mathscr{G}}_1*\times\mathbf{B}_0\delta(t)-\mathbf{E}_0\delta(t)*\cdot\nabla\nabla\mathscr{G}_1$$
$$+\nabla^2\mathscr{G}_1*\mathbf{E}_0\delta(t)-v_1^{-2}\ddot{\mathscr{G}}_1*\mathbf{E}_0\delta(t)-\nabla\dot{\mathscr{G}}_1*\times\mathbf{B}_0\delta(t)$$
$$=-\nabla\cdot\mathbf{E}_0\delta(t)*\nabla\mathscr{G}_1-\delta*\mathbf{E}_0\delta(t)$$
$$=-\mathbf{E}_0\delta(t).\tag{107}$$

We now consider fields $\mathscr{E}$ and $\mathscr{B}$ that propagate in a homogeneous medium of permittivity ϵ and permeability μ, satisfy the homogeneous Maxwell equations and initial conditions, and have jumps equal to ϕ and η, respectively, on the surface S. They are given by

$$\mathscr{E}=-\nabla\mathscr{G}*\hat{n}\cdot\boldsymbol{\phi}\delta(S)+\nabla\mathscr{G}*\times(\hat{n}\times\boldsymbol{\phi})\delta(S)$$
$$-\dot{\mathscr{G}}*\hat{n}\times\boldsymbol{\eta}\delta(S),\tag{108}$$
$$\mathscr{B}=-\nabla\mathscr{G}*\hat{n}\cdot\boldsymbol{\eta}\delta(S)+\nabla\mathscr{G}*\times(\hat{n}\times\boldsymbol{\eta})\delta(S)$$
$$+v^{-2}\dot{\mathscr{G}}*\hat{n}\times\boldsymbol{\phi}\delta(S),\tag{109}$$

and, when we take the appropriate derivatives, we obtain

$$\nabla\cdot\mathscr{E}=-\nabla^2\mathscr{G}*\hat{n}\cdot\boldsymbol{\phi}\delta(S)+\nabla\times\nabla\mathscr{G}*\cdot\hat{n}\times\boldsymbol{\phi}\delta(S)$$
$$-\nabla\dot{\mathscr{G}}*\cdot\hat{n}\times\boldsymbol{\eta}\delta(S)$$
$$=(\delta-v^{-2}\ddot{\mathscr{G}})*\hat{n}\cdot\boldsymbol{\phi}\delta(S)-\dot{\mathscr{G}}*\nabla\cdot[\hat{n}\times\boldsymbol{\eta}\delta(S)]$$
$$=\hat{n}\cdot\boldsymbol{\phi}\delta(S)-v^{-2}\dot{\mathscr{G}}*[\hat{n}\cdot\dot{\boldsymbol{\phi}}\delta(S)+v^2\nabla\cdot(\hat{n}\times\boldsymbol{\eta}\delta(S))],\tag{110}$$
$$\nabla\cdot\mathscr{B}=-\nabla^2\mathscr{G}*\hat{n}\cdot\boldsymbol{\eta}\delta(S)+\nabla\times\nabla\mathscr{G}*\cdot\hat{n}\times\boldsymbol{\eta}\delta(S)$$
$$+v^{-2}\nabla\dot{\mathscr{G}}*\cdot\hat{n}\times\boldsymbol{\phi}\delta(S)$$
$$=(\delta-v^{-2}\ddot{\mathscr{G}})*\hat{n}\cdot\boldsymbol{\eta}\delta(S)+v^{-2}\dot{\mathscr{G}}*\nabla\cdot[\hat{n}\times\boldsymbol{\phi}\delta(S)]$$
$$=\hat{n}\cdot\boldsymbol{\eta}\delta(S)-v^{-2}\dot{\mathscr{G}}*[\hat{n}\cdot\dot{\boldsymbol{\eta}}\delta(S)-\nabla\cdot(\hat{n}\times\boldsymbol{\phi}\delta(S))],\tag{111}$$
$$\nabla\times\mathscr{E}+\dot{\mathscr{B}}=-\nabla\times\nabla\mathscr{G}*\hat{n}\cdot\boldsymbol{\phi}\delta(S)-\nabla^2\mathscr{G}*\hat{n}\times\boldsymbol{\phi}\delta(S)$$
$$+\hat{n}\times\boldsymbol{\phi}\delta(S)*\cdot\nabla\nabla\mathscr{G}-\nabla\dot{\mathscr{G}}*\times(\hat{n}\times\boldsymbol{\eta})\delta(S)$$
$$-\nabla\dot{\mathscr{G}}*\hat{n}\cdot\boldsymbol{\eta}\delta(S)+\nabla\dot{\mathscr{G}}*\times(\hat{n}\times\boldsymbol{\eta})\delta(S)$$
$$+v^{-2}\ddot{\mathscr{G}}*\hat{n}\times\boldsymbol{\phi}\delta(S)$$
$$=\hat{n}\times\boldsymbol{\phi}\delta(S)-\nabla\mathscr{G}*[\hat{n}\cdot\dot{\boldsymbol{\eta}}\delta(S)-\nabla\cdot(\hat{n}\times\boldsymbol{\phi}\delta(S))],\tag{112}$$
$$\nabla\times\mathscr{B}-v^{-2}\dot{\mathscr{E}}$$
$$=-\nabla\times\nabla\mathscr{G}*\hat{n}\cdot\boldsymbol{\eta}\delta(S)-\nabla^2\mathscr{G}*\hat{n}\times\boldsymbol{\eta}\delta(S)$$
$$+\hat{n}\times\boldsymbol{\eta}\delta(S)*\cdot\nabla\nabla\mathscr{G}+\nabla\dot{\mathscr{G}}*\times(\hat{n}\times\boldsymbol{\phi})\delta(S)$$
$$+v^{-2}[\nabla\dot{\mathscr{G}}*\hat{n}\cdot\boldsymbol{\phi}\delta(S)-\nabla\dot{\mathscr{G}}*\times(\hat{n}\times\boldsymbol{\phi})\delta(S)$$
$$+\ddot{\mathscr{G}}*\hat{n}\times\boldsymbol{\eta}\delta(S)]$$
$$=\hat{n}\times\boldsymbol{\eta}\delta(S)+v^{-2}\nabla\mathscr{G}*[\hat{n}\cdot\dot{\boldsymbol{\phi}}\delta(S)$$
$$+v^2\nabla\cdot(\hat{n}\times\boldsymbol{\eta}\delta(S))].\tag{113}$$

Comparison with Eqs. (80)–(83) shows that we obtain the correct result provided we set

$$\hat{n}\cdot\dot{\boldsymbol{\phi}}=-v^2\nabla_s\cdot\hat{n}\times\boldsymbol{\eta},\tag{114}$$
$$\hat{n}\cdot\dot{\boldsymbol{\eta}}=\nabla_s\cdot\hat{n}\times\boldsymbol{\phi},\tag{115}$$

where the surface divergence of a tangential field defined on S satisfies[9]

$$\langle\nabla\cdot[\hat{n}\times\mathbf{f}\delta(S)],\varphi\rangle$$
$$=-\oint_S dS\,\hat{n}\times\mathbf{f}\cdot\nabla\varphi$$
$$=-\oint_S dS\hat{n}\times\mathbf{f}\cdot\nabla_s\varphi=\oint_S dS\,[\nabla_s\cdot(\hat{n}\times\mathbf{f})]\varphi$$
$$=\langle[\nabla_s\cdot\hat{n}\times\mathbf{f}]\delta(S),\varphi\rangle.\tag{116}$$

Equations (114) and (115) are a consequence of Maxwell's equations, and they can be used to determine the jumps in the normal components of $\mathscr{E}$ and $\mathscr{B}$ in terms of the jumps of the tangential components of $\mathscr{B}$ and $\mathscr{E}$, respectively.

As in the previous section, the fields are well defined off the surface S, and we extend the definitions to S by choosing the principal value of the integrals. We find that Eqs. (108) and (109) lead to

$$\mathscr{E}(\mathbf{x},t)=\frac{1}{4\pi}\,\mathrm{P}\oint_S dS'\left[\frac{\hat{n}'\cdot\dot{\boldsymbol{\phi}}_{\rm ret}\times\mathbf{R}}{vR^2}+\frac{\hat{n}\cdot\boldsymbol{\phi}_{\rm ret}\mathbf{R}}{R^3}\right.$$
$$\left.+\frac{(\hat{n}\times\dot{\boldsymbol{\phi}}_{\rm ret})\times\mathbf{R}}{vR^2}+\frac{(\hat{n}'\times\boldsymbol{\phi}_{\rm ret})\times\mathbf{R}}{R^3}-\frac{(\hat{n}'\times\dot{\boldsymbol{\eta}}_{\rm ret}}{R}\right],\tag{117}$$
$$\mathscr{B}(\mathbf{x},t)=\frac{1}{4\pi}\,\mathrm{P}\oint_S dS'\left[\frac{\hat{n}\cdot\dot{\boldsymbol{\eta}}_{\rm ret}\mathbf{R}}{vR^2}+\frac{\hat{n}'\cdot\boldsymbol{\eta}_{\rm ret}\mathbf{R}}{R^3}\right.$$
$$+\frac{(\hat{n}'\times\dot{\boldsymbol{\eta}}_{\rm ret})\times\mathbf{R}}{vR^2}+\frac{(\hat{n}'\times\boldsymbol{\eta}_{\rm ret})\times\mathbf{R}}{R^3}$$
$$\left.+\frac{\hat{n}'\times\dot{\boldsymbol{\phi}}_{\rm ret}}{v^2R}\right].\tag{118}$$

In applications to scattering problems, $\hat{n}\cdot\boldsymbol{\phi}$ and $\hat{n}\cdot\boldsymbol{\eta}$ vanish initially; Eqs. (114) and (115) then show how $\hat{n}\cdot\boldsymbol{\phi}$ is determined by $\hat{n}\times\boldsymbol{\eta}$ and $\hat{n}\cdot\boldsymbol{\eta}$ by $\hat{n}\times\boldsymbol{\phi}$, and the normal components of $\boldsymbol{\phi}$ and $\boldsymbol{\eta}$ can be eliminated from Eqs. (117) and (118) in terms of the tangential components. These equations have the functional form

$$\mathscr{E}=\mathbf{L}\{\hat{n}\times\boldsymbol{\phi}\}+\mathbf{M}\{\hat{n}\times\boldsymbol{\eta}\},\tag{119}$$
$$\mathscr{B}=\mathbf{L}\{\hat{n}\times\boldsymbol{\eta}\}+\mathbf{M}'\{\hat{n}\times\boldsymbol{\phi}\},\tag{120}$$

where

$$\mathbf{L}\{\hat{n}\times\boldsymbol{\phi}\}=\frac{1}{4\pi}\,\mathrm{P}\oint_S dS'$$
$$\times\left[\hat{n}'\times\left(\frac{\dot{\boldsymbol{\phi}}_{\rm ret}}{v}+\frac{\boldsymbol{\phi}_{\rm ret}}{R}\right)\right]\times\frac{\mathbf{R}}{R^2},\tag{121}$$
$$\mathbf{M}\{\hat{n}\times\boldsymbol{\eta}\}=\frac{1}{4\pi}\,\mathrm{P}\oint_S dS'$$
$$\times\left[\hat{n}'\cdot\left(\frac{\dot{\boldsymbol{\phi}}_{\rm ret}}{v}+\frac{\boldsymbol{\phi}_{\rm ret}}{R}\right)\frac{\mathbf{R}}{R^2}-\frac{\hat{n}'\times\dot{\boldsymbol{\eta}}_{\rm ret}}{R}\right],\tag{122}$$
$$\mathbf{M}'\{\hat{n}\times\boldsymbol{\phi}\}=\frac{1}{4\pi}\,\mathrm{P}\oint_S \mathrm{d}S'$$
$$\times\left[\hat{n}'\cdot\left(\frac{\dot{\boldsymbol{\eta}}_{\rm ret}}{v}+\frac{\boldsymbol{\eta}_{\rm ret}}{R}\right)\frac{\mathbf{R}}{R^2}+\frac{\hat{n}'\times\dot{\boldsymbol{\phi}}_{\rm ret}}{v^2R}\right].\tag{123}$$

These functionals depend on v both explicitly and through the retardation effects. From Eqs. (117) and (118) we find the

boundary values

$$\mathscr{E}_{\pm} = \pm\tfrac{1}{2}\phi + \mathscr{E}, \tag{124}$$

$$\mathscr{B}_{\pm} = \pm\tfrac{1}{2}\eta + \mathscr{B}. \tag{125}$$

We now define two sets of auxiliary fields. We assume that the fields $\mathbf{E}_1$ and $\mathbf{B}_1$ propagate in a medium of constants ϵ_1 and μ_1, that they are equal to $\mathbf{E}^{sc}$ and $\mathbf{B}^{sc}$ in V_1, and that the tangential component of $\mathbf{E}_1$ is continuous across S. These fields are then given by Eqs. (119) and (120). We set $\hat{n}\times\phi = 0$, which implies that $\hat{n}\cdot\eta = 0$, and we can write

$$\mathbf{E}_1 = \mathbf{M}_1\{\hat{n}\times\eta\}, \tag{126}$$

$$\mathbf{B}_1 = \mathbf{L}_1\{\hat{n}\times\eta\}; \tag{127}$$

the discontinuity $\hat{n}\times\eta$ of the tangential component of $\mathbf{B}_1$ is the unknown field on S for which we have to find an integral equation. The fields $\mathbf{E}_2$ and $\mathbf{B}_2$ are defined equal to the transmitted fields in V_2 and equal to zero in V_1. The discontinuities in the tangential components of $\mathbf{E}_2$ and $\mathbf{B}_2$ are then given by the boundary conditions (96) and (97), and we have

$$\begin{aligned}\Delta(\hat{n}\times\mathbf{E}_2) &= \hat{n}\times\mathbf{E}_{2+} = \hat{n}\times\mathbf{E}_{+} = \hat{n}\times\mathbf{E}_{-}\\ &= \hat{n}\times\mathbf{E}^{in} + \hat{n}\times\mathbf{E}_{1-},\end{aligned} \tag{128}$$

$$\begin{aligned}\Delta(\hat{n}\times\mathbf{B}_2) &= \hat{n}\times\mathbf{B}_{2+} = \hat{n}\times\mathbf{B}_{+} = \alpha\hat{n}\times\mathbf{B}_{-}\\ &= \alpha\hat{n}\times\mathbf{B}^{in} + \alpha\hat{n}\times\mathbf{B}_{1-}.\end{aligned} \tag{129}$$

Thus, from Eqs. (124)–(129) we find

$$\Delta(\hat{n}\times\mathbf{E}_2) = \hat{n}\times\mathbf{E}_{2+} = \hat{n}\times\mathbf{E}^{in} + \hat{n}\times\mathbf{M}_1\{\hat{n}\times\eta\}, \tag{130}$$

$$\begin{aligned}\Delta(\hat{n}\times\mathbf{B}_2) &= \hat{n}\times\mathbf{B}_{2+}\\ &= \alpha\hat{n}\times\mathbf{B}^{in} - \tfrac{1}{2}\alpha\hat{n}\times\eta + \alpha\hat{n}\times\mathbf{L}_1\{\hat{n}\times\eta\},\end{aligned} \tag{131}$$

which give these jumps in terms $\hat{n}\times\eta$. Then, by Eq. (119) and (120) we have

$$\mathbf{E}_2 = \mathbf{L}_2\{\hat{n}\times\mathbf{E}_{2+}\} + \mathbf{M}_2\{\hat{n}\times\mathbf{B}_{2+}\}, \tag{132}$$

$$\mathbf{B}_2 = \mathbf{L}_2\{\hat{n}\times\mathbf{B}_{2+}\} + \mathbf{M}_2'\{\hat{n}\times\mathbf{E}_{2+}\}. \tag{133}$$

We find an integral equation for $\hat{n}\times\eta$ by setting

$$\hat{n}\times\mathbf{E}_{2-} = 0, \tag{134}$$

which, by Eqs. (119) and (132), becomes

$$-\tfrac{1}{2}\hat{n}\times\mathbf{E}_{2+} + \hat{n}\times\mathbf{L}_2\{\hat{n}\times\mathbf{E}_{2+}\} + \hat{n}\times\mathbf{M}_2\{\hat{n}\times\mathbf{B}_{2+}\} = 0. \tag{135}$$

We substitute the expressions (130) and (131) for $\hat{n}\times\mathbf{E}_{2+}$ and $\hat{n}\times\mathbf{B}_{2+}$ to obtain

$$\begin{aligned}&\hat{n}\times[\tfrac{1}{2}\mathbf{M}_1 - \mathbf{L}_2\mathbf{M}_1 + \alpha(\tfrac{1}{2}\mathbf{M}_2 - \mathbf{M}_2\mathbf{L}_1)]\{\hat{n}\times\eta\}\\ &\quad + \hat{n}\times(\tfrac{1}{2}\mathbf{E}^{in} - \mathbf{L}_2\{\hat{n}\times\mathbf{E}^{in}\} - \alpha\mathbf{M}_2\{\hat{n}\times\mathbf{B}^{in}\}) = 0,\end{aligned} \tag{136}$$

where a composite operator such as $\mathbf{L}_2\mathbf{M}_1$ is defined by the relation

$$\hat{n}\times\mathbf{L}_2\mathbf{M}_1\{\hat{n}\times\eta\} = \hat{n}\times\mathbf{L}_2\{\hat{n}\times\mathbf{M}_1\{\hat{n}\times\eta\}\}. \tag{137}$$

The fields $\mathbf{E}_2$ and $\mathbf{B}_2$ given by Eqs. (132) and (133) obey the homogeneous Maxwell equations in V_1 and homogeneous initial conditions, and Eq. (134) states that $\hat{n}\times\mathbf{E}_2$ vanishes on the boundary when η satisfies the integral equation. To show that $\mathbf{E}_2$ and $\mathbf{B}_2$ actually vanish in V_1, we use the energy relation for fields that obey Maxwell's equations,

$$\nabla\cdot(\mathbf{E}\times\mathbf{H}) = -\mathbf{H}\cdot\dot{\mathbf{B}} - \mathbf{E}\cdot\dot{\mathbf{D}} - \mathbf{j}\cdot\mathbf{E}, \tag{138}$$

whence, if $\mathbf{j} = 0$ and μ and ϵ are constant in V_1,

$$\oint_S dS\,\hat{n}\cdot\mathbf{E}\times\mathbf{B} = -\frac{1}{2}\frac{d}{dt}\int_{V_1} dV\left(\mathbf{B}^2 + \frac{\mathbf{E}^2}{v^2}\right). \tag{139}$$

If $\hat{n}\times\mathbf{E}$ or $\hat{n}\times\mathbf{B}$ vanishes on S, the volume integral is constant, and equal to zero if the fields vanish initially. Since the integrand is nonnegative, the fields have to vanish throughout V_1. Consequently, $\mathbf{E}_2$ and $\mathbf{B}_2$ vanish in V_1 and, in particular,

$$\hat{n}\times\mathbf{B}_{2-} = 0. \tag{140}$$

Thus, once the surface field $\hat{n}\times\eta$ is determined from the integral equation (136), the scattered and transmitted electromagnetic fields can be found from Eqs. (126), (127), (132), and (133), and we have

$$\mathbf{E} = \mathbf{E}^{in} + \mathbf{M}_1\{\hat{n}\times\eta\}, \quad \mathbf{x}\in V_1, \tag{141}$$

$$\mathbf{B} = \mathbf{B}^{in} + \mathbf{L}_1\{\hat{n}\times\eta\}, \quad \mathbf{x}\in V_1, \tag{142}$$

$$\begin{aligned}\mathbf{E} &= \mathbf{L}_2\{\hat{n}\times\mathbf{E}^{in}\} + \alpha\mathbf{M}_2\{\hat{n}\times\mathbf{B}^{in}\}\\ &\quad + (\mathbf{L}_2\mathbf{M}_1 - \tfrac{1}{2}\alpha\mathbf{M}_2 + \alpha\mathbf{M}_2\mathbf{L}_1)\{\hat{n}\times\eta\}, \quad \mathbf{x}\in V_2,\end{aligned} \tag{143}$$

$$\begin{aligned}\mathbf{B} &= \mathbf{M}_2'\{\hat{n}\times\mathbf{E}^{in}\} + \alpha\mathbf{L}_2\{\hat{n}\times\mathbf{B}^{in}\}\\ &\quad + (\mathbf{M}_2'\mathbf{M}_1 - \tfrac{1}{2}\alpha\mathbf{L}_2 + \alpha\mathbf{L}_2\mathbf{L}_1)\{\hat{n}\times\eta\}, \quad \mathbf{x}\in V_2.\end{aligned} \tag{144}$$

These fields satisfy Maxwell's equations with the correct constants in each region, as well as the initial conditions. From Eqs. (128), (129), (134), and (140) we see that the boundary conditions (96) and (97) are also satisfied, and the fields given by Eqs. (141)–(144) satisfy all the conditions of the problem.

There are variations of this method that are equivalent to it, but that might be better suited for numerical calculations in a particular problem. For instance, we might use Eq. (140) instead of (134) to derive the integral equation. We might also define the fields $\mathbf{E}_1$ and $\mathbf{B}_1$ so that $\hat{n}\times\mathbf{B}_1$ is continuous across S, and choose the jump $\hat{n}\times\phi$ in $\hat{n}\times\mathbf{E}_1$ as the unknown field on S.

Another approach to the solution of this scattering problem is through the vector wave equation, which for the distribution $\mathbf{E}$ can be written in the form

$$\begin{aligned}&v^{-2}\ddot{\mathbf{E}} + \nabla\times(\nabla\times\mathbf{E})\\ &\quad = -\mu\frac{\partial\mathbf{j}}{\partial t} + \mathbf{E}_0(\mathbf{x})\delta'(t)\\ &\qquad - \nabla\times\mathbf{B}_0(\mathbf{x})\delta(t) + \Delta[\hat{n}\times(\nabla\times\mathbf{E})]\delta(S)\\ &\qquad + \nabla\times[\Delta(\hat{n}\times\mathbf{E})\delta(S)].\end{aligned} \tag{145}$$

The elementary solution of this equation is the dyadic distribution $\mathbf{Q}$ that satisfies

$$\begin{aligned}&v^{-2}\ddot{\mathbf{Q}}(\mathbf{x},t) + \nabla\times[\nabla\times\mathbf{Q}(\mathbf{x},t)]\\ &\quad = \delta^{(3)}(\mathbf{x})\delta(t)\mathbf{I};\end{aligned} \tag{146}$$

the retarded solution is[7]

$$\begin{aligned}\mathbf{Q}(\mathbf{x},t) &= (4\pi r)^{-1}\delta(t - r/v)(\mathbf{I} - r^{-2}\mathbf{x}\mathbf{x})\\ &\quad + (4\pi r^3)^{-1}v^2 t\theta(t - r/v)(\mathbf{I} - 3r^{-2}\mathbf{x}\mathbf{x})\\ &\quad + \tfrac{1}{3}v^2 t\theta(t)\delta^{(3)}(\mathbf{x})\mathbf{I},\end{aligned} \tag{147}$$

so that the components are defined by the analog of Eq. (12),

$$\langle Q_{ij},\varphi\rangle = \int dV \frac{\delta_{ij}-\alpha_i\alpha_j}{4\pi r}\varphi\left(\mathbf{x},\frac{r}{v}\right) + \int_0^\infty \frac{v^2 t\,dt}{4\pi}\int_0^{vt}\frac{dr}{r^3}\oint_{S(r)} dS(\delta_{ij}-3\alpha_i\alpha_j)\varphi(\mathbf{x},t) + \frac{1}{3}\int_0^\infty v^2 t\,dt\,\delta_{ij}\varphi(0,t), \tag{148}$$

where δ_{ij} is the Kronecker delta and $\alpha_i = x_i/r$ is a direction cosine of $\mathbf{x}$. The singularity at the origin of the second integral is discussed in Ref. 7; this integral is the limit of one where the region about the origin is excluded. We also note that the angular integration of $\delta_{ij} - 3\alpha_i\alpha_j$ over the spherical surface gives a zero result. A field $\mathscr{E}$ that obeys the same homogeneous vector wave equation in V_1 and V_2 can be expressed as

$$\mathscr{E} = \mathbf{Q}*\cdot[\nabla\times(\boldsymbol{\phi}\delta(S)) + \hat{n}\times\boldsymbol{\tau}\delta(S)], \tag{149}$$

where $\boldsymbol{\tau}$ is the discontinuityof $\nabla\times\mathscr{E}$ across S, which is also the discontinuity of $-\partial\mathscr{B}/\partial t$. The normal components of the discontinuities do not appear explicitly in Eq. (146). We can proceed from here and express $\mathscr{E}$ and $\nabla\times\mathscr{E}$ as functionals of $\hat{n}\times\boldsymbol{\phi}$ and $\hat{n}\times\boldsymbol{\tau}$, define the fields $\mathbf{E}_1$ and $\mathbf{E}_2$ as before, and find an integral equation for the discontinuity $\hat{n}\times\boldsymbol{\tau}$ in the tangential component of $\nabla\times\mathbf{E}_1$. We note that Eqs. (76) and (97) imply that

$$\hat{n}\times(\nabla\times\mathbf{E})_+ = \alpha\hat{n}\times(\nabla\times\mathbf{E})_-, \tag{150}$$

a relation we have to use to find the jump in $\nabla\times\mathbf{E}_2$. The energy relation for the vector wave equation is a consequence of

$$\nabla\cdot[\dot{\mathbf{E}}\times(\nabla\times\mathbf{E})] = \nabla\times\dot{\mathbf{E}}\cdot\nabla\times\mathbf{E} + v^{-2}\dot{\mathbf{E}}\cdot\ddot{\mathbf{E}} + \mu\dot{\mathbf{E}}\cdot\frac{\partial\mathbf{j}}{\partial t}, \tag{151}$$

whence, for $\mathbf{j} = 0$,

$$\oint_S dS\,\hat{n}\cdot\dot{\mathbf{E}}\times(\nabla\times\mathbf{E}) = \frac{1}{2}\frac{d}{dt}\int_V dV\,[(\nabla\times\mathbf{E})^2 + v^{-2}\dot{\mathbf{E}}^2]. \tag{152}$$

Thus, if either $\hat{n}\times\mathbf{E}$ or $\hat{n}\times(\nabla\times\mathbf{E})$ vanishes on S, the volume integral is a constant that is zero if $\mathbf{E}$ and $\dot{\mathbf{E}}$ vanish initially. Then both $\nabla\times\mathbf{E}$ and $\dot{\mathbf{E}}$ vanish throughout V. If $\rho = 0$, $\nabla\cdot\mathbf{E}$ vanishes due to the contraints on the initial conditions and to the vector wave equation. Consequently, $\mathbf{E}$ vanishes throughout V for later times if all conditions are satisfied.

IV. MONOCHROMATIC FIELDS

An arbitrary time-dependent field can be written as a superposition of monochromatic fields by means of Fourier integrals, and in many applications it is sufficient to consider such a field.

We can write

$$\psi(\mathbf{x},t) = \psi(\mathbf{x})e^{-i\omega t}, \tag{153}$$

where $\psi(\mathbf{x})$ is complex and it is understood that we have to take the real part of the right-hand side of Eq. (153).

The wave equation reduces to the Helmholtz equation[10]

$$(\nabla^2 + k^2)\psi = 0, \tag{154}$$

where

$$k^2 = \omega^2/v^2. \tag{155}$$

For regions that extend out to infinity, we impose the outgoing wave condition

$$\lim_{r\to\infty} r\left(\frac{\partial\psi}{\partial r} - ik\psi\right) = 0 \tag{156}$$

to the appropriate fields. For instance, we assume that the scattered field obeys the outgoing wave condition, and the corresponding elementary solution of the Helmholtz equation with a source $-\delta$ is

$$\mathscr{G}(\mathbf{x}) = e^{ikr}/4\pi r. \tag{157}$$

We now follow the procedure of the previous two sections to find the scattered and transmitted fields for monochromatic waves.

We have to find a function $\psi(\mathbf{x})$ such that

$$(\nabla^2 + k_i^2)\psi = 0, \quad \mathbf{x}\in V_i,\ i = 1,2, \tag{158}$$

$$k_i^2 = \frac{\omega^2}{v_i^2}, \tag{159}$$

$$\psi_+ = \psi_-, \quad \mathbf{x}\in S, \tag{160}$$

$$\left(\frac{\partial\psi}{\partial n}\right)_+ = \alpha\left(\frac{\partial\psi}{\partial n}\right)_-, \quad \mathbf{x}\in S, \tag{161}$$

$$\psi = \psi^{\text{in}} + \psi^{\text{sc}}, \quad \mathbf{x}\in V_1, \tag{162}$$

where ψ^{in} is a given incident field that satisfies the homogeneous Helmholtz equation and the scattered field obeys an outgoing wave condition. If the region V_2 also extends to infinity, the transmitted field obeys an outgoing wave condition there.

We define the functional operators

$$G\{\eta\} = -\oint_S dS'\,\eta(\mathbf{x}')e^{ikR}/4\pi R, \tag{163}$$

$$N\{\phi\} = \mathbf{P}\oint_S d\mathbf{S}'\cdot\boldsymbol{\phi}(\mathbf{x}')e^{ikR}(1-ikR)\mathbf{R}/4\pi R^3, \tag{164}$$

$$N'\{\eta\} = \hat{n}\cdot\mathbf{P}\oint_S dS'\,\eta(\mathbf{x}')e^{ikR}(1-ikR)\mathbf{R}/4\pi R^3, \tag{165}$$

and fields ψ_1 and ψ_2 such that

$$\psi_1 = \psi^{\text{sc}}, \quad \mathbf{x}\in V_1, \tag{166}$$

$$(\nabla^2 + k_1^2)\psi_1 = 0, \quad \mathbf{x}\in V_2, \tag{167}$$

$$\psi_2 = 0, \quad \mathbf{x}\in V_1, \tag{168}$$

$$\psi_2 = \psi, \quad \mathbf{x}\in V_2. \tag{169}$$

We also assume that ψ_1 is continuous across S; then, if we choose as our unknown function the discontinuity η in $\partial\psi_1/\partial n$, this function obeys the same integral equation (65) with the new definitions of G, N, and N'. The field ψ is then given by Eqs. (70) and (71).

Similarly, for monochromatic electromagnetic fields, we set

$$\mathbf{E}(\mathbf{x},t) = \mathbf{E}(\mathbf{x})e^{-i\omega t}, \tag{170}$$

$$\mathbf{B}(\mathbf{x},t) = \mathbf{B}(\mathbf{x})e^{-i\omega t}, \tag{171}$$

and Maxwell's equations of motion become

$$\nabla\times\mathbf{E} - i\omega\mathbf{B} = 0, \tag{172}$$

$$\nabla\times\mathbf{B} + i(\omega/v^2)\mathbf{E} = \mu\check{\mathbf{j}}. \tag{173}$$

It is simple now to allow the constants ϵ and μ to be functions of ω, thereby introducing dispersion, and to extend the theory to conducting media that have a conductivity σ, which appears in

$$\mathbf{j} = \sigma(\omega)\mathbf{E}. \tag{174}$$

Equation (173) then takes the form

$$i\omega\nabla\times\mathbf{B} - k^2\mathbf{E} = 0, \tag{175}$$

where

$$k^2 = \omega^2\mu(\epsilon + i\sigma/\omega) \tag{176}$$

defines a complex propagation constant k. Equations (114) and (115) now become

$$\hat{n}\cdot\boldsymbol{\phi} = -(i\omega/k^2)\nabla_s\cdot\hat{n}\times\boldsymbol{\eta}, \tag{177}$$

$$\hat{n}\cdot\boldsymbol{\eta} = (i/\omega)\nabla_s\cdot\hat{n}\times\boldsymbol{\phi}, \tag{178}$$

and we use them to find the fields as in Eqs. (108) and (109),

$$\mathscr{E} = (i\omega/k^2)\nabla\mathscr{G}*\nabla_s\cdot\hat{n}\times\boldsymbol{\eta}\delta(S) + \nabla\mathscr{G}*\times(\hat{n}\times\boldsymbol{\phi})\delta(S) + i\omega\mathscr{G}*\hat{n}\times\boldsymbol{\eta}\delta(S), \tag{179}$$

$$\mathscr{B} = -(i/\omega)\nabla\mathscr{G}*\nabla_s\cdot\hat{n}\times\boldsymbol{\phi}\delta(S) + \nabla\mathscr{G}*\times(\hat{n}\times\boldsymbol{\eta})\delta(S) - (ik^2/\omega)\mathscr{G}*\hat{n}\times\boldsymbol{\phi}\delta(S). \tag{180}$$

We can again use Eqs. (119) and (120) to express $\mathscr{E}$ and $\mathscr{B}$ in terms of the operators

$$\mathbf{L}\{\hat{n}\times\boldsymbol{\phi}\} = \mathrm{P}\oint_S dS'\, e^{ikR}(1 - ikR)(\hat{n}'\times\boldsymbol{\phi})\times\mathbf{R}/4\pi R^3, \tag{181}$$

$$\mathbf{M}\{\hat{n}\times\boldsymbol{\eta}\} = -\mathrm{P}\oint_S dS' e^{ikR} i[(1 - ikR)\nabla_s'\cdot(\hat{n}'\times\boldsymbol{\eta})\mathbf{R}/(4\pi k^2 R^3) - \hat{n}'\times\boldsymbol{\eta}/4\pi R], \tag{182}$$

$$\mathbf{M}'\{\hat{n}\times\boldsymbol{\phi}\} = \mathrm{P}\oint_S dS'\, e^{ikR} i[(1 - ikR)\nabla_s'\cdot(\hat{n}'\times\boldsymbol{\phi})\mathbf{R}/4\pi\omega R^3 - k^2\hat{n}'\times\boldsymbol{\phi}/4\pi\omega R]. \tag{183}$$

We choose fields $\mathbf{E}_1$, $\mathbf{B}_1$, $\mathbf{E}_2$, and $\mathbf{B}_2$ as in Sec. III, and, if $\hat{n}\times\boldsymbol{\eta}$ is the discontinuity in the tangential component of $\mathbf{B}_1$, we must solve an integral equation of the form (136) with the definition (181)–(183) of the operators. The fields we seek are then given by Eqs. (141)–(144) in terms of $\hat{n}\times\boldsymbol{\eta}$.

V. CONCLUDING REMARKS

We have shown how a judicious choice of unknown functions can reduce the number of integral equations required to solve the problem of scattering of scalar or electromagnetic waves by three-dimensional bodies. We have done this both for transient and monochromatic fields. We have not analyzed the particular difficulties that might arise in a numerical solution of these integral equations.

The limiting assumption of homogeneous media appears to be inherent to the method of solution of integral equations for functions defined only on the surface. Since all real media are dispersive, it would be useful to extend the methods to the solution of transient scattering problems in dispersive media, as well as allowing for a finite conductivity. Essentially, these problems are already solved if we express a transient field as a superposition of harmonic fields by a Fourier transformation.

We should also be alert to the possibility of reducing the size of other solutions of partial differential equations where boundary conditions in space are involved.

ACKNOWLEDGMENT

I thank Dr. John E. Lagnese for many helpful comments.

[1] A. J. Poggio and E. K. Miller, in *Computer Techniques for Electromagnetics*, edited by R. Mittra (Pergamon, Oxford, 1973), Chap. 4, esp. pp. 168–70.

[2] R. Mittra, in *Transient Electromagnetic Fields*, edited by L. B. Felsen (Springer-Verlag, Berlin, 1976), Chap. 2.

[3] E. Marx, Nat. Bur. Std. (U.S.) Tech. Note 1157, Washington, D.C., February 1982.

[4] D. Maystre, in *Electromagnetic Theory of Gratings*, edited by R. Petit (Springer-Verlag, Berlin, 1980), pp. 63–81.

[5] L. Schwartz, *Théorie des distributions*, nouvelle éd. (Hermann, Paris, 1966); *Mathematics for the Physical Sciences* (Addison-Wesley, Reading, Mass., 1966).

[6] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1.

[7] E. Marx and D. Maystre, "Dyadic Green functions for the time-dependent wave equation," J. Math. Phys. **23**, 1047 (1982).

[8] These and other similar expressions can be obtained by the usual Green function technique, as done in Ref. 3.

[9] See, for instance, L. Brand, *Vector and Tensor Analysis* (Wiley, New York, 1947), p. 223, for a justification of the integration by parts on the surface.

[10] In this section, we use many of the symbols used in previous sections with a different meaning. Here fields are functions of the vector $\mathbf{x}$ for a fixed frequency ω.

On the use of isospectral eigenvalue problems for obtaining hereditary symmetries for Hamiltonian systems

A. S. Fokas
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676

R. L. Anderson
Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602

(Received 15 June 1981; accepted for publication 19 February 1982)

We present an algorithmic method for obtaining an hereditary symmetry (the generalized squared-eigenfunction operator) from a given isospectral eigenvalue problem. This method is applied to the $n \times n$ eigenvalue problem considered by Ablowitz and Haberman and to the eigenvalue problem considered by Alonso. The relevant Hamiltonian formulations are also determined. Finally, an alternative method is presented in the case two evolution equations are related by a Miura type transformation and their Hamiltonian formulations are known.

PACS numbers: 03.65. — w

I. INTRODUCTION AND BASIC NOTIONS

Associated with a given exactly solvable equation there exist at least two interrelated aspects of fundamental interest: (i) The development of a method of solution such as the Inverse Scattering Transform (I.S.T); (ii) The investigation of the "algebraic" properties of the equation. The second aspect, the algebraic one, includes (a) finding the hierarchy of equations which have similar properties to the given equation (e.g., they are solved by the same eigenvalue problem and/or have the same conserved quantities); (b) finding a set of infinitely many conserved quantities of the equation which are in involution; (c) establishing the Hamiltonian and/or bi-Hamiltonian formulation of the equation as well as that of every member of its hierarchy. Fundamental to the solution of an equation via the I.S.T. is the existence of an isospectral eigenvalue problem. In this paper, we show in general that knowledge of this eigenvalue problem yields in principle a complete characterization of the algebraic properties. Moreover, we explicitly answer (a), (b), and (c) above for all equations solvable by (1) the $n \times n$ eigenvalue problem considered in Ref. 1; (2) an eigenvalue problem with complicated dependence on the eigenvalue considered in Ref. 2.

As is well known, AKNS[3] formulated and solved the problem of finding all equations solvable by the Dirac eigenvalue problem, and hence solved (a) above in this particular case. They found that the relevant hierarchy of exactly solvable equations is generated recursively with the aid of an integro–differential operator, the squared-eigenfunction operator. (For the KdV equation this is the well-known Lenard operator.) Subsequently, it was recognized that it is precisely these operators that play a central role in characterizing the algebraic properties of a given equation. They have been given various names in the literature, such as recursion operators,[4] strong symmetries,[5] hereditary symmetries,[5] Kahler operators,[6] and regular operators.[7]

This paper is organized as follows. Below we briefly explain the basic notions needed in this paper. In Sec. II, using ideas of Fuchssteiner,[8] we first prove that a linear isospectral eigenvalue problem gives rise to an hereditary symmetry. (This operator generates the hierarchy of equations linearizable via this eigenvalue problem.) We then discuss how to obtain the Hamiltonian formulation for this hierarchy of equations. In Sec. III we first obtain the hereditary symmetries of the equations associated with the two types of eigenvalue problems mentioned earlier. In connection with this, we make use of the properties of the coadjoint representation of the special linear group. Specifically, we show that the components of the gradients of the eigenvalues can always be identified with coordinates dual to those of the root vectors in the adjoint representation. It then follows that the coordinates dual to those of the Cartan subalgebra can always be eliminated to obtain a system of coupled linear integro–differential equations which immediately yield the hereditary symmetry (or more precisely its adjoint). We then present the relevant Hamiltonian formulations. It is quite interesting that the symplectic operator for the $n \times n$ case is simply the standard symplectic matrix. The existence of these Hamiltonian structures coupled with the hereditary property of the recursion operator implies a set of constants of motion in involution.

The method we present in this paper for obtaining a hereditary symmetry from a given eigenvalue problem is general and algorithmic; however, it is tedious. In Sec. IV we show that if two equations are related by a Miura type transformation and if both their Hamiltonian formulations are known, then there is a much more efficient way of computing their hereditary symmetries. This is illustrated by the Boussinesq and modified Boussinesq equations, where the known recursion operator of the Boussinesq equation[9] is derived in a more efficient way.

It should be noted that while various authors (see, for example, Refs. 3 and 10–13) have developed several methods for deriving recursion operators, they did not notice the hereditary property of these operators. Also, we note that the appearance of coadjoint representation has been recognized by other authors. In particular, it has been used in Ref. 9 for obtaining the recursion operator associated with the third order eigenvalue problem considered in Ref. 14.

 0022-2488/82/061066-08$02.50

A. Basic notions

The basic notions used in this paper can be found in the work of Magri,[6] Gel'fand and Dorfman,[7] and Fokas and Fuchssteiner.[15] In what follows we shall use the formalism of the latter authors, since it is computationally more convenient. The Lie-algebraic structure of this formalism as well as its relation to those of the other authors can be found in Fuchssteiner.[16] The geometric interpretation of some of these notions can be found in Ref. 17.

We consider an evolution equation in its abstract form,

$$u_t = K(u), \tag{1.1}$$

on a normed space M of vector-valued functions on $\mathbb{R}$, where K is a suitable C^∞ vector field on M. We assume that the space of smooth vector fields on M, $\mathfrak{X}(M)$, is some space S of C^∞ functions on the real line vanishing rapidly at $\pm\infty$. (In most applications $M = S$.) Let S^* be the dual of S such that the elements γ of S^* define continuous linear functions on S via

$$\langle \gamma,\sigma\rangle = \int_{-\infty}^{\infty} \gamma(x)\sigma(x)dx, \quad \gamma\in S^*, \ \sigma\in S. \tag{1.2}$$

We deal with functions ϕ (attaining values in S, S^*, or in a space of operators) which are assumed to be differentiable in the sense that the chain rule holds. By $\phi'(u)[v]$ we denote the directional derivative of ϕ at the point u in the direction v. Sometimes we write ϕ and $\phi'[\ ]$ instead of $\phi(u)$ and $\phi'(u)[v]$. We recall that a function $f: M\to S^*$ is said to be a gradient function if it has a potential $I: S\to R$, which means that $I'(u) = f(u)$ for all $u\in M$. It is well known that f is a gradient function iff $f' = (f')^+$ [where $(f')^+$ is the adjoint of f' with respect to $\langle,\rangle$]. Then the potential $I(u)$ is given by

$$I(u) = \int_0^1 \langle f(\lambda u), u\rangle d\lambda. \tag{1.3}$$

A function $\sigma: M\to S$ is called *symmetry* of (1.1) (see Ref. 18 for a motivation of this notion iff),

$$\sigma'[K] - [K']\sigma = 0 \tag{1.4}$$

(i.e., if the Lie-algebra product of K and σ is zero). The infinitely many symmetries $\{\sigma_i\}_{i=1}^{\infty}$ of an exactly solvable evolution equation define the hierarchy

$$u_t = \sigma_i, \quad i = \text{integer} \tag{1.5}$$

associated with a given exactly solvable evolution equation (1.1) (for example, in the KdV the Lax hierarchy).

An operator $\Phi(u): S\to S$, $u\in M$, is a strong symmetry[6] of (1.1) (or a recursion operator in the terminology of Ref. 4) if it generates a new symmetry from a given one. This is the case iff Φ satisfies the operator equation

$$\Phi'[K] - [K',\Phi] = 0. \tag{1.6}$$

The flows defined by the hierarchy (1.5), by definition commute with the flow of (1.1). However, for the exact solvability of (1.1) it seems necessary that the above flows also commute with each other. A sufficient condition for this to be the case is that the strong symmetry possesses a certain property called by Fuchssteiner the hereditary property.[5] An operator $\Phi(u)$, $S\to S$, $u\in M$, is called hereditary (or Kähler,[6] or regular[7]) if $[\Phi'(u),\Phi(u)]$ is a symmetric bilinear operator for all $u\in M$, i.e., iff

$$\Phi'[\Phi v]w - \Phi\Phi'[v]w \text{ is symmetric with respect to } v \text{ and } w. \tag{1.7}$$

If an exactly solvable evolution equation has some additional structure (for example, is a Hamiltonian system) then it will also possess infinitely many *conserved quantities*. There exists a convenient description of the conserved quantities of (1.1) using the notion of the gradient of a conserved quantity: $I(u)$ is a conserved quantity of (1.1) if its gradient γ satisfies.

$$\gamma'[K] + (K')^+[\gamma] = 0, \tag{1.8}$$

where $(K')^+$ denotes the adjoint of K'. Conversely, if γ satisfies (1.8) and if it is also a gradient function, i.e., if $\gamma' = (\gamma')^+$, then the potential of γ, say $I(u)$, is a conserved quantity of (1.1). If Eq. (1.1) is a Hamiltonian system, i.e., if there exists a co-symplectic operator $\theta(u)$ and a gradient function $\gamma(u)$ such that $u_t = K(u) = \theta(u)\gamma(u)$, then the symmetries and the gradients of conserved quantities of (13) are related by $\sigma = \theta\gamma$. (We recall that a skew symmetric operator $\theta(u)$ is co-symplectic if the bracket $\langle a,\theta'(u)[\theta(u)c]b\rangle$ satisfies the Jacobi identity for every a,b,c)

The hereditary symmetries corresponding to Hamiltonian systems admit a symplectic–co-symplectic factorization,[15] i.e., they can be written in the form $\Phi = \theta_1\theta_2^{-1}$, where θ_1, θ_2 define the bi-Hamiltonian formulation of the equation. These latter factorizable hereditary symmetries generate a set of infinitely many constants of motion in involution with gradients given by

$$\gamma^{(j)} = \theta_2^{-1}(\Phi)^j u_x, \quad j = 1,2,\ldots. \tag{1.9}$$

II. LINEAR EIGENVALUE PROBLEMS YIELD HEREDITARY SYMMETRIES FOR HAMILTONIAN SYSTEMS

The algebraic formalism reviewed above can be applied to a given exactly solvable evolution equation independently of whether or not this equation admits an inverse scattering transform formulation. However, if it does admit such a formulation one may ask whether its strong symmetry follows solely from the associated linear eigenvalue problem. In the context of scalar Lax systems (i.e., Gel'fand–Dikii approach), Symes[13] performed the relevant construction, while the question has been answered affirmatively in Ref. 9a for any matrix eigenvalue problem associated with a simple Lie algebra. Here, following ideas of Fuchssteiner,[8] *we give a general approach to the problem and show that the strong symmetries obtained this way (i.e., via linear eigenvalue problems) are necessarily hereditary.*[19] Furthermore, we then discuss the Hamiltonian formulation of the equations generated by the above hereditary symmetries.

Lemma 2.1: Let γ be the gradient of a conserved quantity I of Eq. (1.1) and let $\Phi(u)$ be a strong symmetry of (1.1). Then I is a conserved quantity of the whole hierarchy of equations

$$u_t = \Phi^n K, \quad n = 1,2,\ldots \tag{2.1}$$

iff $(\Phi^+)^n\gamma$, $n = 1,2,\ldots$, are gradient functions.

Proof: The functional $I(u)$ is a conserved quantity of (1.1) iff its gradient γ, where $\gamma = \text{grad } I$, satisfies $\langle \gamma, K \rangle = 0$. Similarly, I is a conserved quantity of (2.1) iff $\langle \gamma, \Phi^n K \rangle = 0$. But

$$\langle \gamma, \Phi^n K \rangle = \langle (\Phi^+)^n \gamma, K \rangle.$$

Since Φ is a strong symmetry of (1.1) it follows that $\gamma_n = (\Phi^+)^n \gamma$ solve Eq. (1.8). However, the right-hand side of the above equation will be zero iff γ_n are gradients, because only then do the solutions of (1.8) give rise to conserved quantities.

Lemma 2.2: Consider the linear eigenvalue problem

$$\Psi \gamma = \mu \gamma, \tag{2.2}$$

where

$$\Psi(u): S^* \to S^*, u \in M, \quad \text{and } \gamma: M \to S^*.$$

Assume that if u satisfies (1.1) then (2.2) is isospectral, i.e., $d\mu/dt = 0$, and also assume that γ solves (1.8). Then

$$(\Psi'[K] + [(K')^+, \Psi])\gamma = 0. \tag{2.3}$$

Proof: The proof follows easily by differentiating (2.2) in the direction u_t.

Lemma (2.2) immediately implies

Proposition 2.1: If (2.2) is satisfied for a sufficiently large number of solutions of (1.8), then $\Phi = \Psi^+$ is a strong symmetry of (1.1).

The constructive approach of obtaining a strong symmetry of (1.1) via its inverse scattering formulation, is based on the above proposition: Given an isospectral eigenvalue problem, say

$$\frac{dV}{dx} = U(u;\lambda)V, \tag{2.4}$$

where λ is the eigenvalue, one can always evaluate the gradient of λ, call it G_λ. If one can then find the eigenvalue equation that G_λ satisfies, then by writing this equation in the form

$$\Psi G_\lambda = \mu(\lambda) G_\lambda, \tag{2.5}$$

one can immediately obtain $\Phi = \Psi^+$. (Assuming of course, that G_λ are in some sense dense. Fuchssteiner, for example, invokes a local-global principle).[8]

In what follows, after giving some preliminary results, we shall prove that if a strong symmetry is obtained by the above approach then it is hereditary.

Lemma 2.3 (see Ref. 8): Assume that the strong symmetry of (1.1) satisfies the isospectral eigenvalue problem

$$\Phi^+ G_\lambda = \mu G_\lambda, \tag{2.6}$$

where G_λ are gradients of conserved quantities of (1.1). Here these conserved quantities are the eigenvalues λ of the linear scattering eigenvalue problem (2.4). Then the λ's and the Φ are conserved quantities and a strong symmetry, respectively, for the whole hierarchy (2.1).

Proof: To prove that λ's are conserved by (2.1) note that

$$\langle G_\lambda, \Phi^n K \rangle = \langle (\Phi^+)^n G_\lambda, K \rangle = \mu^n \langle G_\lambda, K \rangle = 0.$$

Then, it follows from Proposition (2.1) that Φ is a strong symmetry of (2.1), since Φ satisfies (2.6) and the G_λ's are gradients of conserved quantities of (2.1).

Note that this lemma implies that the eigenvalue problem (2.4) is isospectral for the whole hierarchy of equations (2.1).

Proposition 2.2: Assume that the strong symmetry Φ of (1.1) satisfies the isospectral eigenvalue problem (2.6). Then all symmetries of (1.1) generated through Φ commute and Φ is hereditary (assuming that these symmetries are dense).

Proof: Because of Lemma 2.3, Φ is a strong symmetry of the whole hierarchy (2.1) and from this it follows that all symmetries of (1.1) commute. To prove that Φ is hereditary use the following identity:

$$\begin{aligned}\Phi'[\Phi\beta]\alpha - \Phi'[\Phi\alpha]\beta - \Phi\Phi'[\beta]\alpha + \Phi\Phi'[\alpha]\beta \\ = \Phi^2(\alpha'[\beta] - \beta'[\alpha]) + (\Phi\alpha)'[\Phi\beta] - (\Phi\beta)'[\Phi\alpha] \\ + \Phi((\Phi\beta)'[\alpha] - \alpha'[\Phi\beta]) - \Phi((\Phi\alpha)'[\beta] - \beta'[\Phi\alpha]).\end{aligned} \tag{2.7}$$

If α, β are symmetries of (1.1) then the right-hand side of the above equation vanishes. Thus, if the symmetries form a dense set, then the hereditary property follows.

A. On the Hamiltonian formulation

Having obtained the hereditary operator Φ, and assuming that Φ is x-translation invariant [which corresponds to assuming that Eq. (2.4) does not depend explicitly on x] one immediately obtains the hierarchy of exactly solvable evolution equations[20]

$$u_t = \Phi^n u_x, \quad n = 1,2,\ldots . \tag{2.8}$$

The problem we shall discuss now is how to find the Hamiltonian formulation of Eq. (2.8) or, more precisely, how to find a co-symplectic operator θ such that (2.8) can be written in the form $u_t = \theta\gamma_n$ (where γ_n, $n = 1,2,\ldots$, are gradient functions).

We recalled in the basic notions that some hereditary operators Φ admit a certain factorization $\Phi = \theta_2\theta_1^{-1}$, where θ_1, θ_2 are compatible co-symplectic operators. This factorization, although quite useful in many considerations, clearly cannot be used for obtaining a co-symplectic operator from a given hereditary symmetry. However, it turns out that there does exist an explicit relationship between θ and Φ, namely

$$\theta\Phi^+ = \Phi\theta. \tag{2.9}$$

The motivation for the above relationship is quite simple. If an equation possesses a hereditary symmetry Φ and if also it can be written as a Hamiltonian system with θ as the associated co-symplectic operator, then Φ generates σ's, Φ^+ generates γ's, and θ maps γ's into σ's. Thus, if σ_i and γ_i, $i = 1,2$, are solutions of (1.4) and (1.8), respectively, then $\sigma_2 = \Phi\sigma_1 = \Phi\theta\gamma_1$. But $\sigma_2 = \theta\gamma_2 = \theta\Phi^+\gamma_1$. Thus $(\theta\Phi^+ - \Phi\theta)\gamma_1 = 0$, and if this equation is valid for infinitely many γ's then Eq. (2.9) must be valid.

Equation (2.9) is fundamental in our approach for obtaining θ. Although it does not provide an explicit relationship for θ in terms of Φ, it can be very useful for obtaining θ, after a suitable guess for θ has been made. This is best illustrated for the case of the $n \times n$ eigenvalue problem. In this case, after obtaining Φ (and Φ^+) it is clear by inspection

that θ is simply a constant skew symmetric matrix (see Sec. III). This simple form of θ is quite surprising when compared with the known cases where θ is more complicated. For example, θ is D for the KdV and the modified KdV, it is $D^3 + uD + Du$ for the fifth order equations solvable by the third order eigenvalue problem considered in Ref. 14, etc. The above equations can be obtained from the $n \times n$ eigenvalue problem after some reduction (see Sec. III). In this process of reduction the standard symplectic matrix gives rise to very complicated θ's.

III. APPLICATION

A. $n \times n$ eigenvalue problem

In this subsection, we shall determine the explicit form of the hereditary symmetry and the Hamiltonian structure associated with the following isospectral problems:

$$Dv = Uv, \tag{3.1}$$

where

$$U = \lambda H + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} u_{ij} E_{ij}\,, \tag{3.2}$$

v is an n-dimensional vector, the $n(n-1)$ u_{ij}'s are nonzero functionally independent potentials, and the $n \times n$ matrices H and the E_{ij}'s are given by

$$H = \operatorname{diag}(h_1,...,h_n), \quad \text{(no two entries equal)} \tag{3.3}$$

and

$$(E_{ij})_{kl} = \delta_{ik}\delta_{jl}, \tag{3.4}$$

respectively, and D is the operator of total differentiation w.r.t. x.

1. The hereditary symmetry

In light of the results described in Sec. II, we need essentially only present here a derivation of Eq. (2.5) for (3.2). As we shall show, this equation follows from the Lie algebraic structure of (3.1). Therefore, we begin with an elaboration of this Lie structure.

Equation (3.1) for constant u_{ij}'s is the Lie equation for a one-parameter subgroup of the linear transformation group (representation) $\mathrm{GL}(n,K) \times V^n \to V^n$, where $\{v_i\}_{i=1}^n$ is the set of coordinates for the carrier space V^n and the field K is the field of complex numbers $\mathbb{C}$ or the field of real numbers $\mathbb{R}$, in which case $V^n = \mathbb{C}^n$ or $\mathbb{R}^n$, respectively. However, here the u_{ij}'s are functions of x and t and for each t we can rigorously interpret (3.1) as describing a curve in the Lie algebra of vector fields on V^n induced by the $\mathrm{GL}(n,K)$ transformation group structure (see Refs. 21 and 22 for details of this interpretation). Similarly, we can interpret

$$Dv^* = -U^+ v^*, \tag{3.5}$$

where U^+ is the adjoint of U and $\{v_i^*\}_{i=1}^n$ is the set of coordinates for the space V^{n^*} dual to V^n, as describing a curve in the Lie algebra of vector fields on V^{n^*}. Therefore, it directly follows that under the induced group action the coordinates of the direct product space $V^{n^*} \otimes V^n$ evolve w.r.t. x for fixed t according to

$$D(v_i^* v_j) = -U_{ji}(v_j^* v_j - v_i^* v_i) - (U_{ii} - U_{jj}) \times v_i^* v_j - \sum_{k \neq i,j} U_{ki} v_k^* v_j + \sum_{k \neq i,j} U_{jk} v_i^* v_k, \tag{3.6}$$

which immediately yields that the subspace described by the column vector

$$[v_1^* v_1 - v_2^* v_2, ..., v_{n-1}^* v_{n-1} - v_n^* v_n, v_i^* v_j|_{i<j}, v_i^* v_j|_{i>j}]^T \tag{3.7}$$

is an invariant subspace and, in particular, the coordinates of this vector transform like the coordinates of a curve in the coadjoint representation of $\mathrm{gl}(n,K)$ [or, equivalently, to that of $\mathrm{sl}(n,K)$]. Further, fundamental to our result is the fact that it follows from the transformation properties of (3.7) that we can express the x evolution for fixed t of the subset of coordinates $\{v_i^* v_j\}_{i \neq j = 1}^n$ as a matrix integrodifferential eigenvalue equation solely in terms of this set. In particular, in components, we have the following:

$$\begin{aligned}
&[(-D + u_{ji} 2D^{-1} u_{ij}) v_i^* v_j \\
&\quad + \sum_{k \neq i,j} \{(u_{ji} D^{-1} u_{kj} - u_{ki}) v_k^* v_j \\
&\quad - (u_{ji} D^{-1} u_{ki}) v_k^* v_i - (u_{ji} D^{-1} u_{jk}) v_j^* v_k \\
&\quad + (u_{ji} D^{-1} u_{ik} + u_{jk}) v_i^* v_k\} \\
&\quad - 2u_{ji} D^{-1} u_{ji} v_j^* v_i]/(h_i - h_j) \\
&\quad = \lambda v_i^* v_j, \quad i \neq j = 1,...,n.
\end{aligned} \tag{3.8}$$

With (3.8) in hand we now turn to the task of identifying the components of the conserved gradients of the λ's appearing in (3.1) with the coordinates $v_i^* v_j (i \neq j)$.

Consideration of the directional derivative of (3.1) yields this identification and, in particular, it is given by the following lemma.

Lemma 3.1:

$$(G_\lambda)_{ij} = -v_i^* v_j \Big/ \left(\int_{-\infty}^{\infty} v^{*T} H v dx \right), \quad i \neq j = 1,...,n. \tag{3.9}$$

Proof: Consider v determined by (3.1), where U is given by (3.2), then v and λ are functionals of the potentials u_{ij}. Hence, the directional derivative of v in the direction w is an $n \times n(n-1)$ matrix valued operator whose action is given by

$$v'(u)[w] = (v'_{12}(u)[w_{12}], ..., v'_{n\,n-1}(u)[w_{n\,n-1}]), \tag{3.10}$$

where

$$v'_{ij}(u)[w_{ij}] \equiv \frac{\partial}{\partial \epsilon} v(u_{12}, ..., u_{ij} + \epsilon w_{ij}, ...)\Big|_{\epsilon = 0}, \tag{3.11}$$

and similarly for the $n(n-1)$ vector valued operator $\lambda'(u)[w]$.

Further, we now define the functional derivative G_λ of λ as the vector valued function whose components $(G_\lambda)_{ij}$ are given by

$$\lambda'_{ij}(u)[w_{ij}] = \int_{-\infty}^{\infty} (G_\lambda)_{ij} w_{ij} dx \tag{3.12}$$

for arbitrary w_{ij} which tend to zero sufficiently rapidly at $x = \pm \infty$.

Therefore, if we take the directional derivatives of all quantities appearing in (3.1), we obtain the following set of equations:

$$D_x v'_{ij}(u)[w_{ij}] = \left(\left(\int_{-\infty}^{\infty} (G_\lambda)_{ij} w_{ij} dx\right) H + w_{ij} E_{ij}\right) v + U v'_{ij}(u)[w_{ij}], \quad i \neq j = 1,\ldots,n, \tag{3.13}$$

where we have substituted for $\lambda'_{ij}(u)[w_{ij}]$ using (3.12).

Matrix multiplication of (3.13) by the transpose v^{*T} of v^*, given by (3.5), and integration over x from $-\infty$ to $+\infty$ yields

$$\int_{-\infty}^{\infty} v^{*T} D_x v'_{ij}(u)[w_{ij}] dx = \left(\int_{-\infty}^{\infty} (G_\lambda)_{ij} w_{ij} dx\right) \int_{-\infty}^{\infty} v^{*T} H v dx + \int_{-\infty}^{\infty} v_i^* w_{ij} v_j dx + \int_{-\infty}^{\infty} v^{*T} U v'_{ij}(u)[w_{ij}] dx. \tag{3.14}$$

It follows directly from the properties of (3.5) that the lhs of (3.14) is equal to the last term on the rhs of (3.14). Therefore, (3.14) becomes

$$0 = \int_{-\infty}^{\infty} \left\{(G_\lambda)_{ij} \left(\int_{-\infty}^{\infty} v^{*T} H v dx\right) + v_i^* v_j\right\} w_{ij} dx. \tag{3.15}$$

Equation (3.9) then follows from the fact that (3.15) holds for arbitrary w_{ij} and the restriction that

$$\int_{-\infty}^{\infty} v^{*T} H v dx \neq 0.$$

It now follows from (3.9) and the linearity of (3.8) that (3.8), under the replacement $v_i^* v_j \to (G_\lambda)_{ij}$, is (2.5) for the problem at hand. Therefore, we have the following theorem.

Theorem 3.1: The strong symmetry operator Φ for the class of exactly solvable evolution equations described by (3.1) and (3.2) is given by

$$\begin{aligned}\Phi\phi_{ij} = {} & (D - u_{ij} 2D^{-1} u_{ji})\phi_{ij}/(h_i - h_j) \\ & + \sum_{k \neq i,j} \{ -(u_{ij} D^{-1} u_{jk} + u_{ik})\phi_{kj}/(h_k - h_j) \\ & + (u_{ij} D^{-1} u_{ik})\phi_{ki}/(h_k - h_i) \\ & + (u_{ij} D^{-1} u_{kj})\phi_{jk}/(h_j - h_k) \\ & + (-u_{ij} D^{-1} u_{ki} + u_{kj})\phi_{ik}/(h_i - h_k)\} \\ & + 2u_{ij} D^{-1} u_{ij} \phi_{ji}/(h_j - h_i),\end{aligned} \tag{3.16}$$

where the coefficient of ϕ_{rs} on the rhs of the above equation is the matrix element $\Phi_{ij,rs}$ of Φ. Here we have restricted ourselves to $K = \mathbb{R}$ for convenience.

2. The Hamiltonian formulation

In this case θ is simply the constant skew symmetric matrix

$$\theta_1 = \begin{pmatrix} 0 & J \\ -J & 0 \end{pmatrix}, \tag{3.17}$$

where the $n(n-1)/2$ square matrix J is equal to $\mathrm{diag}(h_2 - h_1,\ldots,h_j - h_i,\ldots)$, with ordering of the $n(n-1)$ pairs of subscripts labelling the components of the gradient of λ given by $\{(1,2),\ldots,(i,j),\ldots,(2,1),\ldots,(j,i),\ldots\}_{i<j=1}^{n}$.

This operator can be easily obtained if one starts with the simplest case of 2×2 and then generalizes appropriately.

Example 3.1. The 3×3 eigenvalue problem:

Formula (3.16) yields that Φ is the 6×6 matrix given by

$$\Phi = \left(\frac{\alpha_1}{h_1 - h_2}, \frac{\alpha_2}{h_2 - h_3}, \frac{\alpha_3}{h_1 - h_3}, \frac{\alpha_4}{h_2 - h_1}, \frac{\alpha_5}{h_3 - h_2}, \frac{\alpha_6}{h_3 - h_1}\right), \tag{3.18}$$

where

$$\begin{aligned}\alpha_1 &= (D - 2u_1 D^{-1} u_4, u_2 D^{-1} u_4, u_2 - u_3 D^{-1} u_4, 2u_4 D^{-1} u_4, -u_6 - u_5 D^{-1} u_4, u_6 D^{-1} u_4)^T, \\ \alpha_2 &= (u_1 D^{-1} u_5, D - 2u_2 D^{-1} u_5, -u_1 - u_3 D^{-1} u_5, u_6 - u_4 D^{-1} u_5, 2u_5 D^{-1} u_5, u_6 D^{-1} u_5)^T, \\ \alpha_3 &= (u_5 - u_1 D^{-1} u_6, -u_4 - u_2 D^{-1} u_6, D - 2u_3 D^{-1} u_6, u_4 D^{-1} u_6, u_5 D^{-1} u_6, 2u_6 D^{-1} u_6)^T, \\ \alpha_4 &= (2u_1 D^{-1} u_1, u_3 - u_2 D^{-1} u_1, u_3 D^{-1} u_1, D - 2u_4 D^{-1} u_1, u_5 D^{-1} u_1, -u_5 - u_6 D^{-1} u_1)^T, \\ \alpha_5 &= (-u_3 - u_1 D^{-1} u_2, 2u_2 D^{-1} u_2, u_3 D^{-1} u_2, u_4 D^{-1} u_2, D - 2u_5 D^{-1} u_2, u_4 - u_6 D^{-1} u_2)^T, \\ \alpha_6 &= (u_1 D^{-1} u_3, u_2 D^{-1} u_3, 2u_3 D^{-1} u_3, -u_2 - u_4 D^{-1} u_3, u_1 - u_5 D^{-1} u_3, D - 2u_6 D^{-1} u_3)^T,\end{aligned}$$

and $u_1 = u_{12}$, $u_2 = u_{23}$, $u_3 = u_{13}$, $u_4 = u_{21}$, $u_5 = u_{32}$, $u_6 = u_{31}$. Then the nonzero entries in (3.17) are

$$\theta_{14} = h_2 - h_1 = -\theta_{41}, \quad \theta_{25} = h_3 - h_2 = -\theta_{52}, \quad \theta_{36} = h_3 - h_1 = -\theta_{63}. \tag{3.19}$$

Hence σ_n and γ_n can be calculated similarly as above.

B. A generalized second-order eigenvalue problem

We now consider the following second-order polynominal eigenvalue equation[2]:

$$L\psi = (D^2 + u_0 + \lambda u_1 + \cdots + \lambda^{N-1} u_{N-1})\psi = \lambda^N \psi, \tag{3.20}$$

which we can recast in matrix form as

$$Dv = ((-u_0 - \lambda u_1 \cdots - \lambda^{N-1} u_{N-1} + \lambda^N) E_{12} + E_{21}) v, \tag{3.21}$$

where

$$v = [v_1, v_2]^T = [\psi_x, \psi]^T.$$

1. The hereditary symmetry

Proceeding as before we obtain, by taking the directional derivative of (3.1), the result that the gradient G_λ of λ is proportional to the N-component vector valued function given by

$$[v_1^* v_2, \lambda v_1^* v_2, \ldots, \lambda^{N-1} v_1^* v_2]^T. \tag{3.22}$$

It then follows through the same set of logical steps that the strong symmetry operator Φ is found to be

$$\Phi = \begin{pmatrix} O & \frac{1}{4}D^4 + u_0 + \frac{1}{2}u_{0x} D^{-1} \\ I & u_1 + \frac{1}{2}u_{1x} D^{-1} \\ & u_{N-1} + \frac{1}{2}u_{N-1x} D^{-1} \end{pmatrix}, \tag{3.23}$$

where O is a $1 \times N-1$ matrix and I is the $N-1 \times N-1$ identity matrix.

In this case

$$\theta_1 = \begin{pmatrix} -S_1 & -S_2 \cdots & -S^{N-1} & D \\ -S_2 & & D & 0 \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ -S_{N-1} & & & \\ D & 0 & & 0 \end{pmatrix}; \quad S_j = u_j D + \tfrac{1}{2} u_{jx}. \tag{3.24}$$

Again this operator can be easily obtained if one starts with the simpler case $u_j = 0, j \neq 0,1$, and then generalizes. In the latter case, Eq. (3.23) yields

$$\theta = \begin{pmatrix} 0 & \mathscr{D}_0 \\ 1 & \Delta_1 \end{pmatrix}; \quad \begin{matrix} \mathscr{D}_0 = \tfrac{1}{4} D^2 + u_0 + \tfrac{1}{2} u_{0x} D^{-1}, \\ \Delta_1 = u_1 + u_{1x} D^{-1}. \end{matrix} \tag{3.25}$$

We are now looking for a skew symmetric operator θ, such that (2.9) is valid with Φ given by (3.23). Thus,

$$\begin{pmatrix} 0 & \mathscr{D}_0 \\ 1 & \Delta_1 \end{pmatrix} \begin{pmatrix} A & B \\ -B^+ & C \end{pmatrix} = \begin{pmatrix} A & B \\ -B^+ & C \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \mathscr{D}_0{}^+ & \Delta^+ \end{pmatrix}, \tag{3.26}$$

where $A^+ = -A, C^+ = -C$. Writing the matrix equation (3.26) out one obtains

$$-\mathscr{D}_0 B^+ = B \mathscr{D}_0{}^+, \tag{3.27a}$$

$$B + \Delta_1 C = -B^+ + C\Delta^+, \tag{3.27b}$$

$$\mathscr{D}_0 C = A + B\Delta^+, \tag{3.28a}$$

$$A - \Delta_1 B^+ = -B^+ \mathscr{D}_0{}^+. \tag{3.28b}$$

Equation (3.27a) implies (using the relevant result for the Kdv) that $B = D$. Then (3.27b) implies $C = \alpha D$. Hence (3.28) implies $A = \alpha \mathscr{D}_0 D - D\Delta^+$. Thus $\theta = \theta_1 + \alpha\theta_2$, where

$$\theta_1 = \begin{pmatrix} -s_1 & D \\ D & 0 \end{pmatrix}, \quad \theta_2 = \begin{pmatrix} D^3/4 + u_0 D + u_{0x} & 0 \\ 0 & D \end{pmatrix}. \tag{3.29}$$

Clearly, since α is arbitrary, θ_2 is the second co-symplectic operator associated with Φ. Actually, one can easily verify that $\theta_2 = \Phi\theta_1$.

Generalizing the operator θ_1 given by (3.29a) one obtains (3.24). Then θ_2 follows from the formula $\theta_2 = \Phi\theta_1$.

Having obtained the hereditary operator Φ one can easily obtain a hierarchy of symmetries (and hence of exactly solvable equations) by starting with $\sigma = (u_{0x}, u_{1x}, \ldots, u_{N-1x})^T$

In order to obtain a hierarchy of gradients of conserved quantities, one needs a starting one. However, such a gradient can be obtained by solving the equation $\theta_1 \gamma = \sigma$, where $\gamma = (\gamma_1, \ldots, \gamma_N)^T$. This is straightforward and the details are omitted.

Let us now turn to the problem of constructing Φ when the u_{ij}'s are functionally dependent. In this case (3.9) must be appropriately modified and (3.8), hence (3.16), must be reduced. We shall illustrate the general method of reduction with a simple example. The general problem of reduction has been considered recently by Magri.[23]

Example 3.2: Consider

$$Dv = (\lambda H + uE_{12} - uE_{21})v, \tag{3.30}$$

where H is given by (3.3) for $n = 2$ and E_{12}, E_{21} are given by (3.4) (this corresponds to the case for which u satisfies the modified Korteweg–de Vries (Mod KdV) equation, i.e., $u_t + 6u^2 u_x + u_{xxx} = 0$). Proceeding as we did in the derivation of (3.9), we consider v and λ as functionals of u and take the directional derivatives of the quantities appearing in (3.30) in the direction w. This yields

$$\begin{aligned} Dv'(u)[w] = {} & (\lambda'(u)[w] + wE_{12} - wE_{21})v \\ & + (\lambda H + uE_{12} - uE_{21})v'(u)[w]. \end{aligned} \tag{3.31}$$

It follows, paralleling the proof of (3.9), that in this case

$$G_\lambda = -(v_1^* v_2 - v_2^* v_1) \Big/ \left(\int_{-\infty}^{\infty} v^{*T} H v dx \right), \tag{3.32}$$

where

$$\lambda'(u)[w] = \int_{-\infty}^{\infty} G_\lambda w dx. \tag{3.33}$$

Equation (3.32) illustrates the general feature that arises when the u_{ij}'s in (3.2) are functionally dependent, namely, G_λ is a vector valued function of dimension strictly less than $n(n-1)$ and its components are linear combinations of the elements of the set $\{v_i^* v_j\}_{i \neq j = 1}^n$. Although Eq. (3.8) is valid independently of the functional dependence of the u_{ij}'s, this reduction in dimensionality necessitates the reduction of (3.8). In particular, in this example (3.8) becomes

$$-(D + 2uD^{-1}u)v_1^* v_2 - 2uD^{-1}uv_2^* v_1 = \lambda (h_1 - h_2) v_1^* v_2, \tag{3.34}$$

$$-2uD^{-1}uv_1^* v_2 - (D + 2uD^{-1})v_2^* v_1 = -\lambda (h_1 - h_2) v_2^* v_1. \tag{3.35}$$

Upon taking the sum and difference of (3.34) and (3.35) and eliminating the sum, we find

$$(D + 4uD^{-1}u)D(v_1^* v_2 - v_2^* v_1) = \lambda^2 (h_1 - h_2)^2 (v_1^* v_2 - v_2^* v_1) \tag{3.36}$$

or, employing (3.32) and the linearity of (3.36), we obtain

$$(D + 4uD^{-1}u)DG_\lambda = \lambda^2 (h_1 - h_2)^2 G_\lambda. \tag{3.37}$$

Therefore, (3.37) is (2.5) for this example and we have

$$\Phi^+_{\text{Mod KdV}} = (D + 4uD^{-1}u)D. \tag{3.38}$$

Hence, we recover the known result

$$\Phi_{\text{Mod KdV}} = D^2 + 4u^2 + 4u_x D^{-1} u. \tag{3.39}$$

IV. ON THE USE OF MIURA TYPE TRANSFORMATIONS

In practice, one is usually given an evolution equation which has an obvious Hamiltonian formulation as well as the associated inverse scattering pair, or equivalently[24] the associated Miura type transformation. One is then required to find the algebraic structure of this equation (symmetries, conservative laws, commutativity property, etc.). One approach to the solution of the above problem is to use the results of Sec. III: Given an eigenvalue problem one can always find the corresponding hereditary symmetry. Having obtained this hereditary operator and using the Hamiltonian formulation of the equation, the full algebraic structure is completely determined.

However, although finding Φ is algorithmically possible, it is clearly rather tedious. There exists a very simple alternative approach to the above problem, provided that one makes an additional assumption, namely, that the Hamiltonian formulation of the modified equation is known. In this case one can obtain immediately the hereditary symmetries of both the original equation as well as the modified equation. This is expressed by the following proposition.

Proposition 4.1: Let the two equations

$$u_t = K(u), \tag{4.1}$$

$$s_t = G(s) \tag{4.2}$$

be related through the Miura type transformation

$$B = u + F(s) = 0. \tag{4.3}$$

Assume that (4.1) and (4.2) possess the co-symplectic operators $\theta_1(u)$ and $\hat{\theta}_1(s)$, respectively, and that $\theta_1(u)$ is invertible (i.e., $K = \theta_1 f_1$, $G = \hat{\theta}_1 \hat{f}_1$, where $f_1 \hat{f}_1$ are gradient functions, and θ_1^{-1} exists). Then

(i) the operator

$$\theta_2(u) = B_s \hat{\theta}_1(s) B_s^{+}, \quad \text{where } B = 0, \tag{4.4}$$

is a second co-symplectic operator (4.1)[25];

(ii) the operator

$$\hat{J}_2(s) = B_s^{+} \theta_1^{-1}(u) B_s, \quad \text{where } B = 0, \tag{4.5}$$

is a symplectic operator of (4.2);

(iii) the operators

$$\Phi(u) = \theta_2(u)\theta_1^{-1}(u), \quad \hat{\Phi}(s) = \hat{\theta}_1(s)\hat{J}_2(s), \quad \text{where } B = 0, \tag{4.6}$$

are strong symmetries of Eq. (4.1) and (4.2), respectively.

Proof: In Ref. 15 it was found how symplectic and co-symplectic operators transform under Miura type transformations (see also Refs. 9a and 9b). Applying these results to the case that B is given by (4.3), one obtains that $\hat{\theta}(s)$ is co-symplectic iff $\theta(u) = B_s \hat{\theta}(s) B_s^{+}$ is inverse-symplectic. A similar formula follows for the transformation of symplectic operators. Thus (i) and (ii) follow. The proof of (iii) follows from Remark 2 of Ref. 15.

Remarks: 1. In Ref. 5 it was found how strong and hereditary symmetries transform under Miura type transformations. Applying these results to the case that B is given by (4.3) one obtains that $\Phi(s)$ is a strong symmetry of (4.2) iff

$$\Phi(u) = B_s \hat{\Phi}(s) B_s^{-1} \tag{4.7}$$

is a strong symmetry of (4.1). Also, $\hat{\Phi}(s)$ is hereditary iff $\Phi(u)$ is hereditary. Formulas (4.3), (4.5), and (4.6) imply that $\Phi(u)$ and $\hat{\Phi}(s)$, as defined by Eq. (4.6), satisfy Eq. (4.7). Therefore, if one of these two operators is hereditary, so is the other one.

2. In order to prove that $\Phi(u)$ is hereditary one needs to prove that $\theta_1 + \theta_2$ is a co-symplectic operator. However, in practice this is not necessary, because the Miura type transformation (extended in an obvious way so as to contain also a linear part) actually generates the co-symplectic operator $\Omega(u) = \theta_2(u) + \alpha\theta_1(u)$, where α is a parameter. Therefore, θ_1, θ_2 are compatible.

3. It is also clear that the operator $J_2(u) = \theta_1(u)^{-1}\Phi(u)$ is the second symplectic structure of (4.1) and that the operator $\hat{\theta}_2(s) = \hat{\Phi}(s)\hat{\theta}_1(s)$ is the second co-symplectic structure of (4.2).

Example 4.1. Boussinesq and modified Boussinesq equations: Consider the Boussinesq equation in the form

$$\mathbf{u}_t = \mathbf{K}(\mathbf{u}), \quad \text{where } \mathbf{K}(u) = \begin{pmatrix} 2V_x \\ -\frac{1}{6}(U_{xxx} + 4UU_x) \end{pmatrix},$$

$$\mathbf{u} = \begin{pmatrix} U \\ V \end{pmatrix}. \tag{4.8}$$

It is clear that Eq. (4.8) can be written in the form

$$\mathbf{u}_t = \theta_1 \gamma_1, \quad \text{where } \theta_1 = \begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix},$$

$$\gamma_1 = \begin{pmatrix} -\frac{1}{6}(U_{xx} + 2U^2) \\ 2V \end{pmatrix}. \tag{4.9}$$

Equation (4.9) defines a Hamiltonian system, since θ_1 is a co-symplectic operator and γ_1 is a gradient function:

$$\gamma_1' = \begin{pmatrix} -\frac{1}{6}(D^2 + 4U & 0 \\ 0 & 2 \end{pmatrix} = (\gamma_1')^{+}.$$

The Boussinesq equation is related to the modified Boussinesq equation

$$\mathbf{s}_t = \mathbf{G}(\mathbf{s}), \quad \text{where } \mathbf{G}(\mathbf{s}) = \begin{pmatrix} Z_{xx} + 2(ZY)_x \\ -\frac{1}{3}Y_{xx} - \frac{1}{3}(3Z^2 - Y^2)_x \end{pmatrix},$$

$$\mathbf{s} = \begin{bmatrix} Y \\ Z \end{bmatrix}, \tag{4.10}$$

through the Miura type transformation[9]

$$\mathbf{B} = \begin{pmatrix} U + 2Y_x + 3Z^2 + Y^2 \\ V + Z_{xx} + 3YZ_x + ZY_x + 2Z(Y^2 - Z^2) \end{pmatrix} = 0. \tag{4.11}$$

Using the above facts and Proposition 4.1, we shall now derive the hereditary symmetries of both Eqs. (4.8) and (4.10).

The modified Boussinesq equation is also a Hamiltonian system, since it can be written in the form

$$\mathbf{s}_t = \hat{\boldsymbol{\theta}}_1 \hat{\boldsymbol{\gamma}}_1, \quad \text{where } \theta_1 = \begin{pmatrix} \frac{1}{2}D & 0 \\ 0 & \frac{1}{6}D \end{pmatrix},$$

$$\hat{\gamma}_1(s) = \begin{pmatrix} 2Z_x + 4ZY \\ -2Y_x + 2Y^2 - 6Z^2 \end{pmatrix}, \tag{4.12}$$

and $\hat{\gamma}_1' = (\hat{\gamma}_1')^{+}$. Then, using the results of Proposition 4.1,

one can immediately obtain the hereditary symmetries and the second symplectic and co-inverse structures of both equations:

$$\theta_2(u) = B_s \begin{pmatrix} \frac{1}{2}D & 0 \\ 0 & \frac{1}{6}D \end{pmatrix} B_s^+,$$

$$\Phi(u) = B_s \begin{pmatrix} \frac{1}{2}D & 0 \\ 0 & \frac{1}{6}D \end{pmatrix} B_s^+ \begin{pmatrix} 0 & D^{-1} \\ D^{-1} & 0 \end{pmatrix}, \quad \text{where } B = 0, \tag{4.13}$$

$$\hat{J}_2(s) = B_s^+ \begin{pmatrix} 0 & D^{-1} \\ D^{-1} & 0 \end{pmatrix} B_s,$$

$$\hat{\Phi}(s) = \begin{pmatrix} \frac{1}{2}D & 0 \\ 0 & \frac{1}{6}D \end{pmatrix} B_s^+ \begin{pmatrix} 0 & D^{-1} \\ D^{-1} & 0 \end{pmatrix} B_s, \quad \text{where } B = 0, \tag{4.14}$$

where clearly

$$B_s = \begin{pmatrix} 2D + 2Y & 6Z \\ 3Z_x + ZD + 4ZY & D^2 + 3YD + Y_x + 2Y^2 - 6Z^2 \end{pmatrix},$$

$$B_s^+ = \begin{pmatrix} -2D + 2Y & 3Z_x - DZ + 4ZY \\ 6Z & D^2 - 3DY + Y_x + 2Y^2 - 6Z^2 \end{pmatrix}. \tag{4.15}$$

ACKNOWLEDGMENT

The authors wish to thank H. Pittie for important discussions concerning the coadjoint representation. These discussions came out of a general program of research with one of the authors (R.L.A.).

[1]M. J. Ablowitz and R. Haberman, J. Math. Phys. **16**, 2301 (1975); Phys. Rev. Lett. **35**, 1185 (1975).
[2]L. M. Alonso, J. Math. Phys. **21**, 2342 (1980).
[3]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[4]P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[5]B. Fuchssteiner, Nonlinear Anal. TMA **3**, 849 (1979).
[6]F. Magri, J. Math. Phys. **19**, 1156 (1978); *Nonlinear Evolution Equations and Dynamical Systems*, edited by M. Boiti, F. Pimpinelli, and G. Soliani, *Lecture Noted in Physics*, Vol. 120 (Springer, New York, 1980), p. 233.
[7]I. M. Gel'fand and I. Ya Dorfman, Funct. Anal. Appl. **13**, 13 (1979); **14**, 71 (1980).
[8]B. Fuchssteiner, "Application of Spectral-Gradient methods to nonlinear soliton equations" (preprint).
[9](a) A. P. Fordy and J. Gibbons, J. Math. Phys. **22**, 1170 (1981); (b) B. A. Kuperschmidt and G. Wilson, Invent. Math. **62**, 403 (1981).
[10]I. M. Gel'fand and L. A. Dikii, Funct. Anal. Appl. **10**, 259 (1976).
[11]A. C. Newell, Proc. R. Soc. London Ser. A **365**, 283 (1979).
[12]F.. Calogero and A. Degasperis, Nuovo Cimento B **32**, 201 (1976); Lett. Nuovo Cimento **15**, 65 (1976); **16**, 181 (1976); "Reduction technique for matrix nonlinear evolution equations solvable by the spectral transform" (preprint).
[13]W. Symes, J. Math. Phys. **20**, 721 (1979).
[14]D. J. Kaup, Stud. Appl. Math. **62**, 189 (1980).
[15](1) A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980); (2) B. Fuchssteiner and A. S. Fokas, "Symplectic structures, their Bäcklund transformations and hereditary symmetries," Physica D (Utrecht) **4**, 47 (1981).
[16]B. Fuchssteiner, "The Lie algebra structure of generalized bi-Hamiltonian dynamical systems and their symmetry group" (preprint).
[17]Y. Kosmann-Schwarzback, "Vector fields and Generalized vector fields on fibered manifolds, " Lect. Notes Math. **792**, 307 (1980); Hamiltonian systems on fibered manifolds, to appear in Lett. Math. Phys. (1981).
[18]A. S. Fokas, J. Math. Phys. **21**, 1318 (1980).
[19]This also provides a partial answer to a question posed by J. Greene to one of the authors (A.S.F.), namely, is every strong symmetry hereditary? The answer is affirmative, at least for strong symmetries generated by isospectral problems.
[20]The hierarchy (2.8) is the simplest possible one associated with the hereditary symmetry Φ. One may, for example, replace (2.8) by $u_t + C(\Phi)u_x = 0$, where $C(z)$ is an arbitrary function of z, regular, except possibly at $|z| \to \infty$ and some finite points z_c.
[21]S. Lie and G. Sheffers, *Vorlesungen uber Continuierliche Gruppen* (Teubner, Leipzig, 1893).
[22]R. L. Anderson, J. Harnad, and P. Winternitz, "Systems of ODE's with nonlinear superposition principles," to appear in Physica D (1981); Lett. Math. Phys. **5**, 143 (1981).
[23]F. Magri (private communication).
[24]The Bäcklund transformation corresponds to the projective representation of the pair. See also (1) H. H. Chen, "Relation between Bäcklund transformations and the inverse scattering transform," in *Bäcklund Transformation, the Inverse Scattering Method, Solitons and their Application*, edited by R. Miura, *Lecture Notes in Mathematics*, Vol. 515 (Springer, New York, 1976); (2) A. P. Fordy, "Projective representation and deformation of integrable systems" (preprint).
[25]We have assumed that $B_s \hat{\theta}_1(s) B_s^+ \neq \theta_1(u)$.

Quantum theory as a probability theory on Hilbert space

John F. Cyranski
Physics Department, Hobart and William Smith Colleges, Geneva, New York 14456

(Received 31 August 1981; accepted for publication 8 January 1982)

It is shown how quantum theory (QT) can be expressed as a probability theory on Hilbert space, treated as a measure space. The approach generalizes the work of Bach and clarifies the "generalized trace" of Langerholc. It permits the description of both bounded and unbounded observables as measurable functions. Dynamical evolution can be described in terms of stochastic processes.

PACS numbers: 03.65.Bz

Bach has recently suggested that quantum theory (QT) can be expressed as a theory of probability where Hilbert space is the underlying "event" space.[1-3] While his purpose was to suggest that the elements of Hilbert space can be considered as "hidden variables," the mathematical problem of transcribing the formalism into a quasiclassical context is of intrinsic interest for at least two reasons. The first, as noted by Bach, is that the distinctions between QT and classical physics are more readily apparent if *both* are expressible in the same terms. Secondly, it may prove useful to have an auxiliary formalism available for various calculations. Indeed, among the trivial consequences of our approach is a definition of expectation (in fact, the standard one of probability theory!) which corresponds to the "generalized trace" formula for unbounded observables due to Langerholc.[4] We also note a curious *built-in* "fuzziness" of QT.[5] The description of stochastic processes in QT—a serious problem when viewed from the usual operator formalism[6]—appears to be simplified in our approach.

In this paper we consider the usual Hilbert space over the *complex* field. We thus differ from the standard treatments[7,8] and that of Bach. After a brief description of the Borel algebra, we show how bounded operators—not necessarily self-adjoint (SA)—define measurable functions on the Hilbert space. From this we can immediately do likewise for *any* SA operator. Next, we express density operators as convex integral mixtures defined by a probability measure on Hilbert space. We naturally have a notion of expectation value which we relate to the usual definition. Finally, we define general stochastic processes and show how they are related to the typical Hamiltonian dynamics.

We commence with some remarks on the space $(\mathscr{H}, \mathbb{B}(\mathscr{H}))$. We assume $\mathscr{H}$ to be a general Hilbert space over the complex field $\mathbb{C}$. Its customary topology is based on the metric

$$d(\psi,\phi)\equiv\|\psi-\phi\|, \quad \|\psi\|\equiv(\langle\psi|\psi\rangle)^{1/2}, \tag{1}$$

where $\langle\psi|\phi\rangle$ is the inner product. We denote by $\mathbb{B}(\mathscr{H})$ the σ-algebra generated by the "spheres"

$$S_\phi^\rho=\{\psi\in\mathscr{H}|d(\psi,\phi)\leqslant\rho\}, \quad \rho>0, \quad \phi\in\mathscr{H}. \tag{2}$$

Let $L_N\subset L_{N+1}$ be a sequence of subspaces of $\mathscr{H}$ such that $\cup_{N\geqslant1}L_N=\mathscr{H}$ ($\bar{E}$ denotes the topological closure of $E\subseteq\mathscr{H}$.) Furthermore, choose an orthonormal basis $\{\psi_n\in\mathscr{H}\}_{n\geqslant1}$ such that L_N is spanned by $\{\psi_n\}_{n=1}^N$ and L_{N+1} is spanned by $\{\psi_n\}_{n=1}^{N+1}$ for all $N\geqslant1$. Define the sets

$$\{\psi=\sum_{n=1}^N a_n\psi_n\,|\,a_n\in\mathbb{C};\quad \alpha_n\leqslant\mathrm{Re}(a_n)\leqslant\beta_n;$$
$$\gamma_n\leqslant\mathrm{Im}(a_n)\leqslant\delta_n;\quad -\infty\leqslant\alpha_n<\beta_n\leqslant\infty;$$
$$-\infty\leqslant\gamma_n<\delta_n\leqslant\infty;\quad n=1,2,...,N\}. \tag{3}$$

These sets generate a σ-algebra $\mathbb{B}(L_N)$ and $\mathbb{B}(\mathscr{H})$ is the σ-algebra generated by $\cup_{N\geqslant1}\mathbb{B}(L_N)$. These statements are immediate extensions of results presented in Ref. 7 for a Hilbert space over the real field.

It is necessary to represent linear operators, particularly self-adjoint (SA) "observables" as measurable functions $\mathscr{H}\to\mathbb{C}$. We begin by considering *bounded* operators. If $\hat{T}:D_{\hat{T}}\to\mathscr{H}$ is a linear operator, it is bounded provided there exists $M\in(0,\infty)$ such that $\|\hat{T}\psi\|\leqslant M\|\psi\|$ for all $\psi\in D_{\hat{T}}$. It is well known that this concept is equivalent to continuity:

Lemma 1: The linear operator $\hat{T}:D_{\hat{T}}\to\mathscr{H}$ is bounded iff for each $\psi\in D_{\hat{T}}$ whenever $\{\psi_n\in D_{\hat{T}}\}\to\psi$ ($\|\psi-\psi_n\|\downarrow0$), then $\hat{T}\psi_n\to\hat{T}\psi$. (See Ref. 9, p. 34.)

With each bounded linear operator we can associate a measurable function on $\mathscr{H}$ as follows:

Lemma 2: Let $\hat{T}:D_{\hat{T}}\to\mathscr{H}$ be bounded and let $\hat{T}'$ be its "maximal extension" (i.e., extend $\hat{T}$ to $\hat{T}''$ on $\bar{D}_{\hat{T}}$ by continuity and define $\hat{T}'\psi=0$ if $\psi\in\bar{D}_{\hat{T}}^{\perp}$. For any $\psi\in\mathscr{H}$, $\hat{T}'\psi=\hat{T}''\psi_{\bar{D}_{\hat{T}}}$, where $\psi_{\bar{D}_{\hat{T}}}$ is the component of ψ in $\bar{D}_{\hat{T}}$. (See Ref. 9, p. 34). Then

$$T(\psi)\equiv\langle\psi|\hat{T}'|\psi\rangle/\langle\psi|\psi\rangle, \quad \psi\in\mathscr{H} \tag{4}$$

is a measurable complex-valued function on $\mathscr{H}$.

Proof: Provided $T:\mathscr{H}\to\mathbb{C}$ is continuous, it is measurable (Theorem 1.5, p. 4 of Ref. 8). Note that the operator $\hat{T}'$ on $\mathscr{H}$ is bounded and thus everywhere continuous by Lemma 1. Let $g(\psi)\equiv\langle\psi|\hat{T}'|\psi\rangle$; as $\langle\psi|\psi\rangle$ is obviously continuous, we need only verify the continuity of $g(\psi)$. Let $\{\psi_n\}\to\psi$. Using Schwarz's inequality, we obtain

$$\begin{aligned}|g(\psi_n)-g(\psi)|&\leqslant|\langle\psi_n|[\hat{T}'\psi_n-\hat{T}'\psi]\rangle|+|\langle\psi_n-\psi|\hat{T}'\psi\rangle|\\&\leqslant\|\psi_n\|\cdot\|\hat{T}'\psi_n-\hat{T}'\psi\|+\|\psi_n-\psi\|\cdot\|\hat{T}'\psi\|\\&=[\|\psi_n\|-\|\psi\|]\cdot\|\hat{T}'\psi_n-\hat{T}'\psi\|+\|\psi\|\cdot\|\hat{T}'\psi_n-\hat{T}'\psi\|\\&\quad+\|\psi_n-\psi\|\cdot\|\hat{T}'\psi\|.\end{aligned} \tag{5}$$

Since $\|\psi\|$ and $\|\hat{T}'\psi\|$ are finite and $\{\psi_n\}\to\psi$ implies $\|\hat{T}'\psi_n-\hat{T}'\psi\|\downarrow0$, the right side of (5) goes to zero so $g(\psi)$ is indeed continuous.

Q.E.D.

 0022-2488/82/061074-04$02.50

Our next result shows how we can associate a measurable function on $\mathscr{H}$ with an arbitrary (strictly) self-adjoint (SA) operator. The result depends on the spectral resolution of the operator, so the domain of the operator is necessarily dense—i.e., $\overline{D}_{\hat{T}} = \mathscr{H}$ (Ref. 10, Chap. 6). The spectral representation $\mathbf{B}(\mathbb{R}^1) \to \{\hat{\Pi} \mid \text{projection operator}\}$ characterizes the spectral measure

$$\mu_\psi(\Delta) \equiv \langle\psi|\hat{\Pi}_\Delta|\psi\rangle, \quad \psi \in D_{\hat{T}}, \quad \Delta \in \mathbf{B}(\mathbb{R}^1), \tag{6a}$$

and the quantities (Ref. 9, p. 54)

$$\langle\psi|\hat{T}|\psi\rangle = \int_{-\infty}^{\infty} d\mu_\psi(\lambda)\,\lambda, \quad \psi \in D_{\hat{T}}. \tag{6b}$$

Define for each $z \in [0, \infty)$

$$f_z^+(x) \equiv \begin{cases} x, & \text{if } x \in [0, z], \\ 0, & \text{otherwise,} \end{cases} \tag{7a}$$

$$f_z^-(x) \equiv \begin{cases} x, & \text{if } x \in [-z, 0], \\ 0, & \text{otherwise.} \end{cases} \tag{7b}$$

These functions are "essentially bounded" for each finite z; i.e., for all $\psi \in D_{\hat{T}}$, $\mu_\psi\{\lambda \mid z < |f_z^\pm(\lambda)|\} = 0$. It follows that $f_z^\pm(\hat{T})$ are bounded operators for all finite z (see Ref. 9, p. 56; beware the misprint!). We may therefore define the measurable functions

$$T_z^\pm(\psi) \equiv \langle\psi|f_z'^\pm(\hat{T})|\psi\rangle / \langle\psi|\psi\rangle, \quad \psi \in \mathscr{H}, \quad z \in [0, \infty), \tag{8}$$

where $f_z'^\pm(\hat{T})$ is the maximal extension of $f_z^\pm(\hat{T})$ as defined in Lemma 2. Note that $f_z'^\pm(\hat{T}) = f_z^\pm(\hat{T})$, inasmuch as the latter operators are defined throughout $\mathscr{H}$.

Theorem 1: Let $\hat{T}: D_{\hat{T}} \to \mathscr{H}$ be strictly self-adjoint. Then

$$T(\psi) \equiv \sup\{T_z^+(\psi) \mid z \in [0, \infty)\} + \inf\{T_z^-(\psi) \mid z \in [0, \infty)\} \tag{9}$$

is a measurable function. Its domain excludes only the (measurable) set of ψ's for which both terms on the right side of (9) are unbounded. Moreover,

$$T(\psi) = \langle\psi|\hat{T}|\psi\rangle / \langle\psi|\psi\rangle, \quad \psi \in D_{\hat{T}}. \tag{10}$$

Proof: Since the functions (8) are measurable on $\mathscr{H}$ for all $z \in [0, \infty)$, $\sup\{T_z^+(\psi)\}$ and $\inf\{T_z^-(\psi)\}$ are measurable (Ref. 11, p. 153), and for $\psi \in \mathscr{H}$, such that not both of these functions are unbounded, their sum is measurable. The singular set is also measurable (Ref. 11, pp. 242–3). Formula (10) is immediate from (6), rewritten as

$$\langle\psi|\hat{T}|\psi\rangle = \sup_z\left\{\int_0^z d\langle\psi|\hat{\Pi}_\lambda|\psi\rangle\lambda\right\} + \inf_z\left\{\int_{-z}^0 d\langle\psi|\hat{\Pi}_\lambda|\psi\rangle\lambda\right\}, \quad \psi \in D_{\hat{T}}. \tag{6'}$$

Q.E.D.

We comment that the technique of representing a measurable function as a limit of measurable functions could provide an extension of Theorem 1 to represent any linear operator which can be expressed as a suitably defined limit. We shall not pursue this possibility here as all operators of physical interest are SA.

We note now that any density matrix $\hat{W}$ on $\mathscr{H}$ can be represented most generally as a *convex integral* of "pure states":

$$\hat{W} = \int_{\mathscr{H}} dm(\phi)\,\frac{|\phi\rangle\langle\phi|}{\langle\phi|\phi\rangle}, \quad m \text{ a probability on } \mathbb{B}(\mathscr{H}). \tag{11}$$

In fact we can go somewhat further:

Theorem 2: Let m be a probability measure on $\mathbf{B}(\mathscr{H})$. Then there exists a density matrix $\hat{W}$ defined by (11); i.e., the operator $\hat{W}$ defined by

$$\langle\phi|\hat{W}|\psi\rangle \equiv \int_{\mathscr{H}} dm(\eta)\,\frac{\langle\phi|\eta\rangle\langle\eta|\psi\rangle}{\langle\eta|\eta\rangle}, \quad \phi, \psi \in \mathscr{H}, \tag{12}$$

satisfies

$$\langle\phi|\hat{W}\psi\rangle = \langle\hat{W}\phi|\psi\rangle, \quad \phi, \psi \in \mathscr{H}, \tag{13a}$$

$$\langle\psi|\hat{W}|\psi\rangle \geqslant 0, \quad \psi \in \mathscr{H}, \tag{13b}$$

$$\operatorname{Tr}\hat{W} = 1. \tag{13c}$$

Proof: The operator $|\psi\rangle\langle\phi|$ is clearly bounded:

$$\|(|\psi\rangle\langle\phi|)|\eta\rangle\|^2 = |\langle\phi|\eta\rangle|^2\|\psi\|^2 \leqslant (\|\phi\|^2\|\psi\|^2)\|\eta\|^2, \quad \eta \in \mathscr{H}.$$

Hence, by Lemma 2, the integrand of (12) is measurable with domain $\mathscr{H}$. It follows that $|\langle\phi|\eta\rangle\langle\eta|\psi\rangle/\langle\eta|\eta\rangle|$ is measurable (Ref. 12, p. 11). It is, moreover, integrable:

$$\int dm(\eta)\left|\frac{\langle\phi|\eta\rangle\langle\eta|\psi\rangle}{\langle\eta|\eta\rangle}\right| = \int dm(\eta)\left[\frac{|\langle\phi|\eta\rangle|^2|\langle\eta|\psi\rangle|^2}{\|\eta\|^4}\right]^{1/2} \leqslant \int dm(\eta)[\|\phi\|^2\|\psi\|^2]^{1/2} = [\|\phi\|\,\|\psi\|]m(\mathscr{H}) < \infty \tag{14}$$

so that $|\langle\phi|\hat{W}|\psi\rangle| < \infty$ (Ref. 12, p. 25). Thus, (12) is well defined; properties (13a) and (13b) are trivial, while if $\{\psi_n\}$ is any ON basis for $\mathscr{H}$,

$$\operatorname{Tr}\hat{W} = \sum\langle\psi_n|\hat{W}|\psi_n\rangle = \int dm(\eta)\,\frac{\langle\eta|(\Sigma_n|\psi_n\rangle\langle\psi_n|)|\eta\rangle}{\langle\eta|\eta\rangle} = 1. \tag{15}$$

Q.E.D.

There is an interesting relationship between a probability m on $\mathbb{B}(\mathscr{H})$ and an associated QT probability p on the lattice $L \subset \mathbb{B}(\mathscr{H})$ of closed subspaces of $\mathscr{H}$.[9] Let $\hat{\Pi}_E$ project $\mathscr{H}$ onto $E \in L$.

Lemma 3: Let $q_E(\psi) \equiv \langle\psi|\hat{\Pi}_E|\psi\rangle/\langle\psi|\psi\rangle$, $\psi \in \mathscr{H}$, $E \in L$. Then $q_E: \mathscr{H} \to [0, 1]$ is integrable with

$$p(E) = \int dm(\psi)\,q_E(\psi). \tag{16}$$

Proof: Since $\hat{\Pi}_E$ is bounded, q_E is measurable by Lemma 2. By definition, linearity, and completeness of the basis set $\{\psi_n\}$,

$$\begin{aligned} p(E) &\equiv \operatorname{Tr}\hat{W}\hat{\Pi}_E \\ &= \sum_{n,m}\langle\psi_n|\hat{W}|\psi_m\rangle\langle\psi_m|\hat{\Pi}_E|\psi_n\rangle \\ &= \sum_{n,m}\int dm(\eta)\,\frac{\langle\psi_n|\eta\rangle\langle\eta|\psi_m\rangle\langle\psi_m|\hat{\Pi}_E|\psi_n\rangle}{\langle\eta|\eta\rangle} \\ &= \int dm(\eta)\,\frac{\langle\eta|(\Sigma_{n,m}|\psi_m\rangle\langle\psi_m|\hat{\Pi}_E|\psi_n\rangle\langle\psi_n|)|\eta\rangle}{\langle\eta|\eta\rangle} \\ &= \int dm(\eta)\,q_E(\eta). \end{aligned}$$

Q.E.D.

Note that $q_E(\psi)$ replaces the characteristic function $\chi_E(\psi)$ for $E\in L$. This suggests that QT has an *inherent* "fuzziness" due to the inability to determine the "degree of belonging" of ψ in space E.[5] Note also that while m fixes p uniquely, the converse is not true.

It is natural that the expectation value associated with an SA operator $\hat{T}$ is defined for any probability m on $\mathbb{B}(\mathscr{H})$ by

$$\mathscr{E}(\hat{T}) \equiv \int dm(\phi)\, T(\phi),$$

$$m\{\phi\in\mathscr{H}: |T^+(\phi)| = \infty = |T^-(\phi)|\} = 0, \tag{17}$$

where $T(\phi)$ is defined by (9), which enforces the given restriction on m.

Theorem 3: Let $\hat{T}:\mathscr{H}\to\mathscr{H}$ be bounded, and let m be a probability on $\mathbb{B}(\mathscr{H})$. Then, if $\langle\hat{T}\rangle \equiv \mathrm{Tr}\,\hat{W}\hat{T}$, $\langle\hat{T}\rangle = \mathscr{E}(\hat{T})$

Proof: By definition, if $\hat{T}$ is bounded on $\mathscr{H}$

$$\begin{aligned}\mathscr{E}(\hat{T}) &= \int dm(\psi)\,\frac{\langle\psi|\hat{T}|\psi\rangle}{\langle\psi|\psi\rangle} \\ &= \sum_{n,m} dm(\psi)\,\frac{\langle\psi|\psi_n\rangle\langle\psi_n|\hat{T}|\psi_m\rangle\langle\psi_m|\psi\rangle}{\langle\psi|\psi\rangle} \\ &= \sum_{n,m}\langle\psi_n|\hat{T}|\psi_m\rangle\langle\psi_m|\hat{W}|\psi_n\rangle = \mathrm{Tr}\,\hat{W}\hat{T}.\end{aligned}$$

Q.E.D.

Note that $\hat{T}$ need not be SA in Theorem 3. The definition (17) of expectation—arising naturally from the measure-theoretic description of QT—in fact is identical with an extension of the "trace" operation described in Ref. 4 for unbounded SA operators. This result is important enough to be proven separately.

Theorem 4: Let $\hat{T}$ be a strictly SA operator, and let $\hat{W}$ be defined by (11) where $m\{\phi: |T^+(\phi)| = \infty = |T^-(\phi)|\} = 0$. Then

$$\mathscr{E}(\hat{T}) = \int_{-\infty}^{\infty} \lambda\, d(\mathrm{Tr}\,\hat{W}\hat{\Pi}[-\infty,\lambda]). \tag{18}$$

Proof: Writing $\mathscr{E}(\hat{T})$ explicitly, using (9), we have

$$\begin{aligned}\mathscr{E}(\hat{T}) &= \int dm(\phi)\,\sup\{T_z^+(\phi): z\geqslant 0\} \\ &\quad + \int dm(\phi)\inf\{T_z^-(\phi): z\geqslant 0\} \\ &= \sup_{z\geqslant 0}\int dm(\phi)\,T_z^+(\phi) + \inf_{z\geqslant 0}\int dm(\phi)\,T_z^-(\phi) \\ &= \sup_{z\geqslant 0}\mathscr{E}[f_z^+(\hat{T})] + \inf_{z\geqslant 0}\mathscr{E}[f_z^-(\hat{T})],\end{aligned} \tag{19}$$

where we have used essentially the monotonicity of $T_z^{\pm}(\phi)$ and basic properties of Lebesgue integrals. Explicitly,

$$f_z^+(\hat{T}) = \int_0^z \lambda\, d\hat{\Pi}_{[-\infty,\lambda]}, \tag{20a}$$

$$f_z^-(\hat{T}) = \int_{-z}^0 \lambda\, d\hat{\Pi}_{[-\infty,\lambda]}. \tag{20b}$$

But, application of Theorem 3 in (19) yields

$$\begin{aligned}\mathscr{E}(\hat{T}) &= \sup_{z\geqslant 0}\mathrm{Tr}\,\hat{W}\int_0^z \lambda\, d\hat{\Pi}_{[-\infty,\lambda]} \\ &\quad + \inf_{z\geqslant 0}\mathrm{Tr}\,\hat{W}\int_{-z}^0 \lambda\, d\hat{\Pi}_{[-\infty,\lambda]} \\ &= \sup_{z\geqslant 0}\int_0^z \lambda\, d\,[\mathrm{Tr}\,\hat{W}\hat{\Pi}_{[-\infty,\lambda]}] \\ &\quad + \inf_{z\geqslant 0}\int_{-z}^0 \lambda\, d\,[\mathrm{Tr}\,\hat{W}\hat{\Pi}_{[-\infty,\lambda]}] \\ &= \int_0^{\infty}\lambda d\,[\mathrm{Tr}\,\hat{W}\hat{W}_{[-\infty,\lambda]}] \\ &\quad + \int_{-\infty}^0 \lambda\, d\,[\mathrm{Tr}\,\hat{W}\hat{\Pi}_{[-\infty,\lambda]}].\end{aligned} \tag{21}$$

Q.E.D.

Thus, (18) results.

N.B.: The expectation (19) is defined, of course, only if not both $\mathscr{E}[f_\infty^+(\hat{T})]$ and $\mathscr{E}[f_\infty^-(\hat{T})]$ are unbounded. These are the same conditions required by Ref. 4 in defining the right-hand side of (18), and enforce automatically the assumed restriction on m, again by familiar properties of the Lebesgue integral.

While (17) is valid for *any* measurable function on $\mathscr{H}$, QT is typically concerned with those functions associated with SA operators. Let Γ be the set of measurable functions defined according to Theorem 1 from SA operators. If $T, S\in\Gamma$, we define $T\oplus S\in\Gamma$ as the result of applying Theorem 1 to the SA operator $\hat{T}+\hat{S}$. Similarly, we define $T*S\in\Gamma$ in terms of $\hat{T}\hat{S}$. Note that by definition $\overline{D}_{\hat{T}+\hat{S}} = \mathscr{H} = \overline{D}_{\hat{T}\hat{S}}$—we only admit such SA operators into Γ. This restriction (see Ref. 9, p. 42), nevertheless, leaves us with the most precious physical operators, position and momentum (Ref. 9, p. 43). We use the notation

$$T^{*N} = \underbrace{T*T*T*T*\cdots*T}_{N\text{ times}} \quad\text{and}\quad \bigoplus_{i=1}^{I} T_i = T_1\oplus\cdots\oplus T_I.$$

Note that $T*S\neq S*T$ in general. Obviously, the algebra of QT "random variables" is quite different from that of classical physics.

In particular, $\mathscr{E}(\hat{T}+\hat{S})\neq\mathscr{E}(\hat{T})+\mathscr{E}(\hat{S})$ in general, since $\mathscr{E}(\hat{T}+\hat{S}) = \int dm(T\oplus S)$. The additivity problem rests with the algebra of $\oplus$ and not with the expectation value itself, as suggested by Langerholc.[4]

In general a *stochastic process* is a mapping from $[t_0, t_F]\subseteq\overline{\mathbb{R}}$ into Γ. If a physical system is closed during $[t_0, t_F]$, its unitary evolution $(\hat{U}(t, t_0))$ defines a stochastic process via $T_t(\psi)\equiv T[\hat{U}(t, t_0)\psi]$. In case $\hat{T}$ is bounded on $\mathscr{H}$, this is explicitly evident:

$$T_t(\psi) = \frac{\langle\psi|\hat{U}^\dagger(t, t_0)\hat{T}\hat{U}(t, t_0)|\psi\rangle}{\langle\psi|\psi\rangle} = T[\hat{U}(t, t_0)\psi].$$

More generally any measurable transformation $\hat{V}(t, t_0)$ on $\mathscr{H}$ such that $\hat{V}(t+s, t_0) = \hat{V}(t, t_0)\hat{V}(s, t_0)$ defines a process $T_t(\psi) = T[\hat{V}(t, t_0)\psi]$, $T\in\Gamma$. Note that

$$\begin{aligned}\mathscr{E}[\hat{T}_t] &\equiv \int dm(\psi)\,T[\hat{V}(t, t_0)\psi] \\ &= \int dm[V^{-1}(t, t_0)(\psi)]\,T(\psi),\end{aligned}$$

where $V^{-1}(t, t_0)(\psi) \equiv \{\phi \in \mathscr{H} \mid \hat{V}(t, t_0)\phi = \psi\}$ is not to be confused with an operator inverse.

Most generally, however, the process simply defines a trajectory in Γ which may be an arbitrary function of several variables. Indeed, all physical quantities in QT are Borel functions of coordinates and momenta (and spin) (Ref. 13, Chap. 5). Hence, let $\mathbf{T} \equiv (T_1,...,T_N)$, $T_n \in \Gamma$ $(N < \infty)$ be given, and suppose that the process $[t_0, t_F] \to \Gamma$ admits the representation

$$T_n(t) \equiv \gamma_n[t, t_0; \mathbf{T}], \quad t \in [t_0, t_F], \quad n = 1,2,...,N, \tag{22a}$$

where $\gamma_n(t, t_0; \mathbf{z})$ are Baire functions (not necessarily defined via $\psi \to \psi_t$ maps!) satisfying

$$\gamma_n(t_0, t_0; \mathbf{z}) = z_n, \quad n = 1,2,...,N. \tag{22b}$$

If S is any measurable function on $\mathbb{R}^N$, we can define a process in terms of $\mathbf{T}$ and $\gamma = (\gamma_1,...,\gamma_N)$ by

$$S(t) \equiv S[\gamma(t, t_0; \mathbf{T})]. \tag{23}$$

We shall refer to the functions γ as *dynamical laws*.

We have thus completed the representation of QT on the measure space $(\mathscr{H}, \mathbb{B}(\mathscr{H}))$. That is, we have shown how the (algebra) Γ of admissable QT measurable functions is defined by SA operators, how probabilities and expectations on $(\mathscr{H}, \mathbb{B}(\mathscr{H}))$ are related to QT "states" and expectations, and how "dynamics" can be represented via (more general) stochastic processes. The problem of statistical inference of "states" and dynamical laws based on actual empirical evidence will be treated elsewhere.

ACKNOWLEDGMENTS

This work was supported by the National Hellenic Research Foundation (Greece) and performed at its Institute for Theoretical and Physical Chemistry. I would like to thank Professor C. A. Nicolaides and the members of the Institute for their hospitality and many useful discussions.

[1] A. Bach, "Quantum Mechanics and Integration in Hilbert Space," Phys. Lett. A **73**, 287–8 (1979).
[2] A. Bach, "A probabilistic formulation of quantum theory," J. Math. Phys. **21**, 789–93 (1980).
[3] A. Bach, "Aspects of a Representation of Quantum Theory in Terms of Classical Probability Theory by Means of Integration in Hilbert Space," J. Phys. A **14**, 125–232 (1981).
[4] J. Langerholc, "Trace Formalism for Quantum Mechanical Expectation Values," J. Math. Phys. **6**, 1210 (1965).
[5] A Kaufmann, *Theory of Fuzzy Subsets* (Academic, New York, 1975), Vol. 1.
[6] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[7] A. V. Skorohod, *Integration in Hilbert Space* (Springer-Verlag, Berlin, 1974).
[8] K. R. Parthasarathy, *Probability Measures on Metric Spaces* (Academic, New York, 1967).
[9] J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).
[10] K. O. Friedrichs, *Spectral Theory of Operators in Hilbert Space* (Springer, New York, 1973).
[11] M. E. Munroe, *Introduction to Measure and Integration* (Addison-Wesley, Reading, Mass., 1953).
[12] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966).
[13] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1968), Vol. 1.

A probabilistic formulation of quantum theory. II

Alexander Bach and Thomas Wenning
Institut für Theoretische Physik I, Universität Münster, D-4400 Münster, West Germany

(Received 9 November 1981; accepted for publication 19 February 1982)

An integral representation of the quantum mechanical expectation which recently has been set up for bounded operators is generalized to unbounded operators by means of representation and integration in Gel'fand triplets. For Gaussian measures it turns out that a time translation invariant measure exists which, contrary to classical statistical mechanics, is singular w.r.t. the measures which represent the states.

PACS numbers: 03.65.Ca

I. INTRODUCTION

In Ref. 1 (henceforth cited as I) one of us set up an integral representation of the quantum mechanical expectation by means of integration in Hilbert space H. It was shown that the expectation $E(A;W) = \mathrm{tr}(WA)$ of an operator A can be expressed as

$$E(A;W) = \int_H d\mu_W(\xi) f_A(\xi) \tag{1}$$

if the state of the system is given by the statistical operator W. Mainly to make use of characteristic functions and to define complex Gaussian measures, we introduced a canonical isometry between the complex Hilbert space H and a real one H_0.

If one wants to integrate directly on H, the terms in (1) can be defined as follows. μ_W is a probability measure on the measurable Hilbert space $(H, \mathscr{B}(H))$ which is characterized by the fact that:

(i) Its covariance operator C_{μ_W}, defined by

$$\langle \phi | C_{\mu_W} | \psi \rangle = \int_H d\mu_W(\xi) \langle \phi | \xi \rangle \langle \xi | \psi \rangle, \tag{2}$$

is just the statistical operator W,

(ii) the mean vector m_{μ_W}, defined by

$$\langle \phi | m_{\mu_W} \rangle = \int_H d\mu_W(\xi) \langle \phi | \xi \rangle, \tag{3}$$

vanishes, and

(iii) is arbitrary otherwise.

Finally, the measurable function $f_A: H \to \mathbb{R}$ is given by

$$f_A(\xi) = \langle \xi | A | \xi \rangle. \tag{4}$$

This formula clearly shows that the expression (1) is confined to bounded operators A. For unbounded operators the integrand is not defined on the complement of the domain of the operator. It is the aim of this paper to get rid of these shortcomings and to develop a formulation which is appropriate for that class of operators which usually occur in quantum theory. To this end we make use of a mathematical concept which allows a rigorous application of the Dirac formalism, namely, a Gel'fand triplet $E \subset H \subset E^*$.

The advantage of this procedure is twofold. On the one hand, the restriction of A to E defines a continuous operator such that there is no difficulty in defining a measurable function in analogy to (4). On the other hand, the restriction of W to E defines a bilinear functional on $E \times E$ which induces a Gaussian measure on E^* due to the fact that E is a nuclear space.

In Sec. II we collect the results concerning unbounded operators and Gel'fand triplets and show how to construct complex Gaussian measures on E^*. In Sec. III we present the integral representation of the expectation value. Section IV is devoted to the proof that the dynamics of a quantum system cannot be represented in analogy to classical statistical mechanics by means of invariant measures. Finally, in Sec. V, we collect and discuss the results.

II. CONSTRUCTION OF GAUSSIAN MEASURES ON E^*

To introduce some notations, we recall the basic facts concerning the representation of unbounded operators in Gel'fand triplets (cf., e.g., Ref. 2). Let $A: D(A) \to H$ be an essentially self-adjoint operator defined on a dense domain $D(A) \subset H$. Usually we regard A as an element of a $*$-algebra $\mathfrak{A}$ of operators which have a common maximal invariant dense domain $E \subset H$. This set E is equipped with a topology (the nuclear topology) such that E becomes a nuclear space and the restriction of A to E defines a continuous linear mapping.

We assume in the following that E is a Fréchet space such that its dual (the antilinear continuous functionals) E^* is nuclear too and that E is reflexive, i.e., $E^{**} = E$. The Gel'fand triplet is the sequence $E \subset H \subset E^*$. We denote the pairing of E^* and E by (F, f), $F \in E^*$ and $f \in E$, whereas $\langle \cdot | \cdot \rangle$ denotes the scalar product in H. The restrictions of A and W to E are denoted by A and W again. The adjoint operators which are defined on E^* are denoted by A^+ and W^+, respectively. We remark that in elementary quantum mechanics the Gel'fand triplet can be represented by the Schwartz space S and its dual S^*, the space of tempered distributions, $S(\mathbb{R}^n) \subset L_2(\mathbb{R}^n) \subset S^*(\mathbb{R}^n)$.

Our aim is to represent a bilinear form B_W which is induced on $E \times E$ in terms of the statistical operator W,

$$B_W(f,g) = (f, Wg) = \langle f | W | g \rangle, \tag{5}$$

as an integral on E^*,

$$(f, Wg) = \int_{E^*} d\mu_W(F) \overline{(F,f)}(F,g), \tag{6}$$

where we assume, for simplicity, that B_W is positive definite.

For the construction (cf. Hida[3]) of μ_W as a product measure, we consider H, E, E^* as the complexification of the real Gel'fand triplet $E_r \subset H_r \subset E_r^*$, where H_r, E_r, and E_r^* denote

 0022-2488/82/061078-04$02.50

the associated real spaces (i.e., multiplication with scalars restricted to the reals). Obviously in this representation $E = E_r + iE_r$ etc. such that all elements f, F of E, E^*, respectively, can uniquely be represented as $f = f_1 + if_2$, $F = F_1 + iF_2, f_i \in E_r, F_i \in E_r^*$, $i = 1,2$. The pairing between E^* and E is derived from the canonical pairing of E_r^* and E_r which we denote by $(F,f)_r$ $F \in E_r^*, f \in E_r$ as follows:

$$(F_1 + iF_2, f_1 + if_2) = (F_1,f_1)_r + (F_2,f_2)_r + i[(F_1,f_2)_r - (F_2,f_1)_r]. \quad (7)$$

Let us assume that the operator $W_r : E_r \to E_r$ induces a positive definite continuous bilinear form B_{W_r} on $E_r \times E_r$ via

$$B_{W_r}(f,g) = (f, W_r g)_r \quad (8)$$

and that μ_{W_r} is a probability measure on E_r^* equipped with the Borel algebra generated by the cylinder sets with covariance operator W_r,

$$(f, W_r g)_r = \int_{E_r^*} d\mu_{W_r}(F)\,(F,f)_r (F,g)_r, \quad (9)$$

zero mean, and undetermined on the other hand. On $E^* = E_r^* + iE_r^*$ we define a probability measure μ_W as the product measure

$$\mu_W = \mu_{(W_r/2)} \times \mu_{(W_r/2)}, \quad (10)$$

where E^* is equipped with the σ-algebra generated by the cylinder sets.

As can be seen by explicit calculation, Eq. (6) is fulfilled under the following condition which we formulate as

Proposition 1: The bilinear functional B_W can be represented as the integral (6) on E^* by means of μ_W defined in terms of $\mu_{(W_r/2)}$ if W_r is defined by

$$Wf = W(f_1 + if_2) = W_r f_1 + iW_r f_2, \quad (11)$$

i.e., W_r considered as an operator on H_r has the same spectral representation as W considered on H.

The measures μ_W defined so far are nearly undetermined and in explicit calculations it may be advantageous to have a well-defined measure at one's disposal. In this case we always can use the complex Gaussian γ_W on E^*, which is defined in terms of the Gaussian measure $\gamma_{(W_r/2)}$ on E_r^* with zero mean and covariance operator $W_r/2$. The existence of this measure is assured by the theorem of Minlos (cf. Ref. 4) as E_r is a nuclear space.

III. THE INTEGRAL REPRESENTATION

In the following we assume that an unbounded operator A in H, continuous if restricted to E, and a statistical operator W is given. In order to apply Proposition 1, we assume that W is positive definite. If this is not the case, we confine the integral representation to the Gel'fand triplet generated by the Hilbert space $\overline{\mathrm{ran}(W)}$.

Generalization of the results of I requires, first of all, the definition of the expectation value of an unbounded operator A because the classical formula $E(A;W) = \mathrm{tr}(WA)$ is not well defined in this case. Using the spectral decomposition of A, $A = \int dP_\lambda^A\, \lambda$, we define (cf. Ref. 5)

$$E(A;W) = \int d(\mathrm{tr}(WP_\lambda^A)\,\lambda \quad (12)$$

if the rhs exists and we set $\mathrm{tr}(WA) = \mathrm{tr}(AW) = E(A;W)$ in this case.

If the system of eigenvectors of W, which constitute a c.o.s. in H, is contained in E, i.e., $W = \Sigma_i w_i |w_i\rangle\langle w_i|$ and $|w_i\rangle \in E$ for all $i \in \mathbf{N}$ (we assume for notational convenience, that the eigenspaces are one-dimensional) (12) reduces to the familiar form (cf. Ref. 5)

$$E(A;W) = \sum_i \langle w_i | WA | w_i \rangle, \quad (13)$$

although A is unbounded. We remark that the system of eigenvectors of W cannot, in general, be substituted by some other c.o.s. contained in E as shown in Ref. 5.

In the following we assume that $\mathrm{tr}(WA_+)$ and $\mathrm{tr}(WA_-)$ exist separately, where

$$A_\pm = \int dP_\lambda^A\, (|\lambda| \pm \lambda)/2 \quad (14)$$

and that for $A = A_+ - A_-$

$$\mathrm{tr}(WA) = \mathrm{tr}(WA_+) - \mathrm{tr}(WA_-). \quad (15)$$

This assumption allows us to confine ourselves to positive operators and to rewrite Eq. (13) as follows

Proposition 2: For $A \geq 0$ the expectation (13) can be expressed as

$$E(A;W) = \sum_i \langle A^{1/2} w_i | W | A^{1/2} w_i \rangle, \quad (16)$$

where

$$A^{1/2} = \int dP_\lambda^A\, \lambda^{1/2} \quad (17)$$

is the positive square root of A.

For the proof we remark that $D(A) \subset D(A^{1/2})$ and use the absolute convergence of the sums which result from insertion of the identity in terms of the spectral decomposition of W in $A = A^{1/2} 1 A^{1/2}$.

The functional analogous to (4) is now defined by $f_A : E^* \to \mathbf{R}$:

$$f_A(F) = \sum_i [|(F, A_+^{1/2} w_i)|^2 - |(F, A_-^{1/2} w_i)|^2]. \quad (18)$$

Existence, measurability, and integrability is stated in the following:

Theorem 1: Suppose that $\mathrm{tr}(WA_\pm)$ exists, that (15) is fulfilled, and that the eigenvectors of W are elements of E, then

$$E(A;W) = \int_{E^*} d\mu_W(F) f_A(F) \quad (19)$$

where μ_W is defined in terms of $\mu_{(W_r/2)}$ by Eq. (10) and W_r is characterized by Eq. (11).

Proof: According to the assumptions we can confine ourselves to positive operators and apply Proposition 2:

$$\begin{aligned} \sum_i \langle A^{1/2} w_i | W | A^{1/2} w_i \rangle &= \sum_i (A^{1/2} w_i, WA^{1/2} w_i) \\ &= \sum_i \int_{E^*} d\mu_W(F) |(F, A^{1/2} w_i)|^2 \\ &= \int_{E^*} d\mu_W(F) \sum_i |(F, A^{1/2} w_i)|^2, \end{aligned} \quad (20)$$

where for the last identity the theorem of monotone convergence has been used. This assures the μ_W-a.e. existence of f_A and its integrability.

For operators with discrete spectrum the integrand can be expressed in a more condensed form.

Proposition 3: Suppose that the assumptions of Theorem 1 are fulfilled and that A has a discrete spectrum, e.g., $A = \sum_i a_i |a_i\rangle \langle a_i|$, then (19) holds with $f_A(F)$ given by

$$f_A(F) = \sum_i a_i |(F,a_i)|^2. \tag{21}$$

For the proof we consider $A \geqslant 0$. According to definition (12), we obtain

$$\begin{aligned} E(A;W) &= \sum_i a_i \langle a_i | W | a_i \rangle \\ &= \sum_i \int_{E^*} d\mu_W(F)\, a_i |(F,a_i)|^2 \end{aligned} \tag{22}$$

and use the theorem of monotone convergence to arrive at (21).

Let us conclude this section with a remark concerning the complicated structure of the measurable function f_A for unbounded operators if compared with the simple structure of that of bounded operators in Eq. (4). The structure of f_A in (18) results from the fact that the conditions of evaluation of the trace in a special basis as well as properties of bilinear functionals on E^* must be respected. The fact that we define f_A in terms of the square root of A is technical and can be avoided for operators with discrete spectrum.

IV. SINGULARITY WITH RESPECT TO INVARIANT MEASURES

Quantum dynamics is induced by a one-parameter group U_t, $t \in \mathbb{R}$, where we assume that the generators of U_t are elements of the $*$-algebra $\mathfrak{A}$ under consideration. We assume that the elements of the one-parameter group preserve the Hilbert norm of elements of H and are homeomorphisms of E onto itself. All operators with the latter properties form a group $\mathcal{U}(E)$ referred to as the infinite-dimensional unitary group. The adjoint mappings U^+ are automorphisms of E^* and the collections of all of these operators forms a group which is isomorphic to $\mathcal{U}(E)$ and denoted by $\mathcal{U}^+(E^*)$.

From our viewpoint quantum dynamics is induced by elements of $\mathcal{U}^+(E^*)$ on E^*. The following theorem states that there exists a measure on E^* which is invariant w.r.t. elements of $\mathcal{U}^+(E^*)$.

Theorem *2:* For all $U^+ \in \mathcal{U}^+(E^*)$ we have

$$U^+ \circ \gamma_1 = \gamma_1, \tag{23}$$

where γ_1 is the complex Gaussian measure on E^* which is induced by the identity on E considered as the defining bilinear functional B_1.

For the proof see Ref. 3. We remark that this representation of the identity on E can be considered as a generalized version of a representation in coherent states (cf., e.g., Ref. 6).

The existence of a rotation invariant measure on E^* discerns the present formalism from that one which was established in I as on the measurable Hilbert space there exist neither translation-invariant nor rotation-invariant measures. As the time evolution of quantum mechanics is induced by elements of $\mathcal{U}^+(E^*)$, *Theorem 2 is the quantum analog of the classical Liouville theorem.* It states that there exists a time-translation invariant measure on the state space.

In this context it is interesting to pursue the classical analogy further and to analyze whether it is possible to represent an arbitrary state characterized by a probability measure μ_W by means of a probability density ρ_W w.r.t. the invariant measure γ_1. The existence of a representation of the expectation value by means of the invariant measure γ_1 is assured in

Proposition 4: Under the assumptions of Theorem 1 the expectation value permits a representation

$$E(A;W) = \int_{E^*} d\gamma_1\, \Gamma(F)\ \Phi_{WA}(F), \tag{24}$$

where $\Phi_{WA}: E^* \to \mathbb{R}$ is defined in analogy to f_A in Eq. (18) with $A_\pm^{1/2}$ replaced by $W^{1/2}A_\pm^{1/2}$.

For the proof we set $W = W^{1/2}W^{1/2}$ and use (16). We remark that in general $(WA_\pm)^{1/2} \neq W^{1/2}A_\pm^{1/2}$.

Representation of the dynamics by means of a probability density requires factorization of Φ_{WA} according to

$$\Phi_{WA}(F) = \rho_W(F) f_A(F) \tag{25}$$

with (i) ρ_W independent of A such that it only represents the state but not the observable and (ii) $d\gamma_1\, \rho_W = d\mu_W$ for some μ_W. From the structure of the functions Φ_{WA} and f_A it is evident that Φ_{WA}/f_A still depends on A.

These considerations which seem to indicate that we cannot represent the state by means of a probability density can be sharpened for Gaussian measures.

Theorem *3*: Let W be a positive-definite statistical operator and denote by γ_W the complex Gaussian measure induced by W on E^* then γ_1 and γ_W are orthogonal

$$\gamma_1 \perp \gamma_W\,. \tag{26}$$

For the proof we use a theorem due to Minlos (Proposition 8, VIII in Ref. 4). From this theorem we conclude that the orthogonality of the measures is due to the fact that W is a trace-class operator, a fact which is constitutive for a statistical operator.

Formula (26) implies that integrals w.r.t. γ_W cannot be represented by means of the measure γ_1. From a more physical point of view this means that *the dynamics cannot be represented by a time-dependent probability density w.r.t. an invariant measure.* This precisely is the point where the analogy between the present representation of quantum mechanics and classical statistical mechanics breaks down.

V. DISCUSSION AND CONCLUSION

We have generalized the integral representation set up in I for bounded operators to unbounded ones by means of Gel'fand triplets and integration in nuclear spaces. The restrictions of this representation of the quantum mechanical expectation are of the same kind as those for the generalized trace formula. This can best be seen from Eq. (18) which requires that the eigenvectors of W be elements of $D(A_\pm)$.

The starting point for our consideration was the expression for the expectation by means of the generalized trace formula which is confined to normal states. As singular states are connected with the continuous part of the spectrum of an observable and the generalized eigenvectors are elements of E^* but not of H the representation by means of integration on E^* may indicate that the present formalism is partly based on singular states.

That this is not the case can be seen from the characterization of the supports of the measures involved.

Theorem 4: (i) Denote by

$$E \subset \cdots \subset E_n \subset \cdots \subset E_0 = H = E_0^* \subset \cdots \subset E_n^* \subset \cdots \subset E^*,$$

$n \in \mathbf{N}$, the countable Hilbert space defining the nuclear spaces E, E^* (cf. Ref. 3). Then γ_1 is supported by E_1^*. Moreover, (H, E_1^*) constitute an abstract Wiener space for γ_1.

(ii) Denote by H_W the Hilbert space obtained by completion of H with respect to the norm $\|\xi\|_W = \|W^{1/2}\xi\|$ for $\xi \in H$ and by H_W^* its dual. Obviously, $H_W^* = W^{1/2}H \subset H$. Then the complex Gaussian measure γ_W is supported by H. Moreover, (H_W^*, H) constitute an abstract Wiener space for γ_W.

The proof is a direct application of the theorem of Piech.[7] For the notation of an abstract Wiener space and its properties we refer to Ref. 8. The second part of the theorem allows us to restrict the integrals w.r.t. γ_W to the familiar quantum mechanical Hilbert space H. This clarifies that there is no connection with singular states.

Another important property can be concluded from the restriction of the integral representation to H based on Eq. (21). The theorem of monotone convergence yields in this case that the set of elements $\xi \in H$ for which $\langle \xi | A | \xi \rangle < \infty$ is measurable, i.e., $D(A)$ is an element of the Borel algebra of H, and that the complement of the domain is a set of measure zero. We remind, however, that these results are based on observables with a discrete spectrum and Gaussian measures—they need not hold in general.

From Theorem 4 we conclude that with each complex Gaussian measure γ on E^* there is associated an abstract Wiener space (H_1, H_2) constituted by the two Hilbert spaces $H_1 \subset H_2$. The general properties of the abstract Wiener spaces associated with a Gaussian measure γ imply that $\gamma(H_1) = 0$ and $\gamma(H_2) = 1$. Evidently, the orthogonality of γ_1 and γ_W stated in Theorem 3 is a consequence of the fact that $\gamma_1(H) = 0$ and $\gamma_W(H) = 1$.

To conclude, let us consider the structural analogy of the present representation of quantum theory with classical statistical mechanics and its limitations.

We have set up a probabilistic representation of the quantum mechanical expectation which associates a probability measure γ_W with a statistical operator W and a measurable function f_A with an operator A in H. The fact that the measurable functions defined by Eq. (18) still depend via the eigenvectors of W on the state is structurally connected with the definition of generalized trace. Conceptually this dependence reflects the fact that unbounded observables do not admit finite expectations in arbitrary states.

Just as every system in classical statistical mechanics has a distinct phase space each quantum system, characterized by its algebra of operators $\mathfrak{A}$, has its state space E^* which is determined by this algebra. In analogy to the classical Liouville measure, we have shown the existence of a time-translation invariant measure in state space, namely the rotation-invariant canonical complex Gaussian measure. We remark that in the present representation the Liouville measure is a probability measure whereas classical Liouville measures are in general not even finite measures.

Contrary to classical statistical mechanics it is not possible to describe a state by means of a probability density w.r.t. the invariant measure. This is not primarily connected with the fact that the operator algebra under consideration is noncommutative but is due to the orthogonality of the measures involved which is a typical property of infinite-dimensional analysis.

ACKNOWLEDGMENTS

We would like to thank Professor J. Kamphusmann and Professor D. Plachky for stimulating discussions and valuable suggestions.

[1] A. Bach, J. Math. Phys. **21**, 789 (1980).
[2] A. Böhm, *The Rigged Hilbert Space and Quantum Mechanics* (Springer-Verlag, Berlin, 1978).
[3] T. Hida, *Brownian Motion* (Springer-Verlag, New York, 1980).
[4] K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups* (Polish Scientific Publishers, Warsaw, 1968).
[5] J. Langerholc, J. Math. Phys. **6**, 1210 (1965).
[6] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[7] M. A. Piech, Proc. Am. Math. Soc. **53**, 460 (1975).
[8] H.-H. Kuo, *Gaussian Measures in Banach Spaces* (Springer-Verlag, Berlin, 1975).

Invariant *-quantization associated with the affine group

T. V. Huynh [a)]
Department of Physics, University of California, Los Angeles, California 90024

(Received 9 June 1981; accepted for publication 2 September 1981)

Invariant *-quantization means quantization with the introduction of invariant *-products on C^∞ functions on phase space. A systematic method of constructing invariant *-products for dynamical Lie groups is presented. It involves the notions of invariant Wigner correspondence and *-representation. As an illustration, all possible invariant *-products, in an integral form, are constructed for the affine group of the real line. The Moyal product is recovered as a special case.

PACS numbers: 03.65.Ca

INTRODUCTION

Quantization, according to Bayen *et al.*,[1] is viewed as a deformation of the usual associative algebra of C^∞ functions on phase space which is attained by introducing a deformed product (called *-product) on $C^\infty W$. The *-product is required to be "invariant" with respect to an algebra of "distinguished observables" for the latter to retain their geometrical significance after the deformation. The main concern of this paper is to construct all possible invariant *-products for the affine group of the real line. It is the group of translations and dilations, without reflections, of the real line.

This paper is divided into two main parts: the concept of invariant *-quantization and its application to the affine group. In the first part of the paper we discuss the basic ideas of invariant *-quantization, the concept of invariant Wigner correspondence (or map), the notion of *-representation, and the related invariant *-products. The second part of the paper is devoted to the contruction of invariant *-products for the affine group of the real line. We wish to find all possible invariant Wigner maps and all possible *-representations, with which we obtain an integral representation of all possible invariant *-products for the affine group.

1. INVARIANT *-PRODUCT QUANTIZATION

*-Quantization,[2] a short form of *-product quantization, is a framework for the description of both classical and quantum mechanics, within which the quantization process is continuous—without the introduction of a Hilbert space and with no radical change in the nature of classical observables. Continuity is in the sense of the deformation theory[3]—the mathematical foundation of *-quantization. Quantum mechanics is a deformation of classical mechanics.

The basic mathematical structures of classical mechanics are the symplectic structures attached to phase space W: the symplectic form on W defines the Poisson bracket which induces a Lie algebra structure on the algebra N of C^∞ functions on W with ordinary multiplication. Let a bilinear, associative, internal composition law, called a *-product, be introduced on $C^\infty W$, with respect to which N is closed $[(u,v)\to u*v,\ u,\ v,\ u*v\in N]$ and is a *-algebra, denoted $(N, *)$, with the associated Lie algebra structure defined by

$$(u,v)\to[u*v] = : (u*v - v*u)/i\hbar;$$

$[\ *\ ]$ is called the Moyal bracket.[4] It is this bracket, not the Poisson bracket, that corresponds to the quantum commutator.

The physical aspect of the continuity of the quantization process requires that a family of observables on W, called "good or distinguished observables," be preserved after the deformation. For example, some functions that generate space-time symmetry should continue to do so after quantization. By definition,[5] a good observable generates, by the Poisson bracket, a group of symplectic diffeomorphisms of W. Denote by $\mathbf{g}\subset N$ a finite algebra of good observables with the property that the algebra $\mathbf{g}$ integrates to a group G of diffeomorphisms of W. This group G is sometimes called the dynamical group of a physical system. For $A\in\mathbf{g}$ to remain good after the deformation, it is required that the infinitesimal automorphisms of the *-algebra $(N, *)$ defined by $f\to[A*f]$ generate a group of automorphisms, $f\in N$. The requirement[6] is satisfied simply by

$$A*f - f*A = i\hbar\{A,f\},\quad A\in\mathbf{g},\quad f\in N. \tag{1.1}$$

The *-product is then said to be $\mathbf{g}$-invariant, that is, A-invariant for $A\in\mathbf{g}$.

Furthermore, it will be supposed that $\mathbf{g}$ is "sufficiently large" so that its basis may be used to coordinatize W. If $\{L_i\}$ is a basis of $\mathbf{g}$ and $\{\Gamma^i\}$ is the dual basis of the vector space dual $\mathbf{g}^*$, then the map $J: W\to\mathbf{g}^*$ given by $\xi\in W\to L_i(\xi)\Gamma^i$ is injective,[7] $i = 1,\ldots,\dim\mathbf{g}$. We identify[8] W with an orbit of the coadjoint action $\mathrm{ad}^*\mathbf{g}$ of $\mathbf{g}$ in the dual $\mathbf{g}^*$.

By invariant *-quantization is thus meant quantization on a symplectic space W by means of a *-product defined on $N = C^\infty W$, invariant under a sufficiently large finite subalgebra $\mathbf{g}\subset N$ of good observables. We denote a *-quantization by a triplet $(W,\mathbf{g},*)$.

The problem of quantization of a physical system with a known dynamical Lie group G thus reduces to (i) determining the corresponding Lie algebra $\mathbf{g}$ and its associated good observables, (ii) identifying W with an orbit of the coadjoint action $\mathrm{ad}^*\mathbf{g}$ of $\mathbf{g}$ in $\mathbf{g}^*$, and (iii) selecting an invariant *-product on W. If G is known, (iii) is the crux of the problem.

A. Invariant Wigner correspondence and invariant *-products

A Wigner correspondence $\mathscr{W}$ maps a class of operators in a Hilbert space $L^2(R^n)$ to a class of functions or distribu-

[a)] Research Scientist, Lockheed Palo Alto Research Laboratory, Palo Alto, California 94303.

 0022-2488/82/061082-06$02.50

tions on phase space W. The precise domain of $\mathscr{W}$ will not be determined; it is assumed that, on a proper topological subspace $\mathscr{S}$ of its domain, the linear map $\mathscr{W}$ is a function-valued distribution. Let F be an operator in $\mathscr{S}$, Ker F is its integral kernel. Then the image of F under $\mathscr{W}$, denoted $F_{\mathscr{W}} \in N$, will be assumed to be expressible as

$$F_{\mathscr{W}}(\xi) = \int dk\, d\kappa\, \mathrm{Ker}\, F(k,\kappa)\mathscr{W}(k,\kappa;\xi), \quad \xi \in W, k,\kappa \in R^n, \tag{1.2}$$

where $\mathscr{W}(\cdot)$ is a distribution. Note that $\mathscr{W}$, in contrast to $\mathscr{W}(\cdot)$, refers to the correspondence.

The image $F_{\mathscr{W}}$ is called the Wigner symbol[9] of F. In the Weyl–Wigner–Moyal formalism[10] the Wigner distribution function is the form

$$\mathscr{W}(k,\kappa;p,q) = \exp[q(\kappa - k)/i\hbar]\delta[p - (k+\kappa)/2],$$

where $(q,p) \in W = R \times R$.

The Wigner correspondence defines a $*$-product on N in a natural manner. Let $u,v \in N$ and U,V be operators such that $u = U_{\mathscr{W}}$, $v = V_{\mathscr{W}}$, then

$$u*v =: (UV)_{\mathscr{W}} \in N. \tag{1.3}$$

A Wigner correspondence $\mathscr{W}$ is called g-invariant or simply invariant, if for every U in g

$$i\hbar\{f,u\} = [F,U]_{\mathscr{W}}, \tag{1.4}$$

where F is an operator in the domain[11] of $\mathscr{W}$ and $f = F_{\mathscr{W}}$.

Equation (1.4) readily yields (1.1) and thus guarantees the g-invariance of the $*$-product. This is evident, since $[F,U]_{\mathscr{W}} = i\hbar[f*u]$ according to (1.3). We shall determine all possible $\mathscr{W}(\cdot)$ obeying (1.4) for the affine group. We next discuss the concept of $*$-representation and its relationship to invariant $*$-products.

B. *-Representations and invariant *-products

Let an invariant $*$-product be defined on W and $\mathscr{P}$ be the algebra of formal power series over $\mathbf{g}^*$. By Exp, called $*$-exponential,[12] is meant the function $\mathrm{Exp}: X \in \mathbf{g} \to \mathrm{Exp}(X) \in \mathscr{P}$, defined by

$$\mathrm{Exp}(X) = \sum_{n=0}^{\infty} (1/n!)(i\hbar)^{-n}(X*)^n,$$

$$(X*)^n \equiv X*\cdots*X \quad (n \text{ factors}).$$

We shall establish that $X \mapsto \mathrm{Exp}(X)$ is a $*$-representation of G.

By definition,[13] a $*$-representation of G on W is a distribution $\mathscr{E}$ on G, with values in $C^\infty W$, such that

(i) The domain and the kernel of $\mathscr{E}$ are closed under the convolution in the test function space and

(ii) $\mathscr{E}$ is ad g-invariant in the sense that, for every $Y \in \mathbf{g}$,

$$\mathscr{E} \circ (\mathrm{ad}\, Y) = (\mathrm{ad}^* Y) \circ \mathscr{E}. \tag{1.5}$$

Condition (i) makes it possible to define a $*$-product on the image of $\mathscr{E}$ in $C^\infty W$ by

$$\mathscr{E}(\tilde{f}) * \mathscr{E}(\tilde{f}') = \mathscr{E}(\tilde{f} \circ \tilde{f}'), \tag{1.6}$$

where $\tilde{f}, \tilde{f}'$ are the test functions. Condition (ii) ensures that this $*$-product is invariant in the sense that, for every $Y \in \mathbf{g}$,

$$\{Y, u*v\} = \{Y,u\}*v + u*\{Y,v\},$$

which is equivalent to (1.1).

A $*$-representation $\mathscr{E}$ of G on W is smooth[14] if there exists a smooth function E defined on a neighborhood Λ of the identity of G, which gives rise to $\mathscr{E}(\tilde{f})$ according to

$$\mathscr{E}(\tilde{f}) = \int_{\Lambda} E(x^{-1})\tilde{f}(x)\, d_r x =: (E, \tilde{f}), \quad x \in \Lambda \subset G,$$

in which $d_r x$ is the right-invariant Haar measure on G. Furthermore, $\mathscr{E}$ is called normalized if, in addition, $E|_{id} = 1$ and $l(X)E|_{id} = -X/i\hbar$, where $l(x) \in l(\mathbf{g})$, the space of vector fields on G associated with left translations.

Suppose that $\mathscr{E}$ is a smooth and normalized $*$-representation. We may rewrite (1.6) as

$$E(e^X) * E(e^Y) = E(e^X e^Y), \tag{1.7}$$

where X, Y are in a neighborhood $\mathscr{V}$ of the identity of g.

The cannonical, local diffeomorphisms from $\mathscr{V}$ to Λ allows[15] us to regard $\mathrm{Exp}(X)$ as a function on Λ. All elements of G will be taken to belong to some open sets of Λ, small enough so that their products will be in Λ. With this understanding we define $E(e^X) =: \mathrm{Exp}(X)$; $X \mapsto \mathrm{Exp}(X)$ is thus a $*$-representation. It follows from (1.7) that

$$\mathrm{Exp}(X) * \mathrm{Exp}(Y) = \mathrm{Exp}(Z(X,Y)), \tag{1.8}$$

in which $Z(X,Y)$ is given by the Baker–Hausdorff–Campbell formula.

By virtue of (1.7), the $*$-representation $\mathrm{Exp}(X)$ is invariant in the sense that, for $Y = y^i L_i$ in g,

$$(\mathrm{ad}\, Y + \mathrm{ad}^* Y)\mathrm{Exp}_\xi(X) = 0, \tag{1.9}$$

in which $\mathrm{Exp}_\xi(X)$ is the value of $\mathrm{Exp}(X)$ at $\xi = \xi_i \Gamma^i$ on W. It is understood[16] that ad Y is the extension of the adjoint action of Y in g to formal power series over g and ad$^* Y$ the extension of the coadjoint action of Y in $\mathbf{g}^*$ to formal power series over $\mathbf{g}^*$. Thus,

$$\begin{aligned}(\mathrm{ad}\, Y)\,\mathrm{Exp}_\xi(X) &=: [Y,X]^i \frac{\partial}{\partial x^i}\mathrm{Exp}_\xi(X) \\ &= C^i{}_{jk} y^j x^k \frac{\partial}{\partial x^i}\mathrm{Exp}_\xi(X),\end{aligned} \tag{1.10}$$

$$\begin{aligned}(\mathrm{ad}^* Y)\,\mathrm{Exp}_\xi(X) &=: [\mathrm{ad}^* Y)\xi]_j \frac{\partial}{\partial \xi_j}\mathrm{Exp}_\xi(X) \\ &= -C^i{}_{jk}\xi_i y^j \frac{\partial}{\partial \xi_k}\mathrm{Exp}_\xi(X),\end{aligned}$$

in which $\{C^i{}_{jk}\}$ is the structure tensor of the Lie algebra g, and all the indices reflect the dimension of g. By virtue of (1.10), Eq. (1.9) becomes

$$\left(C^i{}_{jk} y^j x^k \frac{\partial}{\partial x^i} - C^i{}_{jk}\xi_i y^j \frac{\partial}{\partial \xi_k}\right)\mathrm{Exp}_\xi(X) = 0. \tag{1.11}$$

Equation (1.11) or (1.9), like (1.4), is just another expression of the g-invariance of the $*$-product, associated with the $*$-representation $X \mapsto \mathrm{Exp}(X)$.

We have thus far shown that, given an invariant $*$-product on W, we obtain $*$-representations that satisfy (1.9). We now discuss the converse: To construct invariant $*$-products on W with the help of $*$-representations.

We shall consider smooth and normalized $*$-represen-

tations associated with unitary representations of G. By definition,[17] a $*$-representation $\mathscr{E}$ is said to be associated with a unitary representation T of G if there exists an invariant Wigner correspondence $\mathscr{W}$ such that $\mathscr{E} = \mathscr{W} \circ T$. Since $\mathscr{E}$ is smooth, $E(e^X) = T_{\mathscr{W}}(X) =: \mathrm{Exp}(X)$. By virtue of (1.8) the functions Exp form a local group under $*$-multiplication, which is locally homomorphic to G. The group multiplication law of G gives rise to, through (1.8), the group $*$-multiplication law of the group of the $*$-representations Exp. It is (1.8) that aids in the construction of invariant $*$-products. It is now clear as to why we choose $*$-representations that are associated with unitary representations of G.

As a brief summary, once a unitary representation of G and an invariant Wigner correspondence, hence a type of $*$-representation, are established, the group property of Exp expressed by (1.8) allows us to define an invariant $*$-product on W.

2. INVARIANT $*$-PRODUCTS FOR THE AFFINE GROUP

A. Representations of the affine group

We consider the case of a single degree of freedom with the requirement of positive momentum ($p > 0$). This has been applied[18] in a model of quantum theory of gravitation.

Momentum space is the real half-line $p > 0$, and two-dimensional phase space (W) is the real half-plane $p > 0$, $-\infty < q < \infty$. The symplectic structure is given by $\{q,q\} = 0 = \{p,p\}$ and $\{q,p\} = 1$. Conventional Weyl quantization would associate q and p with operators Q and P, respectively, in $L^2(R^+)$, where R^+ is taken to be the half-positive momentum space; in this case, P is multiplication by $p \in R^+$ and Q is of the form $i\hbar\partial/\partial p$. However, from the physical requirement of $p > 0$ it follows that Q is not self-adjoint. Consequently, it is illogical to attempt invariant quantization based on the Heisenberg algebra (p, q, 1 with $\{q,p\} = 1$) as the algebra of distinguished observables. We must point out immediately that we are not rejecting the quantum theory which is discovered by the conventional procedure; on the contrary, we shall adopt it. We just choose our distinguished observables among those that correspond to operators which generate unitary transformations.

As in the case in Ref. 18, we shall let the Lie algebra $\mathbf{g}$ be spanned by two functions p and pq; its Lie-algebra structure is defined by $\{pq, p\} = p$. The Lie algebra $\mathbf{g}$ is recognized as the Lie algebra of the affine group, the two-parameter, non-abelian group of translations and dilations (without reflections) of the real line. The mapping of phase space W into R^2 by $(q,p) \mapsto (\xi_1, \xi_2)$, where $\xi_1 = pq$ and $\xi_2 = p$, may be interpreted as a mapping of W onto an orbit of the coadjoint action of $\mathbf{g}$ in its real vector space dual $\mathbf{g}^*$. Phase space W is thus equipped with coordinated ξ_1 and ξ_2; an element $\xi \in W$ is indicated by $\xi = (\xi_1, \xi_2)$. Furthermore, the Poisson bracket in (1.4) is

$$\{u, v\}(\xi) = \xi_2\left(\frac{\partial u}{\partial \xi_1}\frac{\partial v}{\partial \xi_2} - \frac{\partial u}{\partial \xi_2}\frac{\partial v}{\partial \xi_1}\right).$$

Let G denote the affine group. We write

$$G = \left\{(b,a): z \mapsto az - \eta(a)b;\ a > 1,\ b \in R, \text{ and } \eta(a) =: \frac{a-1}{\ln a}\right\}.$$

Notice that since $a > 1$, $\eta(a) > 0$. The group multiplication law is given by

$$(b,a)(b',a') = \left(\frac{\eta(a)b + a\eta(a')b'}{\eta(aa')}, aa'\right).$$

The affine group G admits two inequivalent, irreducible, unitary representations[19] which we now obtain by the method of induced representations.[20]

Consider the subgroups $G_1 = \{(b,1)\} \subset G$, the group of translations on R, and $G_0 = \{(0,a)\} \subset G$, the multiplicative group on R. The affine group G may be written as $G = G_1 \circledS G_0$, a semidirect product of the two one-parameter groups G_1 and G_2. The character $(b,1) \to \exp(b/i\hbar)$ of G_1 induces the following representations of G:

$$[T(b,a)\phi](k) = a^{1/2}\exp[\eta(a)bk/i\hbar]\phi(ak), \quad \phi \in L^2(R^+). \tag{2.1}$$

The infinitesimal operators of T that correspond to the one-parameter subgroups G_1 and G_0 and form a basis of the Lie algebra $\mathbf{g}$ of G may be obtained as follows. From $[T(0,e^t)\phi](k) = e^{t/2}\phi(e^t k)$ and $[T(t,1)\phi](k) = e^{tk/i\hbar}\phi(k)$, in which $(0,e^t) \in G_1$ and $(t,1) \in G_0$, it follows that

$$\begin{aligned} L_1\phi(k) &\equiv i\hbar\frac{d}{dt}[T(0,e^t)\phi(k)]_{t=0} \\ &= i\hbar\left(\frac{1}{2} + k\frac{\partial}{\partial k}\right)\phi(k) \end{aligned}$$

and

$$L_2\phi(k) \equiv i\hbar\frac{d}{dt}[T(t,1)\phi(k)]_{t=0} = k\phi(k). \tag{2.2}$$

The operators L_1 and L_2 operate on $L_2(R)$, and $[L_1, L_2] = i\hbar L_2$. If we write $[L_j, L_k] = i\hbar C^i{}_{jk} L_i$, then the structure constants of $\mathbf{g}$ are given by

$$C^1{}_{12} = C^1{}_{21} = 0 \quad \text{and} \quad C^2{}_{12} = -C^2{}_{12} = 1. \tag{2.3}$$

Consider a family of unitary operators $\exp[(\ln a\, L_1 + bL_2)/i\hbar]$, which may be expressed with the help of the Baker–Hausdorff–Campbell formula as follows:

$$\exp[(\ln a\, L_1 + bL_2)/i\hbar] = \exp[\eta(a)bL_2/i\hbar]\exp[\ln a\, L_1/i\hbar].$$

By virtue of (2.2), clearly $(b,a) \to \exp[(\ln a\, L_1 + bL_2)/i\hbar]$ is a unitary, irreducible, representation of the affine group G. Thus we write

$$T(b,a) = \exp[\ln a\, L_1 + bL_2)/i\hbar]. \tag{2.4}$$

As already mentioned above, the affine group admits two inequivalent, unitary, irreducible representations: one for which the spectrum of L_2 is positive and one for which it is negative. However, both may be realized in $L^2(R^+)$. In fact, the representation in (2.1) corresponds to the positive spectrum of L_2; the other representation may be obtained from (2.1) with b replaced by $-b$.

Furthermore, the unitary operators $T(b,a)$ form a group with the multiplication law

$$T(b,a)T(b',a') = T\left(\frac{\eta(a)b + a\eta(a')b'}{\eta(aa')}, aa'\right). \tag{2.5}$$

Our ultimate goal is to characterize all possible $\mathbf{g}$-invariant $*$-products on phase space W. It follows from (1.9) that an invariant Wigner correspondence $\mathscr{W}$—that is, a $\mathbf{g}$-invar-

iant Wigner distribution $\mathscr{W}(\cdot)$—is sought. Once an invariant Wigner correspondence is established, we then obtain the $*$-representation associated with the unitary representation of the group. The $*$-representation satisfies the invariance condition (1.10), and the Wigner distribution $\mathscr{W}(\cdot)$ satisfies (1.4). The scheme for constructing a g-invariant product is therefore as follows. All invariant Wigner distributions $\mathscr{W}(\cdot)$ are first obtained by solving (1.4); these yield, by means of (1.2), all $*$-represenatations that satisfy (1.11). All possible g-invariant $*$-products are then constructed with the help of (1.8).

B. Invariant Wigner distribution

We now seek a characterization of all Wigner maps $\mathscr{W}$ that are invariant with respect to the affine Lie algebra g in the sense of (1.4). We impose (1.4). First, let $U = L_1$ and $u = \xi_1$ with the integral kernel of L_1 in (2.2) given by

$$\text{Ker } L_1(k,\kappa) = i\hbar\left(\frac{1}{2} + k\frac{\partial}{\partial k}\right)\delta(k-\kappa).$$

It then follows from (1.4) that

$$\left(\xi_2\frac{\partial}{\partial \xi_2} + k\frac{\partial}{\partial k} + \kappa\frac{\partial}{\partial \kappa}\right)\mathscr{W}(k,\kappa;\xi) = \mathscr{W}(k,\kappa;\xi), \tag{2.6}$$

which implies that $\mathscr{W}(k,\kappa;\xi)$ is a homogeneous function of degree (-1) in k,κ,ξ_1, and ξ_2. Next, taking $U = L_2$ and $u = \xi_2$ with the integral kernel of L_2 in (2.2) given by Ker $L_2(k,K) = k\delta(k-\kappa)$ yields

$$i\hbar\xi_2\frac{\partial \mathscr{W}(k,\kappa;\xi)}{\partial \xi_1} = (\kappa - k)\mathscr{W}(k,\kappa;\xi). \tag{2.7}$$

The general solution to (2.6) and (2.7)—that is, the general invariant Wigner distribution—is of the form

$$\mathscr{W}(k,\kappa;\xi) = \tilde{\mathscr{W}}(k,\kappa;\xi_2)\exp[(\kappa-k)\xi_1/i\hbar\xi_2], \tag{2.8}$$

in which $\tilde{\mathscr{W}}$, by virtue of (2.6), is a homogeneous function of degree (-1) in k,κ, and ξ_2. Recall that the usual Wigner distribution is of the form (2.8) with $\tilde{\mathscr{W}}(k,\kappa;\xi_2) = \delta[\xi_2 - (k+\kappa)/2]$. We have thus arrived at

Proposition 1: If $\mathscr{W}$ is an invariant Wigner correspondence, then the Wigner distribution $\mathscr{W}(\cdot)$ is given by (2.8).

C. $*$-Representation

We now apply (1.11) to the Lie algebra g of the affine group G. By virtue of (2.3) and the arbitrariness of Y in g, we obtain from (1.11) the following system of differential equations:

$$\left(x^2\frac{\partial}{\partial x^2} - \xi_2\frac{\partial}{\partial \xi_2}\right)\text{Exp}_\xi(X) = 0$$
$$\left(x^1\frac{\partial}{\partial x^2} - \xi_2\frac{\partial}{\partial \xi_1}\right)\text{Exp}_\xi(X) = 0, \tag{2.9}$$

in which $x^1 = \ln a$ and $x^2 = b$. The g-invariance condition on a $*$-product on W thus requires that the $*$-representation $X \mapsto \text{Exp}(X)$ satisfy (2.9). The general solution to (2.9) is clearly of the form

$$\text{Exp}_\xi(X) =: [T_{\mathscr{W}}(X)](\xi) = f(x^1, x^1\xi_1 + x^2\xi_2), \tag{2.10}$$

in which $x^1\xi_1 + x^2\xi_2 = \ln a\,\xi_1 + b\xi_2$. From here on we shall use X and (b,a) interchangeably.

We now relate the generality of $f(x^1, x^1\xi_1 + x^2\xi_2)$ to the arbitrariness of the function $\tilde{\mathscr{W}}(k,\kappa;\xi_2)$ in the Wigner distribution $\mathscr{W}(\cdot)$ by

Proposition 2: Let $(b,a) \mapsto T(b,a)$ be the unitary representation (2.4) of the affine group G and W the orbit of the coadjoint action of its Lie algebra g in the dual g*. Then associated with $T(b,a)$ is the $*$-representation $\text{Exp}_\xi(b,a)$, evaluated at $\xi \in W$, which is given by

$$\text{Exp}_\xi(b,a) =: [T(b,a)_{\mathscr{W}}\cdot](\xi)$$
$$= \int d\zeta \exp[\eta(a)(\ln a\,\xi_1 + b\xi_2)\zeta/i\hbar]\,\tilde{\mathscr{W}}(\zeta,a), \tag{2.11}$$

in which $\tilde{\mathscr{W}}$ is a distribution on R, $\xi = (\xi_1,\xi_2)$ with $\xi_2 > 0$, and $-\infty < \xi_1 < \infty$.

Proof: By virtue of (1.2),

$$[T(b,a)_{\mathscr{W}}\cdot](\xi) = \int dk\,d\kappa\,\text{Ker } T(k,\kappa;b,a) \times \exp[(\kappa - k)\xi_1/i\hbar\xi_2]\,\tilde{\mathscr{W}}(k,\kappa;\xi_2), \tag{2.12}$$

in which the invariant Wigner distribution $\mathscr{W}(\cdot)$ in (2.8) has been used. From (2.1) it follows that

$$\text{Ker } T(k,\kappa;b,a) = a^{1/2}\exp[\eta(a)\,bk/i\hbar]\delta(ak - \kappa).$$

Substitution of Ker $T(k,\kappa;b,a)$ in (2.12) leads, with a change of integration variables $(k \mapsto \zeta = k/\xi_2, \kappa \mapsto \pi = \kappa/\xi_2)$ and the use of the homogeneity of degree (-1) of $\tilde{\mathscr{W}}$, to the following expression for $[T(b,a)\mathscr{W}](\xi)$:

$$[T(b,a)_{\mathscr{W}}\cdot](\xi) = \int d\zeta\,d\pi\,a^{1/2}\exp[\eta(a)b\xi_2\zeta/i\hbar] \times \exp[(\pi - \zeta)\xi_1/i\hbar]\delta(a\zeta - \pi)\tilde{\mathscr{W}}(\zeta,\pi)$$

which, upon integration, yields (2.11).

It is obvious from (2.11) that $\text{Exp}_\xi(b,a)$ is of the form (2.10). Equation (2.11) provides a general expression of $\text{Exp}_\xi(b,a)$; an explicit expression of $\text{Exp}_\xi(b,a)$ depends on a particular form of $\tilde{\mathscr{W}}(\zeta,a)$. If we employ the usual Wigner distribution, namely,

$$\mathscr{W}(k,\kappa;\xi) = \exp[(\kappa - k)\xi_1/i\hbar\xi_2]\delta[\xi_2 - (k+\kappa)/2], \tag{2.13}$$

then a simple calculation shows that

$$\text{Exp}_\xi(b,a) = a^{1/2}\exp[2\eta(a)(\ln(a)\xi_1 + b\xi_2)/(a+1)i\hbar]. \tag{2.14}$$

This agrees, except for a normalization constant, with the result reported in Ref. 21, in which Weyl quantization is extended to the metaplectic algebra $h_l *(\oplus^{l}\ \text{sp}(l,R))$ and the Moyal product is used.

Notice that, like (2.11), $\text{Exp}_\xi(b,a)$ in (2.14) is in the form of (2.10). Thus, in effect, we have two independent derivations of the general form of the $*$-representation $X \mapsto \text{Exp}(X)$: one by means of (2.9) and the other via (2.11).

D. Integral representation of invariant $*$-products

The $*$-representation associated with a unitary representation of a group can be used to relate a function on the Lie algebra of the group to a function on its dual.[22] Let $\check{u}$ be a function on g. The $*$-representation $(b,a) \mapsto \text{Exp}(b,a)$ associated with the unitary representation $(b,a) \mapsto T(b,a)$, evaluated

at $\xi \in W$, allows us to define a function u on $\mathbf{g}^*$ by

$$u(\xi) = \int d\mu(b,a)\, \check{u}(b,a)\, \mathrm{Exp}_{\xi}(b,a), \tag{2.15}$$

in which $d\mu(b,a)$ is the Lebesgue measure on $\mathbf{g}$, and Exp is assumed to act on the whole algebra $\mathbf{g}$. Equation (2.15) is a type of generalized Fourier transform: u is said to be the Exp transform of $\check{u}$, and $\check{u}$ is the inverse Exp transform of u.

Recall that the Wigner symbols $T(b,a)_{\mathscr{W}} = \mathrm{Exp}(b,a)$ form a group under $*$-multiplication, which is homomorphic to G. The group $*$-multiplication law is expressd by, via (2.5),

$$\mathrm{Exp}(b,a)*\mathrm{Exp}(b',a') = \mathrm{Exp}\left(\frac{\eta(a)b + \eta(a')ab'}{\eta(aa')}, aa' \right). \tag{2.16}$$

We now show that the Exp transform (2.15) can be used, with the help of (2.16), to obtain an integral representation of invariant $*$-product for the affine group.

By virtue of the Exp transform (2.15) and the group $*$-multiplication law (2.16), we obtain

$$(u*v)(\xi) = \int d\mu(b,a)\, d\mu(b',a')\, \check{u}(b,a)\, \check{v}(b'a') \times \mathrm{Exp}_{\xi}\left(\frac{\eta(a)b + \eta(a')ab'}{\eta(aa')}, aa' \right), \tag{2.17}$$

in which u, v are functions on $\mathbf{g}^*$ and $\check{u}$, $\check{v}$ their inverse Exp transforms, which are functions on $\mathbf{g}$.

The arbitrariness of $\tilde{\mathscr{W}}(\cdot)$ in (2.11) makes it impossible to obtain the inverse Exp transforms $\check{u}$ and $\check{v}$ from (2.15). Equation (2.17) is therefore considered to be the general integral representation of all invariant $*$-products on W for the affine group. Before we obtain a specific invariant $*$-product that corresponds to a particular choice of $\tilde{\mathscr{W}}(\cdot)$, let us summarize the general result thus far obtained in the following proposition:

Proposition 3: Let $(b,a) \mapsto \mathrm{Exp}(b,a)$ in (2.11) be the $*$-representation of the affine group associated with the unitary representation $(b,a) \mapsto T(b,a)$ in (2.1) and C^∞ functions on W be expressible as the Exp transforms in (2.15). Then the general integral representation of all invariant $*$-products on phase space W is given by (2.17).

We now show that (2.17) reduces to, with the use of the usual Wigner distribution (2.13), the integral version[23] of the Moyal product. A simple calculation, using (2.15), (2.14), and $d\mu(b,a) = da\, db\, /a$ (the right-invariant measure), shows that the inverse Exp transform $\check{u}(b,a)$ may be obtained from

$$\check{u}(b,a) = a^{1/2} \int d\xi\, u(\xi) \exp[-2\eta(a)(\ln a\, \xi_1 + b\xi_1)/(a+1)i\hbar], \tag{2.18}$$

where $d\xi$ is the Liouville measure on W.

Substitution of (2.18) for $\check{u}$ and $\check{v}$ in (2.17), followed by integration, leads, with the help of a change in the integration variables, to

$$(u*v)(\xi) = \int d\xi'\, d\xi''\, u(\xi + \xi')v(\xi + \xi'')\exp[2\omega(\xi',\xi'')/i\hbar], \tag{2.19}$$

in which $\xi' = (\xi_1', \xi_2')$, $\xi'' = (\xi_1'', \xi_2'')$, $d\xi' = d\xi_1' d\xi_2'/\xi_2'$, $d\xi'' = d\xi_1'' d\xi_2''/\xi_2''$, and $\omega(\xi',\xi'')$ is defined by

$$\omega(\xi',\xi'') =: \frac{\xi_1'}{\xi_2'}\xi_2'' - \frac{\xi_1''}{\xi_2''}\xi_2'.$$

This is exactly of the form of the integral representation of the Moyal product. As a matter of fact, with $\xi_1 = pq$ and $\xi_2 = p$ we recover from (2.19) the integral version of the Moyal product on phase space $W = R \times R$. We have thus proved

Proposition 4: Let the invariant Wigner map correspond to the invariant Wigner distribution in (2.13), the Wigner symbol of unitary operator $T(b,a)$ in (2.4) be given by (2.14), and C^∞ functions on W be expressible by the Exp transform (2.15). Then the Moyal product is an invariant $*$-product for the affine group and is expressed by the integral representation in (2.19).

3. CONCLUDING REMARKS

In the realm of $*$-quantization, quantizing a physical system of a known dynamical Lie group really means constructing all possible invariant $*$-products for the group. In this work we have demonstrated a systematic method of constructing invariant $*$-products. It involves, mainly, characterizing all possible invariant Wigner maps, identifying all possible $*$-representations associated with unitary representations of the group, and utilizing the group property of Exp (1.8) to construct invariant $*$-products. An invariant Wigner correspondence arises from the invariance condition (1.4) of a $*$-product. The general form of a $*$-representation may be derived in two independent ways: one by imposing the invariance of the $*$-representation (1.9) and the other through the invariant Wigner correspondence (2.11). Furthermore, the $*$-representation associated with a unitary representation of the group may be used to relate, by means of the Exp transform (2.15), functions on the Lie algebra of the group and functions on its dual. This feature of the $*$-representations plays a useful role in the construction of invariant $*$-products in an integral form.

In this work we apply this systematic approach to the case of a single degree of freedom with the requirement of positive momentum, to which the affine group is relevant. We obtain an integral expression (2.17) for all possible invariant $*$-products. We also recover the Moyal product (2.19) when the usual Wigner distribution (2.13) is employed.

We believe that this systematic method can be extended to other dynamic groups, hence providing a systematic way of quantizing physical systems. This allows us to regard $*$-quantization as an alternative paradigm of quantum theory.

ACKNOWLEDGMENTS

It is a pleasure for me to thank Professor Christian Fronsdal, Physics Department, UCLA, for his kind help and guidance and for his careful reading of the manuscript. I also thank Professor Moshe Flato, University of Dijon, France, for suggesting the problem to me, and Professors R. Arens and D. Babbitt, Mathematics Department, UCLA, for helpful discussions. Finally, I am grateful to the NSF and to the UCLA Research committee for partial support of this work.

[1]F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (NY) **111**, 61 (1978), I.
[2]F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (NY) **111**, 111(1978), II.
[3]See Ref. 1, p. 71.
[4]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[5]See Ref. 2, p. 117.
[6]See Ref. 1, p. 89.
[7]D. Arnal and J. C. Cortet, "La Notion de *-Produit et ses Applications aux Représentations de Groupes," Conférence aux Journeés Relativistes 79–Angers (unpublished).
[8]See Ref. 1, p. 89.
[9]A. Vorös, J. Funct. Anal. **29**, 104 (1978).
[10]See Ref. 2, p. 112.
[11]C. Fronsdal, "Invariant *-quantization of the one dimensional Kepler problem," J. Math. Phys. **20**, 2226 (1979).
[12]See Ref. 1, p. 102.
[13]C. Fronsdal, Rep. Math. Phys. **15**, 111 (1978), p. 114.
[14]See Ref. 13, p. 115.
[15]See Ref. 13, p. 122.
[16]See Ref. 1, p. 102.
[17]See Ref. 11, p. 21.
[18]J. R. Klauder and E. W. Aslaksen, Phys. Rev. D **2**, 272 (1970).
[19]I. M. Gel'fand and M. A. Naimark, Dokl. Akad. Nauk. SSSR **55**, 570 (1947).
[20]A. A. Kirillov, *Elements of the Theory of Representations* (Springer, New York, 1976).
[21]G. Burdet and M. Perrin, Lett. Math. Phys. **2**, 93–99 (1977).
[22]See Ref. 1, 107.
[23]T. V. Huynh, Lett. Math. Phys. **4**, 201 (1980), p. 205.

The symmetry group of the harmonic oscillator and its reduction

Toshihiro Iwai
Department of Applied Mathematics, Gifu University, Yanagido, Gifu 501-11, Japan

(Received 14 December 1981; accepted for publication 29 January 1982)

The symmetry of the harmonic oscillator is dealt with in the Hamiltonian formalism. Unitary operators representing the symmetry are studied from this point of view. Of additional interest is reduction of the symmetry group SU(4) for the four-dimensional harmonic oscillator. Subspaces are determined from the representation spaces for SU(4) so as to give those for SO(4).

PACS numbers: 03.65.Ca, 02.20. + b

I. INTRODUCTION

This article deals with a quantum system as a Hamiltonian dynamical system which was formulated by Marsden.[1-3] A purpose of this paper is to discuss the symmetry group of the quantum harmonic oscillator in the Hamiltonian formalism. The harmonic oscillator is a rather simple system whose symmetry is well known. Of particular interest is then the symplectic point of view of symmetry transformations, which has received little attention. Another purpose is to set up a quantum version of a previous paper,[4] in which a reduction of the four-dimensional classical harmonic oscillator was dealt with. The results to be obtained will be utilized in the next paper.[5]

Section II contains Hamiltonian formalism of the quantum harmonic oscillator. Section III is concerned with the symmetry group dealt with in the Hamiltonian formalism. For the harmonic oscillator the usual technique of correspondence for constructing quantum operators from classical first integrals offers no problem, as far as the Cartesian coordinates are concerned. Of central interest is to integrate those operators to give unitary operators which describe the symmetry of the quantum system. Section IV is devoted to the four-dimensional harmonic oscillator. Reduction of eigenspaces of the Hamiltonian operator is discussed together with accompanying reduction of the symmetry group.

II. HAMILTONIAN FORMALISM OF THE HARMONIC OSCILLATOR

The classical harmonic oscillator is described on the space $\mathbb{R}^n \times \mathbb{R}^n$. Let (x_j, p_j) be the Cartesian coordinates. Introducing the coordinates

$$z_j = \lambda x_j + i p_j \quad [i = (-1)^{1/2}], \tag{2.1}$$

where λ is a positive constant, one can equip $\mathbb{R}^n \times \mathbb{R}^n$ with the structure of an n-dimensional complex vector space $\mathbb{C}^n$. The Hamiltonian function is then written in the form

$$H = \tfrac{1}{2} \Sigma\, p_j^2 + (\lambda^2/2) \Sigma\, x_j^2 = \tfrac{1}{2} \Sigma \bar{z}_j z_j, \tag{2.2}$$

where $\bar{z}_j$ is the complex conjugate to z_j. Further properties of the classical harmonic oscillator will be recalled when required.

To formulate the quantum harmonic oscillator as a Hamiltonian system, we have first to designate an infinite dimensional symplectic manifold. According to Marsden,[2] the suitable symplectic manifold and the symplectic form ω are, respectively, $L^2(\mathbb{R}^n)$, the Hilbert space of square integrable complex functions on $\mathbb{R}^n$, and

$$\omega(X, Y) = -\operatorname{Im}\langle X, Y\rangle, \tag{2.3}$$

where $X, Y \in L^2(\mathbb{R}^n)$ and $\operatorname{Im}\langle X, Y\rangle$ denotes the imaginary part of the inner product

$$\langle X, Y\rangle = \int_{\mathbb{R}^n} \bar{X} Y\, dx \tag{2.4}$$

with dx the standard volume element on $\mathbb{R}^n$.

The Hamiltonian function is defined as a quadratic form in $\varphi \in L^2(\mathbb{R}^n)$ with the Hamiltonian operator $\hat{H}$ by

$$H(\varphi) = \tfrac{1}{2}\langle \varphi, \hat{H}\varphi\rangle. \tag{2.5}$$

Here $\hat{H}$ is the self-adjoint extension of the operator determined from (2.2) by the Schrödinger procedure: $p_j = -i\partial/\partial x_j$. For simplicity we use the same letter for the operators formed from (2.1) as in the classical system. Then one has

$$\hat{H} = \tfrac{1}{2}\sum \bar{z}_j z_j + n\lambda/2 = \tfrac{1}{2}\sum z_j \bar{z}_j - n\lambda/2, \tag{2.6}$$

where we do not distinguish the essentially self-adjoint operator, the right-hand sides (2.6), from its self-adjoint extension $\hat{H}$.

The time evolution is governed by Hamilton's equation

$$\frac{d\varphi}{dt} = X_H(\varphi), \tag{2.7}$$

where X_H is the Hamiltonian (or canonical) vector field determined by

$$i(X_H)\omega = -dH, \tag{2.8}$$

$i(\ )$ denoting the interior product.[2] Of course, X_H is defined in the domain of $\hat{H}$, dense in $L^2(\mathbb{R}^n)$. Applying (2.8) to (2.5), we have $X_H(\varphi) = -i\hat{H}\varphi$, so that Eq. (2.7) becomes the Schrödinger equation $d\varphi/dt = -i\hat{H}\varphi$.

We here make a mention of operators z_j and $\bar{z}_j$. The usual commutation relations are given by

$$[z_j, \bar{z}_k] = 2\lambda\delta_{jk}, \quad \text{and the others vanishing.} \tag{2.9}$$

In what follows, instead of dwelling upon the domain of z_j and $\bar{z}_j$, we understand that operators which are polynomials in z_j and $\bar{z}_j$ are defined at least in the linear subspace, dense in $L^2(\mathbb{R}^n)$,

$$\left\{(\text{polynomials in } x_j)\exp\left(-\tfrac{1}{2}\lambda\sum x_j^2\right)\right\}. \tag{2.10}$$

The operators z_j and $\bar{z}_j$ are then conjugate to each other.

In the remainder of this section we review the eigen-

 0022-2488/82/061088-05$02.50

spaces of $\hat{H}$. Let

$$\varphi_0(x_1,\ldots,x_n) = (\lambda/\pi)^{n/4}\exp(-\tfrac{1}{2}\lambda\sum x_j^2) \tag{2.11}$$

be the normalized ground state. All of normalized eigenfunctions are expressed in the form

$$\varphi_{k_1\cdots k_n} = \lambda^{-N/2}C_{k_1\cdots k_n}\bar{z}_1^{k_1}\cdots\bar{z}_n^{k_n}\varphi_0 \tag{2.12a}$$

$$= C_{k_1\cdots k_n}H_{k_1}(\lambda^{1/2}x_1)\cdots H_{k_n}(\lambda^{1/2}x_n)\varphi_0, \tag{2.12b}$$

where $C_{k_1\cdots k_n} = (2^N k_1!\cdots k_n!)^{-1/2}$ and $N = k_1 + \cdots + k_n$ with k_j a nonnegative integer and the H_{k_j} are Hermite's polynomials defined by

$$H_k(t) = (-1)^k\exp(t^2)(d/dt)^k\exp(-t^2). \tag{2.13}$$

The functions (2.12) form a complete orthonormal system in $L^2(\mathbb{R}^n)$. Every eigenspace is assigned by the nonnegative integer $N = k_1 + \cdots + k_n$:

$$\hat{H}\varphi_{k_1\cdots k_n} = \lambda(N + n/2)\varphi_{k_1\cdots k_n}. \tag{2.14}$$

The dimension of the eigenspace, or the degeneracy of the energy level $\lambda(N + n/2)$ is $\binom{N+n-1}{n-1}$, a binomial coefficient.

III. THE SYMMETRY GROUP

We first make a brief review of the symmetry of the classical harmonic oscillator (see Ref. 4). A general first integral is written in the form

$$F = (\lambda/2i)\sum C_{jk}z_k\bar{z}_j, \quad \mathrm{tr}(C_{jk}) = 0, \tag{3.1}$$

where $C = (C_{jk})$ is an anti-Hermitian matrix with vanishing trace. The Hamiltonian vector field X_F associated with F and the symplectic (or canonical) transformations $\exp(tX_F/\lambda)$ generated by X_F are, respectively, given by

$$X_F = -\lambda\sum C_{jk}z_k\frac{\partial}{\partial z_j} - \lambda\sum\overline{C_{jk}z_k}\frac{\partial}{\partial\bar{z}_j}, \tag{3.2a}$$

$$\exp(tX_F/\lambda): \quad z\to\exp(-tC)z, \quad \bar{z}\to\overline{\exp(-tC)z}, \tag{3.2b}$$

where $z = (z_j)$ and $\bar{z} = (\bar{z}_j)$ are column vectors and $\exp(-tC)\in SU(n)$. To quantize the classical observable F, we follow the Schrödinger procedure by writing out F in terms of x_j and p_j and substituting $-i\partial/\partial x_j$ for p_j. We use the symbols z_j and $\bar{z}_j$ as operators in the same way as in (2.6). Let (A_{jk}) and (B_{jk}) be the real and imaginary parts of (C_{jk}), respectively:

$$C_{jk} = A_{jk} + iB_{jk}. \tag{3.3}$$

Then, regardless of the condition $\mathrm{tr}(C_{jk}) = 0$, the quantized operator $\hat{F}$ takes the form

$$\hat{F} = \frac{1}{2}\sum B_{jk}\left(\lambda^2x_jx_k - \frac{\partial^2}{\partial x_j\partial x_k}\right) - \frac{\lambda}{i}\sum A_{jk}x_k\frac{\partial}{\partial x_j} \tag{3.4a}$$

$$= \frac{1}{2i}\sum C_{jk}z_k\bar{z}_j + \frac{i\lambda}{2}\mathrm{tr}(C_{jk}) \tag{3.4b}$$

$$= \frac{1}{2i}\sum C_{jk}\bar{z}_jz_k - \frac{i\lambda}{2}\mathrm{tr}(C_{jk}). \tag{3.4c}$$

It is clear that $\hat{F}$ is a symmetric operator on the domain (2.10). Commutation relations are calculated to give

$$\left[\frac{i}{\lambda}\frac{1}{2i}\sum C_{jk}z_k\bar{z}_j, \frac{i}{\lambda}\frac{1}{2i}\sum D_{jk}z_k\bar{z}_j\right] = \frac{i}{\lambda}\frac{1}{2i}\sum[C,D]_{jk}z_k\bar{z}_j, \tag{3.5}$$

where $[C,D]_{jk}$ denote the (j,k) components of the matrix $[C,D]$. Here the conditions $\mathrm{tr}(C_{jk}) = \mathrm{tr}(D_{jk}) = 0$ have not been required, so that Eq. (3.5) is applicable to the operator $\hat{H} + n\lambda/2$ (see (2.6)). Since $\hat{H} + n\lambda/2$ has the unit matrix as the coefficient matrix, Eq. (3.5) ensures that $\hat{H}$ and $\hat{F}$ commute. Moreover, Eq. (3.5) shows that the mapping

$$(C_{jk})\to(i/\lambda)(1/2i)\sum C_{jk}z_k\bar{z}_j \tag{3.6}$$

is a Lie algebra homomorphism. From now on, we assume that $\mathrm{tr}(C_{jk}) = 0$, so that the ordering of z_j and $\bar{z}_j$ in expressing $\hat{F}$ offers no problem, as is seen from (3.4).

We turn to symplectic transformations associated with $\hat{F}$. Like (2.5), to the operator $\hat{F}$ there corresponds a quadratic form in φ: $F(\varphi) = \frac{1}{2}\langle\varphi,\hat{F}\varphi\rangle$. The Hamiltonian vector field X_F[6] is determined by the same condition as (2.8) to be $X_F(\varphi) = -i\hat{F}\varphi$. The symplectic transformations generated by X_F will be expressible by $\exp(tX_F/\lambda) = \exp(-it\hat{F}/\lambda)$ as in (3.2b). Our problem is then to show the existence and property of the operator $\exp(-it/\hat{F}/\lambda)$.

Before approaching the problem, we touch upon a quantum version of the classical symplectic transformation (3.2b). Let us regard z_j and $\bar{z}_j$ in (3.2b) as operators. Then the linear transformation

$$z_j' = \sum(e^{-tC})_{jk}z_k, \quad \bar{z}_j' = \sum\overline{(e^{-tC})}_{jk}\bar{z}_k \tag{3.7}$$

makes no change in the canonical commutation relations (2.9), as $\exp(-tC)$ belongs to $SU(n)$. In this sense the transformation (3.7) could be called a canonical transformation. However, in the Hamiltonian formalism we are not accurate in calling it so because canonical (or symplectic) transformations must be defined on $L^2(\mathbb{R}^n)$ so as to leave the symplectic structure (2.3) invariant.

We are now going to give a definite meaning to the operator $\exp(it\hat{F}/\lambda)$. In view of (3.7) we first define a one-parameter family of functions $\Phi_{k_1\cdots k_n}(t)$ on $\mathbb{R}^n$ by

$$\Phi_{k_1\cdots k_n}(t) = \left(\sum(e^{tC})_{j_1 1}\bar{z}_{j_1}\right)^{k_1}\cdots\left(\sum(e^{tC})_{j_n n}\bar{z}_{j_n}\right)^{k_n}\varphi_0. \tag{3.8}$$

It is clear that $\Phi_{k_1\cdots k_n}(0) = \bar{z}_1^{k_1}\cdots\bar{z}_n^{k_n}\varphi_0$ and that for all $t\in\mathbb{R}$ the function (3.8) still remains in the eigenspace for $\hat{H}$ assigned by $N = k_1 + \cdots + k_n$. If we look on $\bar{z}_j$ as formal variables, each eigenspaces of $\hat{H}$ is regarded as the space of homogeneous polynomials in $\bar{z}_j$ of degree $N = k_1 + \cdots + k_n$, in which the unitary group $U(n)$ is represented unitarily and irreducibly.[7] By T we mean the representation.[8] Then we have

$$\Phi_{k_1\cdots k_n}(t) = T(e^{tC})\bar{z}_1^{k_1}\cdots\bar{z}_n^{k_n}\varphi_0. \tag{3.9}$$

Let us consider the representations T's for all N at the same time. These define a one-parameter family of invertible linear mappings T_t from the dense subspace (2.10) of $L^2(\mathbb{R}^n)$ onto itself. We recall that the inner product of functions $\bar{z}_1^{k_1}\cdots\bar{z}_n^{k_n}\varphi_0$ is calculated by using the commutation relations (2.9). Since the transformation (3.7) makes no change in the

commutation relations (2.9), the mappings T_t defined above take the complete orthonormal system (2.12a) into another. Therefore, T_t extends uniquely to a unitary operator U_t on $L^2(\mathbb{R}^n)$. Moreover, as T is a representation, the operators U_t, $t\in\mathbb{R}$, form a one-parameter group. Thus we have in particular

$$\Phi_{k_1\cdots k_n}(t) = U_t \bar{z}_1^{k_1}\cdots\bar{z}_n^{k_n}\varphi_0. \tag{3.10}$$

In this respect the domain of $\hat{H}$ is worth noticing.

Remark 3.1: The action of the unitary operator U_t decomposes into unitary transformations of the eigenspaces for $\hat{H}$, so that U_t leaves the domain of $\hat{H}$ invariant, as is readily verified by the spectral decomposition of $\hat{H}$ (see also Ref. 9).

Our next task is to find the infinitesimal generator of U_t. To this end, we calculate the derivative of $\Phi_{k_1\cdots k_n}(t)$ with respect to t to obtain, after a long calculation,

$$\frac{d}{dt}\Phi_{k_1\cdots k_n}(t) = \frac{i}{\lambda}\hat{F}\Phi_{k_1\cdots k_n}(t). \tag{3.11}$$

We notice here only that use has been made of the commutation relations (2.9) together with their immediate consequences and of the fact that $z_j\varphi_0 = 0$.

From Eqs. (3.10) and (3.11) it follows that

$$\frac{d}{dt}U_t f = \frac{i}{\lambda}\hat{F}U_t f \tag{3.12}$$

for any function f in the domain (2.10). Thus we may conclude[10]

Theorem 3.2: The symmetric operator $\hat{F}$ extends to a self-adjoint operator to generate the unitary operator

$$U_t = \exp(it\hat{F}/\lambda), \tag{3.13}$$

where $\hat{F}$ and its extension are not distinguished in notation.

Furthermore, since U_t is a unitary operator and the symplectic form (2.3) is defined through the inner product, one obtains

Theorem 3.3: The operator U_t is a symplectic transformation on $L^2(\mathbb{R}^n)$.

Mention should be here made of angular momentum operators. We discuss the unitary operator (3.13) with the restriction that the coefficient matrix (C_{jk}) of $\hat{F}$ is a real matrix (A_{jk}). Then $\hat{F}$ becomes, from (3.4a), an angular momentum operator

$$\hat{F} = -\frac{\lambda}{i}\sum A_{jk}x_k\frac{\partial}{\partial x_j}. \tag{3.14}$$

Consider the unitary mapping $f\to U_t f$. For simplicity we restrict f within the domain (2.10). From (3.12) and (3.14), U_t satisfies

$$\frac{d}{dt}U_t f = -\sum A_{jk}x_k\frac{\partial}{\partial x_j}U_t f. \tag{3.15}$$

On the other hand, a one-parameter group $\exp(tA)$ acting on $\mathbb{R}^n$ is represented by $f(\exp(-tA)x)$, $x\in\mathbb{R}^n$. Simple calculation yields

$$\frac{d}{dt}f(\exp(-tA)x) = -\sum A_{jk}x_k\frac{\partial}{\partial x_j}f(\exp(-tA)x). \tag{3.16}$$

From (3.15) and (3.16) we see that $U_t f(x)$ and $f(\exp(-tA)x)$ satisfy the same differential equation with the same initial condition at $t = 0$. Therefore, we obtain

$$U_t f(x) = f(\exp(-tA)x), \tag{3.17}$$

which extends so as to hold for all f in $L^2(\mathbb{R}^n)$.

We return to (3.7). The operators z_j and $\bar{z}_j$ are considered as vector fields on the domain (2.10). The mapping U_t induces a mapping U_{t*}, the differential,[11] on these vector fields. Because of linearity, the differential U_{t*} is equal to U_t itself.[11] Then we have by definition

$$U_{t*}z_j(\varphi) = U_t z_j U_{-t}(\varphi), \quad U_{t*}\bar{z}_j(\varphi) = U_t\bar{z}_j U_{-t}(\varphi) \tag{3.18}$$

for φ in the domain (2.10). We note here that the domain (2.10) is invariant under U_t. We now look into the right-hand side of (3.18). The following holds on the domain (2.10):

$$U_t z_j U_{-t} = \sum(e^{-tC})_{jk}z_k, \tag{3.19a}$$

$$U_t\bar{z}_j U_{-t} = \sum\overline{(e^{-tC})}_{jk}\bar{z}_k. \tag{3.19b}$$

For both sides of Eqs. (3.19) satisfy as functions of t the same differential equation with the same initial value. Thus we see that (3.7) are transformations induced on vector fields z_j and $\bar{z}_j$ by U_t. From (3.18), U_{t*} preserves the commutation relations (2.9).

So far we have obtained U_t and U_{t*} whose properties are summed up in

Proposition 3.4: The operators U_t and U_{t*} leave invariant the symplectic structure (2.3) and the commutation relations (2.9), respectively.

We are now to discuss the property of U_t. In the Hamiltonian formalism, a symplectic transformation is called a symmetry transformation (or a symmetry for brevity) if it leaves the Hamiltonian function H invariant. We now show that U_t is indeed a symmetry. Let $\exp(-it\hat{H})$ be the unitary operator generated from the Hamiltonian operator $\hat{H}$, which may be formed in the same method as applied for $U_t = \exp(it\hat{F}/\lambda)$. It is clear that $\exp(-it\hat{H})$ and U_s commute when they are operated with on the dense subspace (2.10). Because of unitarity they commute also on $L^2(\mathbb{R}^n)$. On this account one has

$$\langle U_s\varphi, \exp(-it\hat{H})U_s\varphi\rangle = \langle\varphi, \exp(-it\hat{H})\varphi\rangle. \tag{3.20}$$

Differentiating (3.20) with respect to t, we get $H\circ U_s = H$ for φ in the domain of $\hat{H}$ [see (2.5) and Remark 3.1]. Thus we have

Theorem 3.5: The operator U_t is a symmetry of the harmonic oscillator, that is, it leaves invariant both the symplectic structure (2.3) and the Hamiltonian function (2.5).

The infinitesimal version of this theorem is well known. That is, one has $U_{t*}\hat{H} = U_t\hat{H}U_{-t} = \hat{H}$ and $[\hat{F},\hat{H}] = 0$, as long as the commutator makes sense.

We can easily extend Theorem 3.5 to a theorem for the group $SU(n)$ which acts on the basis $\bar{z}_1^{k_1}\cdots\bar{z}_n^{k_n}\varphi_0$ of $L^2(\mathbb{R}^n)$, as in (3.8), in the form

$$\left(\sum V_{j_1 1}\bar{z}_{j_1}\right)^{k_1}\cdots\left(\sum V_{j_n n}\bar{z}_{j_n}\right)^{k_n}\varphi_0, \tag{3.21}$$

where $(V_{jm})\in SU(n)$.

Theorem 3.6: The group SU(n) is represented unitarily in $L^2(\mathbb{R}^n)$ as a symmetry group for the harmonic oscillator.

This theorem could be interpreted as a quantization of the symmetry group SU(n) for the classical harmonic oscillator.

IV. THE FOUR-DIMENSIONAL HARMONIC OSCILLATOR

We have discussed in Ref. 4 a reduction of the energy surface of the four-dimensional classical harmonic oscillator together with the accompanying reduction of the symmetry group. We consider in this section a quantum analog to these reductions. An eigenspace of the Hamiltonian operator $\hat{H}$ is taken to be associated with an energy surface of the classical system, so that our problem amounts to reducing the eigenspace. The representation (3.21) fits our problem well.

Let N_3 be a 4×4 matrix[12] given by

$$N_3 = \begin{pmatrix} K & \\ & K \end{pmatrix} \quad \text{with } K = \begin{pmatrix} & -\frac{1}{2} \\ \frac{1}{2} & \end{pmatrix}. \tag{4.1}$$

By SO(2) we mean the group $\exp(tN_3)$ acting on $\mathbb{R}^4$. We introduce the complex variables

$$\xi = x_1 + ix_2, \quad \eta = x_3 + ix_4 \tag{4.2}$$

to get a concise expression of SO(2) action on $\mathbb{R}^4 = \mathbb{C}^2$ as

$$(\xi,\eta) \to (e^{it/2}\xi,\ e^{it/2}\eta), \quad t\in\mathbb{R}. \tag{4.3}$$

The action (4.3) yields the orbit space

$$\mathbb{R}^4/\mathrm{SO}(2) = \mathbb{R}^3, \tag{4.4}$$

which has a close relation to the Hopf fibering $S^3/S^1 = S^2$.[13] Introduction of the local coordinates (r,θ,φ,ψ) by

$$\xi = r^{1/2}e^{i(\psi+\varphi)/2}\cos(\theta/2),\ \eta = r^{1/2}e^{i(\psi-\varphi)/2}\sin(\theta/2) \tag{4.5}$$

allows us to have a fair idea of the orbit space; the coordinates (r,θ,φ) becomes the polar spherical coordinates in $\mathbb{R}^3$.

Since N_3 is real and antisymmetric, we can apply (3.14) and (3.16) to N_3. By the help of (4.3), we find that the operator (3.14) with A replaced by N_3, denoted by $\hat{N}_3$, takes the form

$$\begin{aligned}\hat{N}_3 &= \frac{\lambda}{2i}\left(x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2} + x_4\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_4}\right) \\ &= -\frac{\lambda}{i}\frac{\partial}{\partial\psi}.\end{aligned} \tag{4.6}$$

The reduction of the eigenspace of $\hat{H}$ should be carried out by means of the group SO(2) so as to be compatible with (4.4).

Lemma 4.1: A smooth function f on $\mathbb{R}^4$ is SO(2)-invariant if and only if

$$\hat{N}_3 f = 0, \quad f(x) = f(-x). \tag{4.7}$$

Proof: If f is SO(2)-invariant, one has $f(\exp(tN_3)x) = f(x)$ for any $x\in\mathbb{R}^4$. Equation (3.16) with (4.6) then gives $\hat{N}_3 f = 0$. The remaining equation of (4.7) is evident from (4.3) with $t = 2\pi$. Conversely, if f satisfies (4.7), it turns out to be independent of ψ and invariant under the inversion $x \to -x$. The inversion invariance implies that f is periodic in φ with the period 2π, as is seen from (4.5). Therefore f can be looked on as a function on $\mathbb{R}^3$. This is a practical idea of the reduced function f^{red} on $\mathbb{R}^3$ determined for SO(2)-invariant functions by

$$f^{\mathrm{red}}([x]) = f(x) \quad \text{for } x\in\mathbb{R}^4, \tag{4.8}$$

where $[x]$ denotes the equivalence class of x. Clearly any function on $\mathbb{R}^3$ is pulled back through the natural projection associated with (4.4) to an SO(2)-invariant function on $\mathbb{R}^4$. This completes the proof.

We are now to choose SO(2)-invariant subspace of the eigenspace for $\hat{H}$. All we need is to find eigenfunctions which satisfy (4.7). First we consider $\hat{N}_3 f = 0$. An easy access to this equation will be given by finding a basis of the eigenspace which diagonalizes the operator $\hat{N}_3$. The following linear transformation of operators z_j and $\bar{z}_j$, which is a quantum analog to that employed in Ref. 4 for the classical system, will be successful;

$$\begin{aligned} w_1 &= z_1 + iz_2, \quad w_2 = z_3 + iz_4, \\ w_3 &= z_1 - iz_2, \quad w_4 = z_3 - iz_4, \end{aligned} \tag{4.9}$$

together with the adjoint operators $\bar{w}_j$. In terms of w_j and $\bar{w}_j$ the operator $\hat{N}_3$ takes the form

$$\hat{N}_3 = \tfrac{1}{8}(w_1\bar{w}_1 + w_2\bar{w}_2 - w_3\bar{w}_3 - w_4\bar{w}_4). \tag{4.10}$$

Since the transformation (4.9) is nondegenerate, the totality of

$$\bar{w}_1^{k_1}\cdots\bar{w}_4^{k_4}\varphi_0 \quad \text{with } N = k_1 + \cdots + k_4, \tag{4.11}$$

where the k_j are nonnegative integers and N is fixed, form a basis of the eigenspace assigned by the nonnegative integer N. We operate with $\hat{N}_3$ on the basis (4.11) to obtain

$$\hat{N}_3\bar{w}_1^{k_1}\cdots\bar{w}_4^{k_4}\varphi_0 = (\lambda/2)(k_1 + k_2 - k_3 - k_4)\bar{w}_1^{k_1}\cdots\bar{w}_4^{k_4}\varphi_0, \tag{4.12}$$

which shows that $\hat{N}_3$ is diagonalized, as was expected.

The condition $f(x) = f(-x)$ is rather easy to treat. The inversion $x \to -x$ gives rise to the inversion of the operators $\bar{z}_j = x_j - \partial/\partial x_j$, and hence that of $\bar{w}_j$, i.e., $\bar{w}_j \to -\bar{w}_j$. The φ_0 is clearly inversion-invariant. Accordingly, we have under the inversion

$$\bar{w}_1^{k_1}\cdots\bar{w}_4^{k_4}\varphi_0 \to (-1)^N\bar{w}_1^{k_1}\cdots\bar{w}_4^{k_4}\varphi_0. \tag{4.13}$$

From (4.12) and (4.13) it follows that for the eigenfunctions (4.11) the conditions (4.7) read

$$\begin{aligned} &k_1 + k_2 - k_3 - k_4 = 0, \\ &k_1 + k_2 + k_3 + k_4 = N = \text{ an even number}. \end{aligned} \tag{4.14}$$

The SO(2)-invariant eigenfunctions belonging to (4.11) are assigned by the exponents $(k_1,\ldots,k_4)$ satisfying (4.14). Therefore, we find from the number of solutions to (4.14) the following:

Theorem 4.2: In the eigenspace of the four-dimensional harmonic oscillator, there exists a space of SO(2)-invariant eigenfunctions, which is of dimension $(N/2 + 1)^2$, where N is a nonnegative integer assigning the energy level.

We proceed to a reduction of the symmetry group SU(4) acting on the eigenspace of $\hat{H}$. Recall that $\hat{N}_3$ is the operator $\hat{F}$ having the coefficient matrix N_3. We see from (3.13) and (3.17) that a function f is SO(2)-invariant if and only if $\exp(it\hat{N}_3/\lambda)f = f$ for all t. When the domain of $\exp(it\hat{N}_3/\lambda)$ is

restricted to the eigenspace of $\hat{H}$, the operator $\exp(it\hat{N}_3/\lambda)$ becomes $T(\exp(tN_3))$ introduced in (3.9), so that an eigenfunction f is SO(2)-invariant if and only if

$$T(\exp(tN_3))f=f. \tag{4.15}$$

We have obtained in Ref. 4 the subgroup $G=\mathrm{SU}(2)\times\mathrm{SU}(2)$ of SU(4) which commute with U(1), where U(1) is the group $\exp(tN_3)$ acting on $\mathbb{C}^4$. Let g be an arbitrary element of G. Since G and U(1) commute, one can conclude from (4.15) that for an SO(2)-invariant eigenfunction f one has

$$T(\exp(tN_3))T(g)f=f. \tag{4.16}$$

This means that $T(g)f$ is also SO(2)-invariant. In other words, G acts on the space of SO(2)-invariant eigenfunctions. The action of G, describable by (3.21) with $(V_{jm})\in G$, has a simpler form when the basis (4.11) is adopted because elements g of G then get the simpler matrix form[4]

$$g=\begin{pmatrix} u_{jk} & \\ & v_{jk} \end{pmatrix}\in SU(2)\times SU(2), \tag{4.17}$$

where (u_{jk}) and (v_{jk}) are 2×2 matrices. The operators $\bar{w}_j$ transform according to (4.17). Thus we have

$$T(g)\bar{w}_1^{k_1}\cdots\bar{w}_4^{k_4}\varphi_0=\left(\sum_{j=1}^{2}u_{j1}\bar{w}_j\right)^{k_1}\cdots\left(\sum_{j=3}^{4}v_{j4}\bar{w}_j\right)^{k_4}\varphi_0. \tag{4.18}$$

In view of (4.18) we find that inversion $x\to -x$ induces the operator $T(-1)$, where 1 denote the 4×4 unit matrix. Clearly, -1 is an element of (4.17). From this, for an inversion-invariant eigenfunction f one has $T(-1)f=f$ and hence

$$T(-g)f=T(g)f \quad \text{for all } g\in G. \tag{4.19}$$

From (4.19) we conclude that the representation of G induces a representation of $G/\{1,-1\}=\mathrm{SO}(4)$.

Theorem 4.3: The group SO(4) acts on the space of SO(2)-invariant eigenfunctions for the four-dimensional harmonic oscillator.

ACKNOWLEDGMENT

The author wishes to thank Professor H. Wakita for a discussion.

[1] J. E. Marsden, Arch. Rat. Math. Anal. **28**, 362 (1968).
[2] J. Marsden, *Applications of Global Analysis in Mathematical Physics* (Publish or Perish, Boston, 1974).
[3] P. R. Chernoff and J. E. Marsden, *Properties of Infinite Dimensional Hamiltonian Systems*, Lecture Notes in Mathematics, Vol. 425 (Springer-Verlag, Berlin, 1974).
[4] T. Iwai, J. Math. Phys. **22**, 1628 (1981).
[5] T. Iwai, J. Math. Phys. **23**, 1093 (1982).
[6] The same notation as in the classical system will cause no confusion.
[7] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), pp. 139, 299.
[8] G. A. Baker, Jr., Phys. Rev. **103**, 1119 (1956); M. I. Petrashen and E. D. Trifonov, *Applications of Group Theory in Quantum Mechanics* (MIT Press, Cambridge, Mass., 1969), p. 195.
[9] M. H. Stone, *Linear Transformations in Hilbert Space and Their Applications to Analysis* (Amer. Math. Soc., New York, 1932), p. 262.
[10] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic, New York, 1972), p. 269.
[11] S. Lang, *Differential Manifolds* (Addison-Wesley, Reading, Mass., 1972), pp. 8, 106.
[12] We keep the same notation as in Ref. 4. Missing matrix entries are all zero.
[13] J. Dieudonné, *Éléments d'analyse, Tome III*, 2nd ed. (Gauthier-Villars, Paris, 1974), Chap. 16.

Quantization of the conformal Kepler problem and its application to the hydrogen atom [a)]

Toshihiro Iwai
Department of Applied Mathematics, Gifu University, Yanagido, Gifu 501-11, Japan

(Received 14 December 1981; accepted for publication 29 January 1982)

Quantization of the conformal Kepler problem is defined and studied in order that the quantized system, which will be referred to as a conformal hydrogen atom, may associate the harmonic oscillator with the hydrogen atom. The conformal hydrogen atom shares with the harmonic oscillator the eigenspaces of negative energies. The four-dimensional conformal hydrogen atom reduces to the three-dimensional ordinary hydrogen atom. The symmetry group SO(4) of the hydrogen atom is brought out from the symmetry subgroup of the harmonic oscillator. The conformal hydrogen atom gives an example of those quantum systems of which the configuration spaces are curved Riemannian spaces with nonconstant scalar curvatures and of which the Hamiltonian operators depend on the scalar curvatures.

PACS numbers: 03.65.Ca, 02.20. + b

I. INTRODUCTION

Quantization is a process of constructing from a given classical system a quantum system that corresponds to it. However, the word "corresponding" does not here possess a definite meaning. According to various interpretations of the correspondence, various kinds of quantization have been proposed. The Schrödinger[1] and Weyl[2] procedures are ones which give quantum operators corresponding to classical observables. Feynman's quantization by the path integral method[3] is another. Though these quantization procedures are originally applicable when the configuration space of the classical system is the Euclidean space with the Cartesian coordinates, they can be adapted so as to work in a Riemannian manifold. For example, Refs. 4, 5, and 6 are articles which contain generalizations of the Schrödinger, Weyl, and Feynman procedures, respectively.

Generally speaking, Dirac's idea,[7] leading in this subject, that quantization of a classical system is finding a representation of the Lie algebra of classical observables in a Hilbert space has been realized in a series of works by van Hove,[8] Segal,[9] Souriau,[10] and Kostant,[11] which yield the geometric quantization theory.[12]

This article presents another method of quantization which is a quantum analogue to a previous paper.[13] That is, it deals with quantization of the conformal Kepler problem and its application to the hydrogen atom. Here quantization is interpreted as assigning a Hilbert space together with the Hamiltonian operator. That the configuration space of the conformal Kepler problem is actually a curved Riemannian space raises a question as to whether or not the curvature gives an additional term to the Hamiltonian operator which is supposed to be formed of the Laplacian and the potential. This paper shows that a term of the scalar curvature is added through a conformal transformation. Moreover, a Hilbert space this paper adopts is not the standard one that has the inner product defined with the standard volume element on the Riemannian space.

This quantization will need explanation. According to Ref. 13, the conformal Kepler problem is closely related to the harmonic oscillator, and the four-dimensional conformal Kepler problem reduces to the ordinary three-dimensional Kepler problem. Quantization of the conformal Kepler problem should be performed in a manner consistent with the already established quantization of the harmonic oscillator and the Kepler problem. To meet this requirement, the quantized Hamiltonian operator for the conformal Kepler problem is defined through that of the harmonic oscillator on the following principle: The relation between the Hamiltonian operators of the quantum systems has an analogy to that between the Hamiltonian functions of the corresponding classical systems. The Hamiltonian operator thus defined exhibits an unexpected property. The operator has a term depending on the scalar curvature of the configuration space. Moreover, it is not a symmetric operator with respect to the inner product defined with the standard volume element. Of course, there is an adequate inner product which makes the operator symmetric. In the four-dimensional case, the Hamiltonian operator thus obtained reduces to the ordinary Hamiltonian operator of the three-dimensional hydrogen atom.

Section II contains quantization of the conformal Kepler problem. The quantized system will be referred to as the conformal hydrogen atom. Unusual properties of the Hamiltonian operator are examined. Negative energy levels are derived by means of the relation to the harmonic oscillator. Section III is concerned with reduction of the four-dimensional conformal hydrogen atom to the three-dimensional ordinary hydrogen atom. The negative energy levels of the hydrogen atom are brought out from that of the conformal hydrogen atom together with their multiplicities of degeneracy. Section IV deals with the symmetry group of the conformal hydrogen atom with application to that of the hydrogen atom. Full use is made of the results obtained in Ref. 14. In Sec. V the operators which represent the (infinitesimal) symmetry are discussed. A procedure is given for obtaining the momentum and the Runge–Lenz vector for the hydrogen atom from the symmetry generators for the four-dimensional conformal hydrogen atom.

[a)]This work was partially supported by Grant-Aid for Scientific Research 56740018, Ministry of Education.

 0022-2488/82/061093-07$02.50

II. QUANTIZATION

We start with a brief review of the conformal Kepler problem discussed in Ref. 13. Let (x_j, p_j) be the Cartesian coordinates in $\mathbb{R}^n \times \mathbb{R}^n$, and $R^2 = \Sigma x_j^2$. The conformal Kepler problem defined in $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$ has the Hamiltonian function and the canonical 1-form

$$H_c = 4R^2(\tfrac{1}{2}\sum p_j^2) - k/R^2 \quad (k \text{ a positive const}), \tag{2.1}$$

$$\theta_c = 4R^2 \sum p_j dx_j, \tag{2.2}$$

respectively. The configuration space is endowed with the conformally flat metric

$$ds_c^2 = 4R^2 \sum dx_j^2. \tag{2.3}$$

Let $\mathbb{R}^n \times \mathbb{R}^n$ be another space with the Cartesian coordinates (x_j, p_j') on which the harmonic oscillator is defined as follows:

$$H' = \tfrac{1}{2}\sum p_j'^2 + (\lambda^2/2)\sum x_j^2 \quad (\lambda \text{ a positive const}), \tag{2.4}$$

$$\theta' = \sum p_j' dx_j. \tag{2.5}$$

Of course, the configuration space $\mathbb{R}^n$ has the standard flat metric Σdx_j^2.

The mapping

$$(x_j, p_j) \to (x_j, p_j') = (x_j, 4R^2 p_j) \tag{2.6}$$

defines a symplectic diffeomorphism of $(\mathbb{R}^n - \{0\}) \times \mathbb{R}^n$ into $\mathbb{R}^n \times \mathbb{R}^n$, i.e., $d\theta_c = d\theta'$. The Hamiltonian functions (2.1) and (2.4) are connected to each other as follows:

$$4R^2(H_c + \lambda^2/8) = H' - 4k. \tag{2.7}$$

We proceed to construct a quantum system corresponding to the conformal Kepler problem reviewed above. The quantum harmonic oscillator has the Hamiltonian operator

$$\hat{H}' = \tfrac{1}{2}\sum p_j'^2 + \frac{\lambda^2}{2} R^2 \text{ with } p_j' = -i\frac{\partial}{\partial x_j}. \tag{2.8}$$

On the analogy of (2.7) we define an operator $\hat{H}_c$ by

$$4R^2(\hat{H}_c + \lambda^2/8) = \hat{H}' - 4k. \tag{2.9}$$

Then $\hat{H}_c$ has the form

$$\hat{H}_c = -\frac{1}{2}\frac{1}{4R^2}\sum \frac{\partial^2}{\partial x_j^2} - \frac{k}{R^2}. \tag{2.10}$$

In what follows we are going to examine $\hat{H}_c$. A question arises as to whether the operator $(1/4R^2)\Sigma\partial^2/\partial x_j^2$ can be a symmetric operator or not, and as to how it is related to the Laplacian. To work out these questions, we begin with conformal transformations of a Riemannian manifold. Assume we have conformally related Riemannian metrics such that their components are related by the equations

$$g^*_{jk} = e^{2\sigma} g_{jk}. \tag{2.11}$$

The standard volume elements dV^* and dV associated with (g^*_{jk}) and (g_{jk}), respectively, are then related by

$$dV^* = e^{n\sigma} dV. \tag{2.12}$$

When applied to (2.3), Eq. (2.12) gives the standard volume element, denoted by dV_c, to the configuration space of the conformal Kepler problem: $dV_c = (2R)^n dx$, where dx denotes $dx_1 \wedge \cdots \wedge dx_n$ in Cartesian coordinates. In view of the expression of dV_c, we are looking for a volume element in the form $(2R)^m dx$, m being a positive constant, which defines an inner product for wavefunctions so that $(1/4R^2)\Sigma\partial^2/\partial x_j^2$ may be symmetric. Unless stated otherwise, we understand that the operators to be dealt with are defined on the space of smooth functions of compact support on $\mathbb{R}^n$. We denote the space by $C_0^\infty(\mathbb{R}^n)$ as usual.

Let f and h be in $C_0^\infty(\mathbb{R}^n)$. Then one has

$$\begin{aligned}\int_{\mathbb{R}^n} \bar{f}\frac{1}{4R^2}\sum \frac{\partial^2 h}{\partial x_j^2}(2R)^m dx &= \int_{\mathbb{R}^n} \frac{1}{4R^2}\sum \frac{\partial^2 \bar{f}}{\partial x_j^2} h (2R)^m dx \\ &+ 4(m-2)\int_{\mathbb{R}^n}\left(\sum x_j \frac{\partial \bar{f}}{\partial x_j} + (n+m-4)\bar{f}\right) \\ &\times h (2R)^{m-4} dx,\end{aligned} \tag{2.13}$$

where the overbar indicates the complex conjugate. In view of this we employ the inner product

$$\langle f, h \rangle = \int_{\mathbb{R}^n} \bar{f} h (2R)^2 dx, \tag{2.14}$$

and denote by $L^2(\mathbb{R}^n; (2R)^2 dx)$ the Hilbert space of square integrable functions on $\mathbb{R}^n$ with the inner product (2.14). In this Hilbert space, the operator $(1/4R^2)\Sigma\partial^2/\partial x_j^2$ is symmetric. Thus we have

Lemma 2.1: With respect to the inner product (2.14) the Hamiltonian operator $\hat{H}_c$ given by (2.10) is a symmetric operator.

Remark: To make description accurate, we have to verify that the subspace $C_0^\infty(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n; (2R)^2 dx)$. However, the verification is easily made for $n \geqslant 3$ by use of the well-known fact that the subspace $C_0^\infty(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$ with the ordinary volume element.

We turn to the Laplacian. As far as the conformal transformation (2.11) is concerned, the (x_j) are understood to be any local coordinates of the Riemannian manifold. We investigate the relation between the Laplacians Δ^* and Δ which are associated with (g^*_{jk}) and (g_{jk}), respectively. Let g^* denote $\det(g^*_{jk})$, and (g^{*jk}) and (g^{jk}) denote $(g^*_{jk})^{-1}$ and $(g_{jk})^{-1}$, respectively. By definition we obtain, after a calculation,

$$\begin{aligned}\Delta^* f &= \frac{1}{(g^*)^{1/2}}\sum \frac{\partial}{\partial x_j}\left[(g^*)^{1/2} g^{*jk}\frac{\partial f}{\partial x_k}\right] \\ &= \exp\left(-\frac{n+2}{2}\sigma\right)\left\{\Delta\left[f\exp\left(\frac{n-2}{2}\sigma\right)\right]\right. \\ &\left. - f\Delta\exp\left(\frac{n-2}{2}\sigma\right)\right\}\end{aligned} \tag{2.15a}$$

with

$$\begin{aligned}\Delta\exp\left(\frac{n-2}{2}\sigma\right) &= \frac{n-2}{2}\exp\left(\frac{n-2}{2}\sigma\right) \\ &\times\left(\Delta\sigma + \frac{n-2}{2}\sum g^{jk}\sigma_j\sigma_k\right),\end{aligned} \tag{2.15b}$$

where $\sigma_j = \partial\sigma/\partial x_j$. We here employ the formula about scalar curvatures[15]

$$e^{2\sigma}K^* = K - 2(n-1)\left(\Delta\sigma + \frac{n-2}{2}\sum g^{jk}\sigma_j\sigma_k\right), \tag{2.16}$$

where K^* and K are the scalar curvatures formed from (g^*_{jk}) and (g_{jk}), respectively. From (2.15) and (2.16) we conclude:

Lemma 2.2: The conformal transformation (2.11) gives rise to the conformal relation between the Laplacians Δ^* and Δ,

$$\left(\Delta^* - \frac{n-2}{4(n-1)}K^*\right)f = \exp\left(-\frac{n+2}{2}\sigma\right) \times\left(\Delta - \frac{n-2}{4(n-1)}K\right) \times\exp\left(\frac{n-2}{2}\sigma\right)f. \tag{2.17}$$

We apply Eq. (2.17) to our conformally flat metric (2.3) by setting $g^*_{jk} = (2R)^2\delta_{jk}$ and $g_{jk} = \delta_{jk}$. Thus we obtain

$$(2R)^{(n-2)/2}\left(\Delta_c - \frac{n-2}{4(n-1)}K_c\right)(2R)^{-(n-2)/2} = \frac{1}{4R^2}\sum\frac{\partial^2}{\partial x_j^2}, \tag{2.18}$$

where we have written Δ_c and K_c for Δ^* and K^*, respectively, in order to indicate that the metric ds_c^2 is concerned.

From (2.10) and (2.18) it follows that

$$\hat{H}_c = -\frac{1}{2}(2R)^{(n-2)/2}\left(\Delta_c - \frac{n-2}{4(n-1)}K_c\right)(2R)^{-(n-2)/2} - \frac{k}{R^2}. \tag{2.19}$$

Equation (2.19) shows that the operator $\hat{H}_c$ contains the nonvanishing scalar curvature

$$K_c = -12(n-1)(n-2)(2R)^{-4}, \tag{2.20}$$

which can be calculated from (2.16) with $\exp\sigma = 2R$. Our operator (2.19) should be compared in the coefficient of K_c with the correspondents appearing in Refs. 5 and 6; ours has the dimension-dependent coefficient. For $n = 4$ it is $-\frac{1}{6}$. The same operator was treated in Ref. 16 for $n = 4$.

We now come to the following definition.

Definition 2.3: A quantum system corresponding to the classical conformal Kepler problem is defined on the Hilbert space $L^2(\mathbb{R}^n;(2R)^2dx)$ with the Hamiltonian operator $\hat{H}_c$ given by (2.10) or (2.19). Below we refer to this quantum system as the conformal hydrogen atom.

Our definition would require some remarks. We will discuss wavefunctions and momentum operators. First we calculate the mean value of $\hat{H}_c$. For a function f of compact support on $\mathbb{R}^n$ we have from (2.19).

$$\int_{\mathbb{R}^n}\bar{f}\hat{H}_c f(2R)^2dx = \int_{\mathbb{R}^n}\bar{\phi}\left[-\frac{1}{2}\left(\Delta_c - \frac{n-2}{4(n-1)}K_c\right) - \frac{k}{R^2}\right]\phi\, dV_c \tag{2.21a}$$

with

$$\phi = (2R)^{-(n-2)/2}f. \tag{2.21b}$$

This suggests the following: The right-hand side of (2.21a) has an invariant meaning if ϕ is a scalar on the Riemannian manifold $(\mathbb{R}^n,\ ds_c^2)$, so that f which is related to ϕ by (2.21b) could be thought of as a wave "density" rather than a wavefunction. In spite of this consideration we employ Definition 2.3 for convenience's sake.

We next discuss momentum operators. Let a_j^t be the one-parameter group given by

$$a_j^t:\ x_j \to x_j + t, \quad \text{and the others fixed.} \tag{2.22}$$

For the sake of generality we write the volume element as $(2R)^m dx$, which was used in (2.13). Then the a_j^t makes the change in the volume element,

$$(2R)^m dx \to [R(x)^{-1}R(a_j^t(x))]^m(2R)^m dx. \tag{2.23}$$

Here we have adopted the notation $R(x) = (\Sigma x_k^2)^{1/2}$, $x\in\mathbb{R}^n$, for convenience. Because of (2.23), the mapping

$$f(x) \to [R(x)^{-1}R(a_j^{-t}(x))]^{m/2}f(a_j^{-t}(x)) \tag{2.24}$$

is a one-parameter group of unitary operators on $L^2(\mathbb{R}^n;(2R)^m dx)$. When f is smooth, this group can be differentiated with respect to t to give an essentially self-adjoint operator[17]

$$\hat{P}_j = -i\left(\frac{\partial}{\partial x_j} + \frac{mx_j}{2R^2}\right). \tag{2.25}$$

The $\hat{P}_j$ with $m = 2$ is the momentum operator for our quantum system. As a consequence we see that $-i\partial/\partial x_j$ is by no means a symmetric operator.

Angular momentum operators are obtained by the same method. Let $\exp(tA)$ denote an orthogonal matrix with $A = (A_{jk})$ an antisymmetric matrix. Since the volume element $(2R)^m dx$ is left invariant under orthogonal transformations, the mapping

$$f(x) \to f(\exp(-tA)x) \tag{2.26}$$

defines a one-parameter group of unitary transformations on $L^2(\mathbb{R}^n;(2R)^m dx)$. Thus we obtain as $-i$ times the infinitesimal generator of (2.26) the angular momentum operator

$$i\sum A_{jk}x_k\frac{\partial}{\partial x_j}, \tag{2.27}$$

which is essentially self-adjoint.[17]

In the remainder of this section we consider eigenfunctions for the conformal hydrogen atom. If f is an eigenfunction of $\hat{H}_c$ of an negative energy, denoted by $-\lambda^2/8$, it becomes an eigenfunction of $\hat{H}'$, where $\hat{H}'$ is defined through (2.9) from $\hat{H}_c$:

$$4R^2(\hat{H}_c + \lambda^2/8)f = (\hat{H}' - 4k)f = 0. \tag{2.28}$$

As eigenfunctions of $\hat{H}'$ clearly belong to $L^2(\mathbb{R}^n;(2R)^2dx)$, so does the eigenfunction f for $\hat{H}_c$.

Since the eigenvalues of $\hat{H}'$ are $\lambda(N + n/2)$ with N nonnegative integers, λ should be determined by

$$4k = \lambda(N + n/2). \tag{2.29}$$

The eigenvalues of $\hat{H}_c$ are therefore computed to be

$$\frac{-\lambda^2}{8} = -\frac{k^2}{2(\frac{1}{2}N + \frac{1}{4}n)^2}. \tag{2.30}$$

Theorem 2.4: The negative eigenvalues for the conformal hydrogen atom are given by (2.30), and the corresponding eigenfunctions are those for the harmonic oscillator with the eigenvalue (2.29).

III. REDUCTION TO THE HYDROGEN ATOM

In this section we discuss how the conformal hydrogen atom of dimension four reduces to the ordinary hydrogen atom of dimension three. We work in the coordinates defined by

$$\begin{aligned} x_1 &= R\cos[(\psi+\varphi)/2]\cos(\theta/2),\\ x_2 &= R\sin[(\psi+\varphi)/2]\cos(\theta/2),\\ x_3 &= R\cos[(\psi-\varphi)/2]\sin(\theta/2),\\ x_4 &= R\sin[(\psi-\varphi)/2]\sin(\theta/2), \end{aligned} \tag{3.1}$$

and set

$$\sum x_j^2 = R^2 = r. \tag{3.2}$$

Here the range of the angular variables are given by

$$\begin{aligned} &0\leqslant(\psi+\varphi)/2\leqslant 2\pi,\quad -\pi\leqslant(\psi-\varphi)/2\leqslant\pi,\\ &0\leqslant\theta/2\leqslant\pi/2. \end{aligned} \tag{3.3}$$

The volume element $(2R)^2dx$ and the Hamiltonian operator then take the form

$$(2R)^2dx = \tfrac{1}{4}r^2\sin\theta\, dr d\theta d\varphi d\psi, \tag{3.4}$$

$$\hat{H}_c = -\frac{1}{2}\frac{1}{4r}\left(4r\frac{\partial^2}{\partial r^2}+8\frac{\partial}{\partial r}+\frac{1}{r}\Delta_3\right)-\frac{k}{r}, \tag{3.5}$$

respectively, where Δ_3 stands for the Laplacian associated with the unit sphere S^3,

$$\Delta_3 = 4\left[\frac{\partial^2}{\partial\theta^2}+\cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\varphi^2}+\frac{\partial^2}{\partial\psi^2}-2\cos\theta\frac{\partial^2}{\partial\varphi\partial\psi}\right)\right]. \tag{3.6}$$

Let

$$N_3 = \begin{pmatrix} K & \\ & K \end{pmatrix} \quad \text{with } K = \begin{pmatrix} & -\frac{1}{2} \\ \frac{1}{2} & \end{pmatrix}, \tag{3.7}$$

which was used in Refs. 14 and 18. By SO(2) we mean the one-parameter group $\exp(tN_3)$, which acts on $\mathbb{R}^4$. In the coordinates (3.1) the group action is expressed as $\psi\to\psi+t$ with the other coordinates fixed. The SO(2) defines the orbit space

$$\mathbb{R}^4/\mathrm{SO}(2) = \mathbb{R}^3, \tag{3.8}$$

through which the coordinates (r,θ,φ) turn out to be the polar spherical coordinates of $\mathbb{R}^3$. By use of (3.8) we can define for any SO(2)-invariant function f on $\mathbb{R}^4$ a reduced function f^{red} on $\mathbb{R}^3$:

$$f^{\mathrm{red}}([x]) = f(x) \quad \text{for} \quad x\in\mathbb{R}^4, \tag{3.9}$$

where $[x]$ denotes the equivalence class of x. Conversely, any function on $\mathbb{R}^3$ is pulled back to an SO(2)-invariant function on $\mathbb{R}^4$ through the natural projection associated with (3.8).

With N_3 is associated the operator

$$\begin{aligned} \hat{N}_3 &= \frac{\lambda}{2i}\left(x_2\frac{\partial}{\partial x_1}-x_1\frac{\partial}{\partial x_2}+x_4\frac{\partial}{\partial x_3}-x_3\frac{\partial}{\partial x_4}\right)\\ &= -\frac{\lambda}{i}\frac{\partial}{\partial\psi}, \end{aligned} \tag{3.10}$$

which is used for characterizing f^{red}. The following was shown in Ref. 14.

Lemma 3.1: A function f on $\mathbb{R}^4$ is SO(2)-invariant if and only if it satisfies $\hat{N}_3 f = 0$ and is invariant under the inversion $x\to -x$, $x\in\mathbb{R}^4$. In terms of the coordinates (3.1), f is SO(2)-invariant if and only if it is independent of ψ and periodic in φ with the period 2π.

Definition (3.9) provides us with a procedure for reducing the conformal hydrogen atom. We start by considering the inner product of SO(2)-invariant function f and h. By use of (3.4) and (3.8) we integrate $\bar{f}h$ over $\mathbb{R}^4$ to obtain

$$\begin{aligned} &\int_{\mathbb{R}^4}\bar{f}h\,(2R)^2dx\\ &\quad=\int_{\mathrm{SO}(2)}d\psi\int_{\mathbb{R}^4/\mathrm{SO}(2)}\bar{f}h\,\tfrac{1}{4}r^2\sin\theta\, dr d\theta d\varphi\\ &\quad=\pi\int_{\mathbb{R}^3}\bar{f}^{\mathrm{red}}h^{\mathrm{red}}r^2\sin\theta\, dr d\theta d\varphi, \end{aligned} \tag{3.11}$$

where we have employed the fact that SO(2) is identified with each orbit except the origin. From (3.11) we observe that for SO(2)-invariant functions the linear mapping $f\to\pi^{1/2}f^{\mathrm{red}}$ is an isometry from the space of all SO(2)-invariant functions in $L^2(\mathbb{R}^4;(2R)^2dx)$ to the Hilbert space $L^2(\mathbb{R}^3)$. Conversely, by inverting the above discussion, any function in $L^2(\mathbb{R}^3)$ is proved to be pulled back to an SO(2)-invariant function in $L^2(\mathbb{R}^4;(2R)^2dx)$.

We turn to reduction of $\hat{H}_c$. Let f be an SO(2)-invariant function, smooth to some extent. Since $\hat{N}_3$ and $\hat{H}'$ commute, as is proved easily, we obtain from (2.9) and Lemma 3.1

$$\hat{N}_3[4R^2(\hat{H}_c+\lambda^2/8)f] = (\hat{H}'-4k)\hat{N}_3 f = 0. \tag{3.12}$$

As $\hat{N}_3R^2 = 0$, Eq. (3.12) implies that $\hat{N}_3\hat{H}_c f = 0$. Furthermore, from the definition (2.10), $\hat{H}_c$ is invariant under the inversion $x\to -x$. Thus, from Lemma 3.1, $\hat{H}_c f$ proves to be SO(2)-invariant. The reduced operator $\hat{H}_c^{\mathrm{red}}$ is then determined uniquely by

$$\hat{H}_c^{\mathrm{red}}f^{\mathrm{red}} = (\hat{H}_c f)^{\mathrm{red}}. \tag{3.13}$$

We get from (3.5), (3.6), and (3.13) the explicit form of $\hat{H}_c^{\mathrm{red}}$,

$$\hat{H}_c^{\mathrm{red}} = -\frac{1}{2}\left(\frac{\partial^2}{\partial r^2}+\frac{2}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\Delta_2\right)-\frac{k}{r}, \tag{3.14}$$

where Δ_2 is the Laplacian associated with S^2,

$$\Delta_2 = \frac{\partial^2}{\partial\theta^2}+\cot\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}. \tag{3.15}$$

The $\hat{H}_c^{\mathrm{red}}$ is clearly equal to the Hamiltonian operator for the hydrogen atom of dimension three.

The above discussion is summed up to give

Theorem 3.2: The four-dimensional conformal hydrogen atom defined in the Hilbert space $L^2(\mathbb{R}^4;(2R)^2dx)$ reduces to the three-dimensional hydrogen atom defined in $L^2(\mathbb{R}^3)$.

We next discuss eigenfunctions of $\hat{H}_c$ and $\hat{H}_c^{\rm red}$. From (3.13) it follows that for an SO(2)-invariant eigenfunction f of $\hat{H}_c$, if it exists, $f^{\rm red}$ is an eigenfunction of $\hat{H}_c^{\rm red}$. Conversely, the pullback of any eigenfunction of $\hat{H}_c^{\rm red}$ is an eigenfunction of $\hat{H}_c$.

Thus we see that findings SO(2)-invariant eigenfunctions for $\hat{H}_c$ of negative energies suffices to get all the eigenfunctions for $\hat{H}_c^{\rm red}$ of negative energies. Theorem 2.4 says that the conformal hydrogen atom shares eigenfunctions with the harmonic oscillator, so that we have only to show the existence of SO(2)-invariant eigenfunctions of $\hat{H}'$. Incidentally, the existence of such eigenfunctions is already proved in Ref. 14. According to it, for a nonnegative even number N there are $(N/2+1)^2$ linearly independent eigenfunctions of $\hat{H}'$ invariant under SO(2). Thus we come to the following conclusion without solving the eigenvalue problem $\hat{H}_c^{\rm red}\phi = E\phi$.

Theorem 3.3: The negative eigenvalues for the Hamiltonian operator of the three-dimensional hydrogen atom are given by (2.30) with N nonnegative even number and $n=4$. The corresponding eigenspace, which is of dimension $(N/2+1)^2$, is obtained from the SO(2)-invariant subspace of the eigenspace for the Hamiltonian operator of the four-dimensional harmonic oscillator.

We remark here that Ikeda and Miyachi[19] obtained the same conclusion by solving the eigenvalue problem for the four-dimensional harmonic oscillator. In Ref. 20 also the energy levels of the hydrogen atom were derived from those of the harmonic oscillator.

IV. SYMMETRY GROUP

We consider in this section the symmetry group of the hydrogen atom. Theorem 3.3 indicates that the symmetry group for the hydrogen atom may be induced from that for the harmonic oscillator, so that we begin with a brief review of the symmetry for the harmonic oscillator.[14] Let z_j and $\bar{z}_j$ denote the operators given by

$$z_j = x_j + ip_j', \quad \bar{z}_j = x_j - ip_j' \quad \text{with } p_j' = -i\frac{\partial}{\partial x_j}, \tag{4.1}$$

respectively. An infinitesimal symmetry of the harmonic oscillator is expressed by the operator

$$\hat{F} = (\lambda/2i)\sum C_{jk} z_k \bar{z}_j, \tag{4.2}$$

where (C_{jk}) is an anti-Hermitian matrix with vanishing trace. In particular, when (C_{jk}) is N_3 given by (3.7), the operator (4.2) becomes $\hat{N}_3$ given by (3.10). Moreover, the unitary operator $U_t = \exp(it\hat{F}/\lambda)$ gives a global symmetry.

In the previous paper[18] about the classical harmonic oscillator we obtained 4×4 traceless anti-Hermitian matrices commutative with N_3, they were given by linear combinations of matrices N_3, M_α, and B_α, $\alpha = 1, 2, 3$. Here M_α and B_α are antisymmetric real and anti-Hermitian pure imaginary matrices, respectively, and form the Lie algebra of $SU(2)\times SU(2)$. According to Ref. 14, the mapping $(C_{jk}) \to (i/\lambda)\hat{F}$ is a Lie algebra homomorphism. Therefore, the substitution of M_α and B_α for the coefficient matrices (C_{jk}) in (4.2) yields operators commutative with $\hat{N}_3$, which we denote by $\hat{M}_\alpha$ and $\hat{B}_\alpha$, respectively. The commutation relations are given by

$$[(i/\lambda)\hat{M}_\alpha, (i/\lambda)\hat{M}_\beta] = (i/\lambda)\epsilon_{\alpha\beta\gamma}\hat{M}_\gamma, \tag{4.3a}$$

$$[(i/\lambda)\hat{M}_\alpha, (i/\lambda)\hat{B}_\beta] = (i/\lambda)\epsilon_{\alpha\beta\gamma}\hat{B}_\gamma, \tag{4.3b}$$

$$[(i/\lambda)\hat{B}_\alpha, (i/\lambda)\hat{B}_\beta] = (i/\lambda)\epsilon_{\alpha\beta\gamma}\hat{M}_\gamma. \tag{4.3c}$$

With the above review we proceed to symmetry of the hydrogen atom. Let f be an eigenfunction of $\hat{H}_c$ with the eigenvalue $-\lambda^2/8$, so that f is an eigenfunction of $\hat{H}'$ by Theorem 2.4, and let $\hat{F}$ be a symmetry operator (4.2). Then from (2.28), with $\hat{F}f$ substituted for f and the commutativity of $\hat{H}'$ and $\hat{F}$, we obtain

$$4R^2(\hat{H}_c + \lambda^2/8)\hat{F}f = 0. \tag{4.4}$$

This equation means that $\hat{F}f$ is also an eigenfunction of $\hat{H}_c$, so that $\hat{F}$ acts on the eigenspace of $\hat{H}_c$. It is, however, to be noted that $\hat{F}$ and $\hat{H}_c$ do not cummute in general [see (5.2)].

We now assume that f is SO(2)-invariant and that $\hat{F}$ and $\hat{N}_3$ commute. To apply Lemma 3.1, we remark here that the inversion $x \to -x$ induces the mappings $z_j \to -z_j$ and $\bar{z}_j \to -\bar{z}_j$. Since $\hat{F}$ and $\hat{N}_3$ commute and since $\hat{F}$ is quadratic in z_j and $\bar{z}_j$, we see that $\hat{N}_3\hat{F}f = 0$ and that $\hat{F}f$ is invariant under the inversion. Then Lemma 3.1 shows that $\hat{F}f$ is also SO(2)-invariant.

Summing up the above discussion, we have

Lemma 4.1: Assume that $\hat{F}$ and $\hat{N}_3$ commute. If f is an SO(2)-invariant eigenfunction of $\hat{H}_c$, so is $\hat{F}f$.

By Lemma 4.1 the reduced operator $\hat{F}^{\rm red}$ is defined after the same method as in (3.13). Furthermore, one has for an SO(2)-invariant eigenfunction f

$$\begin{aligned}\hat{H}_c^{\rm red}\hat{F}^{\rm red}f^{\rm red} &= (\hat{H}_c\hat{F}f)^{\rm red} = -(\lambda^2/8)(\hat{F}f)^{\rm red} \\ &= -(\lambda^2/8)\hat{F}^{\rm red}f^{\rm red},\end{aligned} \tag{4.5}$$

which shows that $\hat{F}^{\rm red}$ acts on the eigenspace of $\hat{H}_c^{\rm red}$.

We have already obtained the operator $\hat{M}_\alpha$ and $\hat{B}_\alpha$, $\alpha = 1, 2, 3$, commutative with $\hat{N}_3$, so that the reduced operator $\hat{M}_\alpha^{\rm red}$ and $\hat{B}_\alpha^{\rm red}$ are defined and act on the eigenspace of $\hat{H}_c^{\rm red}$. The commutation relations among them are the same as (4.3), since (4.3) remains as it is when applied to the SO(2)-invariant subspace of the eigenspace of $\hat{H}_c$. Hence we have

Theorem 4.2: The eigenspaces of $\hat{H}_c^{\rm red}$ of negative energies are representation spaces of the Lie algebra of SO(4) or of $SU(2)\times SU(2)$.

We conclude this section by discussing the symmetry group of the hydrogen atom. In view of Theorems 3.3 and 4.2 we have only to know which of $SU(2)\times SU(2)$ and SO(4) is a symmetry group acting on the space of SO(2)-invariant eigenfunctions for $\hat{H}'$. Incidentally, the previous paper[14] provides us with the group SO(4) acting on this space. Thus we get

Theorem 4.3: The symmetry group SO(4) acts on the eigenspaces of $\hat{H}_c^{\rm red}$ of negative energies.

Remark: The symmetry of the hydrogen atom, infinitesimal (Theorem 4.2) or finite (Theorem 4.3) is described not in $L^2(\mathbb{R}^3)$ but in each eigenspace for $\hat{H}_c^{\rm red}$, a finite-dimensional subspace of $L^2(\mathbb{R}^3)$. This would deserve notice in contrast with the symmetry of the harmonic oscillator (see Ref. 14).

V. REMARKS ON OPERATORS

In the preceding section we treated the operators $\hat{M}_\alpha$ and $\hat{B}_\alpha$, $\alpha = 1, 2, 3$, in connection with the conformal hydrogen atom. Strange to say, we will show that the B_α is not a symmetric operator in $L^2(\mathbb{R}^4;(2R)^2dx)$. Furthermore, $\hat{B}_\alpha$ and $\hat{H}_c$ do not commute, though the $\hat{B}_\alpha$ serve as infinitesimal symmetries acting on the eigenspaces of $\hat{H}_c$ (see Theorem 4.2). However, these are not contradictions, as will be seen in what follows.

For a while we deal with the n-dimensional conformal hydrogen atom. Let (A_{jk}) and (B_{jk}) be antisymmetric and symmetric real matrices, respectively. Let $\hat{L}$ and $\hat{D}$ denote the operators (4.2) with the coefficient matrices (C_{jk}) replaced by (A_{jk}) and (iB_{jk}), respectively, i being the imaginary unit. $\hat{L}$ and $\hat{D}$ are first- and second-order differential operators, respectively, and cover $\hat{M}_\alpha$ and $\hat{B}_\alpha$, respectively. The $\hat{L}$ causes no problem although the volume element $(2R)^2dx$ is adopted. In fact, as was already seen in Sec. II [see (2.26) and (2.27)], $\hat{L}$ is related to the orthogonal transformation $\exp(-tA)$, and is essentially self-adjoint. Moreover, since any orthogonal transformation leaves the Hamiltonian operator $\hat{H}_c$ invariant, $\hat{L}$ and $\hat{H}_c$ commute.

We proceed to the operator

$$\hat{D} = \tfrac{1}{2}\sum B_{jk}\left(\lambda^2 x_j x_k - \frac{\partial^2}{\partial x_j \partial x_k}\right). \tag{5.1}$$

A careful calculation yields the commutation relation

$$[\hat{D},\hat{H}_c] = R^{-2}\sum B_{jk}\left(\delta_{jk} + x_k\frac{\partial}{\partial x_j} + x_j\frac{\partial}{\partial x_k}\right) \times\left(\frac{\lambda^2}{8} + H_c\right). \tag{5.2}$$

This equation means that $\hat{D}$ and $\hat{H}_c$ do not commute, but do when applied to the eigenspaces of $\hat{H}_c$. Further, because of the volume element $(2R)^2dx$, operators $\partial^2/\partial x_j\partial x_k$ and $\hat{D}$ are not symmetric on a domain, for example, the space of smooth functions of compact support on $\mathbb{R}^n$. Instead of $\partial^2/\partial x_j\partial x_k$ one can obtain a symmetric operator

$$-\hat{P}_j\hat{P}_k + R^{-4}x_jx_k = \frac{\partial^2}{\partial x_j\partial x_k} + R^{-2}\left(\delta_{jk} + x_k\frac{\partial}{\partial x_j} + x_j\frac{\partial}{\partial x_k}\right), \tag{5.3}$$

where $\hat{P}_j$ are the momentum operators (2.25) with $m = 2$. However, this is not applicable to our case.

A good way to obtain appropriate operators is suggested by the fact that only when it is applied to the eigenspaces of $\hat{H}_c$ is the operator $\hat{D}$ thought of as an infinitesimal symmetry for the conformal hydrogen atom. We then replace λ^2 in (5.1) by $-8\hat{H}_c$ to get

$$\hat{D}_c = \tfrac{1}{2}\sum B_{jk}\left(-8x_jx_k\hat{H}_c - \frac{\partial^2}{\partial x_j\partial x_k}\right)$$

$$= \tfrac{1}{2}\sum B_{jk}\left(R^{-2}x_jx_k\sum\frac{\partial^2}{\partial x_m^2} - \frac{\partial^2}{\partial x_j\partial x_k} + 8kR^{-2}x_jx_k\right). \tag{5.4}$$

When restricted to the eigenspaces of $\hat{H}_c$, both $\hat{D}$ and $\hat{D}_c$ have the same action. The $\hat{D}_c$ is a good one for the conformal hydrogen atom. In fact, for functions f and h of compact support, one has

$$\int_{\mathbb{R}^n}\left(R^{-2}x_jx_k\sum\frac{\partial^2\bar{f}}{\partial x_m^2} - \frac{\partial^2\bar{f}}{\partial x_j\partial x_k}\right)h\,(2R)^2dx = \int_{\mathbb{R}^n}\bar{f}\left(-R^{-2}\delta_{jk}h + R^{-2}x_jx_k\sum\frac{\partial^2 h}{\partial x_m^2} - \frac{\partial^2 h}{\partial x_j\partial x_k}\right) \times(2R)^2dx, \tag{5.5}$$

so that with $\mathrm{tr}(B_{jk}) = 0$ the operator $\hat{D}_c$ is symmetric. Further, it is easily shown by a similar computation as in (5.2) that $\hat{D}_c$ and $\hat{H}_c$ commute. Thus we have

Proposition 5.1: The operator $\hat{D}_c$ given by (5.4) with $\mathrm{tr}(B_{jk}) = 0$ is a symmetric operator on $C_0^\infty(\mathbb{R}^n)$ and is commutative with $\hat{H}_c$ on $C_0^\infty(\mathbb{R}^n)$.

For convenience of later discussion, we here give the commutation relation between $\hat{D}_c$'s. Let $\hat{D}_c'$ denote the operator (5.4) with the coefficient matrix (B'_{jk}). Then a straightforward calculation results in

$$[\hat{D}_c,\ \hat{D}_c'] = 8\sum[B,\ B']_{jk}x_k\frac{\partial}{\partial x_j}\hat{H}_c, \tag{5.6}$$

where the $[B, B']_{jk}$ denote the components of the commutator of the matrices $B = (B_{ik})$ and $B' = (B'_{ik})$.

We return to the case of $n = 4$. According to (5.4), the operators $\hat{B}_\alpha$ can be modified to the symmetric ones which we denote by $(\hat{B}_\alpha)_c$, $\alpha = 1, 2, 3$. Let $\hat{D}_c$ be one of $(\hat{B}_\alpha)_c$. Our next task is to make a reduced operator from $\hat{D}_c$. To this end, we have to show that $\hat{N}_3$ and $\hat{D}_c$ commute. A straightforward calculation shows that the commutativity of $\hat{N}_3$ and $\hat{D}_c$ is deduced from that of the matrices N_3 and B_α, and from that of the operators $\hat{N}_3$ and $\hat{H}_c$. (The latter commutativity is covered by the already known fact that $[\hat{L}, \hat{H}_c] = 0$.) In fact, from $[N_3, B_\alpha] = 0$ we get $\hat{N}_3(\Sigma B_{jk}x_jx_k) = 0$ and $[\hat{N}_3, \Sigma B_{jk}\,\partial^2/\partial x_j\partial x_k] = 0$, where (B_{jk}) are the components of B_α, and therefore obtain $[\hat{N}_3, \hat{D}_c] = 0$ from $[\hat{N}_3, \hat{H}_c] = 0$. In addition, it is clear that $\hat{D}_c$ is invariant under the inversion. Thus we are allowed to define the reduced operator $\hat{D}_c^{\mathrm{red}}$. Following a similar method as in (3.11), we can prove that $\hat{D}_c^{\mathrm{red}}$ is a symmetric operator. In effect, for functions f and h with compact support on $\mathbb{R}^3$ and their pullbacks f^* and h^* on $\mathbb{R}^4$, respectively, we have

$$\pi\int_{\mathbb{R}^3}\hat{D}_c^{\mathrm{red}}\bar{f}h\,d\xi = \int_{\mathbb{R}^4}\hat{D}_c\bar{f}^*h^*(2R)^2\,dx = \int_{\mathbb{R}^4}\bar{f}^*\hat{D}_ch^*(2R)^2\,dx = \pi\int_{\mathbb{R}^3}\bar{f}\hat{D}_c^{\mathrm{red}}h\,d\xi, \tag{5.7}$$

where $d\xi$ denotes the standard volume element on $\mathbb{R}^3$.

We are now in a position to describe the angular momentum $\mathbf{L} = (\hat{L}_\alpha)$ and the Runge–Lenz vector $\mathbf{A} = (\hat{A}_\alpha)$ for the hydrogen atom. All we have to do is to find out a concrete form of $(\hat{B}_\alpha)_c$ in the coordinates (3.1). The calculation is too long to write down here. We give a part of the results:

$$\hat{M}_1 = \frac{\lambda}{2i}\left(x_2\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_2} + x_4\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_4}\right) = \frac{\lambda}{i}\left(\sin\varphi\frac{\partial}{\partial\theta} + \cot\theta\cos\varphi\frac{\partial}{\partial\varphi} - \csc\theta\cos\varphi\frac{\partial}{\partial\psi}\right), \tag{5.8}$$

$$\begin{aligned}(\hat{B}_1)_c &= \frac{1}{2}\left[\frac{x_1x_3 + x_2x_4}{R^2}\sum\frac{\partial^2}{\partial x_m^2} - \left(\frac{\partial^2}{\partial x_1\partial x_3} + \frac{\partial^2}{\partial x_2\partial x_4}\right) + \frac{8k(x_1x_3 + x_2x_4)}{R^2}\right] \\ &= \frac{1}{2}\left(\frac{4}{r}\sin\theta\cos\varphi\frac{\partial^2}{\partial\theta^2} + \frac{4\cos\varphi}{r\sin\theta}\frac{\partial^2}{\partial\varphi^2} - 4\cos\theta\cos\varphi\frac{\partial^2}{\partial r\partial\theta}\right. \\ &\quad + 4\frac{\sin\varphi}{\sin\theta}\frac{\partial^2}{\partial r\partial\varphi} + 4\sin\theta\cos\varphi\frac{\partial}{\partial r} + \frac{4}{r}\cos\theta\cos\varphi\frac{\partial}{\partial\theta} + 4k\sin\theta\cos\varphi) \\ &\quad + \frac{1}{2}\left(-4\sin\varphi\cot\theta\frac{\partial}{\partial r} + \frac{4}{r}\sin\varphi\frac{\partial}{\partial\theta} - \frac{4}{r}\cot\theta\cos\varphi\frac{\partial}{\partial\varphi}\right)\frac{\partial}{\partial\psi}, \text{ etc.}\end{aligned} \tag{5.9}$$

$\hat{M}_\alpha^{\text{red}}$ and $(\hat{B}_\alpha)_c^{\text{red}}$ are obtained by getting rid of the terms including $\partial/\partial\psi$. We are eventually led to

$$\hat{M}_1^{\text{red}} = -\lambda\hat{L}_1, \tag{5.10}$$

$$(\hat{B}_1)_c^{\text{red}} = (\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L})_1 + 2k(\mathbf{r}/r)_1 = 2\hat{A}_1. \tag{5.11}$$

The final results are written in the form

$$(\hat{B}_\alpha)_c^{\text{red}} = 2\hat{A}_\alpha, \quad \hat{M}_\alpha^{\text{red}} = -\lambda\hat{L}_\alpha. \tag{5.12}$$

The well-known commutation relations among $\hat{A}_\alpha$ result from (5.6). In fact, by (5.6) and (5.12) one has

$$\begin{aligned}[\hat{A}_\alpha,\hat{A}_\beta] &= \frac{1}{4}[(\hat{B}_\alpha)_c, (\hat{B}_\beta)_c]^{\text{red}} \\ &= 2\sum\left([-iB_\alpha, -iB_\beta]_{jk}x_k\frac{\partial}{\partial x_j}\right)^{\text{red}}\hat{H}_c^{\text{red}} \\ &= i\epsilon_{\alpha\beta\gamma}\hat{L}_\gamma(-2\hat{H}_c^{\text{red}}),\end{aligned} \tag{5.13}$$

where we have employed the matrix commutation relations $[B_\alpha,B_\beta] = \epsilon_{\alpha\beta\gamma}M_\gamma$ (recall that $-i$ times B_α is a real symmetric matrix). When all the operators are applied to the eigenspaces of $\hat{H}_c^{\text{red}}$, the commutation relations (5.13) turn out to be equivalent to (4.3c).

We conclude this section with a remark on the so-called correspondence principle. It is a general thought that the momentum operator should be a symmetric first-order differential operator. However, it is not always true that multiplications of symmetric first-order differential operators are the only method to make higher order differential operators. We now reexamine $\hat{H}_c$ from the viewpoint of quantization of classical observables. We have endowed $\mathbf{R}^n$ with the metric (2.3) with respect to which the covariant momentums are $p_j' = 4R^2p_j$. In terms of p_j' we get an alternative form of the classical Hamiltonian (2.1) as

$$H_c = \tfrac{1}{2}\sum g^{jk}p_j'p_k' - k/R^2 = \tfrac{1}{2}(1/4R^2)\sum p_j'^2 - k/R^2, \tag{5.14}$$

where (g^{jk}) is the inverse of (g_{jk}), the metric tensor given by (2.3). Our operator $\hat{H}_c$ is obtained by substitution of $p_j' = -i\partial/\partial x_j$. However, as was pointed out in Sec. II, $-i\partial/\partial x_j$ is no longer symmetric. If we employ the symmetric momentum operator (2.25) with $m = 2$, we obtain from (5.14) the symmetric operator

$$\tfrac{1}{2}\sum\hat{P}_jg^{jk}\hat{P}_k - k/R^2 = \hat{H}_c - (n-3)/8R^4, \tag{5.15}$$

which is not applicable to our quantum system.

[1]E. Schrödinger, Ann. Phys. **79**, 734 (1926).
[2]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), p. 275.
[3]R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[4]G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963), p. 102; K. R. Parthasarathy, Commun. Math. Phys. **9**, 303 (1968).
[5]J. Underhill, J. Math. Phys. **19**, 1932 (1978).
[6]B. S. DeWitt, Rev. Mod. Phys. **29**, 377 (1957).
[7]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University, New York, 1964).
[8]L. van Hove, Acad. Roy. Belg. Bull. Classe Sci. Mem. **37**, 610 (1951).
[9]I. E. Segal, J. Math. Phys. **1**, 468 (1960).
[10]J. M. Souriau, Commun. Math. Phys. 1, 374 (1966).
[11]B. Kostant, *Quantization and Unitary Representations*, Lecture Notes in Mathematics, Vol. 170 (Springer-Verlag, Berlin, 1970).
[12]D. J. Simms and N. N. J. Woodhouse, *Lectures on Geometric Quantization*, Lecture Notes in Physics, Vol. 53 (Springer-Verlag, Berlin, 1976); N. Woodhouse, *Geometric Quantization* (Oxford U. P., New York, 1980).
[13]T. Iwai, J. Math. Phys. **22**,1633 (1981).
[14]T. Iwai, "The symmetry group of the harmonic oscillator and its reduction," J. Math. Phys. **23**, 1088 (1982).
[15]S. I. Goldberg, *Curvature and Homology* (Academic, New York, 1962), p. 115. We must be careful in the sign in the definition of the curvature.
[16]G. A. Ringwood and J. T. Devreese, J. Math. Phys. **21**, 1390 (1980).
[17]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, Reading, Mass., 1978), 2nd ed., p. 428.
[18]T. Iwai, J. Math. Phys. **22**, 1628 (1981).
[19]M. Ikeda and Y. Miyachi, Math. Jpn. **15**, 127 (1970).
[20]D. Bergmann and Y. Frishman, J. Math. Phys. **6**, 1855 (1965).

Para-Bose and para-Fermi operators as generators of orthosymplectic Lie superalgebras

Tchavdar D. Palev[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 4 August 1980; accepted for publication 29 May 1981)

It is shown that the relative commutation relations between n pairs of para-Fermi operators and m pairs of para-Bose operators can be defined in such a way that they generate the simple orthosymplectic Lie superalgebra $B(n,m)$. In a case of ordinary statistics this leads to mutually anticommuting Bose and Fermi fields.

PACS numbers: 03.65.Fd, 11.30.Pb, 02.20.Sv, 05.30.Ch

In this note we prove a theorem.

Theorem: The relative commutation relations between n pairs of para-Fermi operators (pF operators) and m pairs of para-Bose operators[1] (pB operators) can be defined in such a way that these operators generate the simple orthosymplectic Lie superalgebra $\mathrm{osp}(2n+1, 2m) = B(n,m)$.[2]

The motivation for an investigation along this line stems from the observation that the relative para-Bose para-Fermi commutation relations can be quite arbitrary. Therefore, it is useful to find criterion that limits the choice. One such reasonable requirement is that the creation and annihilation operators (CAO's) of the parafields generate a simple Lie superalgebra (LS).

To justify the last statement, we observe that any set of only pF operators or only pB operators generate a simple Lie superalgebra. More precisely, any n pairs $f_1^{\pm},...,f_n^{\pm}$ of pF operators generate under commutation and linear space operations the simple Lie algebra (and hence a simple LS) B_n of the orthogonal group $\mathrm{SO}(2n+1)$.[3,4] Similarly, m pairs $b_1^{\pm},...,b_m^{\pm}$ of pB operators constitute a basis in the odd part of the LS $B(0,m)$ and generate it.[5] Therefore, if ψ is a parafield describing spin-$\frac{1}{2}$ particles, any subset $f_{i_1}^{\pm},...,f_{i_n}^{\pm}$ of its CAO's $f_1^{\pm}, f_2^{\pm},\cdots$ generate the algebra B_n, or, in other words, the CAO's of ψ generate a simple infinite rank LS B_∞. Similarly, the CAO's $b_1^{\pm}, b_2^{\pm},\cdots$ of a para-Bose field φ, i.e., a field of integer-spin particles generate a LS $B(0,\infty)$.

In view of the above properties, it is natural to assume that the CAO's of an arbitrary parafield φ, i.e., a field describing particles with integer and half-integer spin also generate a simple infinite-rank LS.

The mathematical problem that arises is: Does there exist a simple LS generated by n pairs of pF operators and m pairs of pB operators? We shall show that such an algebra is $B(n,m)$. The commutation relations among parafields were studied by Greenberg and Messiah.[6] On the ground of some natural assumptions they came to the conclusion that for each pair of parafields there can exist at most four types of relative commutation relations: straight commutation, straight anticommutation, relative para-Bose, and relative para-Fermi. It turns out that the requirement of the CAO's to generate a simple LS leads to relative para-Fermi relations among the pF and pB operators.

It is remarkable that one irreducible representation of $B(n,m)$ is given with relatively anticommuting Bose and Fermi operators. Therefore, if one wishes to treat the ordinary statistics on an equal Lie superalgebraical ground, then one must assume that the Bose and Fermi fields anticommute.

We now proceed to prove the theorem.

The Lie superalgebra $B(n,m)$ can be defined as the set of all matrices of the form[2] (T = transposition)

$$\begin{bmatrix} a & b & u & x & x_1 \\ c & -a^T & v & y & y_1 \\ -v^T & -u^T & 0 & z & z_1 \\ y_1^T & x_1^T & z_1^T & d & e \\ -y^T & -x^T & -z^T & f & -d^T \end{bmatrix}. \tag{1}$$

Here a is any $(n\times n)$ matrix; b and c are skew symmetric $(n\times n)$ matrices; d is any $(m\times m)$ matrix; e and f are symmetric $(m\times m)$ matrices; u and v are $(n\times 1)$ matrices; x, x_1, y, y_1 are $(n\times m)$ matrices and z, z_1 are $(1\times m)$ matrices. The even (odd) part of $B(n,m)$ is given with all block diagonal (resp. block off-diagonal) matrices in (1).

Proposition: The algebra $B(n,m)$ is generated from its odd elements

$$\left.\begin{aligned} a_i^-(1) &= \sqrt{2}(e_{0,i} - e_{i+m,0}) \\ a_j^+(1) &= \sqrt{2}(e_{0,j+m} + e_{j,0}) \end{aligned}\right\}, \quad i,j = 1,...m, \tag{2}$$

and from its even elements

$$\left.\begin{aligned} a_i^-(0) &= \sqrt{2}(e_{-i,0} - e_{0,-i-n}) \\ a_j^+(0) &= \sqrt{2}(e_{0,-j} + e_{-j-n,0}) \end{aligned}\right\}, \quad i,j = 1,...n. \tag{3}$$

Proof: Since the element $a_i^{\xi}(\alpha)$ is of order $\alpha = 0, 1$ for the Lie superalgebraical product of any two operators from (2) and (3), we obtain

$$\begin{aligned}
&[a_i^+(0),\ a_j^+(0)] = 2(e_{-j-n,-i} - e_{-i-n,-j}),\\
&[a_i^-(0),\ a_j^-(0)] = 2(e_{-j,-i-n} - e_{-i,-j-n}),\\
&[a_i^-(0),\ a_j^+(0)] = 2(e_{-i,-j} - e_{-j-n,-i-n}),\\
&\{a_i^-(1),\ a_j^+(1)\} = 2(e_{ji} - e_{i+m,j+m}),\\
&\{a_i^+(1),\ a_j^+(1)\} = 2(e_{j,i+m} + e_{i,j+m}),\\
&\{a_i^-(1),\ a_j^-(1)\} = -2(e_{j+m,i} + e_{i+m,j}),\\
&[a_i^-(0),\ a_j^+(1)] = 2(e_{-i,j+m} + e_{j,-i-n}),\\
&[a_i^-(0),\ a_j^-(1)] = 2(e_{-i,j} - e_{j+m,-i-n}),\\
&[a_i^+(0),\ a_j^+(1)] = 2(e_{-i-n,j+m} - e_{j,-i}),\\
&[a_i^+(0),\ a_j^-(1)] = 2(e_{j+m,-i} - e_{-i-n,j}).
\end{aligned} \tag{4}$$

[a)]Permanent address: Institute of Nuclear Research and Nuclear Energy, 1184 Sofia, Bulgaria.

 0022-2488/82/061100-03$02.50

Here and throughout the paper $[x,y] = xy - yx$ and $\{x,y\} = xy + yx$.

One can easily check that the linear envelope of the elements (2)–(4) coincides with (1). ■

Denote by [x,y] the LS product of $x,y \in B(n,m)$, which is a linear extension of the relation

$$[x,y] = xy - (-1)^{(\deg x)(\deg y)}yx. \tag{5}$$

For the third powers of $a_i^\xi(\alpha)$ one obtains after some calculations (throughout $\xi, \eta, \epsilon = \pm$ or ± 1 and $\alpha, \beta, \gamma = 0,1$)

$$[[a_i^\xi(\alpha),a_j^\eta(\beta)],a_k^\epsilon(\gamma)] = \tfrac{1}{2}(\epsilon\gamma + \gamma - 1)\delta_{\alpha\gamma}\delta_{ik}(\xi - \epsilon)^2 a_j^\eta(\beta) + \tfrac{1}{2}(\epsilon\gamma - \gamma + 1)\delta_{\beta\gamma}\delta_{jk}(\eta - \epsilon)^2 a_i^\xi(\alpha). \tag{6}$$

Let $f_i^\xi = a_i^\xi(0)$ and $b_i^\xi = a_i^\xi(1)$. In the case $\alpha = \beta = \gamma = 0$ (6) reduces to

$$[[f_i^\xi,f_j^\eta],f_k^\epsilon] = \tfrac{1}{2}(\eta - \epsilon)^2\delta_{jk} f_i^\xi - \tfrac{1}{2}(\xi - \epsilon)^2\delta_{jk} f_j^\eta, \tag{7}$$

whereas for $\alpha = \beta = \gamma = 1$ it gives

$$[\{b_i^\xi,b_j^\eta\},b_k^\epsilon] = (\epsilon - \xi)\delta_{ik} b_j^\eta + (\epsilon - \eta)\delta_{kj} b_i^\xi. \tag{8}$$

Equations (7) and (8) are the defining relations of the para-Fermi and para-Bose operators, respectively.[1] Thus the operators (2) and (3) are para-Bose and para-Fermi operators, respectively. Hence, according to the proposition, the LS $B(n,m)$ is generated by n pairs of pF and m pairs of pB operators, which proves the theorem.

As we have already remarked ($\approx$ is isomorphism),

$$\text{lin env.} \quad \{f_i^\xi,[f_j^\eta,f_k^\epsilon] \mid i,j,k = 1,\ldots,n\} \approx B_n, \tag{9}$$

$$\text{lin env.} \quad \{b_i^\xi,\{b_j^\eta,b_k^\epsilon\} \mid i,j,k = 1,\ldots,m\} \approx B(0,m). \tag{10}$$

Since $B_n \cap B(0,m) = 0$, the LS $B(n,m)$ is generated from the direct space sum $B_n + B(0,m)$ of its subalgebras B_n and $B(0,m)$.

Equation (6) was derived from the defining representation (1) of $B(n,m)$. Since, however, this equation is written only in terms of Lie superalgebraical operations, it holds also in any other representation. In other words, Eq. (6) defines $B(n,m)$ in an abstract way. It carries all information about the Lie superalgebra. Any (irreducible) representation of the operators $a_i^\xi(\alpha)$ extends to an (irreducible) representation of $B(n,m)$ and vice versa. Therefore, the representation theory of the para-operators (2)–(3) coincides with the representation theory for the LS $B(n,m)$.

In a case of field theory, when $n \to \infty$ and $m \to \infty$ and even if the CAO's are labelled with a continuum number of indices, the trilinear relation (6) remains well defined and determines the Lie superalgebraical products between the generators of an infinite rank orthosymplectic LS.

From (6) one can immediately draw some further conclusions about the CAO's and their representations. The set of all vectors

$$h_i(\alpha) = -\tfrac{1}{2}[a_i^+(\alpha),a_i^-(\alpha)] \tag{11}$$

constitutes a basis in the Cartan subalgebra. From (6) we obtain

$$[h_i(\alpha),a_k^\epsilon(\gamma)] = -\epsilon\delta_{\alpha\gamma}\delta_{ik} a_k^\epsilon(\gamma). \tag{12}$$

Therefore, in the basis (11) the creation ($\epsilon = +$) and annihilation ($\epsilon = -$) operators are negative and positive root vectors of the algebra, respectively.

The Fock representations are defined in the usual for the parastatistics way,[6] namely, with the requirement that the space contains a single vacuum $|0\rangle$ such that: (a) it is an eigenvector of the Cartan subalgebra H; (b) the operators $a_i^-(\alpha)$ annihilate $|0\rangle$; and (c) the whole space is a closure of all vectors generated from the vacuum by means of only creation operators.

From these requirements one can draw consequences about the Lie superalgebraical properties of the Fock representations. Denote by N^+ the subspace of all positive root vectors and let H^* be the dual to H space. From (a)–(c) it immediately follows that

$$h|0\rangle = \Lambda(h)|0\rangle, \quad h \in H, \ \Lambda \in H^*, \tag{13}$$

$$N^+|0\rangle = 0, \tag{14}$$

and hence all Fock representations are contained among the representations with a highest weight.[7] In a more mathematical terminology, they are representations induced from the trivial representations of the subalgebra

$$G = \text{lin. env.}\{a_i^-(a),[a_j^-(\beta),a_k^\xi(\gamma)] \mid \xi = \pm, \alpha,\beta,\gamma = 0,1\}. \tag{15}$$

More explicitly the Lie subalgebra G is described in terms of the defining matrix representaion (1) of $B(n,m)$ by the additional conditions $b = 0$, $u = 0$, $x_1 = 0$, and $e = 0$. The algebra G contains as a subalgebra the general linear Lie superalgebra ($\approx$ denotes an isomorphism)

$$\mathrm{gl}(n,m) \approx \text{lin. env.}\{[a_i^-(\alpha),a_j^+(\beta)] \mid \alpha,\beta = 0,1\} \tag{16}$$

so that the following inclusion holds

$$\mathrm{gl}(n,m) \subset G \subset B(n,m). \tag{17}$$

Note that the Eq. (6) is satisfied if $a_i^\pm(1) \equiv a_i^\pm$, $i = 1,\ldots,m$, and $a_j^\pm(0) \equiv a_{j+m}^\pm$, $j = 1,\ldots,n$, are assumed to be Bose and Fermi operators, respectively, provided that

$$\{a_i^\xi,a_j^\eta\} = 0, \quad i = 1,\ldots,m, \ j = m+1,\ldots,m+n, \tag{18}$$

i.e., the Bose and the Fermi operators mutually anticommute. In this case the Fock space $W(m,n)$ has an orthonormed basis (with respect to the usual for the quantum mechanic scalar product) the vectors[8]

$$|p_1,p_2,\ldots p_r\rangle = \frac{(a_1^+)^{p_1}(a_2^+)^{p_2}\cdots(a_r^+)^{p_r}}{(p_1!p_2!\cdots p_r!)^{1/2}}|0\rangle, \quad r = m+n, \tag{19}$$

where $p_1,\ldots,p_m$ are arbitrary nonnegative integers and for $i > m$, $p_i = 0,1$. If $m > 1$, the representation space $W(m,n)$ is infinite-dimensional. A straightforward calculation gives

$$a_i^+|\ldots,p_i,\ldots\rangle = (g_i)^{p_1+\cdots+p_{i-1}}(1 + g_i p_i)^{1/2}|\ldots,p_i+1,\ldots\rangle,$$
$$a_i^-|\ldots,p_i,\ldots\rangle = (g_i)^{p_1+\cdots+p_{i-1}}(p_i)^{1/2}|\ldots,p_i-1,\ldots\rangle, \tag{20}$$

where $g_1 = g_2 = \cdots = g_m = -g_{m+1} = \cdots = -g_{m+n} = 1$.

To show that the representation of the operators $a_1^\pm,\ldots,a_r^\pm$ in $W(m,n)$ is irreducible, consider two arbitrary vectors from the dense subspace $V(m,n)$, which is a linear envelope of the basis vectors (19). Any two such vectors are of the form $P|0\rangle$ and $Q|0\rangle$, where P and Q are polynomials of

creation operators. Let for definiteness

$$Q = (a_1^+)^{p_1}\cdots(a_r^+)^{p_r} + \sum_{q_1,\ldots,q_r} \alpha_{q_1\cdots q_r}(a_1^+)^{q_1}\cdots(a_r^+)^{q_r}, \qquad (21)$$

where the sum in (21) is over all $(q_1,\ldots,q_r)\neq(p_1,\ldots,p_r)$ such that $q_1+\cdots+q_r<p_1+\cdots+p_r$. From (20) one easily concludes that

$$(a_1^-)^{p_1}\cdots(a_r^-)^{p_r}Q\,|0\rangle = \kappa|0\rangle, \quad \kappa\neq 0.$$

Hence

$$P\,|0\rangle = (1/\kappa)P(a^{-1})^{p_1}\cdots(a_r^-)^{p_r}Q\,|0\rangle$$

and therefore the space $W(m,n)$ is irreducible with respect to the operators $a_1^\pm,\ldots,a_r^\pm$. Since these operators generate a representation of $B(n,m)$, we have the following.

Corollary: Any set of mutually anticommuting Fermi and Bose operators generates one particular infinite-dimensional representation of an orthosymplectic Lie superalgebra.

At the same time it is evident that the so-called normal type commutation relations

$$[f_i^\xi, b_j^\eta] = 0, \qquad i=1,\ldots,n,\ j=1,\ldots,m \qquad (22)$$

of n pairs of para-Fermi operators and m pairs of para-Bose operators close an irreducible representation of the semisimple LS $B_n + B(0,m)$. Therefore, in theories unifying in one irreducible multiplet (one superfield) particles with integer and half-integer spin, it is natural to assume that within one mulitplet the Bose and Fermi fields anticommute, whereas the fields from different multiplets obey the normal commutation relations.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

[1]H. S. Green, Phys. Rev. **90**, 270 (1953).
[2]V. G. Kac, Adv. Math. **26**, 8 (1977). Throughout the paper we use the Kac notation for the simple LS's.
[3]S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1960).
[4]C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[5]A. Gantchev and T. Palev, J. Math. Phys. **21**, 797 (1980).
[6]O. W. Greenberg and A. M. Messiah, Phys. Rev. **138**, B1155 (1965).
[7]V. G. Kac, Lecture Notes Math. **676**, 597 (1978).
[8]T. D. Palev, J. Math. Phys. **22**, 2127 (1981).

Hypervirial-perturbational treatment of the bounded hydrogen atom

Francisco M. Fernández and Eduardo A. Castro [a)]
INIFTA, Sección Química Teórica, Sucursal 4-Casilla de Correo 16, La Plata 1900, Argentina

(Received 22 July 1981; accepted for publication 23 October 1981)

The hydrogen atom model enclosed inside a container with impenetrable walls is studied by means of a method which combines hypervirial theorems and perturbation theory. From hypervirial relations, very useful recursion relationships for diagonal matrix elements of coordinate powers are given. Numerical as well as analytical results are compared with those given in previous works.

PACS numbers: 03.65.Ge

I. INTRODUCTION

The bounded quantal systems such as atoms and molecules enclosed within arbitrary surfaces,[1-28] and the finite harmonic oscillator[29-42] are extremely useful for studying several quite different properties in various physics, chemistry, and astronomy fields. Particular interest has been directed towards the hydrogen atom enclosed in a spherical box with impenetrable[1-7,9-15,17,18,21] and semipermeable walls,[16] or with different shapes,[19,20] during a period of many years.

The mathematical procedures employed in the analysis of the hydrogen atom model enclosed inside a container with impenetrable walls can be classified as follows: (a) application of the confluent hypergeometric equation,[1-3,6,7,13,17,21] (b) use of the perturbation theory,[4,5,10-12] (c) WKB method,[3] (d) employment of variational principles,[9,14] and (e) calculation with finite elements.[15,18]

The purpose of this paper is to display the study of this model by means of a new method recently given by the present authors.[43-45] This method has proved to be powerful enough for the analysis of a wide variety of models, and also has yielded very accurate numerical results. The procedure is based on a combination of the hypervirial theorems and the perturbation theory, and it is described in detail in Sec. II. It has been applied successfully to several models, such as (1) unidimensional harmonic oscillator under Dirichlet and von Neumann boundary conditions (DBC and VNBC, respectively),[43,46] (2) multidimensional isotropic bounded oscillator,[47] and (3) the particle in an unidimensional box in the presence of a uniform electric field.[48] This latter model has been previously treated by other authors.[49,50] In each case we have obtained very accurate numerical results as well as suitable analytical formulas for the coordinate powers matrix elements and energy perturbational corrections.

The main advantage of our method is that perturbational corrections are generated from very simple recursion relations.

The plan of the paper is as follows: Section II deals with the method to be applied, and the fundamental equations are deduced. Then that method is applied to the enclosed hydrogen atom in Sec. III. Although various useful and important relationships are obtained in this section, the method displays its full power in Sec. IV, where the enclosed hydrogen atom model is transformed in a bidimensional harmonic oscillator via an appropriate change of variables. In this section analytical formulas are deduced which allow calculation of any energy level, with a remarkable exactness, in a wide range of values for the radius of the sphere.

II. METHOD

Let us consider the stationary unidimensional Schrödinger equation

$$-\tfrac{1}{2}\phi'' + V(x)\phi = E\phi, \tag{1}$$

where we assume the normalization condition $\langle\phi|\phi\rangle = 1$ and that ϕ is subjected to the DBC in the extreme points of the close interval $[0,b]$,

$$\phi(0) = \phi(b) = 0. \tag{2}$$

It is deduced at once that, owing to these particular boundary conditions (BC), the following relations are satisfied:

$$\langle [H,x^N] \rangle = 0, \tag{3}$$

$$\langle [H,x^N D] \rangle = \tfrac{1}{2} b^N |\phi'(b)|^2, \tag{4}$$

where $D \equiv d/dx$ and $N = 1,2,\cdots$. Also, from Gonda and Gray's results,[51] we know that

$$\partial E/\partial b = -\tfrac{1}{2}|\phi'(b)|^2. \tag{5}$$

The substitution of (5) in (4) allows us to obtain a modified version of the hypervirial theorems introduced by Hirschfelder[52]

$$\langle [H,x^N D] \rangle = -b^N \partial E/\partial b. \tag{6}$$

The expectation value of the commutator in Eq. (6) is not null because $x^N\phi'(x)$, does not fulfill the same BC as $\phi(x)$ does, and consequently does not belong to the domain of the operator H. For the particular case $N = 1$, we arrive to the well-known quantum virial theorem (VT)[53,54]

$$\langle [H,xD] \rangle = 2\langle T \rangle - \langle xV \rangle' = -b\,\partial E/\partial b. \tag{7}$$

The calculation of the commutator in Eq. (6) gives D and D^2 terms, which can be removed by means of Eqs. (1) and (3), in the manner of the previously given procedure presented by Swenson and Danforth[55] and Killingbeck.[56] It enables us to obtain a relationship among expectation values of the coordinate x functions:

$$\tfrac{1}{4}N(N-1)(N-2)\langle x^{N-3}\rangle + 2NE\langle x^{N-1}\rangle - 2N\langle x^{N-1}V\rangle - \langle x^N V'\rangle = -b^N \frac{\partial E}{\partial b}. \tag{8}$$

[a)]Any correspondence regarding this paper should be addressed to E. A. Castro.

 0022-2488/82/061103-09$02.50

In what follows, it will be enough for us to consider that $V(x)$ has the form

$$V(x) = g^2/2x^2 + ax^k. \tag{9}$$

The substitution of (9) in (8) gives us a set of relations among $A^N \equiv \langle x^N \rangle$, the eigenvalues E, and $\partial E/\partial b$

$$(N-1)\{\tfrac{1}{4}N(N-2) - g^2\}A^{N-3} + 2NEA^{N-1} - a(2N+k)A^{N+k-1} = -b^N \partial E/\partial b. \tag{10}$$

The VT yields a new useful equation

$$2E - (k+2)aA^k = -b\,\partial E/\partial b, \tag{11}$$

which permits getting rid of $\partial E/\partial b$ in (10) and arriving at a set of equations which relate exclusively the diagonal matrix elements A^N and E, i.e.,

$$(N-1)\{\tfrac{1}{4}N(N-2) - g^2\}A^{N-3} + 2NEA^{N-1} - a(2N+k)A^{N+k-1} - b^{N-1}\{2E - (k+2)aA^k\} = 0. \tag{12}$$

Expanding A^N and E power series in a,

$$A^N = \sum_s A_s^N a^s, \quad E = \sum_s E^s a^s, \tag{13}$$

and applying the Hellmann–Feynman theorem (HFT)

$$\partial E/\partial a = A^k, \tag{14}$$

Eq. (12) is transformed into a recursion relationship that permits the determination of the whole set of elements A_s^N:

$$A_0^N = \frac{b^N}{N+1} - \frac{N}{2(N+1)E^0}\{\tfrac{1}{4}(N^2-1) - g^2\}A_0^{N-2}, \tag{15}$$

$$A_s^N = \frac{2-(k+2)s}{2(N+1)sE^0} b^N A_{s-1}^k - \frac{N}{2(N+1)E^0}\{\tfrac{1}{4}(N^2-1) - g^2\}A_s^{N-2} + \frac{2N+k+2}{2(N+1)E^0} A_{s-1}^{N+k} - \frac{1}{E^0}\sum_{j=1}^{s} \frac{A_{j-1}^k A_{s-j}^N}{j}, \quad s>0. \tag{16}$$

Equations (11) and (14) determine the form of the perturbative corrections $E^s(b)$,

$$E^s(b) = E^s(1)b^{(k+2)s-2},$$
$$E^s(b) = A_{s-1}^k(b)/s, \quad s>0. \tag{17}$$

On account of Eq. (17), it is only necessary to calculate the perturbation corrections for a single b value, say $b=1$, because

$$E(b) = \sum_{s=0}^{\infty} E^s(1)a^s b^{(k+2)s-2} = \frac{E^0(1)}{b^2} + \sum_{s=1}^{\infty} a^s \frac{A_{s-1}^k(1)b^{(k+2)s-2}}{s}. \tag{18}$$

When $a=0$, the change of variable

$$t = x(2E^0)^{1/2}$$

produces the following differential equation for the zero-order wavefunction:

$$\phi^{0\prime\prime} + (1 - g^2/t^2)\phi^0 = 0. \tag{19}$$

Therefore, $\phi^0(t)$ is related to the Bessel function $J_c(t)$ in the following way:

$$\phi^0 = t^{1/2}J_c(t), \quad c = (g^2 + \tfrac{1}{4})^{1/2}. \tag{20}$$

Denoting by j_{cn} the nth zero of $J_c(t)$, we obtain from the BC(2) a formula for zero-order eigenvalues,

$$E_{cn}^0 = j_{cn}^2/2b^2. \tag{21}$$

The relation (21) and the normalization condition

$$A_s^0 = \delta_{0s} \tag{22}$$

are the starting point for the procedure that allows us to calculate the coefficients A_s^N. From the foregoing deductions, it is clear that Eq. (12) is valid when $N>0$, while Eqs. (15) and (16) are valid when $N > -1$. For $N=0$, Eq. (6) is not applicable owing to

$$\langle [H,D] \rangle = -\langle V' \rangle = -\partial E/\partial b - \tfrac{1}{2}|\phi'(0)|^2. \tag{23}$$

For the potential (9), the preceding equation has the form

$$-g^2A^{-3} + akA^{k-1} = \partial E/\partial b + \tfrac{1}{2}|\phi'(0)|^2. \tag{24}$$

III. FIRST APPLICATION TO THE HYDROGEN ATOM INSIDE A SPHERICAL SURFACE

Because of the spherical symmetry of the potential, the angular dependence of the eigenfunctions ϕ_{nlm} for this model is given by the spherical harmonics $Y_l^m(\theta,\phi)$, while the radial part $R_{nl}(r)$ satisfies the differential equation

$$-\frac{\hbar^2}{2m}\left\{R'' + \frac{2}{r}R' - \frac{l(l+1)R}{r^2}\right\} - \frac{Ze^2R}{r} = ER, \tag{25}$$

$$R(r_0) = 0 \tag{26}$$

where, from now on, to simplify the writing, we omit the indices n and l in the function $R(r)$. The change of variables

$$r = px, \quad p = \hbar^2/me^2, \quad r_0 = px_0 \tag{27}$$

transforms Eq. (25) into

$$-\frac{1}{2}\left\{f''(x) + \frac{2}{x}f'(x) - \frac{l(l+1)}{x^2}f(x)\right\} - \frac{Z}{x}f(x) = \bar{E}f(x), \tag{28}$$

where

$$f(x) = R(px), \quad \bar{E} = \hbar^2 E/me^4. \tag{29}$$

Equation (28) can be changed in such a way that it transforms in a form similar to that discussed in Sec. II. For such a purpose it is only necessary to define a new function $h(x)$ as follows:

$$h(x) = xf(x). \tag{30}$$

Then, substituting (30) in (28), we arrive at

$$-\tfrac{1}{2}h''(x) + V(x)h(x) = \bar{E}h(x), \tag{31}$$

$$V(x) = -Z/x + g^2/2x^2, \quad g^2 = l(l+1). \tag{32}$$

The potential (32) is a particular case of function (9) (when $k=-1$). Besides, the solutions $h(x)$ fulfill the BC (2) in the point $b = x_0$, which assures the validity of equations deduced in Sec. II. The recursion relationships (15) and (16) do not allow calculation of A_s^{-1}, and prevent a direct application of the method for the enclosed hydrogen atom model.

TABLE I. $\langle r^{-1}\rangle$, $\langle r\rangle$, $\langle r^2\rangle^2$, and $\langle T\rangle$ values for $Z = 1$, from the polynomial of Ref. 11, the VT, and Eqs. (40)–(42).

x_0	$\bar{E}$	$-(\partial\bar{E}/\partial x_0)$	$\langle r^{-1}\rangle$	$\langle r\rangle$	$\langle r^2\rangle^2$	$\langle T\rangle$
0.8	4.5433	15.4822	3.2991	0.3802	0.0276	7.8425
1.2	1.2693	4.0345	2.3028	0.5534	0.1261	3.5721
1.6	0.2713	1.4744	1.8163	0.7135	0.3560	2.0877
2.0	−0.1250	0.6426	1.5352	0.8590	0.7605	1.4102
2.4	−0.3064	0.3105	1.3581	0.9884	1.3670	1.0516
2.8	−0.3968	0.1599	1.2414	1.1001	2.1486	0.8446
3.2	−0.4443	0.0859	1.1637	1.1927	3.0383	0.7194
3.6	−0.4703	0.0479	1.1130	1.2648	3.9236	0.6427
4.0	−0.4850	0.0280	1.0822	1.3152	4.6602	0.5971
4.4	−0.4940	0.0179	1.0668	1.3428	5.0931	0.5728
4.8	−0.5001	0.0133	1.0638	1.3470	5.0877	0.5638
∞	−0.5000	0.0000	1.0000	1.5000	9.0000	0.5000

However, the hypervirial theorems (12) are very useful because they provide extremely important formulas for the expectation values A^N which are closely related to various physical properties.

The substitutions $a = -Z, k = -1$, and $b = x_0$ in (10) and (24) produce

$$-g^2A^{-3} + ZA^{-2} = \partial\bar{E}/\partial x_0 + \tfrac{1}{2}|f(0)|^2, \tag{33}$$

$$(N-1)\{\tfrac{1}{4}N(N-2) - g^2\}A^{N-3} + 2NEA^{N-1} + Z(2N-1)A^{N-2} = -x_0^N\,\partial\bar{E}/\partial x_0, \tag{34}$$

and from Eq. (5)

$$\partial\bar{E}_{nl}/\partial x_0 = (-x_0^2/2)|f'_{nl}(x_0)|^2. \tag{35}$$

Equation (35) makes certain that $\bar{E}_{nl}$ decreases for increasing x_0 for any given pair (n,l). This result represents a generalization of a previous theorem for the ground state, given in Ref. 11. When x_0 (or r_0) tends to infinity, $\partial\bar{E}/\partial x_0$ tends to zero, and this allows the fulfillment of the VT in its usual form, i.e.,

$$2\bar{E} + ZA^{-1} = 0.$$

The expansion of $\bar{E}$ in a Z-power series is obtained from Eq. (18),

$$\bar{E}(x_0) = \sum_{s=0}^{\infty} \bar{E}^s(1)Z^s x_0^{s-2} = \frac{\bar{E}^0(1)}{x_0^2} + \sum_{s=1}^{\infty} \frac{A_{s-1}^{-1}}{s} x_0^{s-2}Z^s. \tag{36}$$

If this expansion is cut off, then the simultaneous satisfaction of the VT and HFT is not possible. As an immediate consequence, the correct value will not be gotten when $x_0 \to \infty$. This difficulty was pointed out by Wigner[4] and afterwards it was discussed at full length in several papers.[5,8,10] It is interesting to notice how the employment of the VT and HFT leads to the exact result when the potential is considered as a perturbation.

The combination of both theorems yields

$$2\bar{E} - Z\,\partial\bar{E}/\partial Z = -x_0\,\partial\bar{E}/\partial x_0. \tag{37}$$

Then, when $x_0 \to \infty$, $\bar{E} \to AZ^2$, with A a constant. Accordingly, the variational method used in Ref. 11 must lead to the exact result if one takes a large enough number of terms, because the trial function satisfies both VT and HFT. The method reported in Ref. 11 consists in minimizing the functional energy with respect to the radius of the sphere. This is confirmed in a plain manner, because the value of the radius for which the variational energy is a minimum increases with increasing number of terms in the trial function.

From Eqs. (33) and (34) we can immediately deduce the results

$$ZA^{-2} = \frac{\partial\bar{E}}{\partial x_0} + \tfrac{1}{2}|f(0)|^2 \quad \text{for } l = 0, \tag{38}$$

$$ZA^{-2} - g^2A^{-3} = \frac{\partial\bar{E}}{\partial x_0} \quad \text{for } l \neq 0, \tag{39}$$

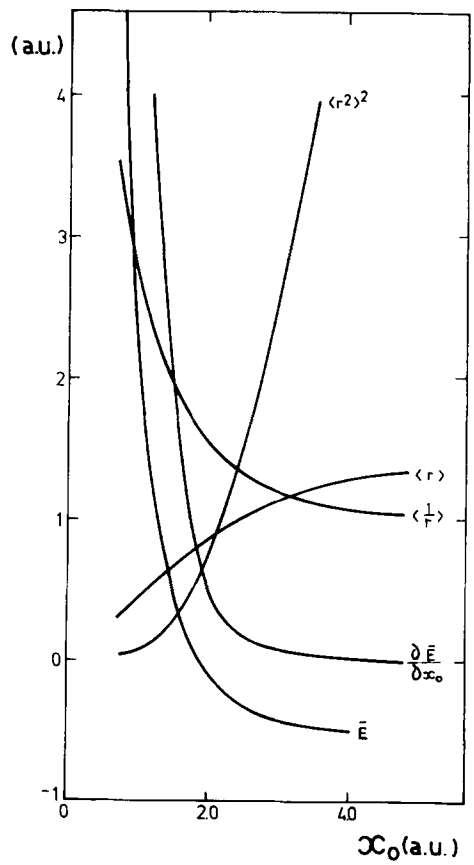


FIG. 1. Matrix elements corresponding to the ground state.

$$A^{-1} = Z^{-1}\left\{ -x_0 \frac{\partial \bar{E}}{\partial x_0} - 2\bar{E} \right\}, \tag{40}$$

$$A^{1} = \frac{1}{4\bar{E}} \left\{ -x_0^2 \frac{\partial \bar{E}}{\partial x_0} + g^2 A^{-1} - 3Z \right\}, \tag{41}$$

$$A^{2} = \frac{1}{6\bar{E}} \left\{ -x_0^3 \frac{\partial \bar{E}}{\partial x_0} + (2g^2 - \tfrac{3}{2}) - 5ZA^{1} \right\}. \tag{42}$$

These formulas can be verified at once through the employment, for example, of the functions $2s$ and $3p$ corresponding to the free hydrogen atom, because they have a node at 2 a.u. and 6 a.u., respectively. The great importance of these relationships rests upon the fact that relevant physical properties, such as the diamagnetic screening constant, the polarizability, and the pressure, are proportional to A^{-1}, $(A^2)^2$, and $\partial \bar{E}/\partial x_0$, respectively.[2,6,13,16,21] It is only necessary to know the polynomial $\bar{E}(x_0)$ in order to be able to calculate those physical magnitudes.

In Table I and Fig. 1 we display the behavior of these quantities as a functions of x_0. For such a purpose, the polynomial presented in Ref. 11 was chosen.

Now we will show a simple method for estimating the $\bar{E}(x_0)$ value within an interval that contains the point x_0^0, when $\langle T\rangle^0 = \langle T\rangle(x_0^0)$ and $\langle V\rangle^0 = \langle V\rangle(x_0^0)$ are exactly known.

As the Hamiltonian operator satisfies the relation

$$H(Z,x) = Z^2 H(1,Zx), \tag{43}$$

it follows at once that

$$\bar{E}(Z,x_0) = Z^2 \bar{E}(1,Zx_0). \tag{44}$$

If $\langle T\rangle^0$ and $\langle V\rangle^0$ are known exactly when $Z = 1$ and $x_0 = x_0^0$, then we can estimate $\bar{E}(Z,x_0^0)$ in an approximate fashion as follows:

$$E(Z,x_0^0) \cong \langle T\rangle^0 + Z\langle V\rangle^0 \tag{45}$$

and this permits us to evaluate $\bar{E}(1,Zx_0^0)$,

$$\bar{E}(1,Zx_0^0) \cong Z^{-2}\{\langle T\rangle^0 + Z\langle V\rangle^0\}, \tag{46}$$

so we have $\bar{E}(1,x_0)$, where $x_0 = Zx_0^0$.

The eigenfunctions corresponding to the free (unbounded) hydrogen atom which have zeros in $(0,\infty)$ may be chosen as exact zero-order functions ϕ^0 (necessary for the calculation). In Table II we show the values for the ground state of the bounded hydrogen atom, using a $2s$ orbital of the free atom ($x_0^0 = 2$). Taking into account the rough approximation involved in Eq. (45), we can properly consider that the results are excellent. A feasible explanation of this success of Eq. (46) is that the VT is satisfied

$$-x_0 \frac{\partial \bar{E}}{\partial x_0}(1,Zx^0) = -Z\frac{\partial \bar{E}}{\partial Z}(1,Zx_0^0) = 2\langle T\rangle + \langle V\rangle, \tag{47}$$

where

$$\langle T\rangle = Z^{-2}\langle T\rangle^0 \text{ and } \langle V\rangle = Z^{-1}\langle V\rangle^0.$$

Equations (43) and (44) are of peculiar interest because they permit a transformation of the BC perturbation into a Hamiltonian perturbation.

When the potential energy has the form (9), the general formula is

$$E(1,a^{1/(k+2)}x_0^0) = a^{-2/(2+k)}E(a,x_0^0). \tag{48}$$

Although this procedure is restricted to the calculation of the energy for a small interval of known value, it possesses the advantage of being very simple and applicable to any state.

TABLE II. Ground state energy of the bounded H atom calculated from Eq. (46) ($x_0 = 2Z$).

x_0	$-\bar{E}$ [Eq. (46)]	$-\bar{E}_{\text{exact}}$ [a]
1.90	0.0535	0.0541
1.92	0.0690	0.0694
1.94	0.0839	0.0841
1.96	0.0982	0.0983
1.98	0.1119	0.1119
2.00	0.1250	0.1250
2.02	0.1376	0.1376
2.04	0.1497	0.1497
2.06	0.1612	0.1614
2.08	0.1723	0.1727
2.10	0.1830	0.1835

[a]Calculated with the polynomial given in Ref. 11.

IV. SECOND APPLICATION

In the preceding section we have seen that a direct application of Eqs. (15) and (16) does not allow us to solve the problem. Generally, these equations are useful when $k > 0$.[43–48] Nevertheless, by means of an approximate treatment of the Schrödinger equation (28) it is possible to make use of our method in order to study the bounded hydrogen atom. Several years ago, Dingle[3] showed that by a redefinition of the variable and the wavefunction, Eq. (28) can be transformed to one corresponding to a two-dimensional harmonic oscillator.

Starting from the definitions

$$q = (8|\bar{E}|)^{1/4}x^{1/2} \tag{49}$$

and

$$h(q) = x^{3/4}f(x), \tag{50}$$

Eq. (28) changes to

$$-\tfrac{1}{2}h'' + \frac{g^2 h}{2q^2} + aq^2 h = Wh, \tag{51}$$

where

$$\bar{E} = \mp 2Z^2/W^2, \tag{52}$$

$$g^2 = (2l+1)^2 - \tfrac{1}{4}, \tag{53}$$

and

$$a = \begin{cases} \tfrac{1}{2} & \text{for } \bar{E} < 0 \\ -\tfrac{1}{2} & \text{for } \bar{E} > 0. \end{cases}$$

Equations (49) and (52) allow us to relate x_0 with q_0 and W, i.e.,

$$x_0 = q_0^2 W/4Z. \tag{54}$$

The solutions $h(q)$ satisfy the DBC in 0 and q_0, so that we can apply freely the method just sketched in Sec. II, using a as the perturbational parameter and W as eigenvalue. The zero-order equation is gotten from Eq. (51) for $a = 0$. In this case,

TABLE III. Eigenvalues for the energy levels $1s$, $2s$, and $2p$ corresponding to the bounded H atom through the perturbational method described in Sec. II.

q_0	x_0	$\bar{E}^a$	$\bar{E}^b$	$\bar{E}^c$	$\bar{E}^d$	$\bar{E}^e$
			$1s$ level ($n=1, l=0$)			
2.1917	0.8100	4.3960	4.935	4.392	4.392	4.4025
2.1392	0.9100	3.1608	3.427	3.159	3.159	3.1710
2.0820	1.0100	2.3025	2.518	2.301	2.301	2.3146
2.0385	1.0800	1.8506	2.193	1.849	1.849	1.8643
1.9916	1.1500	1.4871	1.928	1.486	1.486	1.5023
1.7348	1.4480	0.5400	0.500	0.540	0.540	0.5607
1.3114	1.7110	0.1263	0.125	0.126	0.126	0.1516
1.0815	1.7780	0.0541	0.0566	0.054	0.054	0.0806
~0.0000	1.8352	0.000	0.0000	0.000	0.000	0.0275
1.1264	1.9020	−0.0556	−0.0566	−0.056	−0.056	−0.0269
1.2440	1.9343	−0.0800	−0.0800	−0.080	−0.080	−0.0507
1.4142	2.0000	−0.1250	−0.1250	−0.125	−0.125	−0.0945
1.7043	2.1782	−0.2223	−0.2222	−0.222	−0.222	−0.1884
1.7320	2.2005	−0.2323	−0.2323	−0.232	−0.232	−0.1981
2.0007	2.4720	−0.3277	−0.3273	−0.327	−0.327	−0.2880
2.0999	2.6000	−0.3595		−0.359	−0.359	−0.3170
2.2393	2.8070	−0.3989	−0.3965	−0.398	−0.398	−0.3515
2.2430	2.8130	−0.3999			−0.399	−0.3523
2.3761	3.0413	−0.4307			−0.428	−0.3774
2.4559	3.1920	−0.4463	−0.4417		−0.443	−0.3887
		$2s$ level ($n=2, l=0$)				
3.2713	1.7600	4.6213	4.935			
3.2383	1.9200	3.7289	4.078			
3.2152	2.0300	3.2416	3.427			
3.1895	2.1500	2.7985	2.920			
3.1644	2.2650	2.4432	2.518			
3.1366	2.3900	2.1182	2.193			
3.0722	2.6700	1.5620	1.611			
3.0130	2.9150	1.2123	1.234			
2.8270	3.6090	0.6130	0.6143			
2.7627	3.8230	0.4983	0.5000			
2.2458	5.1110	0.1217	0.1250			
1.9221	5.5890	0.0546	0.0556			
1.6979	5.8080	0.0308	0.0312			
~0.0000	6.1523	0.0000	0.0000			
1.6035	6.4290	−0.0200	−0.0200			
1.8181	6.6110	−0.0312	−0.0312			
1.9689	6.7850	−0.0408	−0.0408			
1.9982	6.8240	−0.0428	−0.0429			
2.1731	7.0960	−0.0554	−0.0566			
2.2360	7.2120	−0.0601	−0.0601			
2.4445	7.6770	−0.0757	−0.0764			
2.4967	7.8150	−0.0795	−0.0800			
2.8074	8.8520	−0.0991	−0.1021			
3.1200	10.3600	−0.1104	−0.1165			
		$2s$ level ($n=2, l=1$)				
2.7758	1.4100	3.7332	3.943			
2.7506	1.5500	2.9784	3.116			
2.7264	1.6800	2.4472	2.524			
2.5472	2.5280	0.8234	0.8261			
2.5057	2.6980	0.6769	0.5000			
2.0168	4.1100	0.1224	0.1250			
1.7374	4.5540	0.0549	0.0556			
1.5279	4.7700	0.0299	0.0312			
~0.0000	5.0823	0.0000				
1.4666	5.3550	−0.0201	−0.0200			
1.6626	5.5280	−0.0313	−0.0312			
1.8042	5.6960	−0.0408	−0.0408			
1.9999	6.0000	−0.0555	−0.0566			
2.2356	6.4970	−0.0740	−0.0740			
2.3150	6.7010	−0.0799	−0.0800			
2.4484	7.0890	−0.0894	−0.0895			
2.8299	8.5400	−0.1099	−0.1097			
3.1625	10.2350	−0.1194	−0.1193			

[a]Present calculation with Eq. (68). [b]Reference 2. [c]Reference 12. [d]Reference 14. [e]Perturbational energy up to the second order (Ref. 11).

TABLE IV. $2p$ level energy of the bounded H atom calculated via numerical integration of the Schrödinger equation.

x_0	1.379	1.412	1.523	1.552	1.661	1.682	2.529
$\bar{E}$	3.943	3.7332	3.116	2.9784	2.524	2.4472	0.8234

the change of variable

$$z = (2W)^{1/2}q$$

produces the following equation:

$$y''(z) + (1 - g^2/z^2)y(z) = 0, \quad y(z) = h\,(z/(2W)^{1/2}), \quad (55)$$

whose solutions are associated with Bessel functions of the first kind,

$$y(z) = z^{1/2}J_{2l+1}(z). \quad (56)$$

If j_{ni} is the nth zero of the ith Bessel function $J_i(x)$, the DBC $y((2W)^{1/2}q_0) = 0$ yields the zero-order eigenvalues

$$W^0_{nl} = j^2_{n\,2l+1}/2q_0^2. \quad (57)$$

Through formula (54) we obtain the corresponding x_0 value

$$(x_0)_{nl} = j^2_{n\,2l+1}/8Z. \quad (58)$$

The quantities j_{ni} have been tabulated to great accuracy.[57] For the procedure shown in Sec. II, it is possible to write W as a power series in q_0

$$W = \sum_{s=0}^{\infty} W^s(q_0 = 1)a^s q_0^{4s-2}. \quad (59)$$

When q_0 is small, $W \cong W^0(q_0 = 1)q_0^{-2}$ is large, $\bar{E}$ tends to zero, and x_0 tends to the values given by the formula (58). For these particular x_0 values, the energy changes sign.[2,3] This transformation, which was recently discussed by Rowley,[58] transforms the hydrogen atom model into a harmonic oscillator ($a = \frac{1}{2}$, $\bar{E} < 0$), or an inverted oscillator[41,42] ($a = -\frac{1}{2}$, $\bar{E} > 0$).

The iterative resolution of Eqs. (15) and (16), with $k = 2$ and $b = 1$, permits us to obtain without further difficulties, the following quantities:

$$A_0^2 = \frac{1}{3} - \frac{1}{3w}(0.75 - g^2), \quad (60)$$

$$A_0^4 = \frac{1}{5} - \frac{2}{15w}\left\{\frac{15}{4} - g^2\right\} + \frac{2}{15w^2}\left\{\frac{15}{4} - g^2\right\}(0.75 - g^2), \quad (61)$$

$$A_1^2 = \frac{2}{45w} + \left\{\frac{1}{3}(0.75 - g^2) - \frac{8}{45}(0.375 - g^2)\right\}\frac{1}{w^2} + \left\{\frac{8}{45}(0.375 - g^2)(0.75 - g^2) - \frac{1}{9}(0.75 - g^2)^2\right\}\frac{1}{w^3}, \quad (62)$$

where A_s^N represents the s-order correction for $\langle q^N\rangle(q_0 = 1)$ and $w = W^0(q_0 = 1)$. With these elements it is possible to deduce an expression for W, corrected up to the second order,

$$W = \frac{w}{q_0^2} + a\left\{\frac{1}{3} - \frac{1}{3w}(0.75 - g^2)\right\}q_0^2 + a^2\left\{\frac{1}{45w} + \left[\frac{1}{6}(0.75 - g^2) - \frac{4}{45}(0.375 - g^2)\right]\frac{1}{w^2} + \left[\frac{4}{45}(0.375 - g^2)(0.75 - g^2) - \frac{1}{18}(0.75 - g^2)^2\right]\frac{1}{w^3}\right\}q_0^6. \quad (63)$$

This formula gives the best results for small q_0 and large w values. For the purpose of determining the accuracy of (63), it is necessary to compare our results with those given by other methods. After a careful search, we found that those published by DeGroot and Ten Seldam[2] are the most complete, so that we have chosen them for comparative purposes. Our results and those calculated in Ref. 2 for the levels $1s$, $2s$, and $2p$ are displayed in Table III. In order to make a rigorous comparison, we have added the $\bar{E}_{1s}(x_0)$ values re-

TABLE V. $1s$, $2s$, and $2p$ level energies calculated from the coth z method.

x_0	$\bar{E}_{1s}$	x_0	$\bar{E}_{2s}$	x_0	$\bar{E}_{2p}$
1.9020	− 0.0556	6.4290	− 0.0200	5.3550	− 0.0201
1.9340	− 0.0800	6.6110	− 0.0312	5.5280	− 0.0312
2.0000	− 0.1250	6.7850	− 0.0407	5.6960	− 0.0408
2.1780	− 0.2223	6.8240	− 0.0427	6.0000	− 0.0555
2.2000	− 0.2322	7.0960	− 0.0553	6.4970	− 0.0739
2.4720	− 0.3279	7.2120	− 0.0600	6.7010	− 0.0799
2.6000	− 0.3595	7.6770	− 0.0758	7.0890	− 0.0893
2.8070	− 0.3985	7.8150	− 0.0797	8.5400	− 0.1100
2.8130	− 0.3995	8.8520	− 0.1005	10.2350	− 0.1196
3.0413	− 0.4295	10.3600	− 0.1151	12.3190	− 0.1236
3.1920	− 0.4442	12.3862	− 0.1223	16.0056	− 0.1249
5.0200	− 0.4972	16.0118	− 0.1248	25.000	− 0.1250
6.1270	− 0.4997	20.2502	− 0.1250		
8.0000	− 0.5000				

TABLE VI. Perturbational calculation of the $1s$ energy level up to convergence.

x_0	0.808 900 8	0.909 250 5	1.009 354 2	1.079 469 1	1.149 620 8	1.447 930 7	1.710 959 3	1.778 018 7
$\bar{E}$	4.408 023 6	3.166 273 8	2.305 397 0	1.852 389 9	1.488 021 7	0.540 022 80	0.126 290 7	0.054 093 298
	1.835 246 4	1.901 923 8	1.934 345 6	2.000 000 0	2.178 319	2.200 577 8	2.472 625 6	
	0.000 000 0	− 0.005 555 555 4	− 0.080 006 158	− 0.125 000 0	− 0.222 254 84	− 0.232 288 71	− 0.327 582 88	
	2.601 147 1							
	− 0.359 231 68							

ported by Gray and Gonda[12] and Ludeña.[14] Gray and Gonda performed the calculation with the polynomial of Ref. 11, which gives the ground state energy corrected up to the fifth order. On the other hand, Ludeña made a diagonalization of the Hamiltonian in a basis set of Slater-type functions which are properly adapted to the BC of the problem. Both methods agree completely in their results for the whole interval of x_0 values, and naturally we can consider that they are exact. A perusal of the third, fourth, fifth, and sixth columns ($n = 1, l = 0$) shows that our values are better than those of Ref. 2, in nearly the complete range of x_0 values studied. DeGroot and Ten Seldam's eigenvalues, calculated from the confluent hypergeometric equation with the help of the method described in Ref. 21, present appreciable errors for the ground state in the interval of small x_0 values. This behavior has been pointed out before by Gray and Gonda.[12] The eigenvalues $\bar{E}_{2s}$ and $\bar{E}_{2p}$ obtained by DeGroot and Ten Seldam are probably subject to marked errors, which precludes using them for comparison with ours.

From the zeros of the eigenfunctions Ψ_{2s}, Ψ_{3s}, and Ψ_{3p} corresponding to the free hydrogen atom, we can deduce the following exact values:

$$\bar{E}_{1s}(1.9019) = -0.0556,$$
$$\bar{E}_{1s}(2.0000) = -0.1250,$$
$$\bar{E}_{2p}(6.0000) = -0.0556.$$

Our values are nearer to these latter values than those reported in Ref. 2. In order to confirm that, we show in Table IV the $\bar{E}_{2p}$ results for different spherical radii calculated via Killingbeck's method[59] with $h = 0.01$. Again we can see that our results are better than those displayed in Ref. 2.

It must be taken into account that the error of Eq. (63) increases with increasing q_0 values, that is to say, when x_0 is very small, or when it is large. For these two extreme cases, Eq. (63) is not credible. The seventh column in Table III, corresponding to the $1s$ level, contains the eigenvalue $\bar{E}_{1s}(x_0)$ corrected in a perturbational way up to the second order, considering the potential term Zr^{-1} as a perturbation.[4,5,10–12] The energy calculated in this manner is subject to large errors, even though x_0 is relatively small. Then, we follow that it is wholly convenient to make the transformation proposed by Dingle[3] before applying the perturbation theory.

From the preceding discussion, we conclude that Eq. (63) yields eigenvalues for the bounded hydrogen atom with an acceptable exactness and for a relatively large range of x_0 values. However, the main significance of Eq. (63) is that it is the first non-numerical expression presented in the current literature that enables us to calculate any eigenvalue for the bounded hydrogen atom in a wide interval of x_0 values. Although Gray[10] has shown how to obtain the different perturbation corrections in a closed form, he did not obtain a unique non-numerical formula for all the levels. When x_0 increases, q_0 also increases, and this establishes failure of Eq. (63) conclusively. Nevertheless, such equation renders another useful service because it constitutes a solid starting point for constructing a valid expression in the whole range where $\bar{E} \leqslant 0$. Such an expression can be obtained, for example, from the method proposed by Vawter[39] and recently discussed by the present authors in relation with the hypervirial theorems.[46] The method consists of representing W_n by means of the function

$$W_n = 2n \coth F(q_0^2), \tag{64}$$

$$F(u) = \sum_{j=0}^{\infty} c_j u^{2j+1}. \tag{65}$$

Thc coefficients c_j are chosen in such a way that the expansion of (64) in q_0 power series must reproduce the perturbational polynomial for W_n. In this work we keep only the first two coefficients of $F(u)$, so that we can only adjust (63) up to the second order in q_0 (first perturbation order). In Table V we present results for $\bar{E}_{1s}(x_0)$, $\bar{E}_{2s}(x_0)$, and $\bar{E}_{2p}(x_0)$. It is immediately clear that this actual procedure gives results practically identical to those obtained from Eq. (63) in the interval of validity of the perturbational polynomial. But, in addition, the correct result is obtained when x_0.

The coth z-method was initially developed for analyz-

TABLE VII. Perturbational calculation of the $2s$ energy level up to convergence.

x_0	1.683 029 2	1.862 196 8	1.982 322 5	2.111 788 5	2.234 504 4	2.366 536 1	2.658 205 4	2.910 305 2
$\bar{E}$	5.053 685 7	3.963 941 0	3.399 330 4	2.900 678 6	2.510 227 5	2.160 331 0	1.575 908 0	1.216 268 6
	3.613 722 9	3.828 447 9	5.112 434 0	5.589 240 5	5.808 080 1	6.152 307 1	6.428 996 2	
	0.611 369 19	0.496 820 44	0.121 657 05	0.054 614 547	0.030 795 925	0.000 000 0	− 0.019 993 993	
	6.610 837 1	6.784 500 4	6.823 331 4	7.098 076 3	7.209 441 6	7.669 240 5		
	− 0.031 251 306	− 0.040 810 217	− 0.042 802 871	− 0.055 555 555	− 0.060 116 537	− 0.758 866 65		
	7.804 823 7	8.806 340 8	10.188 698					
	− 0.079 735 049	− 0.100 123 54	− 0.114 101 42					

TABLE VIII. Perturbational calculation of the $2p$ energy level up to convergence.

x_0	1.421 150 8	1.559 730 4	1.688 819 4	2.531 786 7	2.700 862 0	4.110 174 9
$\bar{E}$	3.674 365 4	2.941 048 5	2.421 597 0	0.820 933 46	0.675 494 02	0.122 416 72
	4.553 991 8	4.770 030 7	5.088 308 2	5.355 431 7	5.527 960 8	5.695 897 8
	0.054 919 274	0.029 939 697	0.000 000 0	− 0.020 163 585	− 0.031 255 843	− 0.040 824 827
	6.000 000 0	6.496 559 5	6.700 280 9	7.087 658 0	8.533 822 4	10.222 783 0
	− 0.055 555 555	− 0.073 980 894	− 0.079 970 161	− 0.089 419 86	− 0.110 079 68	− 0.119 644 81

ing the unidimensional harmonic oscillator symmetrically bounded.[39,46] Then it was applied to multidimensional oscillators enclosed within spheres with impenetrable walls.[47] In this work we have extended successfully its application to the bounded hydrogen atom model. Equations (15) and (16) can be programmed easily, and consequently it is possible to arrive at the desired number of perturbative corrections without any additional difficulty. Tables VI–VIII display the energies for the first three levels and for a considerable number of x_0 values. In each case, the number of perturbative terms necessary to reach constancy in the last numerical digit was added. A comparison with the exact values

$$\bar{E}_{1s}(1.901023789) = -\tfrac{1}{18},$$
$$\bar{E}_{1s}(2.000000000) = -\tfrac{1}{8},$$
$$\bar{E}_{2s}(7.098076211) = -\tfrac{1}{18},$$
$$\bar{E}_{2p}(6.000000000) = -\tfrac{1}{18}$$

allows us to be sure that our results are the most accurate existing up to now in the current literature. The method permits the calculation, without further difficulties, of the expectation values of the x powers, because

$$\langle f|x^{(N-2)/2}f\rangle = 2A^N/(8|\bar{E}|)^{(N+1)/4}, \tag{66}$$

$$\frac{\langle f|x^{(N-2)/2}f\rangle}{\langle f|f\rangle} = \langle x^{(N-2)/2}\rangle = 2^{2-N}\frac{A^N}{A^2}(W/Z)^{(N-2)/2}. \tag{67}$$

As a particular case, for $N=0$ we obtain $\langle 1/x\rangle$,

$$\langle 1/x\rangle = 4Z/A^2W. \tag{68}$$

This last formula is very useful for determining $\partial\bar{E}/\partial x_0$,

$$-x_0\frac{\partial\bar{E}}{\partial x_0} = Z\langle 1/x\rangle + 2\bar{E} = \frac{4Z^2}{W}\left\{\frac{1}{A^2}\mp\frac{1}{W}\right\}. \tag{69}$$

Obviously

$$-x_0\frac{\partial\bar{E}}{\partial x_0}\bigg|_{q_0=0} = 4Z^2/wA_0^2(1). \tag{70}$$

Equations (68)–(70) allow us to calculate the whole set of quantities shown in Table I and Fig. 1, and consequently, the physical properties mentioned in Sec. III.

V. FURTHER DISCUSSION

In this communication we have presented several useful original contributions from a theoretical as well as a practical viewpoint. The method developed in Sec. II,owing to its great generality, permits us to treat a large number of problems with a marked actual physical interest. It has already been applied with success in various cases,[43–48] and it remains under study in our laboratory. As pointed out earlier, this method is preferable to any other when it can be used. The Rayleigh–Schrödinger perturbational theory and the formalism of the Green's function[50] do not lead to so simple and general a formulation as that offered in this work.[48] Our method is equally useful when numerical and/or analytical results are desired. The direct application made to the hydrogen atom model in Sec. III, shows that Eqs. (15) and (16) are of no value in this example. The same occurs for any potential with $k<0$ [see Eq. (9)]. However, hypervirial relationships are intrinsically very useful if one knows $\bar{E}(x_0)$, because there exist several physical properties closely related to the moments $\langle r^N\rangle$. For example, the expectation values $\langle 1/r\rangle$ and $\langle r^2\rangle^2$ calculated in Sec. III are proportional to the diamagnetic screening constant and polarizability, respectively.[2,6,13,16,21] Besides, in Sec. III we briefly displayed a very simple approximate method for determining $\bar{E}(x_0)$ within an interval of x_0^0 whenever one knows in advance $\langle T\rangle(x_0^0)$ and $\langle 1/r\rangle(x_0^0)$. When the kinetic and potential energies are known for a finite set of x_0 values, such a method enables us to obtain $\bar{E}(x_0)$ for in-between points whenever the distance between a pair of points is small.

Undoubtedly, the most significant contribution in this communication is given in Sec. IV. The substitution proposed by Dingle[3] and recently discussed at length by Rowley,[58] transforms the bounded hydrogen atom into a bidimensional bounded harmonic oscillator, and it allows us to solve the problem by means of Eqs. (15) and (16). The two non-numerical expressions discussed in Sec. IV are the sole formulas valid for any energy level that have been published up to now. Besides, the ease of programming Eqs. (15) and (16) has enabled us to obtain excellent numerical results.

It is necessary to recognize that our method is not restricted to the DBC, but is applicable to more general BC.[43–45] Ley-Koo and Rubinstein[16] proposed the utilization of enclosed atoms within boxes with permeable walls for studying certain physical problems. Using that proposal, we have recently deduced the proper formulation of the VT when sectionally-continuous potentials are used.[60]

Our actual interest is directed toward the deduction of appropriate hypervirial relationships and combining them with the perturbation theory for analyzing similar, but more complex, systems of physical interest, in a way similar to that used in this work. Results will be given in a forthcoming paper, which will be published elsewhere.

[1] V. A. Sommerfeld and H. Welker, Ann. Phys. (Leipzig) **32**, 56 (1938).
[2] S. R. De Groot and C. A. Ten Seldam, Physica (Utrecht) **12**, 669 (1946).
[3] R. D. Dingle, Proc. Cambridge Philos. Soc. **49**, 103 (1953).
[4] E. P. Wigner, Phys. Rev. **94**, 77 (1954).

[5]R. E. Trees, Phys. Rev. **102**, 1553 (1956).
[6]A. D. Buckingham and P. Plawley, Mol. Phys. **3**, 219 (1960).
[7]K. K. Singh, Proc. Natl. Inst. Sci. India Sect. A **27**, 86 (1961).
[8]B. F. Gray, J. Chem. Phys. **36**, 1801 (1962).
[9]B. M. Gimarc, J. Chem. Phys. **44**, 373 (1966).
[10]B. F. Gray, J. Chem. Phys. **55**, 2848 (1971).
[11]V. C. Aguilera-Navarro, W. M. Kloet, and A. H. Zimerman, Rev. Bras. Fis. **1**, 55 (1971).
[12]B. F. Gray and I. Gonda, J. Chem. Phys. **62**, 2007 (1974).
[13]D. Suryanarayana and J. A. Weil, J. Chem. Phys. **64**, 510 (1976).
[14]E. V. Ludeña, J. Chem. Phys. **66**, 468 (1977).
[15]M. Friedman and Y. Rosenfeld, J. Comp. Phys. **26**, 169 (1978).
[16]E. Ley-Koo and S. Rubinstein, J. Chem. Phys. **71**, 351 (1979).
[17]J. A. Weil, J. Chem. Phys. **71**, 2803 (1979).
[18]M. Friedman, J. Comp. Phys. **33**, 359 (1979).
[19]E. Ley-Koo and S. Rubinstein, J. Chem. Phys. **73**, 887 (1980).
[20]E. Ley-Koo and S. A. Cruz, J. Chem. Phys. **74**, 4603 (1981).
[21]A. Michels, J. De Boer, and B. Bijl, Physica (Utrecht) **4**, 981 (1937).
[22]A. Michels and S. R. De Groot, Physica (Utrecht) **16**, 183 (1950).
[23]T. L. Cottrell, Trans. Faraday Soc. **47**, 337 (1951).
[24]C. A. Ten Seldam and S. R. De Groot, Physica (Utrecht) **18**, 891, 905 (1952).
[25]K. K. Singh, Physica (Utrecht) **30**, 211 (1964).
[26]B. M. Gimarc, J. Chem. Phys. **47**, 5110 (1967).
[27]E. V. Ludeña, J. Chem. Phys. **69**, 1770 (1978).
[28]E. V. Ludeña and M. Gregori, J. Chem. Phys. **71**, 2235 (1979).
[29]F. C. Auluck, Proc. Natl. Inst. Sci. India **7**, 133 (1941).
[30]F. C. Auluck, Proc. Natl. Inst. Sci. India **8**, 147 (1942).
[31]S. Chandrasekhar, Astrophys. J. **97**, 263 (1943).
[32]F. C. Auluck and D. S. Kothari, Proc. Cambridge Philos. Soc. **41**, 175 (1945).
[33]B. Suryan, Phys. Rev. **71**, 741 (1947).
[34]E. M. Corson and I. Kaplan, Phys. Rev. **71**, 130 (1947).
[35]R. B. Dingle, Proc. Phys. Soc. London, Sect. A **212**, 47 (1952).
[36]J. S. Baijal and K. K. Singh, Prog. Theor. Phys. **14**, 214 (1955).
[37]P. Dean, Proc. Cambridge Philos. Soc. **62**, 277 (1966).
[38]R. Vawter, Phys. Rev. **174**, 749 (1968).
[39]R. Vawter, J. Math. Phys. **14**, 1864 (1973).
[40]A. Consortini and B. R. Frieden, Nuovo Cimento, B **35**, 153 (1976).
[41]F. C. Rotbar, J. Phys. A: Math. Gen. **11**, 2363 (1978).
[42]V. C. Aguilera-Navarro, E. Ley-Koo, and A. H. Zimerman, J. Phys. A: Math. Gen. **13**, 3585 (1980).
[43]F. M. Fernández and E. A. Castro, Int. J. Quantum Chem. **19**, 521 (1981).
[44]F. M. Fernández and E. A. Castro. Int. J. Quantum Chem. **19**, 533 (1981).
[45]F. M. Fernández and E. A. Castro, Int. J. Quantum Chem. **20**, 623 (1981).
[46]F. M. Fernández and E. A. Castro, J. Math. Phys. **22**, 2669 (1981).
[47]F. M. Fernández and E. A. Castro, Phys. Rev. A **24**, 2883 (1981).
[48]F. M. Fernández and E. A. Castro, Physica A (in press).
[49]A. Rabinovitch and J. Zack, Phys. Rev. B **4**, 2358 (1971).
[50]T. Luckes, G. A. Ringwood, and B. Suprapto, Physica A **84**, 421 (1976).
[51]I. Gonda and B. F. Gray, J. Chem. Soc. Faraday Trans. II **71**, 2016 (1975).
[52]J. O. Hirschfelder, J. Chem. Phys. **33**, 1462 (1960).
[53]T. L. Cottrell and S. Paterson, Philos. Mag. **42**, 391 (1951).
[54]W. Byers Brown, J. Chem. Phys. **28**, 522 (1958).
[55]R. J. Swenson and S. H. Danforth, J. Chem. Phys. **57**, 1734 (1972).
[56]J. Killingbeck, Phys. Lett. A **65**, 87 (1978).
[57]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., London, 1966), 2nd ed.
[58]N. Rowley, J. Phys. A: Math. Gen. **12**, L7 (1979).
[59]J. Killingbeck, J. Phys. A: Math. Gen. **10**, L99 (1977).
[60]F. M. Fernández and E. A. Castro, J. Chem. Phys. **75**, 2908 (1981).

Variational principles for resonances. II

A. G. Ramm[a)]
Department of Mathematics, Kansas State University, Manhattan, Kansas 66506

(Received 23 June 1981; accepted for publication 7 August 1981)

Variational principles for calculating the complex poles of Green's function are given. Convergence of the numerical procedure is proved.

PACS numbers: 03.65.Nk

I. INTRODUCTION

This note is a continuation of Ref. 1, where the following problem was considered:

$$(-\nabla^2 - k^2)u = 0 \quad \text{in} \quad \Omega, \; u|_\Gamma = 0. \tag{1}$$

Here Ω is an exterior domain, Γ is its closed smooth boundary, and $D = \mathbb{R}^3 \setminus \Omega$ is bounded. Problem (1) has nontrivial solutions if and only if ($\equiv$iff) k is a complex pole k_q of the Green's function $G(x,y,k)$ of the exterior Dirichlet problem. In Ref. 1 a stationary variational principle for resonances, i.e., complex poles k_q, was given

$$k^2 = \mathrm{st}\{\langle \nabla u, \nabla u\rangle / \langle u,u\rangle\}, \tag{2}$$

where st is the symbol of stationary value,

$$\langle u,v\rangle = \lim_{\epsilon \to +0} \int \exp(-\epsilon r \ln r) u(x) v(x)\, dx,$$

$$\int = \int_\Omega, \quad r = |x|. \tag{3}$$

In Ref. 1 the test functions for (2) were taken in the form

$$u_N = r^{-1}\exp(ikr) \sum_{j=0}^{N} \sum_{|m| \leqslant j} r^{-j} Y_{jm}(n) c_{jm} g(x), \tag{4}$$

where $n = x|x|^{-1}$, Y_{jm} are the spherical harmonics, c_{jm} are constants, k is a parameter, and $g(x) \geqslant 0$ is a fixed smooth function vanishing on Γ and equal to 1 outside of some ball containing D. It was not proved in Ref. 1 that the numerical procedure suggested there converges. The question formulated in Ref. 1 concerning the justification of the numerical approach is still open. The purpose of this note is to formulate another variational principle for calculating the complex poles k_q and to prove the convergence of the numerical procedure. The method in Ref. 1 is similar to Ritz's method. The method suggested in this note is similar to Trefftz's method. The advantage of this method is that one deals with the compact operators, while in Ref. 1 the operator was not compact. Our construction is natural in the framework of the singularity and eigenmode expansion methods.[2] The convergence of the method will be proved. A result which is of general interest, as it seems to the author, is a construction of a stationary variational principle and a proof of convergence for a class of non-self-adjoint symmetric operators ($B^* = \overline{B}$), which occur frequently in the scattering theory.

[a)]Supported by AFOSR 80024; AMS subject classification 47A10, 78A45, 81F05, 35J05.

II. A VARIATIONAL PRINCIPLE

The starting point is the following observation: k is a complex pole of $G(x,y,k)$ iff the equation

$$Af \equiv \int_\Gamma g(s,t,k) f(t)\, dt = 0, \quad \mathrm{Im} k < 0,$$

$$g(s,t,k) = \exp(ik|s-t|)/(4\pi|s-t|), \tag{5}$$

has a nontrivial solution. This observation and some consequences are discussed in Ref. 3. For the convenience of the reader let us note that

$$G = g - \int_\Gamma g(x,t,z) \frac{\partial G(t,y,z)}{\partial N_t}\, dt, \tag{6}$$

where N_t is the unit outer normal to Γ at the point t. If k is a complex pole of G of order r one can multiply (6) by $(z-k)^r$ and take $z \to k$ and $x = s \in \Gamma$. This yields Eq. (5) (see Ref. 3, pp. 290–291) with $f \not\equiv 0$.

Let us formulate the following variational principle

$$F(f) \equiv |Af|_1^2 = \min, \quad \|f\| = 1, \tag{7}$$

where $|f|_p$ is the norm in the Sobolev space $H_p = W_2^p(\Gamma)$, $\|f\| = |f|_0$. From the above observation it follows that (7) has solutions and the min is zero if $k = k_q$, where k_q are the poles of $G(x,y,k)$. If $k \neq k_q$ then $\inf_{\|f\|=1} |Af|_1 > 0$. Indeed, if there exists a sequence $\|f_n\| = 1$, $|Af_n|_1 \to 0$, then $f_n \to f$, $\|f\| = 1$, $Af = 0$, and therefore $k = k_q$ (see Ref. 3, p. 291). The only point which is to be explained is the convergence in H: $f_n \to f$. In Ref. 3 it is explained that A is a pseudodifferential operator of order -1, that is,

$$a_1 |f|_{p-1} \leqslant |Af|_p \leqslant a_2 |f|_{p-1}. \tag{8}$$

Here $a_1, a_2 > 0$ are some constants, $-\infty < p < \infty$ if $\Gamma \subset C^\infty$, and the fact that $k \neq k_q$ was used essentially: if $k \neq k_q$ then $\ker A \equiv \{f: Af = 0\} = \{0\}$ and A maps H_p onto H_{p+1}. If $|Af_n|_1 \to 0$ and $\|f_n\| = 1$, then (8) with $p = 1$ shows that $\|f_n\| \to 0$. This contradicts the equation $\|f_n\| = 1$. Therefore

$$\inf_{\|f\|=1} |Af|_1 > 0 \quad \text{if} \quad k \neq k_q. \tag{9}$$

Consider a numerical method for solving problem (7). Let $\{f_j\}$ be a basis of H,

$$f = f^{(n)} \equiv \sum_{j=1}^{n} c_j f_j. \tag{10}$$

The necessary condition for $F(f)$ to be minimal and $\min F(f^{(n)}) = 0$, $\|f^{(n)}\| = 1$, yields:

$$\sum_{m=1}^{n} a_{jm} c_m = 0 \quad 1 \leqslant j \leqslant n, \tag{11}$$

 0022-2488/82/061112-03$02.50

$$a_{jm} = a_{jm}(k) = (Af_m, Af_j)_1 \quad \sum_{j=1}^{n} |c_j|^2 > 0. \tag{12}$$

Thus

$$\det a_{jm}(k) = 0 \quad 1 \leqslant j,m \leqslant n. \tag{13}$$

Let $k_q^{(n)}$ denote the roots of Eq. (13). Our first result is

Theorem 1: There exists $\lim_{n\to\infty} k_q^{(n)} = k_q$, and k_q are the poles of Green's function $G(x,y,k)$. Every pole k_q is a limit of a sequence $k_q^{(n)}$, where $k_q^{(n)}$ are the roots of (13). Convergence is uniform in q for any finite interval $1 \leqslant q \leqslant Q$.

Proof: We will prove that: (i) Eq. (13) has roots in the circle $|k - k_q| < \epsilon$ for any fixed $\epsilon > 0$ however small if $n > n(\epsilon)$ is large enough. (ii) If $n > n(\epsilon)$ and there are no points k_q in the circle $|k - z| < \epsilon$ then Eq. (13) has no roots in the circle $|k - z| < \epsilon$. An important part of the proof is the reduction of the problem to the problem with the operator $I + T(k)$, where $T(k)$ is compact.

Let us fix $\epsilon > 0$ such that in the circle $\{k: |k - k_q| < \epsilon\} = K_\epsilon$ there are no other poles. The operator $A = A(k)$ can be written as

$$A(k) = A_0[I + T(k)], \tag{14}$$

where

$$A_0 = A(0), \quad A_0^* = A_0 > 0 \quad \text{in} \quad H = L^2(\Gamma),$$

$$A_0 f = \int_\Gamma \frac{f\,dt}{4\pi r_{st}}, \tag{15}$$

and

$$T(k) = A_0^{-1}[A(k) - A_0]. \tag{16}$$

The operator A_0 is a bijection of H_p onto H_{p+1};

$$b_1 |f|_0 \leqslant |A_0 f|_1 \leqslant b_2 |f|_0, b_1, b_2 = \text{const} > 0, \tag{17}$$

while $T(k)$ is compact as a map $H_p \to H_p$ (see Ref. 3 for details) because $A(k) - A_0$ is an operator with a nonsingular kernel. Let us rewrite functional (7) as

$$F(f) = |A_0(I + T)f|_1^2 = \min, \quad \|f\| = 1. \tag{18}$$

From (18) and (17) it follows that the problem (7) is equivalent to

$$F_0(f) = |(I + T)f|_0^2 = \min, \quad \|f\| = 1. \tag{19}$$

The matrix of the system (11) can be written as

$$a_{jm} = ((I + T)f_m, (I + T)f_j), \tag{20}$$

where $(.,.)$ denotes the scalar product which is metrically equivalent to the scalar product in H. This means that $d_1(f,f)_0 \leqslant (f,f) \leqslant d_2(f,f)_0$, where $d_1 > 0$ and d_2 are constants, $f \in H$ is arbitrary. In the sequel we will not discriminate between $(.,.)$ and $(.,.)_0$. This is possible because $((I + T)f, (I + T)f)$ and $((I + T)f, (I + T)f)_0$ attain their zero values simultaneously. The system (11) can be considered as the system which corresponds to the Ritz method for functional (19) with the test functions $\{f_j\}$. This completes the reduction of the original problem to the problem with the operator $I + T(k)$, where $T(k)$ is a compact analytic-in-k operator function on H. To prove (i) let us assume that for a fixed $\epsilon > 0$ and k_q and all n there are no roots $k_q^{(n)}$ of Eq. (13) in the circle $|k - k_q| < \epsilon$. The system (11) with the matrix (20) says that

$$((I + T)f^{(n)}, (I + T)f_j) = 0 \quad 1 \leqslant j \leqslant n, f^{(n)} \neq 0, \tag{21}$$

where

$$f^{(n)} = \sum_{j=1}^{n} c_j f_j. \tag{22}$$

In particular, our assumption means that
$(\tilde{T} \equiv T + T^* + T^*T, I + \tilde{T} = (I + T^*)(I + T))$,

$$(I + P_n \tilde{T}(k))f^{(n)} = 0 \Rightarrow f^{(n)} \equiv 0, \quad |k - k_q| < \epsilon, \tag{23}$$

where P_n denotes the projection in H onto the linear span of $\{f_1, \ldots, f_n\}$. Equation (23) says that $I + P_n \tilde{T}(k)$ is invertible in the circle $|k - k_q| < \epsilon$. If n is large enough this implies that $I + \tilde{T}(k)$ is invertible in the circle $|k - k_q| < \epsilon$, because (*) $\|I + \tilde{T}(k) - (I + P_n \tilde{T}(k))\| \to 0$ as $n \to \infty$. This is a contradiction since $I + \tilde{T}(k_q)$ is not invertible. Let us explain (*). We need to show that $\|(I - P_n)\tilde{T}\| \to 0$ as $n \to \infty$. Since $\tilde{T}(k)$ is compact it can be written as $T_N + B_N$, where $\|B_N\| < d_N$, $d_N \to 0$ as $N \to \infty$, and T_N is a finite-dimensional operator. It is sufficient to prove that $\|(I - P_n)T_N\| \to 0$ as $n \to \infty$. Without loss of generality one can assume that T_N is a one-dimensional operator, $T_N f = (f,v)u$. Then

$$\|(I - P_n)T_N f\| = \|(I - P_n)u\| |(f,v)| \leqslant \|f\| \, \|v\| \, \|u - P_n u\| \to 0 \quad \text{as} \quad n \to \infty, \tag{24}$$

since $P_n \to I$ strongly. Thus the statement (i) is proved. Note that the orthogonality of P_n is not used in (24). In order to prove (ii) we suppose that for any $\epsilon_n > 0$, $\epsilon_n \to 0$, Eq. (13) has a root $k^{(n)}$ in the circle $|k - z| < \epsilon_n$ and show that under this assumption z has to be a pole of the Green's function. The assumption means that

$$[I + T(k^{(n)})]f^{(n)} = 0, \quad \|f^{(n)}\| = 1, \quad k^{(n)} \to z. \tag{25}$$

Since $\|f^{(n)}\| = 1$, one can extract a weakly convergent in H subsequence which is denoted again $f^{(n)}: f^{(n)} \to f$ ($\to$ means weak convergence). Since $T(k)$ is compact the sequence $T(z)f^{(n)}$ converges strongly in H:

$$T(z)f^{(n)} \to T(k)f. \tag{26}$$

On the other hand,

$$\|T(k_n) - T(z)\| \to 0. \tag{27}$$

From (25)–(27) it follows that

$$f^{(n)} \to f, \quad \|f\| = 1, \tag{28}$$

and

$$[I + T(z)]f = 0, \quad \|f\| = 1. \tag{29}$$

The proof is complete.

III. DISCUSSION

The variational principles (19) and (18) can be viewed as the least square method. Let us consider instead of (13) and (20) the following equation:

$$\det b_{jm}(k) = 0, 1 \leqslant j,m \leqslant n, \quad b_{jm} \equiv ((I + T(k))f_m, f_j). \tag{30}$$

Arguments similar to the ones given in Ref. 3, pp. 192–193 show that: (i) For any $\epsilon > 0$ and k_q there exists a root $\tilde{k}_q^{(n)}$ of Eq. (30). such that $|k_q - \tilde{k}_q^{(n)}| < \epsilon$ if $n > n(\epsilon)$. (ii) If $\tilde{k}_q^{(n)}$ is a sequence of the roots of Eq. (30) and $\tilde{k}_q^{(n)} \to k_q$ as $n \to \infty$, then k_q is a pole of the Green's function. Equation (30) can be viewed as a necessary condition for the linear system of the

Galerkin method for the equation $(I + T(k))f = 0$ to have a nontrivial solution. The Galerkin equation is of the form

$$(f^{(n)} + T(k)f^{(n)}, f_j) = 0 \quad 1 \leqslant j \leqslant n, \tag{31}$$

where $f^{(n)}$ is defined in (22). The basic idea is that the poles k_q are the points at which the operator $I + T(k)$ is not invertible. These points can be found by the Galerkin method, by minimizing functional (19) or by some other method. It is interesting to note that the Galerkin equation (31) can be obtained also as a necessary condition for the stationary variational principle

$$((I + T(k))f, f) = \text{st}, \quad \|f\| > 0, \tag{32}$$

where st means stationary value. This is not true for an arbitrary operator, but the operator $B \equiv I + T(k)$ is a symmetric non-self-adjoint operator on $H = L^2(\Gamma)$, that is

$$\overline{B^*} = B \quad \text{or} \quad B(s,t) = B(t,s)\left[\neq \overline{B(t,s)}\right]. \tag{33}$$

Therefore the necessary condition for (32), which can be written as

$$(Bf, h) + (Bh, f) = 0 \quad \text{for all} \quad h \in H, \tag{34}$$

yields

$$0 = (Bf, h) + (\overline{B^* f}, \bar{h}) = (Bf, h) + (B\bar{f}, \bar{h}). \tag{35}$$

Let $h = v$, where $v \in H$ is an arbitrary real-valued function. Then (35) says that

$$0 = B(f + \bar{f}). \tag{36}$$

Let $h = iv$. Then (35) says that

$$0 = B(f - \bar{f}). \tag{37}$$

From (36) and (37) it follows that the equation

$$Bf = (I + T(k))f = 0, \quad \|f\| > 0 \tag{38}$$

is a necessary condition for (32).

Our aim is to show that Eq. (31) is a necessary condition for the problem

$$(Bf, f) = \text{st}, \quad \|f\| > 0. \tag{39}$$

Let us take $f = f^{(n)}$ and rewrite (39) as

$$\sum_{j,m=1}^{n} b_{jm} c_m \bar{c}_j = \text{st}, \quad b_{jm} = (Bf_m, f_j). \tag{40}$$

In general assumption (33) does not imply the equality $b_{jm} = b_{mj}$. Therefore the following lemma is of use.

Lemma 1: Assume (33) and

$$f_j = \bar{f}_j, \quad j = 1,2,\cdots. \tag{41}$$

Then

$$b_{jm} = b_{mj}. \tag{42}$$

The proof is immediate.

Proposition 1: Assume (33) and (41). Then a necessary condition for (40) is the system (31).

Proof: The operator $B = I + T(k)$ satisfies (33). From this and Lemma 1, Proposition 1 follows.

Remark 1: The results of Sec. III give a convergent numerical scheme for a stationary variational principle (32) with a compact operator T satisfying condition (33), i.e., symmetric non-self-adjointness. Such operators occur frequently in the scattering theory. A simple example is problem (1). There are other examples in Ref. 4.

Remark 2: A numerical scheme for calculating the resonances based on theorem 1 is as follows: (1) Calculate matrix a_{jm} by formula (11). (2) Find roots of Eq. (13). The corresponding solutions of (1) can also be calculated by this numerical procedure: Find $f^{(n)}$ by formula (10) and $u^{(n)} = Af^{(n)}$ is the approximate solution of (1), which converges to the exact solution of (1) as $n \to \infty$. This exact solution is of the form $u = A(k_q)f, f = \lim f^{(n)}$ as $n \to \infty$ and lim here means the limit in $H = L^2(\Gamma)$.

Remark 3: For numerical calculations instead of principle (32) one should use the equivalent principle

$$(A(k)f, f) = \text{st}, \quad \|f\| > 0. \tag{43}$$

The equivalence of (43) and (32) follows from the fact that the necessary condition for (43) is the equation

$$A(k)f = A_0(I + T(k))f = 0, \quad \|f\| > 0, \tag{44}$$

which is equivalent to the necessary condition (38) for (32) because $\ker A_0 = \{0\}$. If one takes $f = f^{(n)}$ as in (22), then the analog of (30) is

$$\det(A(k)f_m, f_j) = 0 \quad 1 \leqslant m, j \leqslant n, \tag{45}$$

and the convergence of the numerical procedure follows from the arguments given for Eq. (30).

[1] A. G. Ramm, "A variational principle for resonances," J. Math. Phys. **21**, 2052 (1980).

[2] A. G. Ramm, "Theoretical and practical aspects of the singularity and eigenmode expansion methods," IEEE Trans. Antennas and Propag. **AP-28**, 897 (1980).

[3] A. G. Ramm, *Theory and applications of some new classes of integral equations* (Springer, New York, 1980).

[4] G. Hönl, A. Maue, and K. Westpfahl, *Theorie der Beugung* (Springer, Berlin, 1961).

Scattering by periodic surfaces

A. W. Sáenz [a]
Naval Research Laboratory, Washington, D. C. 20375

(Received 6 October 1981; accepted for publication 29 January 1982)

If $\Omega \subset \mathbf{R}^\nu (\nu \geqslant 2)$ is an exterior domain containing a half-space and contained in a half space, we prove that the wave operators $W_\pm = \text{s-lim}_{t \to \pm\infty} \exp(itH) J \exp(-itH_0)$ are partial isometries and that the invariance principle $W_\pm(\phi) = W_\pm$ holds for suitable real functions ϕ on $\mathbb{R}$ ("admissible" functions). Here H_0 is the negative (distributional) Laplacian in $L^2(\mathbf{R}^\nu)(\nu \geqslant 2)$; H is H_D or H_N, the negative Dirichlet or Neumann Laplacians in $\mathscr{H} = L^2(\Omega)$, respectively; J is an appropriate identification operator; and $W_\pm(\phi)$ are defined as were $W_\pm$, but with H_0 and H replaced by $\phi(H_0)$ and $\phi(H)$, respectively. Suppose, in addition, that Ω has a suitable periodicity property and that when $H = H_N$ it has a certain mild local compactness property. Then we prove: (1) that $W_\pm$ are asymptotically complete, in the sense that $\text{Ran } W_\pm = \mathscr{H}_{\text{scatt}}(H) = \mathscr{H} \ominus \mathscr{H}_{\text{surf}}(H)$; (2) that $\mathscr{H}_{\text{scatt}}(\phi(H)) = \mathscr{H}_{\text{scatt}}(H)$ and $\mathscr{H}_{\text{surf}}(\phi(H)) = \mathscr{H}_{\text{surf}}(H)$ for each "admissible" function ϕ, and hence that $W_\pm(\phi)$ are asymptotically complete in a similar sense for each such ϕ. Here, for any self-adjoint operator A in $\mathscr{H}$, $\mathscr{H}_{\text{scatt}}(A)$ and $\mathscr{H}_{\text{surf}}(A)$ are naturally defined, mutually orthogonal subspaces of scattering and surface states of $\mathscr{H}$, respectively. No smoothness assumptions on $\partial\Omega$ are made in this paper. Its results entail the asymptotic completeness, in a physically very satisfactory sense, of wave operators describing acoustic and certain types of electromagnetic scattering by a very wide class of periodic surfaces.

PACS numbers: 03.65.Nk

1. INTRODUCTION

In a previous paper,[1] we established the existence, partial isometry, and asymptotic completeness (in a suitable sense) of the wave operators $W_\pm = \text{s-lim}_{t \to \pm\infty} \exp(itH) \times J \exp(-itH_0)$ describing quantum-mechanical scattering in an exterior domain Ω bounded by an impenetrable periodic surface of very general type. Here H_0 is the negative (distributional) Laplacian in $\mathscr{H}_0 = L^2(\mathbf{R}^\nu)(\nu \geqslant 2)$, H is H_D, the negative Dirichlet Laplacian acting in $\mathscr{H} = L^2(\Omega)$, and J is an appropriate identification operator. The results of Ref. 1 provide a rigorous foundation for the theory of quantum-mechanical scattering of low-energy atoms by crystal surfaces, modeled as impenetrable periodic surfaces.[2]

In the present paper, H will stand for either H_D or H_N, the negative Neumann Laplacian in $\mathscr{H}$, and the results herein extend those in Ref. 1 to the case of the homogeneous Neumann condition on $\partial\Omega$. In addition, we establish an invariance and an asymptotic completeness result for the wave operators $W_\pm(\phi)$ for a class of real functions ϕ on $\mathbb{R}$ which we call "admissible" (see Sec. 2), where $W_\pm(\phi)$ are defined as were $W_\pm$, but with H_0 and H replaced by $\phi(H_0)$ and $\phi(H)$, respectively. The theorems of the present study have the physically important consequence that certain wave operators, which describe acoustic and certain kinds of electromagnetic scattering by a very wide class of periodic surfaces, are asymptotically complete in a physically natural sense.

The organization of the paper is as follows. In Sec. 2 we prove Theorem 2.1, which asserts that $W_\pm$ exist and are partially isometric, and that $W_\pm = W_\pm(\phi)$ for admissible functions ϕ, provided only that Ω contains a half-space and is itself contained in a half-space (Property (I)). *No smoothness assumptions on $\partial\Omega$ are made in Theorem 2.1 or elsewhere in the paper.* In the proof of this theorem, we used the extension mentioned in the Appendix of results of Chandler and Gibson.[3,4]

Theorems 3.1 and 3.2—our main results—are stated in Sec. 3. Assume that Ω is a domain with Property (I) whose surface is periodic in the sense of Property (II) of Sec. 3, and that, in addition, Ω has the local compactness Property (III) of that section, which is a mild regularity property not implying the smoothness of $\partial\Omega$. Then Theorem 3.1 asserts that $\text{Ran } W_\pm$ equals $\mathscr{H}_{\text{scatt}}(H)$, the subspace of scattering states of $\mathscr{H}$. For any self-adjoint operator A in $\mathscr{H}$, $\mathscr{H}_{\text{scatt}}(A)$ is a subspace of $\mathscr{H}$ whose elements f are such that $\exp(-itA)f$ is evanescent when $t \to \pm\infty$ from the intersection (if nonempty) of Ω with each half-space whose boundary is parallel to that of a half-space containing Ω. When combined with Theorems 2.1 and 3.1, Theorem 3.2 entails the following result: under the assumptions on Ω in this paragraph, the wave operators $W_\pm(\phi)$ are asymptotically complete for admissible functions ϕ, in the sense that $\text{Ran } W_\pm(\phi) = \mathscr{H}_{\text{scatt}}(\phi(H)) = \mathscr{H} \ominus \mathscr{H}_{\text{surf}}(\phi(H))$. If A is a self-adjoint operator in $\mathscr{H}$, the subspace of surface states $\mathscr{H}_{\text{surf}}(A)$ consists of those $f \in \mathscr{H}$ such that $\exp(-itA)f$ remains "close" to $\partial\Omega$, in some sense, for all time. Under the latter assumptions on Ω, Theorem 3.2 also asserts that the subspaces $\mathscr{H}_{\text{scatt}}(\phi(H))$ and $\mathscr{H}_{\text{surf}}(\phi(H))$ are independent of the particular admissible function ϕ considered.

Theorems 3.1 and 3.2 are proved in Secs. 4 and 5, respectively, by using direct-integral techniques and methods related to those of Lyford.[5–7] In particular, a version, due to Lyford,[8] of well-known theorems of Birman and Belopol'skii[9] was used in the proof of Theorem 3.1.

[a] Also at the Department of Physics, Catholic University of America, Washington, D. C. 20064.

After having essentially completed the present study, we learned that Wilcox[10] proved the completeness of wave operators analogous to $W_\pm$ for exterior domains in $\mathbb{R}^2$ with periodic boundaries, under conditions stronger than ours and by an approach different from the present one. However, important results, not obtained in our paper, are derived in Ref. 10 [see Remark (3) after Theorem 3.1 for more detailed comments on Ref. 10].

Remark on Notation: Let A be a self-adjoint operator in a Hilbert space $\mathscr{K}$. The spectrum (respectively, point spectrum) of A will be denoted by $\sigma(A)$ [respectively, $\sigma_p(A)$]. The subspace of absolute continuity of $\mathscr{K}$ with respect to A will be called $\mathscr{K}_{\rm ac}(A)$, and $\mathscr{K}_p(A)$ will denote the closed span of the eigenfunctions of A. We write $P_{\rm ac}(A)$ [respectively, $P_p(A)$] for the projection operator from $\mathscr{K}$ onto $\mathscr{K}_{\rm ac}(A)$ [respectively, $\mathscr{K}_p(A)$]. The spectral measure of A will be denoted by $E(.;A)$.

2. EXISTENCE, PARTIAL ISOMETRY, AND INVARIANCE OF WAVE OPERATORS FOR SCATTERING BY A GENERAL CLASS OF SURFACES WITH THE DIRICHLET OR NEUMANN CONDITIONS

Let Ω be an external domain in $\mathbb{R}^\nu (\nu \geqslant 2)$ having the property

(I) Ω is contained in a half-space and contains a half-space.

We will frequently denote points $x \in \mathbb{R}^\nu$ by $(\tilde{x}, x_\nu)$, with $\tilde{x} \in \mathbb{R}^{\nu-1}$, $x_\nu \in \mathbb{R}$, where the coordinates are chosen so that the half-space $\{(\tilde{x}, x_\nu) \in \mathbb{R}^\nu : x_\nu > 0\}$ contains Ω and so that the boundary $x_\nu = 0$ of this half-space is at a positive distance from $\partial\Omega$.

Let

$$\mathscr{H}_0 = L^2(\mathbb{R}^\nu), \quad \mathscr{H} = L^2(\Omega).$$

The basic self-adjoint operators of our theory are $H_0 = -\Delta$, the negative (distributional) Laplacian in $\mathscr{H}_0$, and H, which acts in $\mathscr{H}$. Here, H denotes the familiar negative Dirichlet or Neumann Laplacians, H_D or H_N, respectively, acting as $-\Delta$ in their respective domains, which are defined by

$$D(H_D) = L^2(\Delta;\Omega) \cap H_0^1(\Omega), \tag{2.1}$$

$$D(H_N) = L^2(\Delta;\Omega) \cap H^1(\Omega) \cap \{u \in H : \langle \nabla v, \nabla u \rangle_\Omega = \langle v, -\Delta u \rangle_\Omega, v \in H^1(\Omega)\}. \tag{2.2}$$

The notation in (2.1) and (2.2) is as follows. For each open subset $M \subset \mathbb{R}^\nu$, $H^1(M)$ and $H_0^1(M)$ are the usual Sobolev spaces, and $L^2(\Delta;M) = \{f \in L^2(M) : \Delta f \in L^2(M)\}$. For any measurable set $M \subset L^2(\mathbb{R}^n)$ $(n = 1,2,...)$, we write $\langle .,. \rangle_M = \langle .,. \rangle_{L^2(M)}$.

We will need the

Definition: Let A, B be self-adjoint operators in the respective Hilbert spaces $\mathscr{K}_0$, $\mathscr{K}$, with A having a purely absolutely continuous spectrum, and let $J: \mathscr{K}_0 \to \mathscr{K}$ be a bounded operator. We then define the wave operators

$$W_\pm(B,A;J) = \operatorname*{s-lim}_{t \to \pm\infty} \exp(itB) J \exp(-itA) \tag{2.3}$$

if they exist.

In this section, we will consider the wave operators

$$W_\pm = W_\pm(H,H_0;J), \tag{2.4}$$

where $J: \mathscr{H}_0 \to \mathscr{H}$ is a bounded operator defined by

$$(Jf)(x) = j(x_\nu) f(x), \quad f \in \mathscr{H}_0, \text{ a.e. } x = (\tilde{x}, x_\nu) \in \Omega. \tag{2.5}$$

Here, $j \in C^\infty(\mathbb{R})$ vanishes for $x_\nu \leqslant a < a_0$ and equals unity for $x_\nu \geqslant b > b_0$, where a_0, b_0, a, and b are constants such that $0 < a < a_0 < b_0 < b < \infty$ and $\partial\Omega \subset \{(\tilde{x}, x_\nu) \in \mathbb{R}^\nu : a_0 < x_\nu < b_0\}$.

Definition: A function ϕ will be said to be admissible if it is an extended real-valued, Borel-measurable function on $\mathbb{R}$ whose values on $[0,\infty)$ are all finite and is such that there exists a finite union $I(\phi) = \bigcup_{n=1}^{N(\phi)} I_n(\phi) \subset \mathbb{R}$ of open intervals $I_n(\phi)$ on each of which ϕ' is continuous, locally of bounded variation, and strictly positive. Moreover, $\overline{I(\phi)} \supset [0,\infty)$.

Theorem 2.1: Let Ω have Property (I). Then

(a) $W_\pm$ exist as partially isometric operators with initial sets

$$\mathscr{H}_\pm = \{f \in \mathscr{H}_0 : \hat{f}(k) = 0 \text{ for a.e. } k = (k_1,...k_\nu) \in \mathbb{R}^\nu \text{ with } k_\nu \lesseqgtr 0\}; \tag{2.6}$$

where $\hat{f}$ is the Fourier transform of f;

(b) the wave operators

$$W_\pm(\phi) = \operatorname*{s-lim}_{t \to \pm\infty} \exp(it\phi(H)) J \exp(-it\phi(H_0)) \tag{2.7}$$

exist and equal $W_\pm$ for each admissible function ϕ.

Remarks: (1) The assertion in part (a) of the theorem concerning the initial sets of $W_\pm$ has a transparent physical meaning.[11]

(2) By virtue of the admissibility of ϕ and since $\sigma_p(H_0) = \emptyset$, and H_0, $H \geqslant 0$, it follows that $\phi(H_0)$ and $\phi(H)$ exist as self-adjoint operators in $\mathscr{H}_0$ and $\mathscr{H}$, respectively.

(3) The obvious extension of part (b) of the theorem to the case when the definition of admissible function in this section is generalized to include functions ϕ whose derivatives can be strictly negative in some of the $I_n(\phi)$'s is of no physical interest here.

Proof of Theorem 2.1: (a) Let $\mathscr{D}_1$ be the set of all g's in $\mathscr{S}(\mathbb{R}^\nu)$ of the form

$$g(x) = g_1(\tilde{x}) g_2(x_\nu), \quad x = (\tilde{x}, x_\nu) \in \mathbb{R}^\nu,$$

where the Fourier transforms $\hat{g}_1$, $\hat{g}_2$ of g_1, g_2 are such that $\hat{g}_1 \in C_0^\infty(\mathbb{R}^{\nu-1} \setminus \{0\})$ and $\hat{g}_2 \in C_0^\infty(\mathbb{R} \setminus \{0\})$. Using, in particular, the fact that $Jg \in D(H)$ for every such g and an estimate of Davies,[12] we conclude that

$$\left\| \frac{d}{dt} \exp(itH) J \exp(-itH_0) g \right\|_\Omega \leqslant \text{const} |t|^{-1-\epsilon},$$

$$\epsilon > 0, \quad |t| > 1, \quad g \in \mathscr{D}_1,$$

whence, since $\mathscr{D}_1$ is dense in $\mathscr{H}_0$ and $t \mapsto \exp(itH) \times J \exp(-itH_0) g$ can be shown to be a strongly continuously differentiable map from $\mathbb{R}$ to $\mathscr{H}$ for each such g by a standard argument, the existence of $W_\pm$ follows. The assertion that $W_\pm$ are partially isometric with initial states $\mathscr{H}_\pm$ follows by the same arguments used earlier.[13]

(b) We remark that $\mathscr{D}_1$ has the two properties

(i) $\mathscr{D}_1 \subset \mathscr{M}(H_0)$;

(ii) for all $u \in \mathscr{D}_1$, there exists a compact interval $\Delta = \Delta(u) \subset (0,\infty)$ such that $u \in E(\Delta;H_0)\mathscr{H}_0$;

where $\mathscr{M}(H_0)$ is defined by (A1) of the Appendix.

Properties (i) and (ii) follow easily from the definition of $\mathscr{D}_1$ and the well-known expressions for the Fourier transform of $E(\Delta;H_0)u$.

Now, $\sigma(H_0) = [0,\infty)$ is purely absolutely continuous and $\mathscr{H}_{ac}(\phi(H_0)) = \mathscr{H}_{ac}(H_0)$, where the second assertion follows by the approach of Kato.[14] By the properties of H_0 in the last sentence, together with (i) and (ii), and the fact that $t \mapsto \exp(itH)J\exp(-itH_0)u$ is a strongly differentiable mapping for each $u \in \mathscr{D}_1$, we may apply Theorem A.1 of the Appendix (with $A = H_0$, $B = H$, $\mathscr{J} = J$, $T = (0,\infty)$, and $\mathscr{D} = \mathscr{D}_1$) and thus conclude that part (b) of the present lemma is true. ■

3. SCATTERING BY PERIODIC SURFACES WITH THE DIRICHLET OR NEUMANN CONDITIONS: MAIN RESULTS

In the remaining Secs. 3–5 of this paper, we will only be concerned with scattering by periodic surfaces. More precisely, henceforth Ω will denote a domain in $\mathbb{R}^\nu$ having Property (I) of Sec. 2 and the additional Properties (II) and (III) stated in this section (unless an explicit statement to the contrary is made).

(II) (Periodicity) For all $l \in L$, $(\tilde{x},x_\nu) \in \Omega \Rightarrow (\tilde{x}+l,x_\nu) \in \Omega, l \in L$.

Here,

$$L = \{l \in \mathbb{R}^{\nu-1} : l = \sum_{i=1}^{\nu-1} a_i n_i, n_i \in \mathbb{Z}, i = 1,\ldots,\nu-1\}, \quad (3.1)$$

with $\{a_i\}_{i=1}^{\nu-1}$ a set of $\nu-1$ linearly independent vectors in $\mathbb{R}^{\nu-1}$.

If A is a subset of $\mathbb{R}^\nu$, we will write

$$A(r) = A \cap \{x \in \mathbb{R}^\nu : |x| < r\}, \quad r \in \mathbb{R}_+, \quad (3.2)$$

where $\mathbb{R}_+ = (0,\infty)$, and

$$A_a = \{x = (\tilde{x},x_\nu) \in \mathbb{R}^\nu : x_\nu < a\}, \quad a \in \mathbb{R}. \quad (3.3)$$

Definition: An open subset $A \subset \mathbb{R}^\nu$ is said to have the local compactness (LC) property for a bounded subset $B \subset H^1(A)$ if B is precompact in $L^2(A(r))$ for all $r \in \mathbb{R}_+$, i.e., if each sequence $\{f_n\} \subset B$ has a subsequence $\{f_{n_k}\}$ such that $\{f_{n_k} | L^2(A(r))\}$ is Cauchy in $L^2(A(r))$ for all such r. (Notice that the subsequence $\{f_{n_k}\}$ can be chosen to be independent of r.)

Thanks to Lemma 4.6 of Ref. 1, it is not necessary to assume in Theorems 3.1 and 3.2 that Ω has properties other than (I) and (II) when $H = H_D$. Unfortunately, it is not known whether this is true when $H = H_N$. We will make the following assumption:

(III) If $H = H_N$, then Ω has the LC property for all bounded subsets of $H^1(\Omega)$.

This is a very weak regularity assumption. Wilcox[15] has shown that (III) is true if Ω has the "finite tiling" property, which holds for all exterior domains likely to arise in applications. Wilcox's result generalizes the corresponding result proved by Agmon[16] for domains possessing the "segment" property.

Definition: If A is a self-adjoint operator in $\mathscr{H} = L^2(\Omega)$, we define the closed and mutually orthogonal subspaces $\mathscr{H}_{scatt}(A)$ and $\mathscr{H}_{surf}(A)$ of scattering and surface states by

$$\mathscr{H}_{scatt}(A) = \{f \in \mathscr{H} : \lim_{t \to \pm\infty} \|\exp(-itA)f\|_{\Omega_a} = 0, \quad a \in \mathbb{R}_+\}, \quad (3.4)$$

$$\mathscr{H}_{surf}(A) = \{f \in \mathscr{H} : \lim_{a \to \infty} \sup_{t \in \mathbb{R}} \|\exp(-itA)f\|_{\Omega \setminus \Omega_a} = 0\}, \quad (3.5)$$

where $\|\cdot\|_M = \|\cdot\|_{L^2(M)}$.

This definition of surface states agrees with that of Davies and Simon.[17] In view of the intuitive meaning of scattering and surface states stated in the Introduction, the following definitions of completeness and asymptotic completeness are very natural for the exterior domain Ω considered here.

Definition: Let A, B be self-adjoint operators in $\mathscr{H}_0, \mathscr{H}$, respectively, and let $\mathscr{J} : \mathscr{H}_0 \to \mathscr{H}$ be a bounded operator. If they exist, the wave operators $W_\pm(B,A;\mathscr{J})$ defined by (2.3) will be said to be complete if

$$\mathrm{Ran}\, W_\pm(B,A;\mathscr{J}) = \mathscr{H}_{scatt}(B) \quad (3.6)$$

and will be said to be asymptotically complete if, in addition to (3.6),

$$\mathscr{H}_{scatt}(B) \oplus \mathscr{H}_{surf}(B) = \mathscr{H}$$

holds.

Theorem 3.1: The wave operators $W_\pm$ in (2.4) are complete:

$$W_\pm^* W_\pm = E_\pm, \quad (3.7)$$

$$W_\pm W_\pm^* = P_{scatt}(H), \quad (3.8)$$

where $E_\pm$ and $P_{scatt}(H)$ are projection operators with domains $\mathscr{H}_0$ and $\mathscr{H}$, respectively, with $E_\pm \mathscr{H}_0 = \mathscr{H}_\pm$ [see (2.6)] and $P_{scatt}(H)\mathscr{H} = \mathscr{H}_{scatt}(H)$.

Theorem 3.2: (a) We have,

$$\mathscr{H}_{scatt}(H) \oplus \mathscr{H}_{surf}(H) = \mathscr{H}. \quad (3.9)$$

(b) If ϕ is an admissible function, then (3.9) holds with H replaced by $\phi(H)$. More specifically, the equations

$$\mathscr{H}_{scatt}(\phi(H)) = \mathscr{H}_{scatt}(H),$$

$$\mathscr{H}_{surf}(\phi(H)) = \mathscr{H}_{surf}(H),$$

holds for each admissible ϕ.

Theorems 2.1(b), (3.1), and (3.2) entail the following

Corollary: If ϕ is an admissible function, then the wave operators $W_\pm(\phi(H))$ are asymptotically complete.

Remarks: (1) Theorems 3.1 and 3.2 with $H = H_D$ apply to the above model of quantum-mechanical scattering of atoms by crystal surfaces.

(2) The above corollary implies the asymptotic completeness of the wave operators $W_\pm(\phi)$ with $\phi(x) = |x|^{1/2}$, pertinent to acoustic scattering in $\mathbb{R}^2$ and $\mathbb{R}^3$ by suitable hard (Neumann) or soft (Dirichlet) periodic surfaces, and electromagnetic scattering in $\mathbb{R}^2$ by appropriate perfectly conducting gratings in the case of TE-(Dirichlet) or TM-polarization (Neumann).

(3) For $\nu = 2$, Wilcox[10] proved a completeness property (more precisely, the unitarity) of wave operators of the type mentioned in the previous Remark, but with H_0 replaced by the specialization of H to the half-plane $x_\nu > 0$. He accomplished this by means of an eigenfunction expansion method,[18] under stronger conditions than those of Theorem 3.1, namely that $\mathscr{H}_{surf}(H) = \emptyset$ and for $H = H_N$ that the bound-

ary of Ω is sufficiently smooth. On the other hand, as stated in the Introduction, Ref. 3 contains important results not derived in the present paper.

(4) Theorem 3.2, and therefore the Corollary, is true if the notion of admissible function is understood in the broader sense of Remark (2) after Theorem 2.1.

Our strategy for proving Theorems 3.1 and 3.2 is similar, *grosso modo*, to that adopted in Ref. 1 to establish the corresponding theorems for $H = H_D$, and hence will not be explicitly outlined here. An acquaintance with the strategy of Ref. 1[19] would perhaps be helpful to the reader in order to gain a panoramic view of our approach in Secs. 4 and 5 of this paper.

4. PROOF OF THEOREM 3.1

In what follows, we will refer to various Fréchet spaces, such as $L^2_{\text{loc}}(\Delta;\overline{\Omega})$, $H^1_{\text{loc}}(\overline{\Omega})$, $H^1_{0,\text{loc}}(\overline{\Omega})$, etc., which are local versions of $L^2(\Delta;\Omega)$, $H^1(\Omega)$, $H^1_0(\Omega)$, etc., where $\overline{\Omega}$ is used to denote that the relevant integrability properties hold *up to the boundary*. Out notation is as stated in Appendix A of Ref. 1 and is similar, e.g., to that used in Ref. 6.

In the special case when Ω is the half-space

$$\Omega^0 = \mathbb{R}^{\nu-1}\times\mathbb{R}_+, \tag{4.1}$$

we will denote H_D by H^0_D and H_N by H^0_N, and H^0 will stand for H^0_D and H^0_N in the respective cases $H = H_D$ and $H = H_N$. The operator H^0 is an auxiliary self-adjoint operator introduced to facilitate the proof of Theorem 3.1.

Since Ω^0 and Ω have the periodicity property (II), it is not surprising that there exist unitary operators U^0 and U, defined previously,[19] which map $\mathscr{H}^0 = L^2(\Omega^0)$ [do not confuse with $\mathscr{H}_0 = L^2(\mathbb{R}^\nu)$] and $\mathscr{H} = L^2(\Omega)$, respectively, into direct integrals of Hilbert spaces

$$U^0\mathscr{H}^0 = \int_{\mathscr{G}}^{\oplus} L^2(\omega^0)\ d\mu,$$

$$U\mathscr{H} = \int_{\mathscr{G}}^{\oplus} L^2(\omega)\ d\mu.$$

Here, ω^0 and ω are the periodicity cells

$$\omega^0 = G\times\mathbb{R}_+,\quad \omega = \omega^0\cap\Omega \tag{4.2}$$

of Ω^0 and Ω, respectively, with

$$G = \{y\in\mathbb{R}^{\nu-1}: y = \sum_{i=1}^{\nu-1} y^{(i)}a_i,\quad y^{(i)}\in(0,1), i = 1,\ldots,\nu-1\}. \tag{4.3}$$

Moreover,

$$\mathscr{G} = \{\theta\in\mathbb{R}^{\nu-1}: \theta = \sum_{i=1}^{\nu-1}\theta^{(i)}b_i, \theta^{(i)}\in(0,2\pi), i = 1,\ldots,\nu-1\},$$

where $\{b_i\}_{i=1}^{\nu-1}$ is a set of linearly independent vectors in $\mathbb{R}^{\nu-1}$ such that $a_i\cdot b_j = \delta_{ij}(i,j = 1,\ldots,\nu-1)$, and $d\mu = |\mathscr{G}|^{-1}\,d\theta$, where $d\theta$ is Lebesgue measure in $\mathbb{R}^{\nu-1}$ and $|\mathscr{G}|$ is the Lebesgue measure of $\mathscr{G}$. We will denote the "component" of a vector $k\in U\mathscr{H}$ in the θ th fiber of $L^2(\omega)$ by k_θ.

Definition (Bloch-Periodicity): A complex-valued function f on Ω is said to have Property (P_θ) for some $\theta\in\mathscr{G}$ if it is of the form

$$f(\tilde{x},x_\nu) = \exp(i\theta\cdot\tilde{x})u(\tilde{x},x_\nu)$$

a.e. on Ω, where $u(\tilde{x}+l,x_\nu) = u(\tilde{x},x_\nu)$ a.e. on Ω for all $l\in L$ [see (3.1)].

We proceed to define families $\{h^0_\theta,\theta\in\mathscr{G}\}$, $\{h_\theta,\theta\in\mathscr{G}\}$ of self-adjoint operators h^0_θ, h_θ. If $H = H_D$, then $h_\theta = h_{D,\theta}$ for all $\theta\in\mathscr{G}$ and if $H = H_N$, then $h_\theta = h_{N,\theta}$ for all such θ. Each $h_{\theta,N}$ is an operator in $L^2(\omega)$ with domain

$$D(h_{N,\theta}) = \{f\in L^2(\Delta;\omega)\cap H^1(\omega): f \text{ has an extension } f\in H^1_{\text{loc}}(\overline{\Omega}) \text{ with Property } (P_\theta) \text{ and } \langle\nabla g,\nabla f\rangle_\omega = \langle g,-\Delta f\rangle^\omega, g\in D_\theta\}, \tag{4.4}$$

where D_θ is a (closed) subspace of $H^1(\omega)$ defined by

$$D_\theta = \{f\in H^1(\omega): f \text{ has an extension } f\in H^1_{\text{loc}}(\overline{\Omega}) \text{ with Property } (P_\theta)\}. \tag{4.5}$$

Each $h_{D,\theta}$ has a domain given by (4.4) and (4.5), but with $H^1_{\text{loc}}(\overline{\Omega})$ replaced by $H^1_{0,\text{loc}}(\overline{\Omega})$ in these two equations. Every h_θ acts by $-\Delta$ on each element in $D(h_\theta)$.

The specialization of each h_θ to $\Omega = \Omega^0$ will be denoted by h^0_θ, and h^0_θ will be denoted by $h^0_{D,\theta}$ and $h^0_{N,\theta}$ in the respective cases $H^0 = H^0_D$ and $H^0 = H^0_N$.

A spectral representation of $h^0_\theta(\theta\in\mathscr{G})$ can be readily obtained,[20] which yields an explicit formula for the spectral function $E(.;h^0_\theta)$ of each h^0_θ.[21] From the latter formula it follows that every h^0_θ has a purely absolutely continuous spectrum.

The next lemma is true if only (I) and (II) hold.

Lemma 4.1: Each $h_\theta(\theta\in\mathscr{G})$ is self-adjoint and

$$UHU^{-1} = \int_{\mathscr{G}}^{\oplus} h_\theta\ d\mu. \tag{4.6}$$

Remarks: (1) $\theta\mapsto h_\theta$ is measurable. This was shown[22] when $h_\theta = h_{D,\theta}$, the proof being similar when $h_\theta = h_{N,\theta}$. Hence (4.6) makes sense.

(2) Specializing Ω to Ω^0, (4.6) becomes

$$U^0H^0(U^0)^{-1} = \int_{\mathscr{G}}^{\oplus} h^0_\theta\ d\mu. \tag{4.7}$$

Proof of Lemma 4.1: It was proved in Ref. 1 for $H = H_D$. Henceforth in this proof, $H = H_N$. The proof that each $h_{N,\theta}$ is self-adjoint is similar to the corresponding proof for $h_{D,\theta}$.[23] We will show that (4.6) holds by proving that its r.h.s and l.h.s coincide on the set of all $g\in Uf$ with $f\in D(H)$ of bounded support,[24] which is a core of UHU^{-1}. Henceforth in the proof, we fix our attention on a g of this type and a fixed $\theta\in\mathscr{G}$. We proceed to show that $g_\theta\in D(h_{N,\theta})$. It is clear that g_θ has all the required properties for this to be true, except perhaps for the partial integration property in (4.4), which we will now show is fulfilled. Arguments of a kind used previously[25] entail that

$$\langle\nabla\psi,\nabla g_\theta\rangle_\omega = \langle\nabla\tilde{\psi},\nabla f\rangle_\Omega, \tag{4.8}$$

where $\psi\in D_\theta$ and $\tilde{\psi}\in H^1_{\text{loc}}(\overline{\Omega})$ is its extension, in the sense of (4.5). Moreover,

$$\langle\nabla\tilde{\psi},\nabla f\rangle_\Omega = \langle\tilde{\psi},-\Delta f\rangle_\Omega = \langle\psi,-\Delta g_\theta\rangle_\omega, \tag{4.9}$$

where the first equality in (4.9) follows by (2.2) and the second equality therein can be proved similarly to (4.8). By (4.8) and (4.9), g_θ has the partial integration property in (4.4), and hence is in $D(h_{N,\theta})$. One now readily shows that

$(UH_N U^{-1}g)_\theta = h_{N,\theta} g_\theta$ by proceeding as before.[23]■

By (III), the condition of the next lemma is satisfied when $H = H_N$, and the lemma will prove to be very useful in this case.

Lemma 4.2: Let Ω have the LC property for all bounded subsets of $H^1(\Omega)$. Then ω has the LC property for each bounded subset $B \subset H^1(\omega)$ which is contained in D_θ for some $\theta \in \mathscr{G}$.

Proof: Let $B \subset H^1(\omega)$ be such a subset and let $\{f_n\}$ be a sequence in B. Let $\gamma \in C_0^\infty(\mathbf{R}^{\nu-1})$ have the property that $\gamma(\tilde{x}) = 1$ for $\tilde{x} \in G$ [see (4.3)] and define $g_n \in H^1(\Omega)$ by

$$g_n(x) = \gamma(\tilde{x}) \tilde{f}_n(x), \quad x = (\tilde{x}, x_\nu) \in \Omega,$$

for each n, where $\tilde{f}_n$ is the extension of f_n to $H^1_{\mathrm{loc}}(\overline{\Omega})$ having Property (P_θ) for some $\theta \in \mathscr{G}$. We then have,

$$\begin{aligned} \|g_n\|^2_{1,\Omega} &= \|\nabla(\gamma \tilde{f}_n)\|^2_\Omega + \|\gamma \tilde{f}_n\|^2_\Omega \\ &\leqslant (\|\tilde{f}_n \nabla\gamma\|_\Omega + \|\gamma \nabla \tilde{f}_n\|_\Omega)^2 + \|\gamma \tilde{f}\|^2_\Omega \\ &\leqslant \mathrm{const}\,[(\|f_n\|_\omega + \|\nabla f_n\|_\omega)^2 + \|f_n\|^2_\omega] \\ &\leqslant \mathrm{const}(\|f_n\|^2_\omega + \|\nabla f\|^2_\omega) = \mathrm{const}\|f_n\|^2_{1,\omega} \end{aligned}$$

for each n, where $\|\cdot\|_{1,M} = \|\cdot\|_{H^1(M)}$ and where we have used, in particular, the (P_θ) property of the $\tilde{f}_n$'s and the stated properties of γ. Whence $\{g_n\}$ is a bounded subset of $H^1(\Omega)$, and hence has a subsequence which is Cauchy in $L^2_{\mathrm{loc}}(\overline{\Omega})$ if Ω has the LC property assumed in the present lemma. By this result and the fact that

$$\|f_n - f_m\|_{\omega(r)} = \|g_n - g_m\|_{\omega(r)} \leqslant \|g_n - g_m\|_{\Omega(r)}$$

for all n,m and all $r \in \mathbf{R}_+$, where we have used the notation (3.2), we see that $\{f_n\}$ has a subsequence which is Cauchy in $L^2_{\mathrm{loc}}(\overline{\Omega})$ if Ω has the stated LC property.■

Let

$$W^\pm_\theta = W_\pm(h_\theta, h^0_\theta; \eta), \quad \theta \in \mathscr{G}, \tag{4.10}$$

$$W'_\pm = W_\pm(H, H^0; J^0), \tag{4.11}$$

where η and J^0 are the respective restrictions of J in (2.5) to $L^2(\omega^0)$ and $L^2(\Omega^0)$ [see (4.1) and (4.2)]. Define the projection operator $P'_{\mathrm{ac}}(H)$: $\mathscr{H} \to \mathscr{H}$ by

$$UP'_{\mathrm{ac}}(H)U^{-1} = \int_{\mathscr{G}}^{\oplus} P_{\mathrm{ac}}(h_\theta)\, d\mu. \tag{4.12}$$

This definition makes sense because the function $\theta \mapsto P_{\mathrm{ac}}(h_\theta)$ from $\mathscr{G}$ to $\mathscr{L}(L^2(\omega))$ is measurable, as follows from a more general result stated in the sentence after (5.3).

As in the case of the less general Theorem 3.1 of Ref. 1, the present Theorem 3.1 can be easily proved once we establish

Lemma 4.3: $W_\pm$ exist as partially isometric operators satisfying

$$W'^*_\pm W'_\pm = I_{L^2(\Omega^0)}, \tag{4.13}$$

$$W'_\pm W'^*_\pm = P'_{\mathrm{ac}}(H). \tag{4.14}$$

Proof of Lemma 4.3: It was proved in Ref. 1 for the case $H^0 = H^0_D$, $H = H_D$. By (4.6), (4.7), (4.10), and (4.11), in particular, the present lemma will follow for the case $H^0 = H^0_N$, $H = H_N$ if each pair $W^\pm_\theta$ with $h^0_\theta = h^0_{N,\theta}$ and $h_\theta = h_{N,\theta}$ exists and is complete in the usual sense:

$$W^{\pm *}_\theta W^\pm_\theta = I_{L^2(\omega^0)}, \quad \theta \in \mathscr{G}, \tag{4.15}$$

$$W^\pm_\theta W^{\pm *}_\theta = P_{\mathrm{ac}}(h_{N,\theta}), \quad \theta \in \mathscr{G}. \tag{4.16}$$

Since each $h^0_{N,\theta}$ has a purely continuous spectrum, a theorem of Lyford[8] entails that (4.15) and (4.16) hold with $W^\pm_\theta$ understood in the present sense if the following three conditions are satisfied for all $\theta \in \mathscr{G}$ and all bounded intervals $\delta \subset \mathbf{R}$:

(i) $\eta D(h^0_{N,\theta}) \subset D(h_{N,\theta})$, $\eta^* D(h_{N,\theta}) \subset D(h^0_{N,\theta})$;

(ii) $(\eta^*\eta - I_{L^2(\omega^0)})E(\delta; h^0_{N,\theta})$ and $(\eta\eta^* - I_{L^2(\omega)})E(\delta; h_{N,\theta})$ are compact;

(iii) $(h_{N,\theta}\eta - \eta h^0_{N,\theta})E(\delta; h^0_{N,\theta})$ is trace class.

In the remainder of the proof, we will consider a fixed $\theta \in \mathscr{G}$ and a fixed bounded interval δ.

(i) We will show that $\eta D(h^0_{N,\theta}) \subset D(h_{N,\theta})$. (The proof that $\eta^* D(h_{N,\theta}) \subset D(h^0_{N,\theta})$ is similar.) It is easily seen that $\eta f \in L^2(\Delta;\omega) \cap H^1(\omega)$ and that ηf has an extension to $H^1_{\mathrm{loc}}(\overline{\Omega})$ with Property (P_θ). There only remains to prove that

$$\langle \nabla\psi, \nabla(\eta f)\rangle_\omega = \langle \psi, -\Delta(\eta f)\rangle_\omega, \quad \psi \in D_\theta,$$

i.e., that

$$\langle \nabla(\eta^*\psi), \nabla f\rangle_{\omega^0} + \langle \eta^*\psi, \Delta f\rangle_{\omega^0} + \int_\omega \frac{\partial(\bar{\psi}(x) f(x) \partial j(x_\nu)/\partial x_\nu)}{\partial x_\nu}\, dx = 0, \ \psi \in D_\theta. \tag{4.17}$$

The sum of the first two terms on the l.h.s of (4.17) vanishes, since $f \in D(h^0_{N,\theta})$ and since $\eta^*\phi$ is in the specialization of D_θ to $\omega = \omega^0$ [recall the definition of $D(h^0_{N,\theta})$]. By, in particular, ψ, $f \in H^1(\omega)$, $\omega \subset \omega^0$, and supp $\partial j/\partial x_\nu \subset [a,b]$ for some $0 < a < b < \infty$, $\bar{\psi}(x) f(x) \partial j(x)/\partial x_\nu$ can be extended in a natural way to $G \times \mathbf{R}$ [see (4.3)] and approximated in the $L^1(G \times \mathbf{R})$ sense by a smooth function of bounded support. A simple limiting argument now shows that the integral over ω in (4.17) vanishes for the stated functions ψ, f.

(ii), (iii): For the case $H^0 = H^0_N$, $H = H_N$, these properties can be proved as before,[26] using, in particular, the LC property of ω for all subsets of $D(h_{N,\theta})$ which are bounded in $H^1(\omega)$ and the explicit formula for $E(.; h^0_{N,\theta})$ mentioned above.[21] This LC property of ω follows from (4.4), (4.5), Property (III), and Lemma 4.2.■

Proof of Theorem 3.1: Since we have already proved it for $H^0 = H^0_D$, $H = H_D$ in Ref. 1, we will only prove it here for $H^0 = H^0_N$, $H = H_N$. By imitating the arguments of Appendix B of that reference, explicit expressions for the wave operators $W^0_\pm = W_\pm(H^0_N, H_0; P^0)$ can be obtained and shown to satisfy

$$W^{0*}_\pm W^0_\pm = E_\pm, \tag{4.18}$$

$$W^0_\pm W^{0*}_\pm = I_{L^2(\Omega^0)}, \tag{4.19}$$

where P^0: $\mathscr{H}_0 \to \mathscr{H}^0$ is defined by

$$(P^0 f)(x) = f(x), \quad f \in \mathscr{H}_0, \quad \text{a.e. } x \in \Omega^0. \tag{4.20}$$

Using in particular, (4.13), (4.14), (4.18)–(4.20), (2.5), and the definition of j immediately after (2.5), $W_\pm = W_\pm(H_N, H_0; J)$ are seen to obey (3.7) and (3.8), but with $P_{\mathrm{scatt}}(H)$ replaced by $P'_{\mathrm{ac}}(H_N)$ [see (4.12)]. However, as follows from the proof of Theorem 3.2 (see Sec. 5),

$$P'_{\mathrm{ac}}(H) = P_{\mathrm{scatt}}(H)$$

if $H = H_D, H_N$, whence the proof of Theorem 3.1 is complete.■

5. PROOF OF THEOREM 3.2

One of the main ingredients in the proof is the next lemma.

Lemma 5.1: h_θ has an empty singular continuous spectrum for all $\theta\in\mathscr{G}$ if $H = H_D$ or $H = H_N$.

This lemma has already been proved[1] for $H = H_D$. Its proof for $H = H_N$ will be sketched after a brief preliminary discussion.

Let

$$\mathscr{D}_\theta = \{f\in L^2_{\rm loc}(\Delta;\bar{\omega})\cap H^1_{\rm loc}(\bar{\omega}): f \text{ has an extension } \tilde{f}\in L^2_{\rm loc}(\Delta;\bar{\Omega}) \cap H^1_{\rm loc}(\bar{\Omega}) \text{ with Property } (P_\theta) \text{ such that } \langle\nabla g,\nabla f\rangle_\omega = \langle g, -\Delta f\rangle_\omega \text{ if } g\in D_\theta \text{ is of bounded support}\},\quad \theta\in\mathscr{G}, \tag{5.1}$$

where D_θ is as in (4.5). Notice that $\mathscr{D}_\theta$ is a local version of $D(h_{N,\theta})$ in (4.4). The terminology "modified Lemma..." will denote the identically numbered lemma of Ref. 1, but with D_θ understood in the sense (5.1) and with h_θ understood as $h_{N,\theta}$.

In the proof of Lemma 5.1, we will use the modified Lemmas 6.2 and 7.5, whose proof, sketched in (c) below, makes use of the auxiliary results (a) and (b). Let $H = H_N$. Then we have,

(a) The modified Lemmas C.2, C.4, and C.5 hold, their proofs being essentially the same as those of the original lemmas.

(b) The modified Lemmas 7.1, 7.3, and 7.4 hold. They follow similarly to the original lemmas by using, in particular, (a) and the spectral representation of $h_{N,\theta}$.[20]

(c) By (a) and (b), Lemma 4.2 of the present paper, and arguments similar to those used in the proof of the original lemmas, the modified Lemmas 6.2 and 7.5 follow.

Proof of Lemma 5.1: For $H = H_N$ the lemma follows from the modified Lemmas 6.2 and 7.5, and arguments of the same type as those used[27] in the case $H = H_D$.■

In the remainder of this section, the notion of admissible function can be understood in the less restrictive sense of Remark (3) after Theorem 2.1.

The next two lemmas will also be needed in the proof of Theorem 3.2. Before stating them, we define the projection operators $P'_{\rm ac}(\phi(H))$ and $P_s(\phi(H))$ from $\mathscr{H}$ into $\mathscr{H}$ when ϕ is an admissible function, as will be assumed in the remainder of this section:

$$UP'_{\rm ac}(\phi(H))U^{-1} = \int_{\mathscr{G}}^{\oplus} P_{\rm ac}(\phi(h_\theta))d\mu, \tag{5.2}$$

$$UP_s(\phi(H))U^{-1} = \int_{\mathscr{G}}^{\oplus} P_p(\phi(h_\theta))d\mu. \tag{5.3}$$

Since $\phi(H)$ and $\phi(h_\theta)(\theta\in\mathscr{G})$ are self-adjoint [as was seen for $\phi(H)$ in Remark (2) after Theorem 2.1 and as follows for the $\phi(h_\theta)$'s similarly, using $h_\theta\geqslant 0(\theta\in\mathscr{G})$ in particular], and since the functions $\theta\mapsto P_{\rm ac}(\phi(h_\theta))$ and $\theta\mapsto P_p(\phi(h_\theta))$ are measurable, definitions (5.2) and (5.3) make sense. The measurability of $\theta\mapsto P_p(\phi(h_\theta))$ follows[28] by that $\theta\mapsto\phi(h_\theta)$ [which itself follows[29] from that of $\theta\mapsto h_\theta$, the equation $E(\Delta;\phi(h_\theta)) = E(\phi^{-1}(\Delta)\cap[0,\infty);h_\theta)$, the admissibility of ϕ, and standard arguments]. The measurability of $\theta\mapsto P_{\rm ac}(\phi(h_\theta))$ is a consequence of that of $\theta\mapsto P_p(\phi(h_\theta))$ and the equation

$$P_{\rm ac}(\phi(h_\theta)) + P_p(\phi(h_\theta)) = I_{L^2(\omega)}, \theta\in\mathscr{G},$$

which is a by-product of results established in the proof of Lemma 5.2.

Lemma 5.2: If ϕ is an admissible function, then

$$\text{Ran } P'_{\rm ac}(H) = \text{Ran } P'_{\rm ac}(\phi(H))\subset\mathscr{H}_{\rm scatt}(\phi(H)) \tag{5.4}$$

and

$$\text{Ran } P_s(H) = \text{Ran } P_s(\phi(H))\subset\mathscr{H}_{\rm surf}(\phi(H)). \tag{5.5}$$

Proof: Since ϕ is admissible and $h_\theta\geqslant 0$ for all $\theta\in\mathscr{G}$, the approach of Kato[14] entails that

$$P_{\rm ac}(h_\theta) = P_{\rm ac}(\phi(h_\theta)) \tag{5.6}$$

for each such θ. Therefore, by (5.2) and (5.6),

$$P'_{\rm ac}(H) = P'_{\rm ac}(\phi(H)).$$

We now show that

$$\text{Ran } P'_{\rm ac}(\phi(H))\subset\mathscr{H}_{\rm scatt}(\phi(H)), \tag{5.7}$$

hence completing the proof of (5.4).

Let $f\in\text{Ran } P'_{\rm ac}(H) = \text{Ran } P'_{\rm ac}(\phi(H))$. Then $g = Uf$ is such that $g_\theta\in\mathscr{H}_{\rm ac}(\phi(h_\theta)) = \mathscr{H}_{\rm ac}(h_\theta)$, a.e. $\theta\in\mathscr{G}$. Consider a fixed $\theta\in\mathscr{G}$ for which this holds and a bounded subset δ or $\mathbb{R}$. We claim that $\chi_a E(\delta;\phi(h_\theta))$ is a compact operator if $a\in\mathbb{R}_+$, χ_a being multiplication by the characteristic function of ω_a (see (3.3)). This follows by $E(\delta;\phi(h_\theta)) = E(\phi^{-1}(\delta)\cap(0,\infty);h_\theta)$, the fact that $\phi^{-1}(\delta\cap[0,\infty))$ is a bounded Borel subset of $\mathbb{R}$, and the LC property of ω for all $H^1(\omega)$-bounded subsets of $D(h_\theta)$ (entailed by Lemma 4.6 of Ref. 1 when $H = H_D$ and established in the proof of Lemma 4.3 of the present paper in the case $H = H_N$). By the compactness of $\chi_a E(\delta;\phi(h_\theta))$, the relation $g_\theta\in\mathscr{H}_{\rm ac}(h_\theta)$, and standard arguments, we infer that

$$\lim_{t\to\pm\infty} ||\exp(-it\phi(h_\theta))g_\theta||_{\omega_a} = 0, \quad a\in\mathbb{R}_+. \tag{5.8}$$

To complete proving (5.7), we need the equation

$$U\exp(-it\phi(H))U^{-1} = \int_{\mathscr{G}}^{\oplus}\exp(-it\phi(h_\theta))d\mu, \tag{5.9}$$

which is implied[30] by Lemma 4.1 and the self-adjointness of $\phi(H)$ and of each $\phi(h_\theta)$. By (5.8), (5.9), and the bounded convergence theorem, we conclude that

$$\lim_{t\to\pm\infty} ||\exp(-it\phi(H))f||_{\Omega_a} = 0$$

for each $a\in\mathbb{R}_+$, and hence that $f\in\mathscr{H}_{\rm scatt}(\phi(H))$. Whence we have proved (5.7).

We now prove (5.5). By Lemma A.1 of the Appendix,

$$P_p(h_\theta)\leqslant P_p(\phi(h_\theta)) \quad \theta\in\mathscr{G}, \tag{5.10}$$

for the admissible function ϕ considered, since , in particular, $h_\theta\geqslant 0$ $(\theta\in\mathscr{G})$.

Now,

$$P_{\rm ac}(h_\theta) + P_p(h_\theta) = I_{L^2(\omega)}, \quad \theta\in\mathscr{G}, \tag{5.11}$$

by Lemma 5.1, whence the symbol $\leqslant$ in (5.10) can be replaced by $=$. Therefore,

$$\text{Ran } P_s(H) = \text{Ran } P_s(\phi(H))$$

is seen to hold by using (5.10) with the equality sign and (5.3).

Finally, one can show that

$$\text{Ran } P_s(\phi(H)) \subset \mathscr{H}_{\text{surf}}(\phi(H)),$$

and hence (5.5), holds by arguments of the same type as those used earlier.[31] ■

Proof of Theorem 3.2: Using Lemma 5.2, (5.2), and (5.3) [with $\phi(H)$ replaced by H], (5.11), and the orthogonality of the subspaces (3.4) and (3.5), we deduce that

$$\begin{aligned}&\mathscr{H}_{\text{scatt}}(\phi(H)) \oplus \mathscr{H}_{\text{surf}}(\phi(H))\\ &\quad \supset \text{Ran } P'_{\text{ac}}(\phi(H)) \oplus \text{Ran } P_s(\phi(H))\\ &\quad = \text{Ran } P'_{\text{ac}}(H) \oplus \text{Ran } P_s(H) = \mathscr{H} \qquad (5.12)\end{aligned}$$

for each admissible function ϕ. By (5.12) and this orthogonality property, the symbol $\subset$ in (5.4) and (5.5) can be replaced by $=$. Since, in addition, (5.4) and (5.5) hold when $\phi(H)$ is replaced by H therein, the proof of Theorem 3.2 is complete. ■

ACKNOWLEDGMENTS

I want to thank my colleagues J. A. DeSanto and W. W. Zachary for useful discussions.

APPENDIX: AUXILIARY RESULTS

In this Appendix, we will state a simple result about the point spectra of suitable functions of self-adjoint operators and an invariance principle of wave operators.

Lemma A.1: Let M be a self-adjoint operator in a Hilbert space and ϕ an extended real-valued function on $\mathbb{R}$ which is M-measurable and finite a.e. with respect to M.[32] Then

$$\mathscr{H}_p(M) \subset \mathscr{H}_p(\phi(M)).$$

Proof: Follows easily from the spectral theorem. ■

Let A,B be self-adjoint operators in the respective separable Hilbert spaces $\mathscr{H}_1$, $\mathscr{H}_2$ and let $\mathscr{J}: \mathscr{H}_1 \to \mathscr{H}_2$ be a bounded operator. For each Borel set $\Delta \subset \mathbb{R}$, we define the local wave operators

$$W_\Delta^{\pm} = \underset{t \to \pm\infty}{\text{s-lim}} \exp(itB)\mathscr{J}\exp(-itA)P_{\text{ac}}(A)E(\Delta;A),$$

$$W_\Delta^{\pm}(\phi) = \underset{t \to \pm\infty}{\text{s-lim}} \exp(it\phi(B))\mathscr{J}\exp(-it\phi(A))P_{\text{ac}}(A)E(\Delta;A)$$

for a suitable real function ϕ when the respective limits exist. Finally, let

$$\begin{aligned}\mathscr{M}(A) = \{u \in \mathscr{H}_1 : &\operatorname*{ess\,sup}_{\lambda \in \mathbb{R}} d\langle u, E((-\infty, \lambda)]; A)u\rangle_{\mathscr{H}_1}/d\lambda\\ &< \infty\}. \qquad (A1)\end{aligned}$$

The following theorem is due essentially to Chandler and Gibson.[33]

Theorem A.1 (Invariance Principle of Wave Operators): Assume that

(1) ϕ is an extended real-valued function which is A-measurable and B-measurable, finite a.e. with respect to A and B, and such that there exists a subset $T \subset \mathbb{R}$ which is a finite or denumerably infinite union of open intervals T_n $(n = 1,...,N(\phi))$ on each of which ϕ' is continuous, locally of bounded variation, and such that one and only one of the inequalities $\phi' > 0$ or $\phi' < 0$ holds;

(2) there exists a dense subset $\mathscr{D} \subset \mathscr{M}(A)$ of $E(T;A)\mathscr{H}_{\text{ac}}(A)$ such that, for all $u \in \mathscr{D}$, $u \in E(\Delta;A)\mathscr{H}_{\text{ac}}(A)$ for some bounded interval $\Delta = \Delta(u) \subset T$;

(3) for each $u \in \mathscr{D}_1$, $t \mapsto w(t) = \exp(itB)\mathscr{J}\exp(-itA)u$ is a strongly differentiable map from $\mathscr{H}_1$ to $\mathscr{H}_2$ such that

$$||dw(t)/dt||_{\mathscr{H}_2} \in L^1(\mathbb{R}_\tau) \cap L^2(\mathbb{R}_\tau),$$

$$|t|^\alpha ||dw(t)/dt||_{\mathscr{H}_2} \in L^1(\mathbb{R}_\tau),$$

for some positive constants τ, α independent of u, where $\mathbb{R}_\tau = (-\infty, -\tau) \cup (\tau, \infty)$. Then $W_T^{\pm}$ and $W_T^{\pm}(\phi)$ exist and

$$W_T^{\pm}(\phi) = W_T^{\pm} E(T_+;A) + W_T^{\mp} E(T_-;A),$$

where T_+ (respectively, T_-) is the union of those intervals $T_n (n = 1,...,N(\phi))$ on which $\phi' > 0$ (respectively, $\phi' < 0$).

Proof: Since it involves only minor changes in the pertinent arguments of Ref. 3, we will omit it. ■

[1] A. W. Sáenz, J. Math. Phys. **22**, 2872 (1981).
[2] For a selective bibliography of mainly formal theoretical work on classical and quantum-mechanical scattering by periodic surfaces, see, e.g., J. A. DeSanto, in *Mathematical Methods and Applications of Scattering Theory*, edited by J. A. DeSanto, A. W. Sáenz, and W. W. Zachary, *Lecture Notes in Physics* **130** (Springer, New York, 1980).
[3] C. Chandler and A. G. Gibson, Indiana Univ. Math. J. **25**, 443 (1976).
[4] C. Chandler and A. G. Gibson, Indiana Univ. Math. J. **28**, 389 (1979).
[5] W. C. Lyford, "Scattering Theory for the Laplacian in Domains with Cylinders," ONR Technical Summary Report No. 23, University of Utah, March 1974 (unpublished).
[6] W. C. Lyford, Math. Ann. **218**, 229 (1975).
[7] W. C. Lyford, Math. Ann. **236**, 255 (1978).
[8] Reference 6, Theorem 3.2. This theorem is a special case of Theorem 2.2.6 of P.A. Deift, "Classical Scattering with a Trace Condition," Ph.D. thesis, Princeton University, 1976. See also W. C. Lyford, Math. Ann. **217**, 257 (1975) [Erratum: Math. Ann. **229**, 96 (1977)].
[9] M. Sh. Birman, Izv. Akad. Nauk SSSR, Ser. Mat. **32**, 914 (1968) [Math. USSR Izv. **2**, 879 (1968)]; A. L. Belopol'skii and M. Sh. Birman, Izv. Akad. Nauk SSSR, Ser. Mat. **32**, 1162 (1968) [Math. USSR Izv. **2**, 1117 (1968)].
[10] C. H. Wilcox, "Scattering Theory for Diffraction Gratings," Preprint 406, University of Bonn, 1980 (unpublished).
[11] Ref. 1, Remark after Theorem 2.1.
[12] We are referring to an obvious modification, appropriate to $L^2(\mathbf{R})$, of Lemma 1.2 of E. B. Davies Math. Proc. Camb. Phil. Soc. **82**, 327 (1977).
[13] Reference 1, proof of Theorem 2.1.
[14] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), p. 545.
[15] C. H. Wilcox, *Scattering Theory for the d'Alembert Equation in Exterior Domains, Lecture Notes in Mathematics* **442** (Springer, New York, 1975), pp. 60–63.
[16] S. Agmon, *Lectures on Elliptic Boundary Value Problems* (Van Nostrand, Princeton, New Jersey, 1965), Theorem 3.8, p. 30.
[17] E. B. Davies and B. Simon, Comm. Math. Phys. **63**, 277 (1978).
[18] C. H. Wilcox, "Raleigh-Bloch Expansions for Diffraction Gratings," I and II, Preprints 363 and 407, respectively, University of Bonn, 1980 (unpublished).
[19] Reference 1, Sec. 3.
[20] The spectral representation in question is given for $h_\theta^0 = h_{D,\theta}^0$ by Lemma 4.1 of Ref. 1 and is given by the same lemma for $h_\theta^0 = h_{N,\theta}^0$ if the definition (4.3) of $w_\tau(x;\xi,\theta)$ in that reference is changed by replacing sin ξx_ν by cos ξx_ν.

[21] $E(.;h^0_\theta)$ is given by Eq. (4.7) of Ref. 1 for $h^0_\theta = h^0_{D,\theta}$ and by the same equation, but with $w_r(x;\xi,\theta)$ modified as stated previously,[20] for $h^0_\theta = h^0_{N,\theta}$.

[22] Reference 1, Lemma C.1.

[23] Reference 1, proof of Lemma 4.5.

[24] If $f \in L^2(A)$, where A is an open set of $\mathbf{R}^\nu$, the support of f is defined as the complement in A of the largest open set $B \subset A$ such that $||f||_B = 0$.

[25] Reference 1, proof of Lemma C.2.

[26] Reference 1, proof of Lemma 5.2.

[27] Reference 1, proof of Lemma 7.6.

[28] Reference 17, Proposition A.2.2.

[29] Reference 17, Lemma A.2.1.

[30] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, IV: Analysis of Operators* (Academic, New York, 1978), Theorem XIII. 85, p. 284.

[31] Reference 17, proof of Proposition 6.1.

[32] M. H. Stone, *Linear Transformations in Hilbert Space*, Am. Math. Soc. Colloq. Pub. **15** (Am. Math. Soc., New York, 1932), Definition 6.3, p. 227.

[33] Reference 3, Theorem 3 and Remark 8; or Ref. 4, Theorem 3. See also M. Reed and B. Simon, *Methods of Modern Mathematical Physics, III: Scattering Theory* (Academic, New York, 1979), Theorem XI. 23, p. 49.

Convergence of the T-matrix approach to scattering theory [a)]

A. G. Ramm
Mathematics Department, Kansas State University, Manhattan, Kansas 66506

(Received 22 July 1981; accepted for publication 13 November 1981)

The T-matrix numerical scheme is widely used in practice. Convergence of this scheme was not proved. A proof of convergence is given in this paper.

PACS numbers: 03.80. + r, 11.20. − e, 02.30.Jr

1. INTRODUCTION

At the international symposium on wave scattering[1] most of the speakers pointed out that the T-matrix scheme needed a justification, its convergence was not proved. In this paper a proof of convergence is given. This proof also clarifies another basic question, namely, convergence of the variational method of finding stationary points of functionals.[6] Many physical problems are formulated as the problems of finding stationary points and/or stationary values of some functionals, and these points are not extremal. A necessary and sufficient condition for a stationary principle to be extremal is given in Ref. 2, p. 90. The standard T-matrix approach is described in Ref. 1, pp. 64. The principal difference between the standard and our approach is as follows. In the standard approach the scattered field is represented as the series in the outgoing spherical waves and the coefficients of the series are found from a linear system. One assumes that the series converges on Γ (the Rayleigh hypothesis) which is not true in general. In our approach one uses a basis of $L^2(\Gamma)$ and no difficulties with convergence arise.

Let us describe a modified T-matrix approach to the problem

$$(-\nabla^2 - k^2)u = f \quad \text{in } \Omega, \quad k > 0 \tag{1}$$

$$u|_\Gamma = 0, \quad r(\partial u/\partial r - iku) \to 0 \quad \text{as } r = |x| \to \infty \tag{2}$$

Here Ω is an exterior domain with a smooth closed boundary Γ and $D = \mathbb{R}^3 \setminus \Omega$ is a bounded domain. From the Green formula it follows that

$$u(x) = v(x) = \int_\Gamma g(x,s)h(s)ds, \quad x \in \Omega \tag{3}$$

$$0 = v(x) - \int_\Gamma g(x,s)h(s)ds, \quad x \in D \tag{4}$$

$$v(x) = \int_\Omega g f dy, \quad g = \frac{\exp(ik|x-y|)}{4\pi|x-y|}, \quad h = \frac{\partial u}{\partial N}. \tag{5}$$

N is the unit normal to Γ directed into Ω. If h is found, then $u(x)$ can be found from (3). Let us rewrite Eq. (4) as

$$Ah = v(s), \quad Ah = \int_\Gamma g(s,s')h ds', \quad s \in \Gamma. \tag{6}$$

Let $\{\phi_j\}, j = 1,2,\ldots$ be a basis of $H_{-1/2}$, where $H = H_0 = L^2(\Gamma)$, $H_q = W_2^q(\Gamma)$ are the Sobolev spaces,[14]

$$h_n = \sum_{j=1}^{n} c_j \phi_j, \tag{7}$$

$$\sum_{m=1}^{n} a_{jm} c_m = v_j, \tag{8}$$

where

$$a_{jm} = (A\phi_m, \phi_j), \quad v_j = (v, \phi_j), \quad (f,v) = (f,v)_H. \tag{9}$$

Let (t_{jm}) be the inverse matrix $a_{jm}^{(-1)}$, $1 \leqslant j,m \leqslant n$. Then c_j, $1 \leqslant j \leqslant n$, can be calculated if v_j, $1 \leqslant j \leqslant n$, are given. From formula (7) one calculates h_n, and from formula (3) with $h = h_n$ one calculates the approximate solution u_n to problem (1)–(2). The problem is to prove that (i) for sufficiently large n the matrix a_{jm} in (8) is invertible, (ii) $\|h_n - h\| \to 0$ as $n \to \infty$, where $\|\cdot\|$ is the norm in H. Actually, convergence will be proved in H_q, where q depends on f and on the smoothness of Γ. Let us assume for simplicity that $\Gamma \subset C^\infty$. Then q depends on the smoothness of f if $\text{meas}(\Gamma \cap \text{supp} f) > 0$ ($\text{supp} f$ is the support of f) and q is arbitrary if $\text{dist}(\text{supp} f, \Gamma) > 0$.

The basic idea of the proof is very simple and is given in Sec. 2. In Sec. 3 some technical details are provided.

2. MAIN RESULT

Theorem 1: System (8) is uniquely solvable for sufficiently large n and $\|h_n - h\| \to 0$ as $n \to \infty$ [without loss of generality we assume that the operator $I + T(k)$ is invertible; see the proof below, $n.1$ of Sec. 3].

Proof: The basic idea is to factorize A in (6) as $A = A_0[I + T(k)]$, where $A_0 = A(0)$, $T(k) = A_0^{-1} \times [A(k) - A(0)]$. The operator $A_0 > 0$ is a bijection of H_q onto H_{q-1}, while $T(k)$ is compact in H_q for any q (see, e.g., Ref. 2, p. 287). The system (8) can be written as $(A_0[I + T]h_n, \phi_j) = (v, \phi_j)$, $1 \leqslant j \leqslant n$. Since $A_0 > 0$ the form $(A_0 u, f)$ is a scalar product which we denote by $[u,f] = (A_0 u, f)$. This scalar product generates a norm $[u,u]^{1/2} = \|A_0^{1/2} u\|$ which is equivalent to the norm in $H_{-1/2}$. This follows from the fact that A_0 is a pseudodifferential elliptic operator of order -1 and therefore $\text{ord} A_0^{1/2} = -\frac{1}{2}$. Thus (8) is of the form $[(I + T)h_n, \phi_j] = (v, \phi_j)$, $1 \leqslant j \leqslant n$. Let $w = A_0^{-1} v$. Then $(v, \phi_j) = [w, \phi_j]$ and $h_n + P_n T h_n = P_n w$, where P_n is the orthoprojection in $H_{-1/2}$ on the linear span of $\{\phi_1, \ldots, \phi_n\}$. The operator $I + T(k)$, $k > 0$, can be assumed invertible (this will be shown in Sec. 3), and $T(k)$ is compact on H_q with arbitrary q, $-\infty < q < \infty$. Therefore $\|(I - P_n)T\|_q \to 0$ as $n \to \infty$, and the norm is the norm of operators on H_q (this will be explained in Sec. 3). Thus $I + P_n T = I + T - (I - P_n)T$ is invertible for sufficiently large n. This means that system (8) is uniquely solvable for sufficiently large n. Furthermore,

$$h - h_n = (I + T)^{-1} w - (I + P_n T) P_n w$$
$$= B[I - (I - P^{(n)} T B)^{-1} P_n] w, \tag{10}$$

[a)] Supported by AFOSR 800204.

 0022-2488/82/061123-03$02.50

where $B = (I + T)^{-1}$, $P^{(n)} = I - P_n$. Thus

$$\|h - h_n\|_q \leqslant c\|(I - Q_n)^{-1} Q_n P_n w\|_q$$

$$\leqslant c_1 \|Q_n\|_q \|P_n w\|_q, \quad c, c_1 = \text{const} > 0, \tag{11}$$

where $Q_n = P^{(n)} TB$, $\|Q_n\|_q \to 0$ as $n \to \infty$. If $\|w\|_q = \|A_0^{-1} v\|_q < \infty$ then (11) shows that $\|h_n - h\|_q \to 0$ as $n \to \infty$ and the rate of convergence is given by the rate of decay of the magnitude $\|P^{(n)} T\|_q$ as $n \to \infty$. In order that $\|A_0^{-1/2} v\|_q < \infty$ it is necessary and sufficient that $v \in H_{q+1/2}$. This is so if $f \in H_{q-1}(\Omega)$ because in this case $v \in H_{q+1}(\Omega)$ and its trace $v|_\Gamma \in H_{q+1/2}$. Our argument shows that if $f \in L^2(\Omega) = H(\Omega)$ the smoothness of $v|_\Gamma$ is even higher than we need. This completes the proof. Theory of the H_q spaces and the trace theorems can be found in Ref. 4.

3. ADDITIONAL DETAILS

(1) Let us show first that $I + T(k)$, $k > 0$, is invertible. Since T is compact, it is sufficient to show that the nullity of this operator is trivial. If $[I + T(k)]h = 0$ then $A(k)h = A_0[I + T(k)]h = 0$. Therefore the function $u(x) = \int_\Gamma g(x, s) h ds$ solves the homogeneous problem (1)–(2). It is well known that the solution of (1)–(2) is unique. Thus $u(x) \equiv 0$ in Ω. If k^2 is not an eigenvalue of the Dirichlet Laplacian in D then $u(x) \equiv 0$ in D, and from the jump relation for the normal derivative of u one derives that $h = 0$. If k^2 is the eigenvalue of the Dirichlet Laplacian in D, then the argument is the same but instead of $g(x, y, k)$ in (3)–(5) one should use the Green function $g_\epsilon(x, y, k)$ of the exterior of a small ball $B_\epsilon \subset D$. This ball is so chosen that k^2 is not an eigenvalue of the Dirichlet Laplacian in $D_\epsilon = D \setminus B_\epsilon$. Obviously such a ball can be found (there are infinitely many such balls).

Remark 1: The idea of applying $g_\epsilon(x, y, k)$ in order to deal with the case when k^2 is an eigenvalue of the interior problem was used in Ref. 3.

(2) Let us show that $\| P^{(n)} T \| \to 0$ as $n \to \infty$. Since $\{\phi_j\}$ is a basis, one has $\| P^{(n)} f \| \to 0$ as $n \to \infty$ for any $f \in H$. Since T is compact it can be written as $T = T_N + d_N$, where $\| d_N \| < \epsilon_N$, $\epsilon_N \to 0$ as $N \to \infty$, and T_N is finite dimensional: $T_N f = \sum_{j=1}^N (f, \psi_j) w_j$. Clearly it is sufficient to prove that $\| P^{(n)} T_N f \| \leqslant \delta_n \| f \|$, where $\delta_n \to 0$ as $n \to \infty$. One has $\| P^{(n)} T_N f \| \leqslant \sum_{j=1}^N \| P^{(n)} w_j (f, \psi_j) \| \leqslant \| f \| \sum_{j=1}^N \| \psi_j \| \, \| P^{(n)} w_j \| \leqslant \delta_n \| f \|$, where $\delta_n \to 0$ as $n \to \infty$ because $\| P^{(n)} w_j \| \to 0$ as $n \to \infty$, $1 \leqslant j \leqslant N$. In this argument $\| \cdot \|$ can denote any norm. What is essential is that $P^{(n)} \to 0$ strongly. In particular one can use the norm of $H_{-1/2}$ provided that the system $\{\phi_j\}$ forms a basis of $H_{-1/2}$. Note that $H_{-1/2} \supset H$, so that if the system $\{\phi_j\}$ forms a basis of $H_{-1/2}$, then every element $f \in H$ can be represented in the form $f = \sum_{j=1}^\infty c_j \phi_j$, where the series converges in $H_{-1/2}$. It does not converge in H, generally speaking, but there exist bases such that if $f \in H$ then the above series converges in H. For example, such a basis is the basis consisting of the eigenfunctions of the operator A_0 (see also Lemma 1 below). If $\{\phi_j\}$ is a basis of H_q then $\{A_0^s \phi_j\}$ is a basis of H_{q+s}. This follows from the fact that A_0^s is a bijection of H_q onto H_{q+s} (A_0^s is an elliptic pseudodifferential operator of order $-s$). Since $|f|_q \leqslant |f|_s$ for $q < s$, it is clear that if the series $\sum_{j=1}^\infty c_j \phi_j$ converges to f in $H = H_0$ it converges to f in $H_{-1/2}$. It is convenient to have a system $\{\phi_j\}$ which forms a basis in any of H_q and if $f \in H_q$ the series $f = \sum_{j=1}^\infty c_j \phi_j$ converges in H_q. For example, if S^1 is the unit circle and $H_q = H_q(S^1)$ then the system $\{\exp(inx)/\sqrt{2\pi}\}$ forms a basis of H_q for any q. The same property has the system $\{\psi_j\}$ of the eigenfunctions of the Laplace–Beltrami operator on Γ, but practically this system is difficult to construct explicitly. Let us prove that for starlike domain D the system $\{Y_j(\xi)\}$, where $\xi = (\theta, \phi)$ is point on a unit sphere S^2 and Y_j are the normalized spherical harmonics, forms a basis in each of H_q. A domain is called starlike if there exists a point 0 inside the domain such that every point of the boundary of the domain can be seen from this point. This means that the equation of the boundary is of the form $r = R(\theta, \phi) = R(\xi)$, where the origin is at the point 0. It is well known that

$$Q_0 Y_n \equiv \int_{S^2} \frac{Y_j(\xi') d\xi'}{4\pi r_{\xi\xi'}} = \frac{Y_j(\xi)}{2n + 1}, \quad j = 0, 1, 2, \ldots$$

where $r_{\xi\xi'} = |\xi - \xi'|$. The system $\{Y_j\}$, $j = 0, 1, 2, \ldots$, forms an orthonormal basis of $H = L^2(S^2)$ and in any $H_q(S^2)$ the scalar product in $H_q(S^2)$ can be defined as $(u, v)_q = (Q_0^{-q} u, Q_0^{-q} v)_0$, $H_0 = L^2(S^2)$, and $(Y_n, Y_m)_q = (2n + 1)^q (2m + 1)^q (Y_N, Y_m) = (2n + 1)^q (2m + 1)^q \delta_{nm}$, where δ_{nm} is the Kronecker delta.

Lemma 1: The system $\{Y_j(\xi)\}$ forms a Riesz basis of $H = H_q(\Gamma)$, provided that D is starlike, $\Gamma \subset C^\infty$, and the elements of H_q are considered as functions of $\xi \in S^2$.

Proof: Consider the eigenfunctions of the equation

$$\int_\Gamma \frac{\psi_n(s') ds'}{4\pi r_{ss'}} = \lambda_n \psi_n(s), \quad r_{ss'} = |s - s'|, \quad s \in \Gamma. \tag{12}$$

Since D is starlike one can rewrite this equation as

$$Q\Phi_n = \int_{S^2} \frac{\Phi_n(\xi') p_0(\xi') d\xi'}{4\pi |R(\xi) - R(\xi')|} = \lambda_n \Phi_n(\xi), \tag{13}$$

where $s = R(\xi)$ is the equation of the surface Γ in the spherical coordinates, $\xi = (\theta, \phi)$, $\Phi_n(\xi) = \psi_n(R(\xi))$, $ds = p_0(\xi) d\xi$, $p_0(\xi) > 0$, and $d\xi = \sin\theta \, d\theta \, d\phi$. The function $p_0(\xi) = |R'_\theta \times R'_\phi|$, where $\times$ denotes the vector product and $r = R(\theta, \phi)$ is the parametric equation of the surface Γ. The system $\{\Phi_j\}$ of the eigenfunctions of the operator Q defined in (13) forms an orthogonal basis of the weighted space $L^2(S^2, p_0(\xi))$. Since (**) $0 < c_1 < p_0(\xi) < c_2$ the normalized system $\{\Phi_j p_0^{-1/2}\}$ forms an orthonormal basis of $L^2(S^2)$. Therefore this system is an image of the system $\{Y_j\}$, $j = 0, 1, 2, \ldots$, under a unitary transformation of $L^2(S^2)$: $UY_j = p_0^{-1/2} \Phi_j$ or $\Phi_j = p_0^{1/2} U Y_j$. The operator $p_0^{1/2} U$ is a bijection of $L^2(S^2)$ onto itself. Let us introduce the operator $J\Phi_j = \psi_j$, $j = 0, 1, 2, \ldots$, where ψ_j are the normalized eigenfunctions of Eq. (12). The operator J defined on the basis elements is isometric and can be extended to the isometric bijection $J: L^2(S^2) \to L^2(\Gamma) = H$. Therefore $\psi_j = J p_0^{1/2} U Y_j$, $j = 0, 1, 2, \ldots$, and Lemma 1 is proved for $H = H_0$. For $q \neq 0$ the system Y_j forms an orthogonal basis of $H_q(S^2)$ and the space $H_q = H_q(\Gamma)$ is metrically equivalent to $H_q(S^2)$ because Γ: C^∞ diffeomorphic to S^2. Thus the system $\{Y_j\}$ forms a basis of $H_q(\Gamma)$ for any q. If instead of C^∞ diffeomorphism one assumes that Γ is C^1 diffeomorphic to S^2, then $\{Y_j\}$ forms

basis of H_q, $q \leqslant l$. Lemma 1 is proved.

(3) Let us consider another projection method of solving Eq. (6) corresponding to the least squares method; namely

$$(Ah_n - v, A\phi_j) = 0, \quad 1 \leqslant j \leqslant n, \tag{14}$$

or

$$\sum_{m=1}^{n} b_{jm} c_m = d_j, \quad 1 \leqslant j \leqslant n, \tag{15}$$

where

$$b_{jm} = (A\phi_j, A\phi_m), \quad d_j = (v, A\phi_j). \tag{16}$$

Since (b_{jm}) is a positive define matrix (if $\ker A = \{0\}$ which we assume for simplicity), the system (15) is uniquely solvable for any n. This system can be obtained from the least squares method as a necessary condition of the minimum of the functional

$$\| Ah_n - v\|^2 = \min, \tag{17}$$

or

$$|(I + T)h_n - w|^2_{-1} = \min, \quad w = A_0^{-1} v. \tag{18}$$

Since $I + T$ is a bijection of H_{-1} onto itself and A_0^{-1} is a bijection of H_q onto H_{q-1} the solution of (18) tends to $(I + T)^{-1} w$ as $n \to \infty$ in H_{-1} if $v \in H_0$, and in H_{q-1} if $v \in H_q$.

[1] *Acoustic, Electromagnetic, and Elastic Wave Scattering-Focus on T-Matrix Approach*, edited by V. Varadan (Pergamon, New York, 1980).

[2] A. G. Ramm, *Theory and Applications of Some New Classes of Integral Equations* (Springer, New York, 1980).

[3] A. G. Ramm, "Investigation of the scattering problem in some domains with infinite boundaries," Vestnik Leningrad Univ. Ser. Math., Mech., Astron., N7, 45–66, (1963); N19, 67–76, (1963). Math. Rev. 27 #483, 28 #374.

[4] J. Lions and E. Magenes, *Problemes aux limits non homogenes et applications* (Dunod, Paris, 1968), Vol. 1.

[5] A. G. Ramm, "Spectral properties of some nonselfadjoint operators and some applications," in *Spectral Theory of Different Operators, Mathematical Studies*, edited by V. I. Knowles and R. Lewis (North Holland, Amsterdam, 1981), pp. 349–354.

[6] A. G. Ramm, J. Math Phys. **23**, 1112 (1982).

Killing vectors in self-dual, Euclidean Einstein spaces

C. P. Boyer[a)]
Instituto de Investigaciones en Matemáticas Applicadas y en Sistemas, Universidad Nacional Autónoma de México, México 20, D.F., México

J. D. Finley, III[b)]
Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131

(Received 29 October 1981; accepted for publication 18 December 1981)

Using the formalism of complex $\mathscr{H}$-spaces, we show that all real, Euclidean self-dual spaces that admit (at least) one Killing vector may be gauged so that only two distinct types of Killing vectors appear; in Kähler coordinates these are the generators of a translational or a rotational symmetry. We give explicit forms both for the Killing vectors and for the constraint on the Kähler potential function Ω which allows for such a Killing vector. In the translational case we show how all such spaces are determined by the general solution of the three-dimensional, flat Laplace's equation and how these are related to the multi-Taub–NUT metrics of Gibbons and Hawking. In the rotational case we simplify the equation determining Ω, but this is not sufficient to obtain the general solution.

PACS numbers: 04.20. – q

I. INTRODUCTION

In this article we make an application of a considerable body of information on (complex) $\mathscr{H}$-spaces, which has been accumulated in the last few years, to the determination of real, self-dual, Euclidean Einstein spaces.[1] There has, of course, been a large amount of interest in such spaces in recent years. Hawking and co-workers[2] have shown that these spaces are very important in the calculations of quantum gravitational effects via the path-integral formalism. They have determined many distinct real, self-dual solutions to the Euclidean Einstein equations and sorted out a subset which are regular and complete, which they call gravitational instantons. As well, some other groups[2] have become interested in these spaces. Tod and Ward[3] have considered the case of complex $\mathscr{H}$-spaces with at least one Killing vector whose covariant derivative is self-dual. They show that all such spaces are just those already given by Gibbons and Hawking,[4] which have real, Euclidean cross sections referred to as generalized multi-Taub–NUT metrics.

We generalize this work by considering all possible types of Killing vectors, but specialize it by considering only *real, Euclidean* self-dual spaces. Any $\mathscr{H}$-space may be described in the Ω formalism of Plebański.[5] In this approach, there is a simple way to pick out real Euclidean cross sections, which are simply manifest real realizations of the space as a complex Kähler manifold (of complex dimension two).

In the next section we first spell out in detail our particular approach to coordinatization of the desired spaces and then discuss the appropriate group of gauge transformations which leaves invariant the general form of this coordinatization, and its physical interpretation. In Sec. III we then solve Killing's equations for such a manifold, obtaining the form which any Killing vector must take in these coordinates and a single (scalar) master equation which gives the constraint which the Kähler scalar Ω must satisfy in order that the space which it determines admit such a Killing vector.

The general form of a Killing vector contains arbitrary functions which simply indicate that there is yet considerable gauge freedom. We demonstrate that the available gauge freedom can always be used in such a way as to reduce any one Killing vector in a real, self-dual Euclidean space to one of two specific forms. (This is similar to the five forms which are necessary[6] to describe all Killing vectors of a complex $\mathscr{H}$-space.) At that point we note that Ω is determined[5] as the solution of a certain nonlinear partial differential equation, whose general solution is not known. Therefore, we introduce the (linear) constraint on Ω introduced by the existence of a Killing vector in the space and use this to simplify the general equation for these two cases. In the one case, where the Killing vector is like a translation (in the Kähler coordinates), we reduce the equation to the three-dimensional (flat) Laplace equation and, thereby, generate the multi-Taub–NUT solutions again. The other case is a rotational Killing vector and, although it does simplify the equation, we have so far been unable to find interesting solutions. (This problem will be discussed in more detail in that section.) Nonetheless, we feel the result is of sufficient interest to merit this explication.

II. BACKGROUND FROM $\mathscr{H}$-SPACES

An $\mathscr{H}$-space admits two distinct congruences of totally null, two-dimensional surfaces,[5,7] which determine complimentary foliations of the entire space. It is this foliation which picks out the natural coordinates to use in describing such a space. We denote the 2-form which is (co-)tangent to one of these congruences by Σ, which must be closed and simple. Therefore, we may pick a pair of coordinates, described as a single spinor q_A, which labels the leaves of this congruence:

$$\Sigma = dq^A \wedge dq_A = 2dq_1 \wedge dq_2. \tag{2.1}$$

[a)]Currently on sabbatical leave at Harvard University, Cambridge, Mass.

[b)]Work supported by the Fomento Educacional, A. C., México 5, D. F., México, and by the C. I. E. A. del Instituto Politécnico Nacional, México 14, D. F., México.

 0022-2488/82/061126-05$02.50

Similarly, the other congruence, $\tilde{\Sigma}$, permits for another pair $\tilde{q}_B$, such that

$$\tilde{\Sigma} = d\tilde{q}^B \wedge d\tilde{q}_B. \tag{2.2}$$

Since the two congruences are complimentary foliations, i.e.,

$$\Sigma \wedge \tilde{\Sigma} \neq 0, \tag{2.3}$$

we may use $q_A, \tilde{q}_B$ as a coordinate system for the entire space.[8] In terms of this geometry it is shown in Ref. 9 that the space is an $\mathscr{H}$-space when

$$g = 2P^{AB} dq_A \underset{s}{\otimes} d\tilde{q}_B, \tag{2.4a}$$

$$P^{AB} = \Omega_{q_A \tilde{q}_B} \quad \text{and} \quad \det P^{AB} = \tfrac{1}{2}\Omega_{q_A \tilde{q}_B}\Omega_{q^A \tilde{q}^B} = 1, \tag{2.4b}$$

where Ω is the Kähler scalar of the manifold and subscripts such as q_A indicate partial differentiation with respect to that variable. One can easily see that the coordinates q_A and $\tilde{q}_B$ are all null and that a real Euclidean space is obtained by looking at the cross section given by setting

$$\tilde{q}_A = \overline{(q_A)} \equiv \bar{q}_A, \quad \Omega \text{ real}. \tag{2.5}$$

Any real, Euclidean self-dual space may have coordinates chosen on it in such a way that it may be described by this approach.[5] After giving a complete description of the complex form of the allowed gauge transformation, and the general Killing vector and master equation, we will proceed to the real, Euclidean case by then insisting on the restriction of the coordinates given by Eq. (2.5).

To determine the allowed transformations we note that Σ and $\tilde{\Sigma}$ are both closed anti-self-dual 2-forms. Since the anti-self-dual curvature vanishes, we may choose a basis of anti-self-dual 2-forms which is closed. It is straightforward to check that a basis composed of

$$S^{\dot{A}\dot{B}} = (\Sigma, P_{AB} dq^A d\tilde{q}^B, \tilde{\Sigma}) \tag{2.6}$$

suffices for this purpose. In fact, choosing Σ and $\tilde{\Sigma}$ as two of the members and requiring the third to be of the form $W_{AB} dq^A \wedge d\tilde{q}^B$, ensures that $W_{AB} \propto \Omega_{q^A \tilde{q}^B} \equiv P_{AB}$. So we require a set of transformations which relabel the sets of null strings,

$$\begin{aligned} q'^R &= q'^R(q^A, \tilde{q}^B), \\ \tilde{q}'^S &= \tilde{q}'^S(q^A, \tilde{q}^B), \end{aligned} \tag{2.7}$$

but which preserve the closed nature of our basis $S^{\dot{A}\dot{B}}$— which is equivalent to vanishing of the anti-self-dual connections, $\Gamma_{\dot{A}\dot{B}}$, in this basis. Since the $S^{\dot{A}\dot{B}}$ are true spinorial quantities, they will transform as

$$S'^{\dot{R}\dot{S}} = \ell^{\dot{R}}{}_{\dot{A}} \ell^{\dot{S}}{}_{\dot{B}} S^{\dot{A}\dot{B}}, \quad \ell^{\dot{R}}{}_{\dot{A}} \in \mathrm{SL}(2,C). \tag{2.8}$$

This requirement is quite strong and has the solution

$$\begin{aligned} dq'^R &= (d^R{}_A)(\Delta dq^A - i\tau \Omega_{q_A \tilde{q}^B} d\tilde{q}^B), \\ d\tilde{q}'^S &= (\tilde{d}^S{}_B)(\tilde{\Delta} d\tilde{q}^B + i\tilde{\tau} \Omega_{q^A \tilde{q}_B} dq^A), \end{aligned} \tag{2.9a}$$

where $d^R{}_A$, $\tilde{d}^S{}_B$ are independent arbitrary elements of SL(2,C) and Δ, $\tilde{\Delta}$, τ, $\tilde{\tau}$ are constant scalars related by

$$1 = \Delta\tilde{\Delta} + \tau\tilde{\tau}, \tag{2.9b}$$

which is the requirement that the volume form on the manifold is preserved. (Since everything is complex the factors of i are not essential but are inserted to make simpler the transition to the real, Euclidean case.) The inverse transformations are

$$\begin{aligned} \partial'_R &\equiv \frac{\partial}{\partial_{q'^R}} = (d_R{}^A)(-\tilde{\Delta}\partial_A + i\tau \Omega_{q^A \tilde{q}_B} \tilde{\partial}_B), \\ \tilde{\partial}'_S &\equiv \frac{\partial}{\partial_{q'^S}} = (\tilde{d}_S{}^B)(-\Delta\tilde{\partial}_B - i\tilde{\tau} \Omega_{q_A \tilde{q}^B} \partial_A), \end{aligned} \tag{2.10}$$

while

$$\ell^{\dot{R}}{}_{\dot{A}} = \begin{pmatrix} \Delta, & +i\tau \\ +i\tilde{\tau}, & \tilde{\Delta} \end{pmatrix}, \quad \ell^R{}_A = d^R{}_A, \tag{2.11}$$

where the matrices $\ell^R{}_{\dot{A}}$ transform undotted spinorial indices. In particular,

$$S^{AB} \equiv 2dq^{(A} \wedge dq^{B)} \tag{2.12}$$

transform as

$$S'^{RS} = \ell^R{}_A \ell^S{}_B S^{AB}. \tag{2.13}$$

Lastly, we want to know how Ω itself transforms. In order to specify this transformation, it is necessary to introduce a pair of potentials the existence of which is guaranteed by the constraint equation on Ω,

$$(\partial^A \tilde{\partial}^B \Omega)(\partial_A \tilde{\partial}_B \Omega) = 2. \tag{2.14}$$

This equation implies the existence of functions F and $\tilde{F}$ such that

$$(\tilde{\partial}^B \Omega)\partial_A \tilde{\partial}_B \Omega = q_A + \partial_A F, \quad (\partial^A \Omega)\partial_A \tilde{\partial}_B \Omega = \tilde{q}_B + \tilde{\partial}_B \tilde{F}. \tag{2.15}$$

Then we find that

$$\Omega' = \Omega + i\frac{\tilde{\Delta}\tau(\tilde{F} + P) - \Delta\tilde{\tau}(F + \tilde{P})}{\Delta\tilde{\Delta} - \tau\tilde{\tau}} + A(q) + \tilde{A}(\tilde{q}), \tag{2.16}$$

where P, $\tilde{P}$ are new arbitrary functions of all four variables while A and $\tilde{A}$ are new functions of q_A only and $\tilde{q}_B$ only, respectively.

All the transformation equations given retain the form of the metric, the coordinates as labels for a pair of sets of null strings which, together, span the space, as well as preserving the closed property of our choice of basis of anti-self-dual 2-forms. Upon restriction to the real, Euclidean case, q^A should be interpreted as $\bar{q}^A$ with Ω real, and, as well, all quantities with a tilde are interpreted as the complex conjugates of the similar symbols without the tilde.

III. KILLING'S EQUATIONS

In order to determine the form of an allowed Killing vector, we write it in the form

$$K = L^A \partial_A + \tilde{L}^A \tilde{\partial}_A. \tag{3.1}$$

The homothetic Killing equations

$$K_{(\mu;\nu)} = \chi_0 g_{\mu\nu} \tag{3.2}$$

then become the three sets[10]

$$P_{(A}{}^B \partial_{C)} L_B = 0, \tag{3.3a}$$

$$P^{C(A} \tilde{\partial}^{B)} L_C = 0, \tag{3.3b}$$

$$\begin{aligned} L^D \partial_D P^{AB} + \tilde{L}^D \tilde{\partial}_D P^{AB} + P^{AD} \tilde{\partial}^B \tilde{L}_D + P^{DB} \partial^A L_D \\ = 2\chi_0 P^{AB}. \end{aligned} \tag{3.4}$$

The solutions of Eqs. (3.3) are given by

$$L^A = B\partial^A\Omega + J^A(q),$$
$$\tilde{L}^A = \tilde{B}\,\tilde{\partial}^A\Omega + \tilde{J}^A(\tilde{q}), \tag{3.5}$$

where $B, \tilde{B}$ are arbitrary, while J^A is a function of q_A only and $\tilde{J}^A$ a function of $\tilde{q}_A$ only.

Insertion of Eqs. (3.5) into (3.4) gives

$$\partial^B\tilde{\partial}^C(\tilde{J}^D\tilde{\partial}_D + J^D\partial_D - 2\chi_0)\Omega + \partial^B[B(\partial^A\Omega)\partial_A\tilde{\partial}^C\Omega] + \tilde{\partial}^C[\tilde{B}(\tilde{\partial}^D\Omega)\tilde{\partial}_D\partial^B\Omega] = 0,$$

from which it follows that B and $\tilde{B}$ must be constants, which are therefore written from now on as b_0 and $\tilde{b}_0$. Therefore, we have simply

$$\partial^B\tilde{\partial}^C\left[\underset{K}{\pounds}\,\Omega - 2\chi_0\Omega + b_0\tilde{F} + \tilde{b}_0\tilde{F}\right] = 0.$$

This allows us to infer the existence of functions $H = H(q_A)$ and $\tilde{H} = \tilde{H}(\tilde{q}_A)$ and a master equation

$$K\Omega = 2\chi_0\Omega - \tilde{b}_0F - b_0\tilde{F} + H + \tilde{H}, \tag{3.6}$$

which describes the relation which a Killing vector K and the Kähler function Ω must have in order for the space in question to admit a Killing vector. There are, however, yet some integrability conditions which such a solution must satisfy.[11] These can be shown to require the existence of a constant c_0 and functions $\zeta = \zeta(q_A)$, $\tilde{\zeta} = \tilde{\zeta}(\tilde{q}_A)$ such that

$$L^A = b_0\partial^A\Omega + (\chi_0 - \tfrac{1}{2}ic_0)q^A + \partial^A\zeta,$$
$$\tilde{L}^A = \tilde{b}_0\tilde{\partial}^A\Omega + (\chi_0 + \tfrac{1}{2}ic_0)\tilde{q}^A + \tilde{\partial}^A\tilde{\zeta}. \tag{3.7}$$

The particular integrability condition involved here is that

$$\ell^{\dot{A}\dot{B}} \equiv \nabla_A{}^{\dot{A}}K_B{}^{\dot{B}}\epsilon^{AB} \tag{3.8}$$

must be constant, where $\nabla_A{}^{\dot{A}}$ is the usual spinorial covariant derivative and $K_B{}^{\dot{B}}$ is the spinorial form of the components of the Killing vector. (In the vector picture, $\ell^{\dot{A}\dot{B}}$ is the spinorial image of the anti-self-dual part of the covariant derivative of K. Therefore, $\ell^{\dot{A}\dot{B}} = 0$ is the case considered by Tod and Ward.[3]) In general, the covariant derivative of $\ell^{\dot{A}\dot{B}}$ would be proportional to the anti-self-dual part of the curvature.[11] However, in an $\mathscr{H}$-space, viewed in a basis where $dS^{\dot{A}\dot{B}} = 0$, this simply means that $\ell^{\dot{A}\dot{B}}$ must be constant. As a summary, we see that the Killing vector is determined by three constants, $b_0, c_0, \tilde{b}_0$, two pair of functions of two variables only ζ, $\tilde{\zeta}$, H, $\tilde{H}$, and the homothetic constant χ_0. In the real, Euclidean case all that is necessary is, again, to replace all tildes by complex conjugates.

We now want to consider a single, specific, allowed Killing vector and to use the group of coordinate transformations to simplify its form as much as possible. In doing this we will restrict our attention to the real, Euclidean case. Also, we restrict attention to the pure Killing vector case—$\chi_0 \equiv 0$. We first focus attention on the matrix

$$\ell^{\dot{A}\dot{B}} = -2\begin{pmatrix} b_0 & ic_0 \\ ic_0 & \bar{b}_0 \end{pmatrix}, \quad c_0 \text{ real}. \tag{3.9}$$

The determinant of $\ell^{\dot{A}\dot{B}}$ is just $4(|b_0|^2 + c_0{}^2)$. The determinant is invariant under the transformation group. Therefore we may split the discussion into only two cases: (1) $\det\ell^{\dot{A}\dot{B}} = 0$—$b_0 = 0 = c$, i.e., $\ell^{\dot{A}\dot{B}} = 0$; (2) $\det\ell^{\dot{A}\dot{B}} \neq 0$.

In case (1), $\ell^{\dot{A}\dot{B}} = 0$ and therefore the Killing vector has a purely self-dual covariant derivative. Then we have simply

$$L^A = \partial^A\zeta. \tag{3.10}$$

We write

$$q_A = \begin{pmatrix} q \\ p \end{pmatrix} \tag{3.11}$$

and then perform a transformation

$$\left.\begin{aligned} q' &= \zeta = \zeta(q,p) \\ p' &= j(q,p) \end{aligned}\right\} \det d^R{}_A = \begin{vmatrix} \zeta_q & \zeta_p \\ j_q & j_p \end{vmatrix} = 1. \tag{3.12}$$

Here j is a function chosen so as to satisfy the determinant condition and ζ is assumed nonconstant since a constant ζ generates only 0 as a Killing vector. We also choose an Ω transformation ($\Delta = 1$, $\tau = 0$) with $A_P = -H$, resulting in the form

$$K\Omega = \Omega_p + \Omega_{\bar{p}} = 0, \tag{3.13}$$

as the simplest possible form for this type of Killing vector and the constraint it imposes on Ω. This Killing vector simply generates a translation in the variable Rep.

In case (2), $\ell^{\dot{A}\dot{B}} \neq 0$, so we note that it satisfies

$$\ell'^{\dot{R}\dot{S}} = d^{\dot{R}}{}_{\dot{A}}d^{\dot{S}}{}_{\dot{B}}\,\ell^{\dot{A}\dot{B}}, \tag{3.14}$$

which results in

$$b_0' = \Delta^2 b_0 - 2\Delta\tau c_0 - \tau^2\bar{b}_0,$$
$$c_0' = c_0(|\Delta|^2 - |\tau|^2) + \bar{b}_0\bar{\Delta}\tau + b_0\Delta\bar{\tau}, \tag{3.15}$$

The procedure to follow to simplify this Killing vector is somewhat involved, but we give an outline here. First, if b_0 happens to vanish, choose a transformation with nonzero $\Delta\tau$ so that $b_0' \neq 0$. With b_0 surely nonzero, then perform an Ω transformation with $\tau = 0$ and $A = \zeta/b_0$, which leaves (dropping primes)

$$K = (\partial^A b_0\Omega - \tfrac{1}{2}ic_0q^A)\partial_A + \text{c. c.}$$

At this point we note that there is also some additional gauge freedom in the potential F. From Eq. (2.15) it is clear that F is undefined to within the addition of an arbitrary function $\bar{h}(\bar{q}_A)$, only. Therefore, by choosing such an $\bar{h} = -\bar{H}/\bar{b}_0$, the master equation is left in the form (dropping primes)

$$K\Omega = -\bar{b}_0F - b_0\bar{F}.$$

At this point, we see that we may always tranform to $b_0' = 0$ by choice of τ, causing thereby $c_0' = \frac{1}{2}[\det\ell^{\dot{A}\dot{B}}]^{1/2}$, which gives

$$K\Omega = -\tfrac{1}{2}ic_0'(q^A\partial_A - \bar{q}^A\bar{\partial}_A) = 0. \tag{3.16}$$

At this point we have not yet used the freedom of $d^R{}_A$. By choosing $q' = q/p, p' = \frac{1}{2}p^2$, ($\det d^R{}_A = 1$), and $c_0' = -2$, we finally have (again dropping primes)

$$K\Omega = i(p\partial_p - \bar{p}\partial_{\bar{p}})\Omega = 0 \tag{3.17}$$

as an optimal form for those Killing vectors whose covariant derivative has a nonzero anti-self-dual part. (By setting $p = |p|e^{i\theta}$, we can rewrite K as ∂_θ and see that K generates rotation of the $p, \bar{p}$ plane.)

IV. RELATION TO SOLUTIONS OF THE HEAVENLY EQUATION

In this section we impose the existence of a Killing vector on the manifold and use this to simplify the determination of solutions of the Ω equation, (2.4b). First, consider Killing vectors for which $\ell^{A\dot{B}} \equiv 0$. Then Eq. (3.13) tells us that

$$\Omega = \Omega(p,q,\bar{p},\bar{q}) = L[i(\bar{p}-p),q,\bar{q}]. \tag{4.1a}$$

Setting $u = i(\bar{p}-p)$, we reduce Eq. (2.4b) to

$$L_{uu}L_{q\bar{q}} - L_{uq}L_{u\bar{q}} = 1. \tag{4.1b}$$

By setting $U = L_u$ and $Q = L_q$ we may write this partial differential equation as the vanishing of a pair of 3-forms[12]:

$$dQ \wedge dU \wedge dq - dq \wedge d\bar{q} \wedge du = 0,$$
$$dU \wedge du \wedge d\bar{p} + dQ \wedge dq \wedge d\bar{q} = 0. \tag{4.2}$$

Then, choice of $U, q, \bar{q}$ as new independent variables can be made explicit by the Legendre transformation

$$U = L_u, \quad V \equiv Uu - L = V(U,q,q), \tag{4.3a}$$

which transforms Eq. (4.1b) into

$$V_{uu} + V_{q\bar{q}} = 0, \tag{4.3b}$$

thereby making the problem linear. Equation (4.3b) is equivalent to the original (4.1b) provided that $dU \wedge dq \wedge d\bar{q} \neq 0$. However, if $dU \wedge dq \wedge d\bar{q} \neq 0$, it is easy to show that the space so generated is flat. Using the Legendre transformation, we have that

$$\Omega = \Omega(p,q,\bar{p},\bar{q}) = uU - V[U(u,q,\bar{q}),q,\bar{q}], \tag{4.4}$$

with $u = i(\bar{p}-p)$, and with $u = V_U(U,q,\bar{q})$ determining $U = U(u,q,\bar{q})$ while $V(U,q,\bar{q})$ is any solution of the three-dimensional Laplace equation. However, it is also useful to rewrite the metric in terms of V. Therefore, setting $p = \frac{1}{4}(t+iu)$, $q = \frac{1}{2}(v+iw)$ changes the canonical form of Eq. (2.4) to

$$ds^2 = (u_U)(dv^2 + dw^2 + dU)^2 + (u_U)^{-1}(dt + u_v dw - u_w dv)^2. \tag{4.5}$$

By setting $u_U = \Phi$, defining $x^i = (u,w,U)$ as coordinates, and setting $w_i = (u_w, -u_v, 0)$, we can write this in the form

$$ds^2 = \Phi\delta_{ij}dx^i dx^j + \Phi^{-1}(dt - w_i dx^i)^2, \tag{4.6a}$$

where $\epsilon^{jki}\partial_k w_i = \delta^{jr}\partial_r \Phi$, while, of course, Eq. (4.3b) becomes

$$\delta^{ij}\partial_i\partial_j\Phi = 0, \tag{4.6b}$$

which is the form of metric first given by Gibbons and Hawking[4] and first shown by Tod and Ward[3] to be the general form of (real, Euclidean, self-dual) metric which permits a self-dual Killing vector, i.e., $\ell^{A\dot{B}} = 0$.

It is also worth writing down the general form of the curvature tensor for this case, which does not appear to be in the literature. Selecting the obvious orthonormal tetrad

$$\hat{e}^i = \Phi^{1/2}dx^i, \quad \hat{e}^4 = (dt - w_i dx^i)/\Phi^{1/2}, \tag{4.7a}$$

and then going to a null basis by

$$\vec{e}^1 = (\hat{e}^1 + i\hat{e}^2)/2^{1/2} = (\vec{e}^2)^*,$$
$$\vec{e}^3 = (\hat{e}^3 + i\hat{e}^4)/2^{1/2} = (\vec{e}^4)^*, \tag{4.7b}$$

we may display the usual Penrose forms of the conformal curvature in the form

$$C^{(5)} = \tfrac{1}{2}\Phi\partial_+\partial_+\Phi^{-2}, \quad C^{(4)} = \tfrac{1}{2}\Phi\partial_+\partial_0\Phi^{-2},$$
$$C^{(3)} = \tfrac{1}{2}\Phi\partial_0\partial_0\Phi^{-2}, \tag{4.8}$$
$$C^{(2)} = \tfrac{1}{2}\Phi\partial_-\partial_0\Phi^{-2}, \quad C^{(1)} = \tfrac{1}{2}\Phi\partial_-\partial_-\Phi^{-2},$$

where $\partial_\pm = \mp(\partial_{x^1} \pm i\partial_{x^2})$, $\partial_0 = \partial_{x^3}$ are basis vectors for a helicity space. The invariants of this curvature are, of course, of algebraically general type except for very special choice of Φ.

The other case, where we permit $\ell^{A\dot{B}}$ to be different from zero is considerably more complicated. To consider it, we begin by integrating Eq. (3.17) (which ensures the existence of such a Killing vector) in the optimal gauge discussed in Sec. III. The general solution is given by a function $H = H(q,\bar{q},r)$ such that

$$\Omega = H(q,\bar{q},p\bar{p}). \tag{4.9}$$

The heavenly equation then takes the form

$$(rH_r)_r H_{q\bar{q}} - (rH_r)_q H_{r\bar{q}} = 1. \tag{4.10}$$

In order to simplify the form of this equation, again write it in terms of an ideal of differential forms. Set $J \equiv rH_r$, $Q \equiv H_q$ and we have

$$dJ \wedge dQ \wedge dq - dr \wedge dq \wedge dq = 0,$$
$$dJ \wedge \frac{dr}{r} \wedge dq + dQ \wedge dq \wedge dQ = 0. \tag{4.11}$$

Another choice of independent variables, $J, q, \bar{q}$, allows Eq. (4.10) to be transformed into

$$F_{q\bar{q}} + (e^F)_{JJ} = 0, \tag{4.12}$$

where $F = \ln r$. (The degenerate case where $dJ \wedge dq \wedge d\bar{q} = 0$ again only generates flat manifolds and is therefore not of interest here.) This can be accomplished by means of the (Legendre-like) transformation

$$W = W(J,q,\bar{q}) = FJ - H, \tag{4.13}$$

with $\ln r = F = F(J,q,\bar{q})$ determined by $J = rH_r$.

The problem now is to solve Eq. (4.12) [or the equivalent Eq. (4.10)]. Any such solution is a real, Euclidean self-dual space with at least one rotational Killing vector. However, it is possible that any particular solution so determined will also have another Killing vector. In that case it is rather easier to find the solution by using the solution to the linear equation (4.3b) given explicitly by Eq. (4.6). Therefore we characterize a nontrivial solution to Eq. (4.12) as one which generates a space without any translational symmetries. As yet we have found no nontrivial solutions.

A plausible method to attempt such a solution is to consider the case when both terms in Eq. (4.12) vanish separately. In the complex case this actually generates interesting solutions,[6] but they have no real Euclidean cross sections. An attempt to generalize this to the case where each term is a constant (which cancel) generates only the well-known Eguchi–Hansen metric[2] which is given in this formalism by $\Omega = N + b\ln[(N-b)/(N+b)]$, with $N^2 = b^2 - 4(q+\bar{q})^2 p\bar{p}$ [see Eq. (3.10)] and b an arbitrary constant.

In an attempt to understand why these nontrivial solutions are so difficult to obtain, we considered the set of such

spaces which have two "rotational" Killing vectors where the Lie algebra so generated closes upon itself. That is, let K_1 and K_2 be two distinct pure Killing vectors with nonzero matrix $\ell_{A\dot{B}}$, permitted by some (class of) spaces, and two constants α and β (possibly zero) such that

$$[K_1,K_2] = \alpha K_1 + \beta K_2. \tag{4.14}$$

We gauge K_1 [as in Eq. (3.17)] to have the form

$$K_1 = q^A \frac{\partial}{\partial_{q^A}} - q^A \frac{\partial}{\partial_{\bar{q}^A}}, \tag{4.15a}$$

and write K_2 as

$$K_2 = (b_0\Omega_{q_A} + \zeta_{q_A}) \frac{\partial}{\partial_{q^A}} + (\bar{b}_0\Omega_{\bar{q}_A} + \zeta_{\bar{q}_A}) \frac{\partial}{\partial_{\bar{q}^A}}, \tag{4.15b}$$

where we have taken the liberty of adding a term $\frac{1}{2}c_0K_1$ to the most general form K_2 could have had, thus ensuring that K_2 and K_1 are independent. Using $K_1\Omega = 0$, it is then a matter of straightforward algebra to show that $\alpha = 0$, β is pure imaginary, and $b_0 \equiv 0$! Therefore $\ell_{A\dot{B}}$ vanishes for K_2, which is then "translational," contrary to assumption. The conclusion, then, is that there are no real, Euclidean, self-dual spaces with two rotational Killing vectors whose algebra is closed upon itself. In fact, when one tries to find such a space, the equations insist that one of the Killing vectors be "translational," thereby generating what we have referred to as a trivial solution of Eq. (4.12).

It is not difficult to see that the only possible symmetry algebras which do not contain a two-dimensional, solvable subalgebra are the one-dimensional algebra and SU(2). A subclass of solutions belonging to the latter case [when SU(2) acts transitively on 3-surfaces] was given in Ref. 1(c).

V. CONCLUSIONS

We have reduced the problem of finding all real, Euclidean, self-dual spaces with one Killing vector to two cases. These correspond to the possibilities that the Killing vector be either of a "translational" type (with respect to Kähler coordinates) or of a "rotational" type. Killing vectors of the "translational" type all have self-dual covariant derivative and are therefore of the type already considered by Tod and Ward.[3] All spaces which admit such a Killing vector are determined by the different solutions of the (flat) three-dimensional Laplace equation (4.4a) and are therefore easily studied further. Regular, complete solutions of this type are given by the multi-Taub–NUT metrics of Gibbons and Hawking.[4]

Spaces which admit only Killing vectors of the "rotational" type may be determined by finding solutions of Eq. (4.12). (These Killing vectors have the anti-self-dual part of their covariant derivative nonzero.) Since this equation is difficult to solve, whereas the Laplace equation is quite straightforward, it is relevant to look only at solutions which generate spaces not having any "translational" Killing vectors. As yet we have been unable to find any solutions of this type. As an indication of why this should be expected to be difficult, we have shown that there are no spaces which admit two and only two Killing vectors both of which are of "rotational" type.

[1]As samples we note S. W. Hawking, Phys. Lett. A **60**, 81 (1977); G. W. Gibbons and C. N. Pope, Commun. Math. Phys. **61**, 239 (1978); **66**, 267 (1979); S. W. Hawking, Nucl. Phys. B **144**, 349 (1978). See C. N. Pope, "The Role of Instantons in Quantum Gravity," to appear in the proceedings of the Second Oxford Conference on Quantum Gravity, April, 1980 (Cambridge U. P., Cambridge) for a much more complete bibliography. See. C. P. Boyer, Hadronic J. **4**, 2 (1981) for a review of some mathematical properties of gravitational instantons.

[2]T. Eguchi and A. J. Hanson, Ann. Phys. (N.Y.) **120**, 82 (1979); S. -T. Yau, Proc. Natl. Acad. Sci. USA **74**, 1798 (1977); N. J. Hitchin, Math. Proc. Camb. Phil. Soc. **85**, 465 (1979); T. Eguchi and P. G. O. Freund, Phys. Rev. Lett. **37**, 1251 (1976).

[3]K. P. Tod and R. S. Ward, Proc. R. Soc. London Ser. A **368**, 411 (1979).

[4]G. W. Gibbons and S. W. Hawking, Phys. Lett. B **78**, 430 (1978).

[5]J. F. Plebański, J. Math. Phys. **16**, 2396 (1975). See also M. Ko, M. Ludvigsen, E. T. Newman, and K. P. Tod, Phys. Rep. **71**, 52 (1981) for an account of the relation of various approaches to these spaces.

[6]J. D. Finley, III, and J. F. Plebański, J. Math. Phys. **20**, 1938 (1979).

[7]See, for example, R. Penrose and M. A. H. MacCallum, Phys. Rep. **6**, 241 (1972).

[8]We use the convention $q_A = \epsilon_{AB}q^B$, $q^A = \epsilon^{BA}q_B$, while $\Omega_{q_A} = \epsilon^{AB}\Omega_{q^B}$, where the subscripts indicate partial derivatives. See J. D. Finley, III, and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976) or Ref. 9 for details of $\mathscr{H}$-spaces in spinorial coordinates.

[9]C. P. Boyer, J. D. Finley, III, and J. F. Plebański, in "Complex General Relativity, $\mathscr{H}$ and $\mathscr{H}\mathscr{H}$ Spaces—A Survey of One Approach," in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).

[10]See Ref. 5 for the components of the affine connection. In Ref. 6 the same problem (in different coordinates) is discussed in rather more detail and some examples are given in Appendix B.

[11]J. F. Plebański and S. Hacyan, J. Math. Phys. **17**, 2203 (1976).

[12]See C. P. Boyer and J. F. Plebański, J. Math. Phys. **18**, 1022 (1977) and references therein to the work of E. Cartan.

Linear deformation problems for the Ernst equation

R. K. Dodd and H. C. Morris
School of Mathematics, Trinity College, Dublin 2, Ireland

(Received 28 January 1981; accepted for publication 24 April 1981)

In the last few years several papers have appeared which give linear deformation problems for the Ernst equation. In this paper we investigate the relationships between the different problems and show that essentially there are two distinct problems up to a gauge transformation corresponding to a finite or infinite dimensional realization of the algebraic prolongation structure associated with the Ernst equation.

PACS numbers: 04.20.Cv

1. THE PROLONGATION STRUCTURE AND DEFORMATION PROBLEMS FOR THE ERNST EQUATION

The stationary axially symmetric empty space–times are those solutions of the empty Einstein field equations which admit two commuting Killing vectors, one spacelike and the other timelike. The determination of the metrics can be reduced to solving a complex equation in two variables,[1]

$$(\mathrm{Re}E)\nabla^2 E = \nabla E\cdot\nabla E \tag{1.1a}$$

where

$$\nabla^2 E \equiv x_1^{-1}\frac{\partial}{\partial x_1}\frac{(x_1\partial E)}{\partial x_1} + \frac{\partial^2 E}{\partial x_2}$$

and

$$\nabla E \equiv \frac{\hat{x}_1\partial E}{\partial x_1} + \frac{\hat{x}_2\partial E}{\partial x_2}. \tag{1.1b}$$

The complex function ϵ is defined in terms of the metric functions which occur in Lewis's canonical form for the metrics,[2]

$$ds^2 = f(dt+\omega d\phi)^2 - f^{-1}(e^{2\gamma}(dx_1^2+dx_2^2) + x_1^2 d\phi^2), \tag{1.2}$$

namely, $E = f + ig$, where g is the "twist potential". This is conveniently defined using the Hodge $*$ operator by

$$*dg = -x_1^{-1}f^2 d\omega$$

with

$$*dz = -idz, \quad *d\bar{z} = id\bar{z}, \quad z = x_1 + ix_2, \tag{1.3}$$

where $\bar{}$ denotes complex conjugation. Considerable progress has been made in obtaining solutions to the Ernst equation (1.1) by exploiting the internal symmetry group of the equation. In particular the Kinnersley–Chitre representation of the Geroch group K has been especially useful.[3–8]

A more recent alternative approach has been to associate the Ernst equation with a linear deformation problem. In this method the Ernst equation arises as the integrability conditions of the deformation problem. One can then generate new solutions from known "seeding solutions" by means of the Bäcklund transformation deduced from the deformation problem[9,10] or by using the technique of Zakharov and Mikhailov.[11]

The first linear problems of this type were derived by Belinskii and Zakharov,[12] Maison,[13] and Harrison[9]; although Kinnersley–Chitre[6] contains an earlier example it was not developed by them in this fashion. In this paper we present in Sec. 1 a unified derivation of these problems as well as those due to the authors[14] and Kramer and Neugebauer.[10] In Sec. 2 the inter-relations between the different problems are considered in detail as well as the corresponding Bäcklund transformations relating solutions of the Ernst equation. In particular, we show that there is a commuting diagram whereby one can pass from the problem in Ref. 14 to the Belinskii–Zakharov problem.[12] This provides an example of a transformation from a deformation problem in which the seeding operators do not commute to one in which they do.

The method we use to obtain the deformation problems is the Wahlquist–Estabrook prolongation technique,[15] which yields an incomplete Lie-algebraic structure which can be associated with the equation in a manner explained below. The structure is closed by imposing constraints on the algebra which do not destroy its nonabelian nature. The linear deformation problems result from choosing some appropriate representation of the resulting Lie algebra.

By putting $\epsilon = f + ig$, the complex equation (1.1) can be written as the first order system,

$$P_j = f^{-1}f_{,x_j}, \quad Q_j = f^{-1}g_{,x_j}, \quad j = 1,2, \tag{1.4a}$$

$$x_1^{-1}P_1 + P_{1,x_1} + P_{2,x_2} = -Q_1^2 - Q_2^2, \tag{1.4b}$$

$$x_1^{-1}Q_1 + Q_{1,x_1} + Q_{2,x_2} = Q_1P_1 + Q_2P_2, \tag{1.4c}$$

and if $\Phi_i = f^{-1}\phi_{,i}$, then

$$\Phi_{i,j} - \Phi_{j,i} = P_i\Phi_j - P_j\Phi_i, \quad i,j = 1,2, \tag{1.5}$$

where $\phi_{,j} \equiv \partial\phi/\partial x_j$. Putting $\phi = f$ or $\phi = g$ then Eqs. (1.5) represent the integrability conditions on P_i, Q_i which, together with (1.4), can be used to generate an exterior system $E(M)$ on the manifold $M \subset R^8$ with local coordinates (x_i, f, g, P_i, Q_i) equivalent to the Ernst equation (1.1) provided $dx_1 \wedge dx_2 \neq 0$ on a solution manifold.[16,17] The generators of $E(M)$ are $(\alpha_1,\alpha_2,\beta_1,\beta_2,\gamma_1,\gamma_2,\gamma_3,\gamma_4)$, where

$$\alpha_1 = dP_1\wedge dx_1 + dP_2\wedge dx_2, \tag{1.6a}$$

$$\alpha_2 = dQ_1\wedge dx_1 + dQ_2\wedge dx_2 - 1_1 dx_1\wedge dx_2, \tag{1.6b}$$

$$\beta_1 = dP_1\wedge dx_2 - dP_2\wedge dx_1 - 1_2 dx_1\wedge dx_2, \tag{1.7a}$$

$$\beta_2 = dQ_1\wedge dx_2 - dQ_2\wedge dx_1 - 1_3 dx_1\wedge dx_2, \tag{1.7b}$$

$$\gamma_1 = df\wedge dx_1 + P_1 f dx_1\wedge dx_2, \tag{1.8a}$$

$$\gamma_2 = df\wedge dx_2 - P_2 f dx_1\wedge dx_2, \tag{1.8b}$$

$$\gamma_3 = dg\wedge dx_2 - Q_1 f dx_1\wedge dx_2, \tag{1.8c}$$

 0022-2488/82/061131-06$02.50

$$\gamma_4 = dg \wedge dx_1 + Q_2 f dx_1 \wedge dx_2, \tag{1.8d}$$

and

$$1_1 = P_1Q_2 - P_2Q_1, \quad 1_2 = -(x_1^{-1}P_1 + Q_1^2 + Q_2^2),$$
$$1_3 = -x_1^{-1}Q_1 + Q_1P_1 + Q_2P_2. \tag{1.9}$$

It is easy to check that Eqs. (1.6)–(1.9) generate an exterior system, that is, a differential ideal of the exterior algebra on M which is closed under the operation of exterior differentiation, $dE(M) \subset E(M)$. If we then use the standard ideas on the existence of a linear Wahlquist–Estabrook prolongation of $E(M)$,[15] we require to determine matrix valued functions F,G on M such that the vector valued one form

$$\omega = dy + F\,dx_1 + G\,dx_2 \tag{1.10}$$

prolongs $E(M)$ to an exterior system $E(N)$ with generators $(\alpha_1,\alpha_2,\beta_1,\beta_2,\gamma_1,\gamma_2,\gamma_3,\gamma_4,\omega)$ on the trivial vector bundle $\pi_N : N = M \times Y \to M$. It is straightforward to show that ω determines a Wahlquist–Estabrook prolongation, provided

$$F = X_0 + P_1X_1 + P_2X_2 + Q_1X_3 + Q_2X_4, \tag{1.11a}$$

$$G = X_5 - P_1X_2 + P_2X_1 - Q_1X_4 + Q_2X_3, \tag{1.11b}$$

where the X_i's ($i = 1,\ldots,5$) *are matrix valued functions of* x_1, x_2, f, and g, which satisfy the following incomplete Lie algebraic or prolongation structure:

$$X_{5,x_1} - X_{0,x_2} + [X_0,X_5] = 0, \tag{1.12a}$$

$$X_{1,x_2} + X_{2,x_1} - x_1^{-1}X_2 + [X_0,X_2] - [X_1,X_5] - fX_{5,f} = 0, \tag{1.12b}$$

$$X_{3,x_2} + X_{4,x_1} - x_1^{-1}X_4 + [X_0,X_4] - [X_3,X_5] - fX_{5,g} = 0, \tag{1.12c}$$

$$X_{2,x_2} - X_{1,x_1} - [X_0,X_1] - [X_2,X_5] + fX_{0,f} = 0, \tag{1.12d}$$

$$X_{4,x_2} - X_{3,x_1} - [X_0,X_3] - [X_4,X_5] + fX_{0,g} = 0, \tag{1.12e}$$

$$fX_{2,f} + [X_1,X_2] = 0, \tag{1.13a}$$

$$X_4 + [X_1,X_4] + [X_3,X_2] - f(X_{4,f} + X_{2,g}) = 0, \tag{1.13b}$$

$$X_3 - [X_2,X_4] - [X_1,X_3] + f(X_{3,f} - X_{1,g}) = 0, \tag{1.13c}$$

$$X_2 - [X_3,X_4] - fX_{4,g} = 0, \tag{1.13d}$$

where the bracket $[\,,\,]$ denotes matrix commutation, $[A,B] = AB - BA$.

Equations (1.12) give the x_1, x_2 dependency whereas Eqs. (1.13) are functionally simpler and determine the prolongation structure for the subalgebra generated by (X_1,X_2,X_3,X_4). To obtain a representation we have to close this structure so that the resulting Lie algebra is nonabelian and consistent with the prolongation structure. This condition is satisfied if we assume that X_i's are independent of f and g. In this case Eqs. (1.11) reduce to

$$[X_1,X_2] = 0,\ X_4 + [X_1,X_4] - [X_2,X_3] = 0,$$
$$X_2 - [X_3,X_4] = 0,\ X_3 - [X_2,X_4] - [X_1,X_3] = 0. \tag{1.14}$$

Two apparently nonequivalent representations are possible, one in which X_1 is poroportional to X_2 and the other in which X_1 and X_2 are distinct. In fact, these two cases do not yield distinct deformation problems, as there exist gauge transformations of the prolongation (y) variables which transform between the two cases. For this reason we shall deal solely with the dependent case,

$$X_1 = \lambda X_2, \quad \lambda = \lambda(x_1,x_2), \tag{1.15a}$$

$$X_2 - [X_3,X_4] = 0, \quad X_4 + [X_2,\lambda X_4 - X_3] = 0,$$
$$X_3 - [X_2,X_4 + \lambda X_3] = 0. \tag{1.15b}$$

A 2×2 matrix representation of this algebra in terms of the SL (2, C) basis,

$$[Z_0,Z_1] = Z_2, \quad [Z_0,Z_2] = -Z_1, \quad [Z_1,Z_2] = Z_0 \tag{1.16a}$$

with

$$Z_0 = -\tfrac{1}{2}i\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad Z_1 = -\tfrac{1}{2}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix},$$
$$Z_2 = -\tfrac{1}{2}i\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \tag{1.16b}$$

is given by

$$X_1 = i(\sin\theta)Z_0, \quad X_2 = i(\cos\theta)Z_0,$$
$$X_3 = Z_1 + i(\sin\theta)Z_2,\ X_4 = i(\cos\theta)Z_2, \quad \text{where } \tan\theta = \lambda. \tag{1.17}$$

It is convenient at this stage to introduce half-angles into the representation. Putting $\cot\theta = -2v/(1-v^2)$, we find that

$$X_1 = \frac{i(1-v^2)}{(1+v^2)}Z_0, \quad X_2 = \frac{-2iv}{(1+v^2)}Z_0,$$
$$X_3 = Z_1 + i\frac{(1-v^2)}{(1+v^2)}Z_2, \quad X_4 = \frac{-2iv}{(1+v^2)}Z_2. \tag{1.18}$$

To complete the representation we have to determine X_0,X_5, and v to be compatible with Eqs. (1.12). Herein lies a radical departure from the usual prolongation technique as applied to the Korteweg–de Vries equation, for example,[15] in that these equations can be satisfied by an *infinite dimensional representation* of the Lie algebra. Geometrically this amounts to erecting a line bundle over N, $P = N \times \Xi$ so that X_0 and X_5 are vertical vector fields on this trivial bundle, $\pi_P : P \to N$. Equations. (1.12) and the representations (1.18) are now defined on P and consequently in the future we shall write $v = v(x_1,x_2,\eta)$, where η parametrizes Ξ. Let

$$X_0 = h(v)\frac{\partial}{\partial\eta}, \quad X_5 = l(v)\frac{\partial}{\partial\eta}, \tag{1.19}$$

then we find that Eqs. (1.12) become, after some rearrangement,

$$-(1+v^2)v_{,x_1} + x_1^{-1}(1-v^2) - h(1+v^2)v_{,\eta} = 0, \tag{1.20a}$$

$$-(1+v^2)v_{,x_2} + 2x_1^{-1}v^2 - l(1+v^2)v_{,\eta} = 0 \tag{1.20b}$$

$$2v\dot{h} - \dot{l}(1-v^2) = 0, \quad \text{where } \dot{m} = \frac{dm}{dv}. \tag{1.20c}$$

If $h \neq 0$, $l \neq 0$ then a solution to (1.20) is given by

$$h = \frac{2v}{(1+v^2)}, \quad l = \frac{-2}{(1+v^2)}, \quad \text{and } v = \frac{x_1}{\eta}. \tag{1.21}$$

It is easy to check that (X_0,X_5) do not form an involutive distribution on P,

$$[X_0,X_5] = X_6, \quad \text{where } X_6 = \frac{-4v}{\eta(1+v^2)^2}\frac{\partial}{\partial\eta}. \tag{1.22}$$

If we now piece the information from (1.18), (1.21), and (1.11) together, we obtain the deformation problem given earlier by the authors,[14]

$$y_{,x_1} + \frac{2v}{(1+v^2)}y_{,\eta} = \frac{-f^{-1}}{(1+v^2)}[i(1-v^2)Z_0 f_{,x_1} - 2ivZ_0 f_{,x_2} + ((1+v^2)Z_1 + i(1-v^2)Z_2)g_{,x_1} - 2ivZ_2 g_{,x_2}]y, \tag{1.23a}$$

$$y_{,x_2} - \frac{2}{(1+v^2)}y_{,\eta} = \frac{f^{-1}}{(1+v^2)}[2ivZ_0 f_{,x_1} + i(1-v^2)Z_0 f_{,x_2} + 2ivZ_2 g_{,x_1} + ((1+v^2)Z_1 + i(1-v^2)Z_2)g_{,x_2}]y, \tag{1.23b}$$

where $v = x_1/\eta$ and (Z_0,Z_1,Z_2) are given by (1.16).

In the case $h = 0, l = 0$, then Eqs. (1.20) can be integrated to yield

$$\tilde{v} = 1 + \epsilon(1+m^2)^{1/2}, \quad m(x_1,x_2) = x_1^{-1}(x_2 - k), \tag{1.24}$$

where k is an arbitrary constant of integration $\epsilon = \pm 1$ and we have used $\sim$ to distinguish this case from the previous one. If we introduce the complex function

$$\gamma = \frac{x_1 - i(x_2 - k)}{x_1 + i(x_2 - k)}, \tag{1.25a}$$

then we find that

$$(1-\tilde{v}^2)/(1+\tilde{v}^2) = -\epsilon m(1+m^2)^{-1/2} \equiv -\tfrac{1}{2}i\epsilon(\gamma^{1/2} - \gamma^{-1/2}) \tag{1.25b}$$

and

$$\tilde{v}/(1+\tilde{v}^2) = \tfrac{1}{2}(1+m^2)^{-1/2} \equiv \tfrac{1}{4}\epsilon(\gamma^{1/2} + \gamma^{-1/2}). \tag{1.25c}$$

Using (1.25) and introducing the complex variable $z = x_1 + ix_2$, the deformation problem can be written in this case as

$$w_{,z} = -\epsilon f^{-1}[\gamma^{1/2}(f_{,z}Z_0 + g_{,z}Z_2) + \epsilon g_{,z}Z_1]w, \tag{1.26a}$$

$$w_{,\bar{z}} = -\epsilon f^{-1}[-\gamma^{-1/2}(f_{,\bar{z}}Z_0 + g_{,\bar{z}}Z_2) + g_{,\bar{z}}Z_1]w. \tag{1.26b}$$

Equation (1.26) is the linear deformation problem introduced recently by Kramer and Neugebauer.[10] Their representation results from choosing the SL (2,C) basis,

$$\tilde{Z}_0 = -\epsilon Z_1, \quad \tilde{Z}_1 = -Z_0, \quad \tilde{Z}_2 = -\epsilon Z_2 \tag{1.27a}$$

and rescaling the dependent variable,

$$w = f^{-1/2}\tilde{w}. \tag{1.27b}$$

This completes the derivation of two of the different representations which can be associated with the Ernst equation. In order to investigate the correspondence more closely, rewrite (1.23) in the form given in Ref. 14,

$$L_1 y = (A\eta + B)y, \quad L_2 y = (C\eta^{-1} + D)y, \tag{1.28a}$$

where

$$L_1 \equiv x_1\frac{\partial}{\partial x_1} + 2\eta\frac{\partial}{\partial\eta} - \eta\frac{\partial}{\partial x_2}, \tag{1.28b}$$

$$L_2 \equiv x_1^{-1}\frac{\partial}{\partial x_1} + \eta^{-1}\frac{\partial}{\partial x_2},$$

and

$$A \equiv f^{-1}(f_{,x_1}Z_0 - x_1^{-1}g_{,x_1}Z_-),$$
$$B \equiv f^{-1}(if_{,x_2}Z_0 + g_{,x_2}Z_+),$$
$$C \equiv f^{-1}(-ix_1^{-1}f_{,x_1}Z_0 - x_1^{-1}g_{,x_1}Z_+),$$
$$D \equiv f^{-1}(if_{,x_2}Z_0 - g_{,x_2}Z_-), \quad Z_\pm = Z_1 \pm Z_2. \tag{1.28c}$$

The solution of the homogeneous system,

$$L_1 u = 0, \quad L_2 u = 0, \tag{1.29}$$

obtained by the method of characteristics, is

$$u = u(k), \quad \text{where } k = x_2 + \tfrac{1}{2}(\eta - \eta^{-1}x_1^2). \tag{1.30}$$

The expression in (1.30) can be inverted to give

$$\eta = -[(x_2 - k) + \epsilon(x_1^2 + (x_2-k)^2)^{1/2}], \quad \epsilon = \pm 1 \tag{1.31a}$$

which also defines the relationship between the functions v, $\tilde{v}$,

$$\tilde{v}(x_1,x_2,k) \equiv -v^{-1}(x_1,\eta). \tag{1.31b}$$

It follows that if we introduce the vector valued functions

$$w_\epsilon(z,\bar{z},k) \equiv y(x_1,x_2, -[(x_2-k) + \epsilon(x_1^2 + (x_2-k)^2)^{1/2}]), \tag{1.32}$$

then (1.23) is transformed into (1.26). Notice that the two cases $\epsilon = \pm 1$ in (1.26) are related by the transformation

$$\gamma^{1/2} \to \epsilon\gamma^{1/2} \tag{1.33}$$

which corresponds to the choice of root in (1.25a). Since $\gamma^{1/2}$ is single valued the map defined by (1.32) is double valued, $y \to (w_+, w_-)$.

The Harrison deformation problem results from a gauge transformation of (1.16). If we write this system as

$$w_{,z} = Hw, \quad w_{,\bar{z}} = \bar{H}w \tag{1.34}$$

and then apply the gauge transformation

$$w = Uy, \; U = \exp(hZ_0) \tag{1.35a}$$

with

$$h = -i\ln\left(\frac{1 - i\epsilon\gamma^{1/2}}{1 + i\epsilon\gamma^{1/2}}\right), \quad \epsilon = \pm 1 \tag{1.35b}$$

and use the representation (1.16), we find that

$$\exp(hZ_0) = \begin{pmatrix} \exp(-ih/2) & 0 \\ 0 & \exp(ih/2) \end{pmatrix} \tag{1.35c}$$

and that (1.26) transforms to the system given by restricting the prolongation form

$$0 = dy + \epsilon f^{-1}[(\gamma^{1/2}f_{,z} - \tfrac{1}{2}fx_1^{-1}\gamma^{1/2})Z_0 + g_{,z}(\epsilon Z_1 - \gamma^{1/2}Z_2)]y\,dz + \epsilon f^{-1}[(-\gamma^{-1/2}f_{,\bar{z}} + \tfrac{1}{2}fx_1^{-1}\gamma^{-1/2})Z_0 - g_{,z}(\epsilon Z_1 + \gamma^{-1/2}Z_2)]y\,d\bar{z} \tag{1.36}$$

to a solution manifold on which $dz \wedge d\bar{z} \neq 0$. Using the same kernel letter to denote the homomorphic image of the SL (2,C) basis in the set of vertical vector fields on N, we find that

$$Z_0 = -\tfrac{1}{2}i\left(y_1\frac{\partial}{\partial y_1} - y_2\frac{\partial}{\partial y_2}\right),\quad Z_1 = -\tfrac{1}{2}\left(y_2\frac{\partial}{\partial y_1} - y_1\frac{\partial}{\partial y_2}\right), Z_2 = -\tfrac{1}{2}i\left(y_2\frac{\partial}{\partial y_1} + y_1\frac{\partial}{\partial y_2}\right). \tag{1.37}$$

If we introduce the homogeneous coordinate $q = y_1/y_2$ then the vector fields (1.37) become

$$Z_0 = -iq\frac{\partial}{\partial q},\quad Z_1 = -\tfrac{1}{2}(1+q^2)\frac{\partial}{\partial q},$$
$$Z_2 = -\tfrac{1}{2}i(1-q^2)\frac{\partial}{\partial q}, \tag{1.38}$$

which are vertical vector fields on the trivial bundle $\pi_{N'}$ $:N' = M \times Q \to M$, where Q is a one-dimensional fiber parametrized by q. This can be obtained from the action of elements of SL (2,C) on Q as fractional linear transformations. Harrison's prolongation form[9] is given by the SL (2,C) representation

$$\tilde{Z}_0 = \frac{-i\epsilon}{2}(\tilde{q}^2-1)\frac{\partial}{\partial\tilde{q}},\quad \tilde{Z}_1 = i\tilde{q}\frac{\partial}{\partial\tilde{q}},$$
$$\tilde{Z}_2 = \frac{-\epsilon}{2}(\tilde{q}^2+1)\frac{\partial}{\partial\tilde{q}} \tag{1.39}$$

upon introducing the variables

$$t = 2f^{-1}(f_{,z} + ig_{,z}) - x_1^{-1}, \tag{1.40a}$$
$$v = 2f^{-1}(f_{,z} - ig_{,z}) - x_1^{-1}, \tag{1.40b}$$
$$w = \bar{t},\quad v = \bar{u},\quad x = \tfrac{1}{2}z,\quad y = \tfrac{1}{2}\bar{z},\quad \zeta = i\gamma^{1/2}. \tag{1.40c}$$

It is clear from (1.26), using the basis (1.16) for SL (2,C), that if w is a solution of the equation then $\tilde{w} = \bar{w}$ is also a solution. For the deformation problem (1.36) a second solution is given by

$$\tilde{y} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\bar{y}.$$

In either case one can therefore introduce the fundamental solution matrix, Φ, whose columns are these solutions and then (1.26) or (1.36) can be written as

$$\Phi_{,z} = (A\gamma^{1/2} + B)\Phi, \tag{1.41a}$$
$$\Phi_{,\bar{z}} = (\bar{A}\gamma^{-1/2} + \tilde{B})\Phi, \tag{1.41b}$$

with appropriately defined A, B, and $\tilde{B}$. If we now apply a gauge transformation to (1.41),

$$U = G\Phi, \tag{1.42}$$

then the transformed system can be written as

$$U_{,z} = (1 + i\gamma^{1/2})CU, \tag{1.43a}$$
$$U_{,\bar{z}} = (1 - i\gamma^{-1/2})CU, \tag{1.43b}$$
$$\text{where } C \equiv -iGAG^{-1}, \tag{1.43c}$$

provided G^{-1} satisfies

$$G^{-1}_{,z} = (iA + B)G^{-1}, \tag{1.44a}$$
$$G^{-1}_{,\bar{z}} = (-i\bar{A} + \tilde{B})G^{-1}. \tag{1.44b}$$

This system is precisely (1.41) with $\gamma^{1/2} = i$. Consequently, given a solution of the Ernst equation, the complete integrability of (1.41) ensures the existence of G^{-1} since $\gamma^{\lambda}_{,\bar{z}} = \gamma^{\lambda}_{,z}$ in general. However, in this case there is not necessarily any relationship between the column vectors comprising G^{-1}. The problem (1.43) has the same form as the problem given by Maison[18] [Eq. (2.19), his [γ is our $i\gamma^{-1/2}$]. The deformation problem (1.43) can be transformed into the Belinskii–Zakharov form[12] by reintroducing the variable η into the problem using (1.31). Thus defining

$$\Psi(z,\bar{z},\eta) \equiv U(z,\bar{z},\tfrac{1}{2}\{(\eta - \tfrac{1}{4}\eta^{-1}(z+\bar{z})^2) - i(z-\bar{z})\}), \tag{1.45}$$

we find that

$$\Psi_{,z} - \frac{2\eta}{(2i\eta - (z+\bar{z}))}\Psi_{,\eta} = \frac{-2(z+\bar{z})C\Psi}{(2i\eta - (z+\bar{z}))}, \tag{1.46a}$$
$$\Psi_{,\bar{z}} + \frac{2\eta}{(2i\eta + (z+\bar{z}))}\Psi_{,\eta} = \frac{2(z+\bar{z})\bar{C}\Psi}{(2i\eta + (z+\bar{z}))}, \tag{1.46b}$$

which is of the required form. One further point remains to be answered; this is the relationship between C and the corresponding coefficient in Refs. 12 or 18. The coefficients appearing in these last two papers are defined in terms of the components of the Lewis metric tensor (1.1),

$$h = \begin{pmatrix} f & \omega f\\ \omega f & \omega^2 f - x_1^2 f^{-1}\end{pmatrix}. \tag{1.47}$$

A calculation shows that corresponding to the deformation problem (1.36) for which

$$A = -\epsilon f^{-1}\{(f_{,z} - \tfrac{1}{2}fx_1^{-1})Z_0 - g_{,z}Z_2\}, \tag{1.48a}$$
$$B = -f^{-1}g_{,z}Z_1,\quad \tilde{B} = -\bar{B}, \tag{1.48b}$$

we have the relationship, when $\epsilon = +1$,

$$C = \tfrac{1}{2}(h_{,z}h^{-1} - \tfrac{1}{2}x_1^{-1}I), \tag{1.49a}$$
$$C \equiv -iGAG^{-1}$$

where

$$G = p(z,\bar{z})\begin{pmatrix} -ix_1 f^{-1}\omega^{-1} & 0\\ -1, & \omega^{-1}\end{pmatrix}. \tag{1.49b}$$

If we define $\tilde{A} = A + (i/4)x_1^{-1}I$, (1.50a)

then $\bar{C} \equiv -iG\tilde{A}G^{-1} = \tfrac{1}{2}h_{,z}h^{-1}$. (1.50b)

The introduction of $\tilde{A}$ corresponds to rescaling Φ in (1.36), $\tilde{\Phi} = ((1 - i\gamma^{1/2})/(1 + i\gamma^{1/2}))^{1/2}\Phi$. Equation (1.46) is then identical to the Belinskii–Zakharov problem.[12] The apparently arbitrary function p appearing in the definition of G^{-1}, (1.49b) is determined from (1.44) with A replaced by $\tilde{A}$. In fact G^{-1} can be determined for both of the cases $\epsilon = \pm 1$ and it is found that Eq. (1.50) is again valid with

$$G_{\epsilon}^{-1} = -\tfrac{1}{2}f^{-1/2}x_1^{-1}\begin{pmatrix}\omega f(\epsilon-1) + ix_1(\epsilon+1), & -f(\epsilon-1)\\ \omega f(\epsilon+1) - ix_1(\epsilon-1), & -f(\epsilon+1)\end{pmatrix}. \tag{1.51}$$

Maison's coefficient is obtained by introducing

$$\mu = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} h, \tag{1.52}$$

where upon we have the relationship

$$h_{,z} h^{-1} = -x_1^{-2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} (\mu\mu_{,z} - x_1 I) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{1.53}$$

Then defining

$$U' = (z+\bar{z})^{-1} \begin{pmatrix} 1 + i\gamma^{1/2} \\ 1 - i\gamma^{1/2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} U, \tag{1.54}$$

Eqs. (1.43) transform to Maison's deformation problem,

$$U'_{,z} = (-x_1^{-2}/2)(1 + i\gamma^{1/2})\mu\mu_{,z} U', \tag{1.55a}$$

$$U'_{,\bar{z}} = (-x_1^{-2}/2)(1 - i\gamma^{-1/2})\mu\mu_{,z} U'. \tag{1.55b}$$

2. THE RELATIONSHIP BETWEEN THE FINITE AND INFINITE DIMENSIONAL DEFORMATION PROBLEMS

In this section we give a geometrical interpretation of the mapping between the finite and infinite dimensional realizations of the prolongation algebra as defined by (1.32). To this end we introduce the complex variables $z = x_1 + ix_2, \bar{z}$ into (1.23), whereupon it assumes the form

$$y_{,z} - \frac{2\eta}{(2i\eta - (z+\bar{z}))} y_{,\eta} = \left[\left(\frac{(z+\bar{z}) + 2i\eta}{(z+\bar{z}) - 2i\eta} \right) A + B \right] y, \tag{2.1a}$$

$$y_{,\bar{z}} + \frac{2\eta}{(2i\eta + (z+\bar{z}))} y_{,\eta} = \left[i \left(\frac{(z+\bar{z}) - 2i\eta}{(z+\bar{z}) + 2i\eta} \right) \bar{A} + \tilde{B} \right] y, \tag{2.1b}$$

where $A = -f^{-1}(f_{,z} Z_0 + g_{,z} Z_2)$, $B = -f^{-1} g_{,z} Z_1$,

$$\tilde{B} = -f^{-1} g_{,\bar{z}} Z_1. \tag{2.1c}$$

This is easily obtained from (1.41) upon noting that

$$i\gamma^{1/2} = \left(\frac{(z+\bar{z}) + 2i\eta}{(z+\bar{z}) - 2i\eta} \right) \tag{2.2a}$$

and writing

$$y(z,\bar{z},\eta) \equiv w(z,\bar{z},\tfrac{1}{2}\{(\eta - \tfrac{1}{4}\eta^{-1}(z+\bar{z})^2) - i(z-\bar{z})\}), \tag{2.2b}$$

where w is the first column vector of Φ in (1.41).

Consider now the eight-dimensional manifold $\tilde{M}$ with local coordinates $(m^i) \equiv (z, \bar{z}, f, g, P_z, P_{\bar{z}}, Q_z, Q_{\bar{z}})$, where $H_z = \frac{1}{2}(H_1 - iH_2), H_{\bar{z}} = \frac{1}{2}(H_i + iH_2)$. Then corresponding to the exterior system $E(M)$ introduced in Sec. 1 there is an exterior system $E(\tilde{M})$ with generators $\{\theta_j\} \equiv \{\tilde{\alpha}_a, \tilde{\beta}_b, \tilde{\gamma}_c\}$ on $\tilde{M}$ which is equivalent to the original Ernst equation in the sense that $E(\tilde{M})$ is annulled when restricted to a solution manifold of the Ernst equation on which $dz \wedge d\bar{z} \neq 0$. The $\{\theta_j\}$ are easily calculated from the generators for $E(M)$. Thus, for example,

$$\tilde{\alpha}_1 = \tfrac{1}{2} d(P_z + P_{\bar{z}}) \wedge (dz + d\bar{z}) + \tfrac{1}{2} d(P_z - P_{\bar{z}}) \wedge (dz - d\bar{z}) = dP_z \wedge dz + dP_{\bar{z}} \wedge d\bar{z}. \tag{2.3}$$

Corresponding to the finite dimensional and infinite dimensional representations of the prolongation structure there are prolongations $E_k(N)$ and $E(P)$ of the exterior system $E(\tilde{M})$ defined on the manifolds $N = \tilde{M} \times W$, $P = \tilde{M} \times Y \times \Xi$, with local coordinates (m^i, w^1, w^2), (m^i, y^1, y^2, η), respectively. $E_k(N)$ is generated by $\{\theta_j, \omega_k\}$, where

$$\omega_k = dw + (\gamma_k^{1/2}(P_z Z_0 + Q_z Z_2) + Q_z Z_1) w\, dz + (-\gamma_k^{-1/2}(P_{\bar{z}} Z_0 + Q_{\bar{z}} Z_2) + Q_{\bar{z}} Z_1) w\, dz \tag{2.4a}$$

and

$$\gamma_k = (\bar{z} + ik)/(z - ik). \tag{2.4b}$$

That is, there is a one parameter family of prolongation forms. A solution manifold S of $E(\tilde{M})$ conveniently parametrized by $(z,\bar{z})$ such that $dz \wedge d\bar{z} \neq 0$ annuls $E(\tilde{M})$. In this case we see that when $E_k(N)$ is over S that ω_k is completely integrable,

$$d\omega_k = \theta \wedge \omega_k, \quad \theta \in \Lambda^1(\tilde{M}) \tag{2.5}$$

and we have the usual interpretation of ω_k as an SL $(2,\mathbb{C})$ connection on N which is flat when over S.

In the infinite dimensional case things are more complicated. For this purpose it is more natural to replace N by the manifold $Q = \tilde{M} \times W \times K$ with local coordinate (m^i, w^1, w^2, k) on which is defined the exterior system $E(Q)$ generated by $\{\theta_j, w\}$, where

$$w = dw \wedge dk + (\gamma^{1/2}(P_z Z_0 + Q_z Z_0) + Q_z Z_1) w\, dz \wedge dk + (-\gamma^{-1/2}(P_{\bar{z}} Z_0 + Q_{\bar{z}} Z_2) + Q_{\bar{z}} Z_1) w\, d\bar{z} \wedge dk. \tag{2.6}$$

The two form (2.6) pulls back to a two-form ϕ on P via the bundle map

$$\begin{array}{ccc} Q = \tilde{M} \times W \times K & \xleftarrow{\ h\ } & P = \tilde{M} \times W \times \Xi \\ \downarrow \pi_q & & \downarrow \pi_p \\ \tilde{M} \times W & \xleftarrow{\ Id\ } & \tilde{M} \times W \end{array}$$

which is given in local coordinates by $(m^i, w^1, w^2, \eta) \to (m^i, w^1, w^2, -(i/2)(z-\bar{z}) + (1/8\eta)(4\eta^2 - (z+\bar{z})^2))$. A simple calculation shows that

$$\phi = h^* w \equiv a\, dw \wedge dz + b\, dw \wedge d\bar{z} + c\, dw \wedge d\eta + (bC - aD) w\, dz \wedge d\bar{z} + Cw\, dz \wedge d\eta + cDw\, d\bar{z} \wedge d\eta, \tag{2.7a}$$

where

$$a = -(1/4\eta)(2i\eta + (z+\bar{z})), \quad b = (1/4\eta)(2i\eta - (z+\bar{z})),$$

$$c = (1/8\eta^2)(4\eta^2 + (z+\bar{z})^2)$$

and

$$C = \gamma^{1/2}(P_z Z_0 + Q_z Z_2) + Q_z Z_1,$$

$$D = -\gamma^{-1/2}(P_{\bar{z}} Z_0 + Q_{\bar{z}} Z_2) + Q_{\bar{z}} Z_1. \tag{2.7b}$$

If we restrict (2.7) to a solution manifold parametrized by z, $\bar{z}$ $dz \wedge d\bar{z} \neq 0$, which is also a solution manifold of $E(\tilde{M})$, then (2.7) gives rise to the deformation problem (2.7). When over S we see that ϕ satisfies a two-form version of the Frobenius complete integrability condition

$$d\phi = \phi \wedge \theta, \quad \theta \in \Lambda'(\tilde{M}). \tag{2.8}$$

It is interesting to note that h^{-1} is double valued,

$$\eta_{\pm} = k + (i/2)(z - \bar{z}) \pm \tfrac{1}{2}((z+\bar{z})^2 + (k + i(z-\bar{z}))^2)^{1/2}. \tag{2.9}$$

[1] F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[2] T. Lewis, Proc. R. Soc. London Ser. A **136**, 176 (1932).
[3] R. Geroch, J. Math. Phys. **12**, 918 (1971).
[4] R. Geroch, J. Math. Phys. **13**, 394 (1972).

[5]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977).
[6]W. Kinnersley and D. Chitre, J. Math. Phys. **18**, 1538 (1977).
[7]W. Kinnersley and D. Chitre, J. Math. Phys. **19**, 1926 (1978).
[8]W. Kinnersley and D. Chitre, J. Math. Phys. **19**, 2037 (1978).
[9]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978).
[10]D. Kramer and G. Neugebauer, Phys. Lett. A **75**, 259 (1980).
[11]V. E. Zakharov and A. V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1979) [Zh. Eksp. Teor. Fiz. **74**, 1953 (1978)].
[12]V. E. Zakharov and V. A. Belinskii, Sov. Phys. JETP **48**, 985 (1978) [Zh. Eksp. Teor. Fiz. **75**, 1955 (1978)].
[13]D. Maison, Phys. Rev. Lett. **41**, 521 (1978).
[14]H. C. Morris and R. K. Dodd, Phys. Lett. A **75**, 20 (1979).
[15]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975).
[16]R. Hermann, *Interdisciplinary Mathematics*, Vol. XIIA (Mathematics Science Press, Brookline, MA, 1976).
[17]R. K. Dodd and H. C. Morris, *Geometrical Approaches to Differential Geometry*, edited by R. Martini, *Lecture Notes in Mathematics*, Vol. 810 (Springer, Berlin, 1980), p. 63.
[18]D. Maison, J. Math. Phys. **20**, 871 (1979).

Spinor sources in cosmology [a)]

Robert T. Jantzen
Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138

(Received 1 December 1981; accepted for publication 7 January 1982)

An alternative Hamiltonian formulation is presented for the spatially homogeneous Einstein–Dirac system which in the nondegenerate case enables the number of gravitational degrees of freedom to be explicitly reduced to two.

PACS numbers: 04.20.Fy,98.80.Dr

1. INTRODUCTION

Teitelboim and Nelson[1] and Henneaux[2] have given a Hamiltonian formulation of the coupled Einstein–Dirac system in which the basic geometrical canonical variable is an orthonormal frame or "tetrad." However, an alternative formulation exists in which the usual gravitational canonical variables are retained and a fixed tetrad depending on those variables is introduced only to define the Dirac spinor fields. This can be useful when the symmetry of the class of spacetimes under consideration is such that a "natural" candidate for the tetrad is available. Such a formulation has recently been used by Ryan and Obregón[3] to study the symmetric case Bianchi type IX Einstein–Dirac system.

For spatially homogeneous spacetimes of all of the Bianchi types except I, II, and V, a natural parametrization of the spatial metric exists which is closely related to a preferred class of orthonormal frames adapted to the symmetry properties of these spacetimes.[4,5] These spatially homogeneous tetrads consist of the unit vector field normal to the family of homogeneous hypersurfaces and a triad tangent to that family (a "time gauge" tetrad[1] or "suited tetrad"[2]). The triad is characterized by the fact that the Lie brackets of its elements are in "standard diagonal form"[5] and represent a compromise between the desire to maintain the simplicity of the group properties of the triad while introducing as little rotation of the triad as possible with respect to one parallely propagated along the normal congruence, the latter choice being favored by Henneaux.[6]

For the Bianchi types I, II, and V, there is some freedom involved in choosing a parametrization of the spatial metric adapted to the symmetry. One may arbitrarily fix this freedom as is done here and hence single out an associated triad, but the resulting discussion is complicated by the symmetries which have been ignored. In particular, one cannot easily exploit the constants of the motion which arise for these special types. Here the approach of Henneaux[6] is perhaps more suitable.

The natural parametrization of the metric divides the gravitational variables into two sets, the scale variables and automorphism variables, the latter being associated with the symmetry of the dynamics. In the Einstein–Dirac system the automorphism "coordinates" do not explicitly appear due to the symmetry of the system, leaving only their velocities (and momenta) to influence the remaining variables. This is slightly different from the Einstein-perfect fluid and Einstein–Maxwell systems, where a change of source variables is first required to eliminate the automorphism "coordinates".[4] In this sense the Einstein–Dirac case is more similar to the vacuum case, where the automorphism symmetry is not broken. However, due to the supermomentum constraints it is only in the degenerate Bianchi types I, II, V, and $VI_{-1/9}$ that any of the associated constants of the motion are allowed to be nonvanishing. Unfortunately, it is precisely in the first three cases that the choice of variables is not well suited to the exploitation of these constants of the motion.

The aim of the present paper is to extend the Lagrangian/Hamiltonian formulation of spatially homogeneous dynamics described in Refs. 4 and 5 to the Einstein–Dirac case, in which a spatially homogeneous classical Dirac spinor field acts as the source of the spatially homogeneous gravitational field, the notable new feature here being the occurrence of derivative coupling. The derivative coupling involves only the automorphism velocities, which are responsible for the rotation of the triad. In the class B case the spinor field contributes to the nonpotential force[4,5,7] which drives the Langrangian/Hamiltonian system for the gravitational variables, while its own equation of motion obtained from the variational principle differs from the Dirac equation by a single term. The Dirac equation must therefore be imposed separately in this case.

In Sec. 2, the metric parametrization and preferred tetrad are introduced together with Dirac spinor fields. In Sec. 3, the Lagrangian/Hamiltonian analysis is carried out and its relation to the more conventional approach is described, while Sec. 4 studies the allowable special cases of the general system of equations. Section 5 discusses the possibility of considering "twisted" Dirac spinor fields as sources. The notation and results established in Ref. 5 are assumed in the present paper, which is intended as a sequel to that earlier work. In the nondegenerate case here, as described in Ref. 5 for a general nonderivatively coupled source, the number of gravitational degrees of freedom may also be reduced to two by explicitly solving the supermomentum and super-Hamiltonian constraints.

Since the spatially homogeneous Dirac field is taken to be an ordinary commuting classical field, it does not satisfy any of the usual energy conditions required of reasonable matter. In particular the energy–momentum tensor vanishes identically when the axial current vanishes, leading to

[a)]Part of this work was done while the author was a NATO postdoctoral fellow at the University of Rome and the remainder while a visitor at the Max-Planck-Institut für Astrophysik, München, Federal Republic of Germany. Partially supported by NSF Grant No. PHY-80-07351.

 0022-2488/82/061137-10$02.50

"ghost solutions" of the Einstein–Dirac system[3,8] consisting of a vacuum solution of the Einstein equations on which a Dirac field with vanishing energy-momentum tensor propagates. This is a rather clear sign that the classical Einstein–Dirac system, although very interesting mathematically, cannot be taken seriously from the physical point of view.

2. THE PREFERRED TETRAD

The metric of a spatially homogeneous spacetime $(R \times G, {}^4\mathrm{g})$ may be written in the following form:

$$ {}^4\mathrm{g} = -dt \otimes dt + g_{ab}\omega^a \otimes \omega^b, \tag{2.1}$$

where $\{\omega^a\}$ is the basis of left invariant 1-forms dual to a basis $e = \{e_a\}$ of the Lie algebra $\mathfrak{g}$ of left invariant vector fields on a 3-dimensional (for simplicity, simply connected) Lie group G and the positive-definite matrix $\mathbf{g} = g_{ab}\mathbf{e}^b{}_a$ depends only on the time t. The frame $\{e_\alpha\} = \{e_0 = \partial/\partial t = \partial_0, e_a\}$ is a comoving ADM frame for the spatially homogeneous slicing, which simply means that the "reduced frame" e is tangent to the hypersurfaces of constant t (the group orbits of the natural left action of G on $R \times G$) and has vanishing Lie brackets with e_0, which in this case happens to be the unit normal $e_\perp$ to those hypersurfaces. Let $\omega^\perp = \omega^0 = dt$.

The remaining Lie brackets define the components of the structure constant tensor of $\mathfrak{g}$ in the basis e:

$$ [e_a, e_b] = C^c{}_{ab} e_c, \tag{2.2}$$

which may be written in the well-known form[9]

$$ C^c{}_{ab} = \epsilon_{abd} n^{cd} + a_f \delta^{fc}_{ab},$$
$$ a_f = \tfrac{1}{2} C^b{}_{fb} \quad n^{ab} = C^{(a}{}_{cd}\epsilon^{b)cd}, \quad a_f n^{fa} = 0. \tag{2.3}$$

When the symmetric matrix $\mathbf{n} = n^{ab}\mathbf{e}^b{}_a$ is diagonal, i.e., $\mathbf{n} = \mathrm{diag}\,(n^{(1)}, n^{(2)}, n^{(3)})$, and $a_f = a\delta^3{}_f$, the components $C^a{}_{bc}$ are said to be in standard diagonal form. They are called canonical components and e a canonical basis of $\mathfrak{g}$ when in addition they assume the canonical values given in Ref. 5 for each Bianchi type Lie group G. The matrix representation $\mathrm{Aut}_e(\mathfrak{g})$ with respect to the basis e of the automorphism group $\mathrm{Aut}(\mathfrak{g})$ of the Lie algebra $\mathfrak{g}$ is just that subgroup of $\mathrm{GL}(3,R)$ which leaves $C^a{}_{bc}$ fixed under the natural action of $\mathrm{GL}(3,R)$ on these components. $\mathrm{Aut}_e(\mathfrak{g})$ for any canonical basis is called the canonical automorphism matrix group and $\mathrm{SAut}_e(\mathfrak{g})$ designates its unimodular subgroup. Unless otherwise stated, e is assumed to be a canonical basis.

As described in Refs. 4 and 5, the special automorphism matrix group $\mathrm{SAut}_e(\mathfrak{g})$ is the symmetry group of the ordinary differential equations satisfied by the metric matrix $\mathbf{g}$ when no sources are present. For all but the degenerate Bianchi types I, II, and V, the canonical special automorphism matrix group $\hat{G}$ is 3-dimensional and has off-diagonal generators $\{\boldsymbol{\kappa}_a\}$ which permit the following parametrization of the metric matrix in terms of diagonal matrix $\mathbf{g}' = e^{2\boldsymbol{\beta}}$ and a general element $\mathbf{S}$ of $\hat{G}$:

$$ \mathbf{g} = \mathbf{S}^T\mathbf{g}'\mathbf{S} = \mathbf{H}^T\mathbf{H}, \quad \mathbf{H} = e^{\boldsymbol{\beta}}\mathbf{S},$$
$$ \boldsymbol{\beta} = \mathrm{diag}(\beta^1, \beta^2, \beta^3) = \beta^A \mathbf{e}_A = \beta^0 \mathbf{e}_0 + \beta^+ \mathbf{e}_+ + \beta^- \mathbf{e}_-,$$
$$ \{\mathbf{e}_0, \mathbf{e}_+, \mathbf{e}_-\} = \{\mathbf{1}, \mathrm{diag}(1,1,-2), \mathrm{diag}(\sqrt{3}, -\sqrt{3}, 0)\}. \tag{2.4}$$

This parametrization can be interpreted in terms of two successive transformations of the canonical basis e:

$$ e'_a = S^{-1}{}^b{}_a e_b, \quad \omega'^a = S^a{}_b \omega^b,$$
$$ e''_a = (e^{-\boldsymbol{\beta}})^b{}_a e'_b = H^{-1}{}^b{}_a e_b, \quad \omega''^a = (e^{\boldsymbol{\beta}})^a{}_b \omega'^b = H^a{}_b \omega^b. \tag{2.5}$$

$e' = \{e'_a\}$ is an orthogonal time-dependent canonical basis of $\mathfrak{g}$ which can be completed to a comoving ADM frame $\{e'_0, e'_a\}$ by adjoining a certain not necessarily spatially homogeneous vector field $e'_0 = e_\perp + N'^a e'_a$ as described in Refs. 4 and 5. $\mathbf{g}'$ is the metric matrix in the new comoving frame while $N'^a e'_a$ is the shift vector field for the frame and is determined algebraically by the "automorphism velocities" $\{\dot{\tilde{\omega}}^a\}$, which are defined by

$$ \dot{\mathbf{S}}\mathbf{S}^{-1} = \boldsymbol{\kappa}_a \dot{\tilde{\omega}}^a. \tag{2.6}$$

On the other hand, $\{e_0, e'_a\}$ is an orthogonal spacetime frame which may be normalized to produce the tetrad $\{e''_\alpha\} = \{e_0, e''_a\}$. $e'' = \{e''_a\}$ is a triad with time-dependent structure constant tensor components:

$$ C''^a{}_{bc} = (e^{\boldsymbol{\beta}})^a{}_d C^d{}_{fg} (e^{-\boldsymbol{\beta}})^f{}_b (e^{-\boldsymbol{\beta}})^g{}_c, \tag{2.7}$$

which are still in standard diagonal form since they differ from canonical components only by a scaling (hence the term "scale variables" for $\boldsymbol{\beta}$). Since $\hat{G}$ is a symmetry group, the components of all geometrical quantities in the frame $\{e''_\alpha\}$ do not explicitly involve $\mathbf{S}$, except implicitly through the automorphism velocities. In fact, by the introduction of the equivalent shift vector field they determine, one can obtain the spacetime metric without knowledge of $\mathbf{S}$.

The components of the metric connection in the tetrad $\{e''_\alpha\}$ are given in Appendix A. The components Γ''_{a0b} determine the angular velocity of the triad e'' relative to a triad parallely transported along $e_\perp$:

$$ De''_a/dt = \nabla_{e_0} e''_a = \Gamma''_{b0a} e''_b = -\epsilon_{abc} \Sigma''^b e''_c,$$
$$ \Sigma''^a = -\tfrac{1}{2}\kappa''_{d[bc]}\epsilon^{abc}\dot{\tilde{\omega}}^d \equiv \Sigma''^a{}_d \dot{\tilde{\omega}}^d. \tag{2.8}$$

Clearly, the automorphism velocities are responsible for the rotation of the triad.

Any other spatially homogeneous triad e''' is related to e'' by some (generally time-dependent) rotation $\mathbf{R} \in \mathrm{SO}(3,R)$:

$$ e'''_a = \mathscr{R}^{-1}{}^b{}_a e''_b. \tag{2.9}$$

The angular velocity of the triad then becomes

$$ \Sigma'''^a = \mathscr{R}^a{}_b (\Sigma''^b + \dot{\sigma}^b),$$
$$ \mathbf{R}^{-1}\dot{\mathbf{R}} = \mathbf{k}_a{}^{\mathrm{IX}} \dot{\sigma}^a, \quad \mathbf{k}_a{}^{\mathrm{IX}} = \epsilon_{bac}\mathbf{e}^c{}_b. \tag{2.10}$$

However, unless the spacetime metric has additional symmetry, the new structure constant tensor components $C'''^a{}_{bc}$ will no longer have the simple standard diagonal form and will depend explicitly on the matrix $\mathbf{R}$. If there is additional symmetry, this form may be preserved but will generally increase the angular velocity of the triad, as discussed below.

For Bianchi types I, II, and V, $\mathrm{SAut}_e(\mathfrak{g})$ has dimension 8,5, and 5, respectively. For Bianchi type I, $\mathrm{SAut}_e(\mathfrak{g}) = \mathrm{SL}(3,R)$, while in the other two cases, respectively, $\mathrm{SAut}_e(\mathfrak{g})$ is the semidirect product group $(T_3)^T \times_s \mathrm{SL}(2)_3$ and $(T_3) \times_s \mathrm{SL}(2)_3$, where $\mathrm{SL}(2)_3$ is the $\mathrm{SL}(2,R)$ subgroup of $\mathrm{SL}(3,R)$ which leaves the third axis of R^3 invariant under its

natural action on R^3 while T_3 is the 2-dimensional abelian subgroup generated by $\{\mathbf{e}^3{}_1,\mathbf{e}^3{}_2\}$. A suitable but nonunique candidate $\hat{G}$ for the parametrization (2.4) is obtained by replacing the special linear groups SL(3,R) and SL(2)$_3$, respectively, by their special orthogonal subgroups SO(3,R) and SO(2)$_3$. However, all rotations $\mathbf{R}$ belonging to the special orthogonal subgroup occurring in the parametrization leave the structure constant tensor components in standard diagonal form and so may be used to decrease the rotation of the triad while preserving that form. For example, one may always choose $\Sigma'''^a = 0$ in the type I case and $\Sigma'''^3 = 0$ in the others. This freedom leads to the problem that improper choice of initial data can introduce "spurious time-dependence",[4,5] masking the fact that fewer independent functions of time may be involved in the solution than seem to appear.

Using the tetrad $\{e''_\alpha\}$, Dirac spinor fields are introduced in the usual way. Let $\{\gamma_\alpha\}$ be a set of Dirac matrices satisfying $\gamma_{(\alpha}\gamma_{\beta)} = \eta_{\alpha\beta}$ with $\boldsymbol{\eta} = (\eta_{\alpha\beta}) = \mathrm{diag}(-1,1,1,1)$ and introduce the notation $\gamma_{\alpha\cdots\beta} = \gamma_{[\alpha\cdots\beta]}, \gamma_5 = \gamma_{0123}$. Let ψ be the column vector of components with respect to $\{e''_\alpha\}$ of a spatially homogeneous Dirac spinor field, i.e., $e''_a\psi \equiv \partial''_a\psi = 0$, and let $\bar{\psi} = \psi^\dagger\gamma_0$ be the conjugate spinor. The weighted spinor $\Psi = g^{1/4}\psi = e^{3/2\beta^0}\psi$ satisfies simpler differential equations and enters the Hamiltonian formalism more naturally. The covariant derivatives of ψ and $\bar{\psi}$ are defined by

$$\nabla''_\alpha\psi = \nabla_{e''_\alpha}\psi = (\partial''_\alpha + \tfrac{1}{4}\Gamma''_{\beta\alpha\gamma}\gamma^{\beta\gamma})\psi,$$
$$\nabla''_\alpha\bar{\psi} = \partial''_\alpha\bar{\psi} - \tfrac{1}{4}\bar{\psi}\gamma^{\beta\gamma}\Gamma''_{\beta\alpha\gamma}. \tag{2.11}$$

It is also convenient to introduce the following standard notation for derivatives of spinor expressions:

$$A\overset{\leftrightarrow}{\nabla}B = A(\nabla B) - (\nabla A)B. \tag{2.12}$$

Recall that the Dirac matrices behave like covariant constants.

The two equivalent forms for the Dirac equation may be evaluated using the expressions for the connection components given in Appendix A:

$$0 = g^{1/4}\gamma_0(\gamma^\alpha\nabla''_\alpha - m)\psi = \dot{\Psi} + \tfrac{1}{2}(\Sigma'''^a\gamma_a{}^0\gamma_5 - \Gamma'''^a{}_a\gamma^0\gamma_5 + 2a''_c\gamma^{0c} + 2m\gamma^0)\Psi,$$
$$0 = g^{1/4}(-\nabla''_\alpha\bar{\psi}\gamma^\alpha - m\bar{\psi})\gamma_0 = -\dot{\bar{\Psi}} + \tfrac{1}{2}\bar{\Psi}(\Sigma'''^a\gamma_a{}^0\gamma_5 + \Gamma'''^a{}_a\gamma^0\gamma_5 + 2a''_c\gamma^{0c} + 2m\gamma^0). \tag{2.13}$$

The current density $\mathscr{J}$ and axial current $4\mathscr{A}$ are defined by

$$\mathscr{J}'''^\alpha = \bar{\Psi}\gamma^\alpha\Psi, \quad 4\mathscr{A}'''^\alpha = i\bar{\Psi}\gamma^\alpha\gamma_5\Psi. \tag{2.14}$$

By choosing a particular representation of the Dirac matrices, one can show that if $\mathscr{A}'''^a = 0$ at any time, the quantities $\mathscr{A}'''^0$, $\bar{\Psi}\Psi$, and $\bar{\Psi}\gamma_5\Psi$ will also vanish.[3] The equations of Appendix B then show that they remain zero for all times. The next section shows that such solutions have a vanishing energy–momentum tensor and hence correspond to the "ghost solutions" mentioned in the Introduction.

Define $\hat{\sigma}_a = -\frac{1}{2}i\epsilon_{abc}\gamma^{bc}$. The matrices $\{\hat{\sigma}_a\}$ have the same commutation relations as the standard Pauli matrices $\{\sigma_a\}$ and $\{-\frac{1}{2}i\hat{\sigma}_a\}$ is a basis of the real Lie subalgebra of the Dirac algebra which is isomorphic to the Lie algebra of SU(2). This basis has the same structure constant tensor components as the canonical basis $\{\mathbf{k}_a{}^{\mathrm{IX}}\}$ of the Lie algebra of SO(3,R). If the rotation of (2.9) is given by $\mathscr{R} = \exp\theta^a\mathbf{k}_a{}^{\mathrm{IX}}$, the components of ψ transform in the following way:

$$\psi''' = \mathscr{U}\psi, \quad \mathscr{U} = \exp(-\tfrac{1}{2}i\theta^a\hat{\sigma}_a). \tag{2.15}$$

The matrix $\mathscr{U}$ is one of the two matrices (namely $\mathscr{U}$ and $-\mathscr{U}$) which satisfy

$$\mathscr{U}\gamma_a\mathscr{U}^{-1} = \gamma_b\mathscr{R}^b{}_a. \tag{2.16}$$

The parametrization (2.4) has singularities when the subgroup $\hat{G}\cap$SO(3,R) is nontrivial, which occurs for Bianchi types I, II, V, VII$_0$, VII$_h$, VIII, and IX. For all types but I and IX, $\boldsymbol{\kappa}_3 = \mathbf{k}_3{}^{\mathrm{IX}}$ generates this subgroup and the singularity occurs for $e^{2\beta} \in \mathscr{M}_{T(3)}$, i.e., $\beta^- = 0$, and $\mathbf{S}$ is determined only up to left multiplication by an arbitrary element of the subgroup. This presents difficulties only if β^- remains zero for all time (if β^- passes through zero, $\mathbf{S}$ is determined by continuity), the case of local rotational symmetry when there exists an additional spacetime Killing vector field which occurs only when $\mathbf{g}$ is at least diagonal for all times. Any $\mathbf{S}\in\hat{G}\cap$SO(3,R) then satisfies

$$\mathbf{S}^T e^{2\boldsymbol{\beta}}\mathbf{S} = e^{2\boldsymbol{\beta}}, \quad \boldsymbol{\beta}^- = 0. \tag{2.17}$$

The velocity $\dot{\omega}^3$ is therefore arbitrary. However, it is logical to choose $\mathbf{S} = \mathbf{1}$, i.e., $\dot{\omega}^3 = 0$, leading to a parallely propagated triad. Any other choice of triad (except those differing by a time-independent rotation) will then have nonzero angular velocity. For types I and IX, $\hat{G} = $ SO(3,R) and singularities occur for $e^{2\boldsymbol{\beta}}$ lying in each of the Taub submanifolds $\mathscr{M}_{T(a)}$ [where $\beta^{bc} \equiv \beta^b - \beta^c = 0$ and (a,b,c) is a cyclic permutation of (1,2,3)]. When there is additional symmetry, the freedom in $\mathbf{S}$ may again be fixed by choosing $\mathbf{S} = \mathbf{1}$.

For Bianchi type IX, $\mathbf{S}^T = \mathbf{S}^{-1}$ and so

$$\mathbf{g} = \exp 2\mathbf{S}^{-1}\boldsymbol{\beta}\mathbf{S} = \mathbf{H}_R{}^T\mathbf{H}_R, \quad \mathbf{H}_R = \mathbf{S}^{-1}\mathbf{H}. \tag{2.18}$$

The triad $\{e'''_a = H_R{}^{-1b}{}_a e_b\}$ has $\mathscr{R} = \mathbf{S}^{-1}$ and $\dot{\sigma}^a = -\dot{\tilde{\omega}}^a$ so that

$$\Sigma'''^a = S^{-1a}{}_b(\Sigma''^b{}_c - \delta^b{}_c)\dot{\tilde{\omega}}^c. \tag{2.19}$$

Using the expressions for $\Sigma'''^a{}_b$ given in Appendix A (with $\mathbf{n} = \mathbf{1}$), one sees that $\Sigma'''^a\Sigma'''_a < \Sigma''^a\Sigma''_a$, i.e., the angular velocity of e''' is less than that of e''. However, unless Σ'''^a is itself invariant under the rotation $\mathbf{S}$, which is the case when $\mathbf{S}$ is confined to a 1-dimensional subgroup, it will depend explicitly on $\mathbf{S}$ and therefore $\mathbf{S}$ appears explicitly in the Einstein-Dirac system, preventing its elimination as discussed in the next section. Ryan and Obregón have chosen the triad e''' for the type IX symmetric case, where $\mathbf{S}$ is confined to a 1-

dimensional subgroup, and this problem does not occur.[3]

Note that in the class A case, the structure constant tensor enters the Dirac equation only through the rotationally invariant expression $2\Gamma''^{a}{}_{a} = g^{-1/2}\,\mathrm{Tr}\,\mathbf{n}''$, which therefore is independent of the choice of triad. However, unless $S \in SAut_e(\mathfrak{g})$, this value will itself depend explicitly on S as well as the scale variables.

3. LAGRANGIAN/HAMILTONIAN ANALYSIS

The Lagrangian/Hamiltonian analysis of the dynamical Einstein equations given in Ref. 5 using the parametrization (2.4) is easily extended to the Einstein–Dirac case by first evaluating the spinor contribution to the ADM Lagrangian and then passing to the Hamiltonian formulation. In the class B case the Lagrangian and Hamiltonian equations must of course be checked against the Einstein–Dirac equations to see what modifications are required to make them correct. The Dirac Lagrangian, super-Hamiltonian, supermomentum, and spatial energy–momentum tensor must therefore be evaluated. The gravitational constant κ defined by $G_{\alpha\beta} = \kappa T_{\alpha\beta}$ is left unspecified, although other treatments of the Einstein–Dirac system choose $2\kappa = 1$. The choice of $L_{ADM} = 2\kappa L^{\text{true}}_{ADM}$ as the gravitational Lagrangian[10] requires that the usual matter Lagrangian be multiplied by the factor 2κ as well.

The components of the energy–momentum tensor of the spinor field are given by

$$T''_{\alpha\beta} = -\tfrac{1}{2}i\bar{\psi}\gamma_{(\alpha}\overleftrightarrow{\nabla}''_{\beta)}\psi\,. \tag{3.1}$$

The spatial components are found to be

$$g^{1/2}T''_{ab} = -\Gamma''_{(ab)}\mathscr{A}''^{0} + 2K''^{c}{}_{(a}\epsilon_{b)cd}\mathscr{A}''^{d}\,. \tag{3.2}$$

The supermomentum and super-Hamiltonian are evaluated using the Dirac equation to trade time derivatives for spatial derivatives:

$$\begin{aligned}
\mathscr{H}^{1/2}_{a''} &= -2\kappa g^{1/2}T^{\perp}{}_{a''} = 2\kappa(\tfrac{1}{2}i\bar{\Psi}\gamma^{0}\overleftrightarrow{\nabla}''_{a}\Psi + \nabla''_{c}\mathscr{H}''^{c}_{1/2\,a})\\
&= -2\kappa\,\mathrm{Tr}\,\boldsymbol{\delta}''_{a}\mathscr{H}''_{1/2}\,,\\
\mathscr{H}''^{ab}_{1/2} &= \epsilon^{abc}\mathscr{A}''_{c}\,,\quad \mathscr{H}''_{1/2} = \mathscr{H}''^{a}_{(1/2)b}\mathbf{e}^{b}{}_{a}\,,\\
\mathscr{H}^{1/2} &= -2\kappa g^{1/2}T^{\perp}{}_{\perp} = -i\kappa(\bar{\Psi}\gamma^{c}\overleftrightarrow{\nabla}''_{c}\Psi - 2m\bar{\Psi}\Psi)\\
&= 2\kappa(im\bar{\Psi}\Psi - \Gamma''^{a}{}_{a}\mathscr{A}''^{0})\,.
\end{aligned} \tag{3.3}$$

The matrix notation used here is explained in Appendix A. The components of the supermomentum with respect to the triad e' are needed for the discussion of the supermomentum constraints:

$$\mathscr{H}^{1/2}_{a'} = -2\kappa\,\mathrm{Tr}\,\boldsymbol{\delta}_{a}\mathscr{H}'_{1/2}\,. \tag{3.4}$$

The following Dirac Lagrangian is chosen:

$$\begin{aligned}
L^{1/2} &= 2\kappa i g^{1/2}\bar{\psi}(\tfrac{1}{2}\gamma^{\alpha}\overleftrightarrow{\nabla}''_{\alpha} - m)\psi\\
&= i\kappa\bar{\Psi}\gamma^{0}\overleftrightarrow{\partial}_{0}\Psi + 4\kappa\Sigma''^{a}\mathscr{A}''_{a} - \mathscr{H}^{1/2}\,.
\end{aligned} \tag{3.5}$$

By inspection one can see that the Lagrange derivatives of $L^{1/2}$ with respect to Ψ and $\bar{\Psi}$ produce the Dirac equations (2.13) only in the class A case where a''_{c} vanishes. The term involving a''_{c} is missing from each of the Dirac equations because of the choice of $\frac{1}{2}\overleftrightarrow{\nabla}$ in the Lagrangian; however, either of the choices ∇ or $-\overleftarrow{\nabla}$ will lead to at least one wrong Dirac equation as well as the wrong expressions for the derivatives of $L^{1/2}$ with respect to the gravitational variables. One is therefore forced to impose the Dirac equations by hand in the class B case.

Let the total Lagrangian, super-Hamiltonian, and supermomentum be the sum of the gravitational expressions given in Ref. 5 and the Dirac expressions given above:

$$\begin{aligned}
&L = L^{G} + L^{1/2}\,,\quad \mathscr{H} = \mathscr{H}^{G} + \mathscr{H}^{1/2}\,,\\
&\mathscr{H}_{a'} = \mathscr{H}^{G}_{a'} + \mathscr{H}^{1/2}_{a'}\,.
\end{aligned} \tag{3.6}$$

The presence of the automorphism velocities in the Dirac Lagrangian changes the relationship between the canonical automorphism momenta and the velocities; this does not occur for the scale variables.

$$\begin{aligned}
&P_{a} = \partial L^{G}/\partial\dot{\tilde{\omega}}^{a} = 2e^{3\beta^{0}}\mathscr{G}_{ab}\dot{\tilde{\omega}}^{b}\,,\\
&\Pi_{a} = \partial L/\partial\dot{\tilde{\omega}}^{a} = P_{a} + X_{a}\,,\\
&X_{a} = 2\kappa\,\mathrm{Tr}\,\boldsymbol{\kappa}''_{a}\mathscr{H}''_{1/2} = 4\kappa\mathscr{A}''_{b}\Sigma''^{b}{}_{a}\,.
\end{aligned} \tag{3.7}$$

Using (A9) and (A10) leads to another useful expression for Π_{a}.

$$\Pi_{a} = 2\,\mathrm{Tr}\,\boldsymbol{\kappa}_{a}(\boldsymbol{\pi}' + \kappa\mathscr{H}'_{1/2}) \equiv 2\,\mathrm{Tr}\,\boldsymbol{\kappa}_{a}\mathbf{P}'\,. \tag{3.8}$$

The noncanonical "coordinates" $\{\mathbf{S},\dot{\tilde{\omega}}^{a}\}$ and $\{\mathbf{S},P_{a}\}$ on the velocity and momentum phase spaces are employed here. These may be used to perform calculations by first imagining a local coordinate system $\{\theta^{i}\}$ on $\hat{G}$ with velocities $\dot{\theta}^{i}$ and mechanical momenta p_{i}, in terms of which one has[4]

$$\begin{aligned}
&\dot{\tilde{\omega}}^{a} = \tilde{\omega}^{a}{}_{i}\dot{\theta}^{i}\,,\quad P_{a} = \tilde{e}^{i}{}_{a}p_{i}\,,\quad \Pi_{a} = \tilde{e}^{i}{}_{a}\Pi_{i}\,,\\
&p_{i} = \partial L^{G}/\partial\dot{\theta}^{i}\,,\quad \Pi_{i} = \partial L/\partial\dot{\theta}^{i}\,,
\end{aligned} \tag{3.9}$$

where $\{\tilde{e}_{a} = \tilde{e}^{i}{}_{a}\partial_{i}\}$ and $\{\tilde{\omega}^{a} = \tilde{\omega}^{a}{}_{i}d\theta^{i}\}$ are a basis and corresponding dual basis of the Lie algebra of right invariant vector fields on $\hat{G}$ determined by the basis $\{\boldsymbol{\kappa}_{a}\}$ of the matrix Lie algebra:

$$\begin{aligned}
&[\boldsymbol{\kappa}_{a},\boldsymbol{\kappa}_{b}] = \hat{C}^{c}{}_{ab}\boldsymbol{\kappa}_{c}\,,\quad [\tilde{e}_{a},\tilde{e}_{b}] = -\hat{C}^{c}{}_{ab}e_{c}\,,\\
&d\mathbf{S}\,\mathbf{S}^{-1} = \boldsymbol{\kappa}_{a}\tilde{\omega}^{a}\,.
\end{aligned} \tag{3.10}$$

Since the ordinary canonical momenta Π_{i} have vanishing Poisson brackets and $\{\theta^{i},\Pi_{j}\} = \delta^{i}_{j}$, one finds

$$\{\Pi_{a},\Pi_{b}\} = \hat{C}^{c}{}_{ab}\Pi_{c}\,,\quad \{\mathbf{S},\Pi_{a}\} = \boldsymbol{\kappa}_{a}\mathbf{S}\,. \tag{3.11}$$

Furthermore, since only the automorphism velocities $\dot{\tilde{\omega}}^{a}$ enter the Lagrangian, one may introduce the generalized Lagrange derivative

$$\begin{aligned}
\delta L/\delta\tilde{\omega}^{a} &= -e^{i}{}_{a}[(\partial L/\partial\dot{\theta}^{i})^{\cdot} - \partial L/\partial\theta^{i}]\\
&= -(\partial L/\partial\dot{\tilde{\omega}}^{a})^{\cdot} + \partial L/\partial\dot{\tilde{\omega}}^{c}\hat{C}^{c}{}_{ab}\dot{\tilde{\omega}}^{b}\\
&= -\dot{\Pi}_{a} + \Pi_{c}\hat{C}^{c}{}_{ab}\dot{\tilde{\omega}}^{b}\,.
\end{aligned} \tag{3.12}$$

By direct calculation one finds

$$\begin{aligned}
&\partial L^{1/2}/\partial g'_{aa} = \kappa g^{1/2}T'^{aa}\,,\qquad \text{(no sum on } a\text{)}\\
&-\delta L^{1/2}/\delta\dot{\tilde{\omega}}^{a} = 2\kappa g^{1/2}\,\mathrm{Tr}\,\boldsymbol{\kappa}_{a}\mathbf{T}' + Q^{1/2}_{a}\,,\\
&Q^{1/2}_{a} = 8\kappa a''_{b}\Sigma''^{b}{}_{a}\mathscr{A}''^{0}\,.
\end{aligned} \tag{3.13}$$

Since the Lagrange derivatives of the gravitational Lagrangian are given by[4]

$$-\delta L^{G}/\delta\beta^{A} = -2g^{1/2}\,\mathrm{Tr}\,\dot{\mathbf{e}}_{A}{}^{4}\mathbf{G}' + Q^{*}_{A},$$
$$-\delta L^{G}/\delta\dot{\tilde{\omega}}^{a} = -2g^{1/2}\,\mathrm{Tr}\,\boldsymbol{\kappa}_{a}{}^{4}\mathbf{G}' + Q^{*}_{a},$$
$$Q^{*}_{A} = \delta^{+}_{A}Q^{*}_{+}, \quad Q^{*}_{a} = -2g^{1/2}\,\mathrm{Tr}\,\boldsymbol{\kappa}_{a}{}^{3}\mathbf{G}' = \delta^{3}{}_{a}Q^{*}_{3}, \tag{3.14}$$

the correct dynamical Einstein equations are

$$-\delta L/\delta\beta^{A} = Q_{A} \equiv Q^{*}_{A}, \quad -\delta L/\delta\dot{\tilde{\omega}}^{a} = Q_{a} \equiv Q^{*}_{a} + Q^{1/2}_{a}. \tag{3.15}$$

The momentum canonically conjugate to the Dirac spinor field Ψ is

$$\pi_{\Psi} = \partial L/\partial\dot{\Psi} = 2\kappa i\bar{\Psi}\gamma^{0}. \tag{3.16}$$

However, the passage from the Lagrangian to the Hamiltonian is most clearly accomplished following Nelson and Teitelboim[1] and assuming a real representation of the Dirac matrices and splitting Ψ into its real and imaginary parts:

$$\Psi = \phi + i\chi, \quad \pi_{\phi} = 2\kappa\chi^{T}, \quad \pi_{\chi} = -2\kappa\phi^{T},$$
$$\pi_{\phi}\dot{\phi} + \pi_{\chi}\dot{\chi} = \kappa i\,\bar{\Psi}\gamma^{0}\overset{\leftrightarrow}{\partial}_{0}\Psi,$$
$$\pi_{\phi}\dot{\phi} + \pi_{\chi}\dot{\chi} - L^{1/2} = \mathscr{H}^{1/2} - (\Pi_{a} - P_{a})\,\dot{\tilde{\omega}}^{a}. \tag{3.17}$$

The Hamiltonian is then defined in the usual way:

$$H = p_{A}\,\dot{\beta}^{A} + \Pi_{a}\,\dot{\tilde{\omega}}^{a} + \pi_{\phi}\dot{\phi} + \pi_{\chi}\dot{\chi} - L$$
$$= (p_{A}\,\dot{\beta}^{A} + P_{a}\,\dot{\tilde{\omega}}^{a} - L^{G}) + H^{1/2}$$
$$= \mathscr{H}^{G}(\beta^{A}, p_{A}, P_{a}) + \mathscr{H}^{1/2} = \mathscr{H}. \tag{3.18}$$

The result is just the total super-Hamiltonian, where the gravitational super-Hamiltonian is the same function of the mechanical momenta P_a as in the vacuum case. However, P_a must be interpreted as shorthand for $\Pi_a - X_a$ when it appears in the Hamiltonian which is a function on the momentum phase space where Π_a and not P_a are generalized canonical coordinates. For example, $\{\Pi_a, \Psi\}$ vanishes while $\{P_a, \Psi\}$ does not. Taking into account the nonpotential force which drives the Lagrangian equations, the Hamiltonian equations for the gravitational variables are given by

$$\begin{bmatrix}\dot{\beta}^{A}\\ \dot{\mathbf{S}}\end{bmatrix} = \left\{\begin{bmatrix}\beta^{A}\\ \mathbf{S}\end{bmatrix}, \mathscr{H}\right\}\begin{bmatrix}\dot{p}_{A}\\ \dot{\Pi}_{a}\end{bmatrix} = \left\{\begin{bmatrix}p_{A}\\ \Pi_{a}\end{bmatrix}, \mathscr{H}\right\} + \begin{bmatrix}Q_{A}\\ Q_{a}\end{bmatrix}. \tag{3.19}$$

The Dirac equations (2.13) follow from $\mathscr{H}$ in the class A case if the spinor field has the following nonzero Poisson brackets:

$$2\kappa i\{\Psi^{\mu}, \Psi^{\dagger}{}_{\nu}\} = \delta^{\mu}{}_{\nu}, \quad \mu,\nu = 1,2,3,4. \tag{3.20}$$

The Einstein–Dirac system is subject to the constraints $\mathscr{H} = 0 = \mathscr{H}_{a'}$. The supermomentum components are easily evaluated using the expression for the gravitational supermomentum[5]:

$$\mathscr{H}^{G}_{a'} = -2\,\mathrm{Tr}\boldsymbol{\delta}_{a}\,\boldsymbol{\pi}', \tag{3.21}$$

together with (3.4), (3.8), (A10), and (A13):

$$\mathscr{H}_{a'} = -2\,\mathrm{Tr}\boldsymbol{\delta}_{a}\,\mathbf{P}' = -\Pi_{b}\,\rho^{b}{}_{a} - a_{a}\,p_{+}. \tag{3.22}$$

These are the expressions appearing in Ref. 5 except for the replacement of the mechanical momenta P_a by the "canonical" momenta Π_a (once an incorrectly omitted factor of $\frac{1}{2}$ is reinserted before $\mathscr{H}^{G}_{a'}$ there).

In the nondegenerate class A case, the supermomentum constraints are simply $\Pi_a = 0$, leaving an effective potential behind in the Hamiltonian:

$$U^{\mathrm{eff}} = \tfrac{1}{4}g^{-1/2}\,\mathscr{G}^{ab}X_{a}\,X_{b}. \tag{3.23}$$

At this point none of the automorphism variables appear in the Hamiltonian and the geometrical degrees of freedom have been reduced in number to three, those associated with the scale variables which exhibit no derivative coupling (i.e., mechanical and canonical momenta coincide). As in Ref. 5, one may eliminate another degree of freedom (associated with β^{0}) by using the super-Hamiltonian constraint, leading to an unconstrained Hamiltonian system for the two degrees of freedom $\{\beta^{\pm}\}$, driven of course by the matter variables whose equations of motion also follow from the Hamiltonian.

In the nondegenerate class B case, the supermomentum constraints are $\Pi_1 = \Pi_2 = \Pi_3 + ap_{+} = 0$, so Π_3 must be retained in the effective potential:

$$U^{\mathrm{eff}} = \tfrac{1}{4}g^{-1/2}(\mathscr{G}^{11}X_{1}{}^{2} + \mathscr{G}^{22}X_{2}{}^{2} + \mathscr{G}^{33}(\Pi_{3} - X_{3})^{2}), \tag{3.24}$$

until the equations of motion for the scale variables have been derived, at which time it may be replaced by $-ap_{+}$. With the energy reduction, one again arrives at an unconstrained Hamiltonian system for the two geometrical degrees of freedom $\{\beta^{\pm}\}$, driven by the remaining component Q_{+} of the nonpotential force and the matter variables, whose equations of motion do not follow from the Hamiltonian.

In both cases the supermomentum constraints may be used to determine the automorphism velocities in terms of the remaining variables. One may either use these to integrate (2.6) for $\mathbf{S}$ or to define an equivalent shift vector field algebraically, thus eliminating the need to integrate (2.6). Since the super-Hamiltonian constraint reduction may also be viewed as defining an equivalent lapse function,[5] one sees that the solution of the four constraints may be interpreted as leading to the introduction of particular lapse and shift variables.

The degenerate Bianchi types I, II, V, and $\mathrm{VI}_{-1/9}$ require more detailed discussion similar to that given for the perfect fluid case,[4] since the degeneracy of the supermomentum constraints allows one to eliminate fewer degrees of freedom while permitting nonzero values for some of the constants of the motion associated with the symmetry group $\mathrm{SAut}_{e}(\mathfrak{g})$.

To discuss the constants of the motion it is useful to introduce the variables $\mathbf{H} = e^{\boldsymbol{\beta}}\mathbf{S}$ with conjugate momenta $\mathscr{P}$, which may be evaluated using the expression (A7) for Γ''_{a0b}:

$$\mathscr{P}^{a}{}_{b} = \partial L/\partial\dot{H}^{b}{}_{a} = \partial L^{G}/\partial\dot{g}_{cd}\partial\dot{g}_{cd}/\partial\dot{H}^{b}{}_{a}$$
$$+ 2\kappa\mathscr{H}^{ncd}_{1/2}\,\partial\Gamma''_{c0d}/\partial\dot{H}^{b}{}_{a}$$
$$= (\mathbf{P}\mathbf{H}^{-1})^{a}{}_{b},$$
$$\mathbf{P} = 2\boldsymbol{\pi} + \kappa\mathscr{H}_{1/2}. \tag{3.25}$$

The only nonzero Poisson brackets of these canonical variables are

$$\{H^{a}{}_{b}, \mathscr{P}^{c}{}_{d}\} = \delta^{a}{}_{d}\,\delta^{c}{}_{b}. \tag{3.26}$$

The action of $\mathrm{SAut}_e(\mathfrak{g})$ on the metric variables (the spinor density Ψ is invariant),

$$\mathbf{g}\to B^{-1T}\mathbf{g}B^{-1},\quad B\in\mathrm{SAut}_e(\mathfrak{g}), \tag{3.27}$$

reflecting the change of basis $e_a\to e_b\, B^{-1b}{}_a$ of $\mathfrak{g}$ induces the following canonical transformation of the gravitational variables $(\mathbf{H},\mathscr{P})$:

$$(\mathbf{H},\mathscr{P})\to\mathbf{B}\cdot(\mathbf{H},\mathscr{P})=(\mathbf{H}\mathbf{B}^{-1},\mathbf{B}\mathscr{P}). \tag{3.28}$$

If $\mathbf{A}\in\mathfrak{saut}_e(\mathfrak{g})$, then

$$d/dt\,|_0\, e^{t\mathbf{A}}\cdot(\mathbf{H},\mathscr{P})=(\{P(\mathbf{A}),\mathbf{H}\},\{P(\mathbf{A}),\mathscr{P}\}) \tag{3.29}$$

shows that

$$P(\mathbf{A})=-\operatorname{Tr}\mathbf{A}\mathscr{P}=-\operatorname{Tr}\mathbf{A}\mathscr{P}\mathbf{H} \tag{3.30}$$

is the canonical generator of the 1-parameter subgroup of canonical transformations corresponding to the action of the 1-parameter subgroup of $\mathrm{SAut}_e(\mathfrak{g})$ generated by $\mathbf{A}$. When $\Psi=0$ this reduces to the expression (2.9) of Ref. 7.

For example, using the definition

$$\mathbf{S}\kappa_a\mathbf{S}^{-1}=\kappa_b\, R^b{}_a \tag{3.31}$$

of the adjoint matrix[4] $\mathbf{R}$ on $\hat{G}$ together with (3.8), one has

$$P(\kappa_a)=-\operatorname{Tr}\kappa_a\mathbf{P}=-\operatorname{Tr}\mathbf{S}\kappa_a\mathbf{S}^{-1}\mathbf{P}'=-\Pi_b R^b{}_a. \tag{3.32}$$

In the class A case these are constants of the motion. However, in the nondegenerate case where $\mathbf{R}=\mathbf{S}$, the supermomentum constraints require that they vanish. For Bianchi type II, where $R^3{}_a=\delta^3{}_a$ (see Ref. 5), the supermomentum constraints $\Pi_1=\Pi_2=0$ lead to $P(\kappa_a)=-\Pi_3\delta^3{}_a$, so that only the constants of the motion $P(\mathbf{e}_-)$, $P(\mathbf{e}^1{}_2+\mathbf{e}^2{}_1)$, and $P(\kappa_3)=-\Pi_3$ associated with the subgroup $\mathrm{SL}(2)_3\subset\mathrm{SAut}_e(\mathfrak{g})$ are allowed to be nonzero. For Bianchi type I the supermomentum constraints are identically satisfied and eight linearly independent constants of the motion associated with $\mathrm{SL}(3,R)$ exist. This case was completely solved by Henneaux[6]; the choice of variables made here is not adapted to the exploitation of these constants.

In the class B case, the time derivative of a function F on the phase space depending only on the canonical gravitational variables is given by

$$\dot{F}=\{F,\mathscr{H}\}+Q_a\,\partial F/\partial\Pi^a+Q_+\,\partial F/\partial p_+, \tag{3.33}$$

hence the generators $\mathbf{P}(\kappa_a)$ satisfy

$$P(\kappa_a)^{\cdot}=-Q_3\,\partial(\Pi_b R^b{}_3)/\partial\Pi_3=-Q_3\,\delta^3{}_a, \tag{3.34}$$

since $R^3{}_a=\delta^3{}_a$ here. Thus $P(\kappa_1)$ and $P(\kappa_2)$ are constants of the motion but since $\Pi_1=\Pi_2=0$ for all types except $\mathrm{VI}_{-1/9}$, one again has $P(\kappa_a)=-\Pi_3\,\delta^3{}_a$, so they must vanish. For Bianchi type $\mathrm{VI}_{-1/9}$ the constraints only require $\Pi_1-\Pi_2=0$, so the constant of the motion $P(\kappa_1+\kappa_2)$ is allowed to be nonzero. For Bianchi type V the situation is similar to type II with only the canonical generators associated with $\mathrm{SL}(2)_3\subset\mathrm{SAut}_e(\mathfrak{g})$ allowed to be nonzero. However, the connection of this subgroup with constants of the motion is unclear.

The conventional Hamiltonian approach of Henneaux as applied in the class A spatially homogeneous case[6] may be obtained from the present approach by introducing the time-dependent orthogonal matrix $\mathscr{R}$ of (2.9) as an arbitrary gauge variable (involving three degrees of freedom). If $\mathscr{U}$ is a spinor transformation which covers $\mathscr{R}$, i.e., satisfies (2.16), then the following may be considered as a time-dependent transformation of the configuration space variables:

$$\mathbf{h}=\mathscr{R}\mathbf{H},\quad \chi=\mathscr{U}\Psi. \tag{3.35}$$

Choosing $\mathbf{h}$ and χ as new configuration space variables, one is led to the Hamiltonian approach of Henneaux, who uses the notation

$$h_{(b)a}=h^b{}_a,\quad \mathscr{H}_{1/2}{}^{(ab)}=\mathscr{H}_{1/2}{}^{ab},\quad \Omega_{(ab)}=-\Gamma'''_{a0b},$$

and the opposite sign for the components of the structure constant tensor. The present formulation is simply a particular way of fixing the rotational gauge freedom possessed by the triad variables of his approach. The class B case may also be handled with the Henneaux approach by expressing the nonpotential force in terms of the new variables.

4. CLASSIFICATION OF SOLUTIONS BY DEGREE OF GENERALITY

Let $\mathscr{M}_{S(a)}\supset\mathscr{M}_D\supset\mathscr{M}_{T(a)}\supset\mathscr{M}_I$ be the symmetric case, diagonal, Taub, and isotropic submanifolds, respectively, of the manifold $\mathscr{M}$ of metric matrices as defined in Ref. 5. [For example, $\mathbf{g}\in\mathscr{M}_{S(3)}$ has $g_{13}=g_{23}=0$, $\mathbf{g}\in\mathscr{M}_{T(3)}$ has $g_{11}=g_{22}$, and $\mathbf{g}\in\mathscr{M}_I$ has $\mathbf{g}=\alpha\mathbf{1}$.] One can classify solutions of the Einstein–Dirac equations according to which gravitational degrees of freedom are nontrivial, i.e., modulo the action of constant automorphisms, to which of these submanifolds (or others) does a given solution curve $\mathbf{g}_t$ belong? One is thus led to ask which gravitational degrees of freedom may be frozen out, or more precisely, what are the invariant submanifolds of the phase space on which a Hamiltonian system of fewer degrees of freedom is therefore induced?

Let (a,b,c) be a fixed cyclic permutation of $(1,2,3)$. The first possible specializations in the class A case are the symmetric cases $\mathscr{M}_{S(a)}$ for each value of the index a characterized by the conditions $\mathscr{A}'_b=\mathscr{A}'_c=0=\dot{\tilde{\omega}}^b=\dot{\tilde{\omega}}^c$. From Appendix B one sees that $\mathscr{A}'_a$ is then constant. Without loss of generality one may assume that the metric matrix is confined to $\mathscr{M}_{S(a)}$. In the class B case only the symmetric case $\mathscr{M}_{S(3)}$ is possible since $a'_f=a''\delta^3{}_f\neq0$; $\mathscr{A}'_3$ is not constant for this case, however. If in addition one sets the one remaining spatial tetrad component of the axial current to zero in either the class A or class B case, the spinor energy–momentum tensor vanishes identically and the gravitational equations reduce to the vacuum case discussed in Ref. 4. It is therefore assumed that the remaining spatial tetrad component of the axial current does not vanish. However, in the nondegenerate case as well as the Bianchi type $\mathrm{VI}_{-1/9}$, no further specialization is possible except for the type VI_0 Taub-like subcase $\mathscr{M}_{T(3)}$ of the symmetric case $\mathscr{M}_{S(3)}$. This case is characterized by the conditions $\mathscr{A}'_3\neq0$ but $\beta^-=\mathrm{p}_-=0=\dot{\tilde{\omega}}^a$ (for all a), which implies $\Gamma''{}^a{}_a=0=\Sigma''{}^3{}_3$. Since $\partial\Gamma''{}^a{}_a/\partial\beta^-|_{\beta^-=0}\neq0$ one must have $\mathscr{A}''0=0$, so that p_- remains zero, but the conditions $\mathscr{A}'_1=\mathscr{A}'_2=\mathscr{A}'_0=0$ imply $\bar{\Psi}\gamma_5\Psi=0$ and $|\bar{\Psi}\Psi|=|4\mathscr{A}'_3|$, as easily seen by choosing a particular representation of the Dirac matrices. The only nonzero part of

the spinor energy–momentum tensor is then $\mathscr{H}^{1/2} = 2\kappa im\bar{\Psi}\Psi$, which is a constant. One thus obtains solutions of the vacuum type VI_0 Taub-like case for which the gravitational super-Hamiltonian is allowed to be a nonvanishing constant, i.e., dust (pressure free perfect fluid) solutions with 4-velocity $u = e_\perp$ and energy density $\rho = im\bar{\Psi}\Psi$.[6] Unfortunately, this energy density may assume nonpositive values.

For the degenerate Bianchi types I,V, and II, the specializations corresponding to local rotational symmetry and, in the first two cases isotropy (of the metric), are possible. For Bianchi type II there is only the locally rotationally symmetric (LRS) case $\mathscr{M}_{T(3)}$ characterized by the conditions $\mathscr{A}_3' \neq 0$ and $\beta^- = p_- = 0 = \dot{\tilde{\omega}}^a$. An additional constant of the motion in the massless case enabled Henneaux to integrate all of the equations explicitly.[6] For Bianchi type V one has the same LRS case. However, in the general type V case the third supermomentum constraint $p_+ = 0 = \beta^+$ is a simple holonomic constraint which requires β^+ to be a constant which may be set equal to zero by the action of a constant automorphism, as in the vacuum case.[4] Thus the LRS case coincides with the isotropic case. The only nonvanishing part of the spinor energy–momentum tensor is the super-Hamiltonian $\mathscr{H}^{1/2} = 2\kappa im\bar{\Psi}\Psi = 2\kappa\rho g^{1/2}$, which is a constant since $\Gamma''^a{}_a = 0$ but not necessarily a positive constant. One thus obtains the open Friedman dust solutions, although the energy density ρ may assume nonpositive values. When the mass m is set equal to zero, one arrives at flat spacetime in the hyperboloidal time slicing.[8] Bianchi type I exhibits the same LRS case $\mathscr{M}_{T(3)}$ [the others are equivalent under the action of $SAut_e(\mathfrak{g}) = SL(3,R)$] for which p_+ is a constant of the motion. Setting it to zero yields the isotropic case. The general Bianchi type I case has been solved exactly by Henneaux.[6]

It is worth noting that although the LRS metrics of Bianchi types I and VII_0 and V and VII_h, respectively, coincide, the corresponding spatially homogeneous Dirac spinor fields differ in their symmetry properties. Those of types VII_0 and VII_h undergo a space-dependent rotation relative to the spatially homogeneous spinor fields of types I and V.

5. "TWISTED" DIRAC FIELDS

Let $\{\tilde{e}_a\}$ now stand for the basis of the Lie algebra $\tilde{\mathfrak{g}}$ of right invariant vector fields on G which coincides with the basis $\{e_a\}$ of $\mathfrak{g}$ at the identity. Interpreted as fields on $R \times G$, the elements of $\tilde{\mathfrak{g}}$ are spacetime Killing vector fields. The most general spinor field ψ whose energy–momentum tensor is spatially homogeneous satisfies[11]

$$\pounds_{\tilde{e}_a}\psi = i\lambda_a\psi\,, \tag{5.1}$$

where λ_a are real constants which satisfy $\lambda_c C^c{}_{ab} = 0$, since

$$-i\lambda_c C^c{}_{ab}\psi = \pounds_{[\tilde{e}_a,\tilde{e}_b]}\psi = [\pounds_{\tilde{e}_a},\pounds_{\tilde{e}_b}]\psi = 0\,. \tag{5.2}$$

In other words $\lambda = \lambda_c\omega^c$ is a time-independent exact left invariant and, therefore, bi-invariant 1-form on G. The classification of possible symmetries for such spinor fields therefore corresponds exactly to the classification of spatially self-similar spacetimes.[12,13] All Bianchi types except the semisimple types VIII and IX admit bi-invariant 1-forms.

Let $\psi_{t\omega} = F\psi$, where $|F|^2 = 1$, $dF = i\lambda F$, and ψ is spatially homogeneous[11] and say that $\psi_{t\omega}$ is obtained from ψ by "twisting", the analog of the conformal scaling of spatially homogeneous metrics to obtain spatially self-similar metrics.[12] The energy–momentum tensor of $\psi_{t\omega}$ and its Dirac equation may be rewritten in terms of ψ using the relations

$$\partial''_a\,\Psi_{t\omega} = F(i\lambda''_a\,\Psi),\bar{\Psi}_{t\omega}\,\gamma\partial''_a\,\Psi_{t\omega} = i\lambda'_a\,\bar{\Psi}\gamma\Psi\,,$$
$$\lambda''_a = \lambda_b H^{-1b}{}_a\,. \tag{5.3}$$

One must therefore add the terms $i\lambda''_a\,\gamma^{a0}\Psi$ and $-i\lambda''_a\,\bar{\Psi}\gamma^{a0}$ to the r.h.s.'s of the Dirac equations (2.13), the term $-i\lambda''_a\,\psi(\gamma\gamma^{a0} + \gamma^{a0}\gamma)\Psi$ to the r.h.s. of (B1) and the term $g^{-1/2}\mathscr{J}''_{(a}\delta^b{}_{\beta)}\lambda_b$ to the expression (3.1) for the energy–momentum tensor of $\psi_{t\omega}$. The twisted super-Hamiltonian and supermomentum become

$$\mathscr{H}^{1/2}_{t\omega} = \mathscr{H}^{1/2} + 2\kappa\lambda''_a\,\mathscr{J}''^a\,,$$
$$\mathscr{H}^{1/2}_{t\omega a'} = \mathscr{H}^{1/2}_{a'} - 2\kappa\mathscr{J}''^0\lambda'_a\,. \tag{5.4}$$

The presence of the constant 1-form λ in the Einstein–Dirac system reduces the symmetry group from $SAut_e(\mathfrak{g})$ to the subgroup which leaves λ_a invariant. However, the space of bi-invariant 1-forms is 1-dimensional and hence automatically invariant under the action of $SAut_e(\mathfrak{g})$ for all Bianchi types except I,II, and III = VI_{-1}, where this space has dimension 3, 2, and 2, respectively. For type III, only the 1-dimensional subspace of "exceptional" bi-invariant 1-forms is not invariant under $SAut_e(\mathfrak{g})$, corresponding to Eardley's exceptional type $^*_1$III.[12]

For the Bianchi types other than I, II, VIII, and IX and for the nonexceptional type III case, one has

$$\lambda'_a = \lambda_b S^{-1b}{}_a = \lambda_a = \lambda_3\delta^3{}_a,\ \lambda_3 \neq 0. \tag{5.5}$$

Thus $SAut_e(G)$ remains a symmetry group. Only the third momentum constraint is changed here, becoming

$$(1 - \delta^V_z)\Pi_3 = -2\kappa\lambda_3\mathscr{J}''^0 - ap_+. \tag{5.6}$$

For the class A types of this class, $\mathscr{J}''^0$ is a constant and Π_3 remains a constant of the motion but which is now allowed to be nonvanishing. Only the symmetric cases $\mathscr{M}_{S(3)}$ are now permitted, requiring $\mathscr{A}''_a = \mathscr{A}''_3\delta^3{}_a$ and $\mathscr{J}''_a = \mathscr{J}''_3\delta^3{}_a$ (for the class A types, $\mathscr{A}''_3$ is again constant). However, the consistency of these conditions with the equations of motion (B2) and (B3) requires that Ψ be an eigenvector of $\sigma_3 = i\gamma^{12}$, i.e., a spin eigenstate:

$$\sigma_3\Psi = \tau\Psi,\ \tau^2 = 1, \tag{5.7}$$

which implies

$$4\mathscr{A}''_3/\mathscr{J}''^0 = -\tau,\quad \mathscr{J}''^3 = -\tau\,4\mathscr{A}''^0, \tag{5.8}$$

results obtained by choosing an explicit representation of the Dirac matrices. Excluding Bianchi type V, the one nontrivial momentum constraint is

$$P_3 + ap_+ = -2\kappa(\lambda_3\mathscr{J}''^0 + 2\Sigma''^3{}_3\mathscr{A}''_3). \tag{5.9}$$

For types VII_0 and VII_h, the LRS case $\mathscr{M}_{T(3)}$ requires $\beta^- = 0 = P_3$, in which case $\Sigma''^3{}_3 = 1$ and (5.9) becomes

$$ap_+ = -2\kappa\mathscr{J}''^0(\lambda_3 - \tfrac{1}{2}\tau). \tag{5.10}$$

For type VII_0, this is equivalent to

$$\lambda_3 = \tfrac{1}{2}\tau. \tag{5.11}$$

If this condition is imposed in the type VII_h case as well, then p_+ also vanishes, automatically implying isotropy as in the spatially homogeneous type V case. In fact, (5.11) is exactly the condition which undoes the twisting of the spatially homogeneous type VII_0/VII_h spin eigenstate fields relative to their spatially homogeneous type I/V counterparts. To see this, identify the manifolds of the type $VII_0/VIII_h$ and type I/V Lie groups by identifying canonical coordinates $\{x^a\}$ of the second kind with respect to canonical bases of their Lie algebras. Using the explicit expressions given in Ref. 5 for the canonical left invariant 1-forms in these coordinates, one finds in an obvious notation

$$\omega''^a(\mathrm{VII}_0/\mathrm{VII}_h) = [\exp(-x^3\mathbf{k}_3^{\mathrm{IX}})]^a{}_b\omega''^b(\mathrm{I/V}), \quad \mathbf{g}\in\mathscr{M}_{T(3)}. \tag{5.12}$$

Since $\{-\frac{1}{2}i\sigma_a = -\frac{1}{4}\epsilon_{abc}\gamma^{bc}\}$ is the image basis of $\{\mathbf{k}_a^{\mathrm{IX}}\}$ in the Lie subalgebra of the Dirac algebra, the following relation gives the components of the twisted type VII_0/VII_h spinor fields in terms of the type I/V tetrad:

$$\begin{aligned}\Psi(\mathrm{I/V}) &= [\exp(-\tfrac{1}{2}ix^3\sigma_3)]\ \Psi_{t\omega}(\mathrm{VII}_0/\mathrm{VII}_h)\\ &= [\exp(\lambda_3 - \tfrac{1}{2}\sigma_3)ix^3]\ \Psi(\mathrm{VII}_0/\mathrm{VII}_h),\end{aligned} \tag{5.13}$$

where the choice $F = \exp i\lambda_3 x^3$ has been made. The twisted type VII_0/VII_h spin eigenvector fields satisfying (5.11) coincide with the corresponding spatially homogeneous type I/V fields. Thus, the LRS twisted type VII_0/VII_h spin eigenstate case satisfying (5.11) is equivalent to the spatially homogeneous spin eigenstate subcase of Bianchi type I/V.

On the other hand, if (5.11) is not imposed in the LRS type VII_h case, one sees from (5.13) that the spinor field is equivalent to a twisted type V field with twisting parameter $\lambda_3^{\mathrm{V}} = \lambda_3 - \frac{1}{2}\tau$, i.e., it is sufficient to consider only the LRS type V case when Ψ is a spin eigenstate.

For Bianchi type V the momentum constraint is

$$P_+ = -2\kappa\lambda_3\mathscr{J}''^0. \tag{5.14}$$

The LRS case $\mathscr{M}_{T(3)}$ now imposes no condition on λ_3 but this constraint shows that isotropy is no longer possible for nonzero λ_3, since $\mathscr{J}''^0 = \Psi^\dagger\Psi$ vanishes only when Ψ does. The massless LRS case was solved by Michalik and Melvin,[14] who mistakenly thought it represented a nontrivial type VII_h case, and later by Ray using a different method.[8]

For Bianchi type I, λ_a are arbitrary but since the spatially homogeneous supermomentum vanishes identically, the supermomentum constraint requires $\lambda_a = 0$. For Bianchi type II, $\lambda_3 = 0$ is sufficient for bi-invariance and using the freedom in the choice of a canonical basis one may assume $\lambda_a = \lambda_1\delta^1{}_a$, which breaks the $SL(2)_3$ invariance. However, the symmetric case $\mathscr{M}_{S(1)}$ is still possible provided that $\mathscr{A}''_a = \mathscr{A}''_1\delta^1_a$ and $\mathscr{J}''_a = \mathscr{J}''_1\delta^1{}_a$. For this case $\lambda'_a = \lambda_a$ and $\mathscr{A}''_1$ and $\Pi_1 = -2\kappa\lambda_1\mathscr{J}''^0$ are constants of the motion while $\Pi_2 = 0 = \Pi_3$.

6. CONCLUSION

The conventional Hamiltonian formulation of the Einstein–Dirac system was developed for a general spacetime where no class of preferred tetrads exists. Additional gravitational variables subject to new constraints must be introduced to describe the most general tetrad,[1] which of course is not adapted to a given slicing of the spacetime. However, the mathematical machinery is much simpler when the timelike member of the tetrad is the unit normal to the chosen slicing,[1,2] and hence for a spacetime where a natural slicing exists, it is therefore appropriate to use a reduced formulation in which only the rotational freedom in the spatial triads remains. For a space-time in which a preferred class of spatial triads also exists, it is logical to suppress the rotational freedom as well, thereby removing completely the additional gauge variables required to consider arbitrary tetrads.

For spatially homogeneous spacetimes (excluding those of Bianchi types I, II, and V), the symmetry picks out such a preferred class of tetrads, permitting a reduced Hamiltonian formulation from which simpler differential equations follow and which allows the explicit solution of the remaining constraints, thus leading to a system of equations for the minimum number of unconstrained variables. Moreover, the decomposition of the gravitational variables which accomplishes this has a direct geometrical interpretation which aids in the understanding of the system. For the Bianchi types I, II, and V, the symmetry allows a certain rotational gauge freedom to remain which complicates the discussion somewhat but does not prevent a treatment analogous to that of the other types.

ACKNOWLEDGMENTS

The author thanks M. P. Ryan, Jr. for a discussion leading to this work and both M. P. Ryan, Jr. and M. Henneaux for their helpful preprints.

APPENDIX A

The frame $\{e_\alpha\}$ has structure functions $C^\alpha{}_{\beta\gamma} = \omega^\alpha([e_\beta, e_\gamma]) = \delta^\alpha_a C^a{}_{bc}\delta^b{}_\beta\delta^c{}_\gamma$. The metric connection components are given by

$$\begin{aligned}\Gamma^\alpha{}_{\beta\gamma} &= \omega^\alpha(\nabla_{e_\beta}e_\gamma)\\ &= \tfrac{1}{2}g^{\alpha\delta}(e_\gamma g_{\delta\beta} + e_\beta g_{\delta\gamma} - e_\delta g_{\beta\gamma}) + \tfrac{1}{2}C^\alpha{}_{\beta\gamma} + C_{(\beta}{}^\alpha{}_{\gamma)}.\end{aligned} \tag{A1}$$

The only nonvanishing components are

$$\Gamma^0{}_{ab} = -K_{ab}, \quad \Gamma^a{}_{0b} = -K^a{}_b = \Gamma^a{}_{b0},$$
$$\Gamma^c{}_{ab} = \tfrac{1}{2}C^c{}_{ab} + C_{(a}{}^c{}_{b)}, \tag{A2}$$

where $K_{ab} = -\frac{1}{2}\dot{g}_{ab}$ are the components of the extrinsic curvature tensor.

If W is any two index spatial geometric object, let $\mathbf{W}$, $\mathbf{W}'$, and $\mathbf{W}''$ be the matrices of mixed components of W with respect to the spatial frames e, e', and e'', respectively, i.e., $\mathbf{W} = W^a{}_b\,\mathbf{e}^b{}_a$, etc. For example, for each a let $\kappa_a{}^b{}_c$ be the components of a tensor field with respect to e', so that one has

$$\mathscr{K}'_{abc} = \kappa'_{a(bc)} = \kappa_a{}^d{}_{(c}\,g'_{b)d}, \quad \mathscr{K}''_{abc} = \kappa''_{a(bc)}, \tag{A3}$$

or in matrix form

$$\mathscr{K}'_a = \tfrac{1}{2}(\boldsymbol{\kappa}_a + e^{-2\beta}\boldsymbol{\kappa}_a{}^T e^{2\beta}),$$

$$\mathscr{K}''_a = \tfrac{1}{2}(e^{\beta}\kappa_a e^{-\beta} + [e^{\beta}\kappa_a e^{-\beta}]^T). \tag{A4}$$

The primed extrinsic curvature matrix was evaluated in Ref. 5:

$$-\mathbf{K}' = \dot{\beta} + \mathscr{K}'_a \dot{\tilde{\omega}}^a. \tag{A5}$$

Using the formula

$$\Gamma^{\alpha}{}_{\beta\gamma}\omega^{\beta} = H^{\alpha}{}_{\delta}(\Gamma^{\delta}{}_{\beta\epsilon}\omega^{\beta}H^{-1\epsilon}{}_{\gamma} + dH^{-1\delta}{}_{\gamma}) \tag{A6}$$

for the frame transformation $e''_{\alpha} = H^{-1\beta}{}_{\alpha}e_{\beta}$ of (2.5), one finds the connection components in the orthonormal frame $\{e''_{\alpha}\}$, where Latin indices may be raised and lowered at will:

$$\Gamma''{}_{a0b} = -\kappa''_{c[ab]}\dot{\tilde{\omega}}^c = -\tfrac{1}{2}[\mathbf{H}\mathbf{H}^{-1} - (\mathbf{H}\mathbf{H}^{-1})^T]^a{}_b,$$

$$\Gamma''{}_{ab0} = -\Gamma''{}_{0ba} = -K''_{ab}, \quad -\mathbf{K}'' = \dot{\beta} + \mathscr{K}''_a\dot{\tilde{\omega}}^a,$$

$$\Gamma''{}^{a}{}_{bc} = \tfrac{1}{2}C''{}^{c}{}_{ab} + C''{}^{c}{}_{(a\ b)}. \tag{A7}$$

Since $\Gamma''{}_{\alpha}{}^{\beta}{}_{\gamma} = \Gamma''{}_{[\alpha}{}^{\beta}{}_{\gamma]}$, one introduces the following objects:

$$\Sigma''{}^{a} = \tfrac{1}{2}\epsilon^{abc}\Gamma''{}_{b0c} = -\tfrac{1}{2}\epsilon^{abc}\kappa''_{d[bc]}\dot{\tilde{\omega}}^d \equiv \Sigma''{}^{a}{}_{d}\dot{\tilde{\omega}}^d,$$

$$\Gamma''_{ab} = \tfrac{1}{2}\epsilon_{bcd}\Gamma''{}^{c}{}_{a}{}^{d},$$

$$\Gamma''{}_{(ab)} = \tfrac{1}{2}(n''{}_{ab} - \tfrac{1}{2}\delta_{ab}n''{}^{c}{}_{c}), \quad \Gamma''{}_{[ab]} = 2\epsilon_{abc}a''{}^{c},$$

$$\Gamma''{}^{a}{}_{a} = -\tfrac{1}{2}n''{}^{a}{}_{a} = -\tfrac{1}{2}g^{-1/2}n^{ab}g_{ab}$$

$$= -\tfrac{1}{2}(n^{(1)}e^{\beta^1-\beta^2-\beta^3} + n^{(2)}e^{\beta^2-\beta^3-\beta^1}$$

$$+ n^{(3)}e^{\beta^3-\beta^1-\beta^2}). \tag{A8}$$

As in Ref. 5 let $\mathscr{G}_{ab} = \mathrm{Tr}\,\mathscr{K}'_a\mathscr{K}'_b = \mathrm{Tr}\,\mathscr{K}''_a\mathscr{K}''_b$ (with inverse $\mathscr{G}^{ab}$) and let $p_A = \partial L_G/\partial\dot{\beta}^A$ and $P_a = \partial L_G/\partial\dot{\tilde{\omega}}^a$ be the mechanical momenta. Using the explicit velocity–momentum relation given there, one finds the following expression for the gravitational mechanical momentum matrix:

$$\boldsymbol{\pi}' = -g^{1/2}(\mathbf{K}' - \mathbf{1}\,\mathrm{Tr}\,\mathbf{K}')$$

$$= \tfrac{1}{12}(\eta^{AB}p_A\mathbf{e}_B + 3p_0\mathbf{e}_0) + \tfrac{1}{2}\mathscr{G}^{ab}P_a\mathscr{K}'_b, \tag{A9}$$

with the immediate consequence

$$p_{\pm} = 2\,\mathrm{Tr}\,\mathbf{e}_{\pm}\boldsymbol{\pi}', \quad P_a = 2\,\mathrm{Tr}\,\kappa_a\boldsymbol{\pi}'. \tag{A10}$$

Introduce also the following matrices:

$$\mathbf{k}_a = C^{b}{}_{ac}\,\mathbf{e}^{c}{}_{b},$$

$$\boldsymbol{\delta}_a = (C^{b}{}_{ac} - 2\delta^{b}{}_{a}a_c)\mathbf{e}^{c}{}_{b}, \quad \mathbf{k}^{0}{}_{a} = \epsilon_{cad}n^{db}\mathbf{e}^{c}{}_{b}, \tag{A11}$$

with similar definitions for the primed and double primed components.

The following choice of matrices $\{\kappa_a\}$ generating $\hat{G}$ is made:

$$\textit{CLASS A:}\quad \kappa_a = \mathbf{k}_a + \delta^{\mathrm{II}}{}_{Z}\,\delta^{3}{}_{a}\,\mathbf{k}^{\mathrm{IX}}{}_{3} + \delta^{\mathrm{I}}{}_{Z}\,\mathbf{k}^{\mathrm{IX}}{}_{a},$$

$$\boldsymbol{\delta}_a = \mathbf{k}_a, \tag{A12}$$

$$\textit{CLASS B:}\quad \kappa_1 = \mathbf{e}^{3}{}_{1}, \qquad \boldsymbol{\delta}_1 = -3a\kappa_1 + n^{(2)}\kappa_2,$$

$$\kappa_2 = \mathbf{e}^{3}{}_{2}, \qquad \boldsymbol{\delta}_2 = n^{(1)}\kappa_1 - 3a\kappa_2,$$

$$\kappa_3 = \mathbf{k}^{0}{}_{3} + \delta^{\mathrm{V}}{}_{Z}\,\mathbf{k}^{\mathrm{IX}}{}_{3}, \quad \boldsymbol{\delta}_3 = (1 - \delta^{\mathrm{V}}{}_{Z})\kappa_3 + a\mathbf{e}_{+}.$$

$\{\mathbf{k}^{\mathrm{IX}}{}_{a}\}$ are the canonical Bianchi type IX adjoint matrices [generating SO(3,R)] while $\delta^{Z}{}_{Z'}$ is the Bianchi type Kronecker delta. Introduce also the notation

$$\boldsymbol{\delta}_a = \kappa_b\,\rho^{b}{}_{a} + a_a\,\mathbf{e}_{+}, \quad [\kappa_a,\kappa_b] = \hat{C}^{c}{}_{ab}\,\kappa_c. \tag{A13}$$

One finds $\hat{C}^{a}{}_{bc} = \epsilon_{bcd}n^{da}$ for all Bianchi types except type I, where $\hat{C}^{a}{}_{bc} = C^{a}{}_{bc}(\mathrm{IX})$, and types II and V, where $\hat{C}^{a}{}_{bc} = C^{a}{}_{bc}(\mathrm{VII}_0)$; the Roman numeral in parentheses indicates the canonical components of that Bianchi type.

The basis $\{\kappa_a\}$ has been chosen so that $\mathscr{G}_{ab}$ is diagonal and $\Sigma''{}^{b}{}_{a}$ has only one nonvanishing component for each value of a. When $\kappa_a = \mathbf{k}^{0}{}_{a}$, as occurs in the nondegenerate class A case for all values of a, for $a = 1,2$ in the Bianchi type II case and for $a = 3$ in the class B case (except for type V), then $\Sigma''{}^{a}{}_{a}$ (no sum on a) is the nonvanishing component and the appropriate cyclic permutations of the following formulas hold, given here for $a = 3$

$$\Sigma''{}^{3}{}_{3} = -\kappa''_{3[12]} = \tfrac{1}{2}(n^{(1)}e^{\beta^{12}} + n^{(2)}e^{-\beta^{12}}),$$

$$\kappa''_{3(12)} = -\tfrac{1}{2}(n^{(1)}e^{\beta^{12}} - n^{(2)}e^{-\beta^{12}}), \tag{A14}$$

$$\mathscr{G}_{33} = 2[\kappa''_{3[12]}]^2.$$

In the class B case (including type V), the remaining components are given by

$$\Sigma''{}^{2}{}_{1} = \tfrac{1}{2}e^{\beta^{13}}, \quad -\Sigma''{}^{1}{}_{2} = \tfrac{1}{2}e^{\beta^{23}},$$

$$\mathscr{G}_{11} = \tfrac{1}{2}e^{2\beta^{13}}, \quad \mathscr{G}_{22} = \tfrac{1}{2}e^{2\beta^{23}}. \tag{A15}$$

The remaining components in Bianchi types I, II, and V are given by the appropriate version of (A14) evaluated at $\mathbf{n} = \mathbf{1}$. (Recall that $\beta^{ab} = \beta^a - \beta^b$.)

APPENDIX B

An immediate consequence of (2.13) is that if γ is any product of Dirac matrices, the following equation is obeyed by $\bar{\Psi}\gamma\Psi$:

$$(\bar{\Psi}\gamma\Psi)^{\cdot} = -\tfrac{1}{2}\bar{\Psi}\{\Sigma''{}^{a}[\gamma,\gamma_a\gamma^0\gamma_5] + \Gamma''{}^{a}{}_{a}(\gamma\gamma_5 + \gamma_5\gamma)$$

$$+ 2a''_c[\gamma,\gamma^{0c}] + 2m[\gamma,\gamma^0]\}\Psi. \tag{B1}$$

Applying this to the basis elements of the Dirac algebra one finds the following equations of motion:

$$\dot{\mathscr{A}}''{}^{0} = 2a''{}_{c}\mathscr{A}''{}^{c} + \tfrac{1}{2}m\bar{\Psi}\gamma_5\Psi,$$

$$\dot{\mathscr{A}}''{}^{a}\epsilon_{abc}\Sigma''{}^{b}\mathscr{A}''{}^{c} + 2a''_a\mathscr{A}''{}^{0},$$

$$\dot{\mathscr{J}}''{}^{0} = 2a''_c\mathscr{J}''{}^{c},$$

$$\dot{\mathscr{J}}''{}^{a} = \epsilon_{abc}\Sigma''{}^{b}\mathscr{J}''{}^{c} + 2a''_a\mathscr{J}''{}^{0} + m\bar{\Psi}\gamma^{0a}\Psi,$$

$$(\bar{\Psi}\Psi)^{\cdot} = \Gamma''{}^{a}{}_{a}\bar{\Psi}\gamma_5\Psi,$$

$$(\bar{\Psi}\gamma_5\Psi)^{\cdot} = \Gamma''{}^{a}{}_{a}\bar{\Psi}\Psi + 4m\mathscr{A}''{}^{0},$$

$$(\bar{\Psi}\gamma^{a0}\Psi)^{\cdot} = \epsilon_{abc}\Sigma''{}^{b}\bar{\Psi}\gamma^{c0}\Psi - \Gamma''{}^{c}{}_{c}\bar{\Psi}\gamma^{a0}\gamma_5\Psi$$

$$+ 2\epsilon_{abc}a''_b\bar{\Psi}\gamma^{c0}\gamma_5\Psi + 2m\mathscr{J}''{}^{a},$$

$$(\bar{\Psi}\gamma^{a0}\gamma_5\Psi)^{\cdot} = \epsilon_{abc}\Sigma''{}^{b}\bar{\Psi}\gamma^{c0}\gamma_5\Psi - \Gamma''{}^{c}{}_{c}\bar{\Psi}\gamma^{a0}\Psi$$

$$+ 2\epsilon_{abc}a''_b\bar{\Psi}\gamma^{c0}\Psi. \tag{B2}$$

The addition of the term $-2i\lambda_a\bar{\psi}(\gamma^{a0}\gamma + \gamma\gamma^{a0})\psi$ to the r.h.s. of (B1) leads to the addition of the following terms to the r.h.s.'s of (B2):

$$\Delta\dot{\mathscr{A}}''{}^{0} = 0 = \Delta\mathscr{J}''{}^{0},$$

$$\Delta\dot{\mathscr{A}}''{}^{a} = \tfrac{1}{2}\epsilon_{abc}\lambda''{}^{b}\mathscr{J}''{}^{c},$$

$$\Delta\dot{\mathscr{J}}''{}^{a} = -8\epsilon_{abc}\lambda''{}^{b}\mathscr{A}''{}^{c},$$

$$\Delta(\bar{\Psi}\Psi)^{\cdot} = -2i\lambda''_a\bar{\Psi}\gamma^{a0}\Psi,$$

$$\Delta(\bar{\Psi}\gamma_5\Psi)^{\cdot} = -2i\lambda''_a\bar{\Psi}\gamma^{a0}\gamma_5\Psi,$$

$$\Delta(\bar{\Psi}\gamma^{a0}\Psi)^{\cdot} = -2i\lambda''{}^{a}\bar{\Psi}\Psi,$$

$$\Delta(\bar{\Psi}\gamma^{a0}\gamma_5\Psi)^{\cdot} = -2i\lambda''_a\bar{\Psi}\gamma_5\Psi. \tag{B3}$$

ACKNOWLEDGMENTS

The author thanks M. P. Ryan, Jr. for a discussion leading to this work and both M. P. Ryan, Jr. and M. Henneaux for their helpful preprints.

[1]J. E. Nelson and C. Teitelboim, Ann. Phys. (NY) **116**, 86 (1978).
[2]M. Henneaux, Gen. Relativ. Gravit. **9**, 1031 (1978).
[3]O. Obregón and M. P. Ryan, Jr., J. Math. Phys. **22**, 623 (1981).
[4]R. T. Jantzen, "Spatially Homogeneous Cosmology: Background and Dynamics", in *Bianchi Universes and Relativistic Cosmology*, edited by R. Ruffini (Princeton U.P., Princeton, N.J., to appear).
[5]R. T. Jantzen, Commum. Math. Phys. **64**, 211 (1979).
[6]M. Henneaux, Phys. Rev. D **21**, 857 (1980); "Bianchi Universes and Spinor Fields," to appear.
[7]R. T. Jantzen, Nuovo Cimento B **55**, 161 (1980).
[8]T. M. Davis and J. R. Ray, J. Math Phys. **16**, 75 (1975), **17**, 1049 (1976); J. R. Ray, Prog. Theor. Phys. **63**, 1213 (1980).
[9]F. B. Estabrook, H. D. Wahlquist, and C.G. Behr, J. Math. Phys. **9**, 497 (1968); G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[10]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).
[11]M. Henneaux, Gen. Relativ. Gravit. **12**, 137 (1980).
[12]D. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[13]R. T. Jantzen, Ann. Phys. (NY) **127**, 302 (1980).
[14]T. R. Michalik and M. A. Melvin, J. Math. Phys. **21**, 1952 (1980).

Self-dual Kerr–Schild metrics and null Maxwell fields [a)]

K. P. Tod
Mathematical Institute, University of Oxford, Oxford, England

(Received 16 January 1981; accepted for publication 3 April 1981)

A construction for self-dual solutions of Einstein's equations with metric in Kerr–Schild form from null electromagnetic fields in Minkowski space is described. Two examples are given.

PACS numbers: 04.20.Jb

In this paper, I wish to show that any null Maxwell field in Minkowski space determines a solution of the self-dual Einstein equations for which the metric has Kerr–Schild (KS) form. Such Maxwell fields can then be generated by twistor methods and depend essentially on two free functions of two variables. This situation is contrasted with real KS metrics and some examples are given.

The self-dual Einstein equations have been reduced by Plebański[1] to a single nonlinear equation on a scalar potential $\Theta(u, v, X, Y)$ which can be written

$$\Theta_{,uv} - \Theta_{,XY} = -\Theta_{,vv}\Theta_{,XX} + (\Theta_{,vX})^2. \tag{1}$$

The corresponding metric is then

$$ds^2 = 2dudv - 2dXdY - 2 \times(\Theta_{,XX}du^2 + 2\Theta_{,vX}dudY + \Theta_{,vv}dY^2). \tag{2}$$

The KS metrics have the form[2]

$$g_{ab} = \eta_{ab} + 2HL_aL_b, \tag{3}$$

where η_{ab} is the Minkowski metric, L_a is a null vector field on Minkowski space subject to certain conditions, and H is related to the derivative of L_a in a way described below.

The metric (2) takes the KS form (3) if the second part is a perfect square, i.e., if

$$\Theta_{,vv}\Theta_{,XX} - (\Theta_{,vX})^2 = 0 \tag{4}$$

so that, to find self-dual KS metrics, we must set both sides in (1) separately equal to zero, obtaining (4) and the Minkowski-space wave equation

$$\Box\Theta \equiv 2(\Theta_{,uv} - \Theta_{,XY}) = 0. \tag{5}$$

Rather than solving (4) and (5) directly, we shall obtain the same equations as determining null Maxwell fields. First, choose a constant normalized spinor dyad $(O_A, 1^A; O_A 1^A = 1)$ on Minkowski space and introduce the coordinates (u, v, X, Y) in the conventional way

$$x^{AA'} = u1^A\bar{1}^{A'} + Y1^A\bar{O}^{A'} + XO^A\bar{1}^{A'} + vO^A\bar{O}^{A'}.$$

Now define

$$\delta_{A'} = O^A\nabla_{AA'}$$

so that

$$\delta_{A'} = \bar{O}_{A'}\frac{\partial}{\partial X} - \bar{1}_{A'}\frac{\partial}{\partial v}.$$

If $\phi_{A'B'}$ is a solution of Maxwell's equations in Minkowski space,

$$\nabla_A{}^{A'}\phi_{A'B'} = 0,$$

then by transvecting with O^A we find

$$\delta^{A'}\phi_{A'B'} = 0.$$

Therefore, we can find a potential $\phi_{A'}$ with

$$\phi_{A'B'} = \delta_{A'}\phi_{B'}$$

and by the symmetry of $\phi_{A'B'}$

$$\delta^{A'}\phi_{A'} = 0$$

so that

$$\phi_{A'} = \delta_{A'}\Theta$$

for some scalar potential Θ. Now

$$\phi_{A'B'} = \delta_{A'}\delta_{B'}\Theta \tag{6}$$

and the remaining Maxwell equations on $\phi_{A'B'}$ imply that Θ can be chosen to satisfy the Minkowski space wave equation:

$$\Box\Theta \equiv 2(\Theta_{,uv} - \Theta_{,XY}) = 0.$$

Further, if $\phi_{A'B'}$ is null, then

$$\phi_{A'B'}\phi^{A'B'} = \delta_{A'}\delta_{B'}\Theta\delta^{A'}\delta^{B'}\Theta = 0,$$

which is just

$$\Theta_{,vv}\Theta_{,XX} - (\Theta_{,vX})^2 = 0.$$

Thus if Θ is a scalar potential for a null Maxwell field, then Θ also satisfies Plebański's equation for a KS metric. However, there is a gain in viewing the KS metrics in this way in that it is not necessary actually to solve (4) and (5) and find Θ explicitly. Rather, if $\phi_{A'B'}$ is any null Maxwell field, then the metric

$$ds^2 = \eta_{ab}dx^a dx^b + \lambda O_A O_B \phi_{A'B'}dx^{AA'}dx^{BB'} \tag{7}$$

for arbitrary λ automatically satisfies the self-dual Einstein equations.

To generate null Maxwell fields, we recall from Robinson's theorem[3] that if the spinor field $\pi_{A'}$ on Minkowski space is geodesic and shear-free (GSF), then a scaling of $\pi_{A'}$ can be found such that the spinor $\phi_{A'B'} = \pi_{A'}\pi_{B'}$ satisfies Maxwell's equations. Further, the spinor $\psi_{A'B'} = \phi\pi_{A'}\pi_{B'}$ will also satisfy Maxwell's equations provided

$$\pi^{A'}\nabla_{AA'}\phi = 0. \tag{8}$$

Finally, GSF spinors $\pi_{A'}$ can be generated in Minkowski space via the Kerr theorem[4] by choosing an arbitrary homogeneous twistor function $f(Z^\alpha)$ and solving the equation

$$f(ix^{AA'}\pi_{A'}, \pi_{A'}) = 0 \tag{9}$$

for $\pi_{A'}$ as a function of $x^{AA'}$.

[a)] Work supported by S. R. C. Post-Doctoral Research Fellowship.

 0022-2488/82/061147-02$02.50

(Null Maxwell fields may also be obtained directly by means of twistor integrals.[4] The functional freedom is the same.)

The general self-dual KS metric is therefore defined by an arbitrary twistor function f and a solution ϕ of (8). This corresponds to two free functions of two variables.

From the work of Plebański,[1] the Weyl spinor of the metric (7) is just

$$\psi_{A'B'C'D'} = \delta_{A'}\delta_{B'}\phi_{C'D'} = \delta_{A'}\delta_{B'}\delta_{C'}\delta_{D'}\Theta \tag{10}$$

suitably interpreted. (Basically, O_A is now thought of as a constant spinor in the self-dual space with $\nabla_{AA'}$ in the definition of $\delta_{A'}$ being the self-dual connection).

If $\phi_{A'B'} = \pi_{A'}\pi_{B'}$, then (6) implies

$$\delta_{A'}\pi_{B'} = \tfrac{1}{6}(\beta_{A'}\pi_{B'} + 2\pi_{A'}\beta_{B'})$$

for some $\beta_{A'}$ and then (10) is just

$$\psi_{A'B'C'D'} = \pi_{(A'}\pi_{B'}\beta_{C'}\beta_{D')} + \pi_{(A'}\pi_{B'}\delta_{C'}\beta_{D')}.$$

As anticipated, the Weyl spinor is algebraically special and can be of any type. However, a function-counting argument indicates that not every algebraically special metric can be put into KS form.

This situation may be contrasted with real KS metrics which are defined by special GSF spinors and have no freedom in the scaling. The GSF spinors arise from a specialized version of (9):

$$g(\pi_{A'}) + t_A{}^{A'}x^{AB'}\pi_{A'}\pi_{B'} = 0, \tag{11}$$

where $t^{AA'}$ is a constant vector on Minkowski space and $g(\pi_{A'})$ is homogeneous of degree 2 in $\pi_{A'}$. The real KS metric is then defined by

$$ds^2 = \eta_{ab}dx^a dx^b + 2H\pi_{A'}\pi_{B'}\bar{\pi}_A\bar{\pi}_B dx^{AA'}dx^{BB'}, \tag{12}$$

where

$$H = m(K + \bar{K}),$$

$$K = (\pi^{B'}\bar{\pi}^A t^{A'C}\bar{\pi}_C\nabla_{AA'}\pi_{B'})(t^{DD'}\bar{\pi}_D\pi_{D'})^{-4}.$$

[Observe that, although (11) does not fix the scale of $\pi_{A'}$, (12) is independent of this scale. Equations (11) and (12) are obtained from (5.66) and (5.70) in Ref. 2.]

The real KS metrics have a Killing vector given by t^a and can only be type (2,1,1) or D [not (3,1) or N]. The self-dual KS metric need not have Killing vectors, although they will have the same Killing vector t^a if based on a special GSF spinor given by (11). Unfortunately, there seems to be no natural way of associating a self-dual KS metric with a given real one. Instead a whole family can be built around the same GSF spinor.

We finish with a couple of examples:

$$\text{(a)}\phi_{A'B'} = O_A x^A{}_{A'}O_B x^B{}_{B'}/(x^a x_a)^3,$$

which leads to

$$ds^2 = 2\,dudv - 2\,dXdY + \lambda\,(Y\,du - u\,dY)^2/(uv - XY)^3;$$

$$\text{(b)}\phi_{A'B'} = 1_A x^A{}_{A'}1_B x^B{}_{B'}/(x^a x_a)^3,$$

which leads to

$$ds^2 = 2\,dudv - 2\,dXdY + \lambda\,(v\,du - X\,dY)^2/(uv - XY)^3.$$

These two metrics, respectively type N and D, are two of the $\mathscr{H}$-spaces found in Ref. 5. In fact, the second of them is the Eguchi–Hanson metric in unfamiliar coordinates.[6]

[1] J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[2] G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[3] I. Robinson, J. Math. Phys. **2**, 290 (1960).
[4] R. Penrose and M. A. H. MacCallum, Phys. Rep. C **6**, 241 (1973).
[5] G. A. J. Sparling and K. P. Tod, "An example of an $\mathscr{H}$ space" J. Math. Phys. **22**, 331 (1981).
[6] G. Burnett-Stuart, in *Twistor Newsletter* **9** (Mathematical Institute, Oxford, 1979).

Spacetimes admitting a vector field whose inner product with the Riemann tensor is zero

C.B.G. McIntosh and E. H. van Leeuwen
Department of Mathematics, Monash University, Clayton, Victoria, 3168 Australia

(Received 15 July 1980; accepted for publication 3 April 1981)

The equation $v_\mu R^\mu{}_{\nu\alpha\beta} = 0$ arises in various places in general relativity, in particular as the integrability conditions of the equations $\mathcal{L}_{\mathbf{v}}\mathbf{g} = 2\phi\mathbf{g}$, where ϕ is a constant. These are the equations of a homothetic vector field $\mathbf{v}$, with a zero homothetic bivector ($d\mathbf{v} = 0$) in some spacetime with metric $\mathbf{g}$, and $R^\mu{}_{\nu\alpha\beta}$ are the components of the Riemann tensor of that metric in some frame. In this paper the equation $v_\mu R^\mu{}_{\nu\alpha\beta} = 0$ is examined and the components of the Riemann tensor for the spacetimes which admit nonzero solutions v_μ of this equation are given. The Petrov types of the Weyl tensors of these spacetimes are listed and, as a result, a correction is then made to a theorem in a paper by Collinson and Fugère about the Petrov types of spacetimes, which admit the type of separation that they require of the Hamilton–Jacobi equation for these spacetimes.

PACS numbers: 04.20.Jb, 02.40.Ky

1. INTRODUCTION

In various places in general relativity the condition arises that, for a given metric tensor $\mathbf{g}$, there is a vector field $\mathbf{v}$ which satisfies

$$v_\mu R^\mu{}_{\nu\alpha\beta} = 0, \tag{1.1}$$

where the $R^\mu{}_{\nu\alpha\beta}$ are the components of the Riemann tensor of $\mathbf{g}$ in some frame. For example, this equation arises as the integrability conditions of the equations

$$v_{\mu;\nu} = \phi g_{\mu\nu}, \quad \phi \text{ constant}. \tag{1.2}$$

These are the equations satisfied by the homothetic vector field $\mathbf{v}$ with zero homothetic bivector (HBV) ($v_{[\mu;\nu]} = 0$) or, when $\phi = 0$, by a Killing vector field with zero bivector. McIntosh[1] has shown, by using results of Ehlers and Kundt[2] and Debney,[3] that if a vacuum spacetime admits a vector field $\mathbf{v}$ with zero HBV, as in (1.2), then $\mathbf{v}$ is a null Killing vector field and the spacetime metric is the type N, *pp* wave metric.

In this paper, Eq. (1.1) is discussed in a general situation and the equation is examined to see what restrictions are placed on the Petrov types of the spacetimes in which there is a vector $\mathbf{v}$ which satisfies (1.1). The following result is obtained:

Theorem 1.1: If a spacetime g admits a vector field $\mathbf{v}$ which satisfies (1.1), then the Weyl tensor of $\mathbf{g}$ is

(i) algebraically special if $\mathbf{v}$ is null,

(ii) of Petrov type I or D if $\mathbf{v}$ is timelike,

(iii) of any Petrov type if $\mathbf{v}$ is spacelike.

Part of the motivation for studying (1.1) is to correct a claim by Collinson and Fugère,[4] who discuss the separability of the Hamilton–Jacobi equations for geodesics in n-dimensional Riemannian or pseudo-Riemannian manifolds. They state that in spacetime the Hamilton–Jacobi equation can only separate, in the way they require, when the Weyl tensor of these spacetimes is of Petrov type I, D or, provided the separable coordinate is spacelike, type II. It is shown here that it follows from Theorem 1.1 that their statement is not correct. There is, in fact, no restriction on the Petrov type if the separable coordinate is spacelike, since the existence of a spacelike separable coordinate is equivalent to the existence of a spacelike vector $\mathbf{v}$ which satisfies (1.1) and theorem 1.1 holds.

2. FORMALISM AND BASIC EQUATIONS

It is easiest to examine Eq. (1.1) by using differential forms and the language of Newman and Penrose (NP).[5] The conventions of Debever, McLenaghan, and Tariq[6] are followed. The spacetime manifold is spanned by four null vectors $\mathbf{l}$, $\mathbf{n}$, $\mathbf{m}$, and $\bar{\mathbf{m}}$, where $\mathbf{l}$ and $\mathbf{n}$ are real and $\mathbf{m}$ is complex, and the metric takes the form

$$ds^2 = \mathbf{l}\otimes\mathbf{n} - \mathbf{m}\otimes\bar{\mathbf{m}} = g_{ab}\theta^a\theta^b, \tag{2.1}$$

where

$$\theta^1 = \mathbf{n}, \quad \theta^2 = \mathbf{l},$$
$$\theta^3 = -\bar{\mathbf{m}}, \quad \theta^4 = \bar{\theta}^3 = -\mathbf{m}. \tag{2.2}$$

The curvature two-forms are defined by

$$\Theta^a_b = \tfrac{1}{2}R^a{}_{bcd}\theta^c\wedge\theta^d. \tag{2.3}$$

Then, in the language of NP, where $\Psi_A (A = 0\text{–}4)$ and Φ_{AB} $(A = 0\text{–}2)$ are the tetrad components of the Weyl tensor and trace-free Ricci tensor, respectively, and Λ is proportional to the trace of the Ricci tensor, the independent curvature two-forms are

$$\Theta^4{}_1 = \Theta^2{}_3 = \Psi_0 Z^1 + \Psi_1 Z^2 + (\Psi_2 + 2\Lambda)Z^3 + \Phi_{00}\bar{Z}^1 + \Phi_{01}\bar{Z}^2 + \Phi_{02}\bar{Z}^3, \tag{2.4a}$$

$$\Theta^2{}_2 = -\Theta^1{}_1 = (\Psi_1 + \Phi_{01})Z^1 + (\Psi_2 - \Lambda + \Phi_{11})Z^2 + (\Psi_3 + \Phi_{21})Z^3 + (\bar{\Psi}_1 + \Phi_{10})\bar{Z}^1 + (\bar{\Psi}_2 - \Lambda + \Phi_{11})\bar{Z}^2 + (\bar{\Psi}_3 + \Phi_{12})\bar{Z}^3, \tag{2.4b}$$

$$\Theta^4{}_4 = -\Theta^3{}_3 = (\Psi_1 - \Phi_{01})Z^1 + (\Psi_2 - \Lambda - \Phi_{11})Z^2 + (\Psi_3 - \Phi_{21})Z^3 - (\bar{\Psi}_1 - \Phi_{10})\bar{Z}^1 - (\bar{\Psi}_2 - \Lambda - \Phi_{11})\bar{Z}^2 - (\bar{\Psi}_3 - \Phi_{12})\bar{Z}^3, \tag{2.4c}$$

$$\Theta^1{}_4 = \Theta^3{}_2 = -(\Psi_2 + 2\Lambda)Z^1 - \Psi_3 Z^2 - \Psi_4 Z^3 - \Phi_{20}\bar{Z}^1 - \Phi_{21}\bar{Z}^2 - \Phi_{22}\bar{Z}^3. \tag{2.4d}$$

 0022-2488/82/061149-04$02.50

The other nonzero $\Theta^a{}_b$ can be found from these equations by taking the complex conjugates of (2.4a) and (2.4d)

Equation (1.1) can now be written in tetrad form as

$$v_a \Theta^a{}_b = 0 \tag{2.5}$$

and it is in this form that the equation can be investigated to see what relationships are imposed on the Ψ_A, Φ_{AB}, and Λ by taking various forms of the nonzero v_a for the cases where v is null, timelike, and spacelike.

3. SOLUTIONS OF (1.1)

(*a*) v *null*: When v is a null vector field, the θ^a can be chosen so that, for

$$\mathbf{v} = v_a \theta^a, \tag{3.1}$$

θ^2 and v are aligned. Then

$$\mathbf{v} = v_2 \theta^2, \quad v_1 = v_3 = v_4 = 0. \tag{3.2}$$

In this case (2.5) gives

$$\Theta^2{}_1 = \Theta^2{}_3 = \Theta^2{}_4 = 0. \tag{3.3}$$

Then (2.4) gives

$$\begin{aligned} &0 = \Psi_0 = \Psi_1 = \Phi_{00} = \Phi_{01} = \Phi_{02}, \\ &\Psi_2 = -2\Lambda, \quad \Phi_{11} = 3\Lambda, \quad \Psi_3 = -\Phi_{21}, \\ &\Psi_4 \text{ and } \Phi_{22} \text{ arbitrary.} \end{aligned} \tag{3.4}$$

Thus, only spacetimes whose Riemann tetrad components satisfy these relationships can admit a null vector v which satisfies (2.5). In vacuum, as mentioned in the Introduction, this means that the spacetime must have a Weyl tensor of Petrov type N. In all cases the Weyl tensor is algebraically special and l is a repeated principal null congruence.

(*b*) v *timelike*: When v is timelike, the θ^a can be chosen so that

$$\mathbf{v} = v(\theta^1 + \theta^2), \quad \mathbf{v}\cdot\mathbf{v} = 2v^2 > 0. \tag{3.5}$$

Then (2.4) and (2.5) give the requirement that the nonzero NP quantities satisfy

$$\begin{aligned} &\Psi_0 = \overline{\Psi}_4 = \Phi_{02}, \quad \Psi_1 = -\overline{\Psi}_3 = \Phi_{12} = -\Phi_{01}, \\ &\Psi_2 = \Lambda - \Phi_{11}, \quad \Psi_2 + 2\Lambda = \Phi_{00} = \Phi_{22} = 3\Lambda - \Phi_{11}. \end{aligned} \tag{3.6}$$

It is shown in the Appendix that these restrictions mean that the Weyl tensor has Petrov type I or D. There is no vacuum solution which satisfies these restrictions.

(*c*) v *spacelike*: Similarly with v spacelike, the tetrad can be chosen so that

$$\mathbf{v} = v(\theta^2 - \theta^1), \quad \mathbf{v}\cdot\mathbf{v} = -2v^2 < 0 \tag{3.7}$$

and (2.4) and (2.5) now give

$$\begin{aligned} &\Psi_0 = \overline{\Psi}_4 = -\Phi_{02}, \quad \Psi_1 = \overline{\Psi}_3 = -\Phi_{12} = -\Phi_{01}, \\ &\Psi_2 = \Lambda - \Phi_{11}, \quad \Psi_2 + 2\Lambda = -\Phi_{00} = -\Phi_{22} \\ &\qquad\qquad = 3\Lambda - \Phi_{11}. \end{aligned} \tag{3.8}$$

The Weyl tensor can now be of any Petrov type (see Appendix). Again there is no vacuum solution which satisfies these restrictions.

There is another very simple possibility for v in the case where it is spacelike, and this is

$$\mathbf{v} = v(\theta^3 + \overline{\theta}^3). \tag{3.9}$$

This would have led to

$$\begin{aligned} &\Psi_0 = -\Phi_{00}, \quad \Psi_4 = -\Phi_{22}, \\ &\Psi_1 = -\overline{\Psi}_1 = \Phi_{01}, \quad \Psi_3 = -\overline{\Psi}_3 = \Phi_{21}, \\ &\psi_2 = \Lambda + \Phi_{11}, \\ &\psi_2 + 2\Lambda = -\Phi_{02} = 3\Lambda + \Phi_{11}, \end{aligned} \tag{3.10}$$

rather than (3.8), but a tetrad rotation can be found which maps (3.9) and (3.10) into (3.7) and (3.8).

From these results the statement of Theorem 1.1 in the Introduction now follows immediately.

4. CLASSIFICATION OF THE RIEMANN TENSORS

McIntosh and Halford[7] discuss the equation

$$x_{\mu\nu} R^{\mu}{}_{\lambda\alpha\beta} + x_{\mu\lambda} R^{\mu}{}_{\nu\alpha\beta} = 0 \tag{4.1}$$

in general relativity; in particular in relationship with curvature collineations and with those spacetimes whose Riemann tensors have solutions for $x_{\mu\nu}$ in (4.1) which are not propotional to the metric tensor components $g_{\mu\nu}$.

Hall and McIntosh[8] have shown that almost all nontrivial solutions of (4.1) are of the form

$$x_{\mu\nu} = \phi g_{\mu\nu} + \alpha v_\mu v_\nu, \tag{4.2}$$

where ϕ and α are aribtrary scalar fields and where v_μ satisy

$$v_\mu R^{\mu}{}_{\nu\alpha\beta} = 0. \tag{4.3}$$

Indeed Collinson[9] showed that in vacuum the only nontrivial solutions of (4.1) were of this form where α is not zero the metric has its Weyl tensor of Petrov type N, and l is tangent to the four-fold repeated principal null direction of the Weyl tensor of such a metric. Examples can easily be found of spacetimes where (4.1) does have nontrivial solutions which are not of the form (4.2) and (4.3), but this paper incidentally examines the spacetimes which do have solutions of the form (4.2) and (4.3). Indeed, every spacetime in which there exists a vector v which satisfies (4.3) automatically has nontrivial solutions of (4.1) of the form (4.2).

There is in McIntosh and Halford[7] a list of the classification of all nontrivial solutions $x_{\mu\nu}$ of (4.1), where the classfication is in various Plebański classes as outlined by Plebański[10] and as given in NP canonical form by McIntosh, Foyster, and Lun.[11]

With v null, the restrictions (3.4) on the Riemann tensor components are the same as those when $x_{\mu\nu}$ in (4.1)–(4.3) has type $[4N]_2$ in Plebański's classification. When v is timelike, the main type is $[T\text{-}3S]_2$ and when v is spacelike, the main type is $[3S\text{-}T]_2$. Again there are degenerate cases in all these forms of nonzero components of the Riemann tensor tetrad components and corresponding higher types of the $x_{\mu\nu}$.

5. HAMILTON-JACOBI SEPARABILITY

Collinson and Fugère[4] discuss the Hamilton-Jacobi equation for geodesics, namely

$$g^{\mu\nu} S_{,\mu} S_{,\nu} - m^2 = 0, \tag{5.1}$$

and look for solutions S of this equation which separate with respect to the coordinate x^1 such that S can be written

$$S = S_1(x^1) + S(x^2,x^3,x^4). \tag{5.2}$$

Their definitions lead them to study a metric $\tilde{\mathbf{g}}$ which is conformal to $\mathbf{g}$ and which admits a non-null hypersurface orthogonal Killing vector $\tilde{\mathbf{v}}$. This Killing vector then has to satisfy

$$\tilde{v}_{\mu;\nu} = 0 \tag{5.3}$$

and therefore the integrability conditions

$$\tilde{v}_\mu \tilde{R}^\mu{}_{\nu\alpha\beta} = 0. \tag{5.4}$$

They claim that in this case $\tilde{\mathbf{v}}$ can be written as

$$\tilde{\mathbf{v}} = \psi(\theta^2 + \epsilon\theta^1), \quad \epsilon = \pm 1, \tag{5.5}$$

where ψ is an arbitrary scalar field and θ^2 is a principal null vector of the Weyl tensor in the spacetime under discussion. They then discuss the Weyl tensor of this spacetime and list the results which are mentioned in the Introduction of this paper. Their results would be true if it were always possible to choose $\mathbf{v}$ to satisfy (5.5) with θ^2 a principal null vector. However, it is not possible in general to satisfy both of these requirements simultaneously. If $\mathbf{v}$ is chosen to satisfy (5.5) then, since (5.4) still holds, the results of Sec. 3(b) and 3 (c) hold and the restrictions on the Weyl tensor as listed in Theorem 1.1 hold.

These changes mean that the theorem listed in their paper as Theorem 1.1 should thus read:

Theorem 5.1: The Hamilton-Jacobi equation can only separate in space-times of Petrov type I or D if the separable coordinate is timelike, or of any Petrov type if the separable coordinate is spacelike.

6. DISCUSSION

One spacetime in particular can be shown to always possess a vector $\mathbf{v}$ which satisfies (1.1) and this is the one which satisfies the conditions of the following theorem.

Theorem 6.1: A spacetime which possesses a hypersurface orthogonal homothetic vector field $\mathbf{v}$ which is of nonzero constant length is such that the vector field is covariantly constant and thus the homothetic vector field is a Killing vector field.

The proof follows the working of the first part of Sec. 3 of Collinson and Fugère,[4] except that $\mathbf{v}$ satisfies $\mathcal{L}_\mathbf{v}\mathbf{g} = 2\psi\mathbf{g}$ to begin with rather than just $\mathcal{L}_\mathbf{v}\mathbf{g} = 0$. Here ψ is a constant. The equations from the assumptions soon give $\psi = 0$ and

$$v_{\mu;\nu} = 0 \tag{6.1}$$

and hence the theorem is proved and (1.1) is satisfied by $\mathbf{v}$.

The following, related, result is worth noting.

Theorem 6.2: If a spacetime admits a homothetic vector field $\mathbf{v}$ with zero homothetic bivector, and if $\mathbf{v}$ has constant length, then $\mathbf{v}$ is a Killing vector field.

The proof just uses $v^\mu v_\mu = \text{const}$, which implies $v^\mu v_{\mu;\nu} = 0$ and (1.2). These equations give $\phi = 0$ and the theorem is proved. Equation (1.1) is automatically satisfied. Notice that this holds for null vectors $\mathbf{v}$ as well as for other constant length, non-null vectors $\mathbf{v}$.

This result was used in the proof of Theorem 3.2 in McIntosh.[1] This is the result discussed in the Introduction. However, that theorem was only concerned with vacuum spacetimes.

Equation (1.1) arises in other spacetimes. It is the general case which is discussed in this paper. The main result perhaps is that there are very few spacetimes which admit a vector field $\mathbf{v}$ which satisfies (1.1). The tetrad components of the Riemann tensors of these spacetimes satisfy the strong restrictions (3.4), (3.6), and (3.8) for $\mathbf{v}$ null, spacelike and timelike. Some spacetimes which satisfy these conditions (or at least subcases of these conditions) are discussed by McIntosh and Halford[7] in their discussion of (4.1). One of the results of the restriction in (3.4), (3.6), and (3.8) is that the Weyl tensor in these spacetimes satisfy the restrictions imposed by Theorem 1.1. This result leads to a change of a result by Collinson and Fugère.[4]

APPENDIX

When a spacetime metric admits a timelike vector $\mathbf{v}$ which satisfies $v_\mu R^\mu{}_{\nu\alpha\beta} = 0$, it is stated in Sec. 3(b) that a null tetrad $\mathbf{l}$, $\mathbf{n}$, $\mathbf{m}$, and $\bar{\mathbf{m}}$ can be chosen such that the possible nonzero NP components of the Weyl tensor are

$$\Psi_0 = \bar{\Psi}_4, \quad \Psi_1 = -\bar{\Psi}_3, \quad \Psi_2 = \bar{\Psi}_2. \tag{A1}$$

It will be shown that these restrictions mean that the Weyl tensor has Petrov type I or D.

Consider the null rotation which leaves $\mathbf{l}$ fixed:

$$\begin{aligned} &\mathbf{l}' = \mathbf{l}, \quad \mathbf{m}' = \mathbf{m} + a\mathbf{l}, \\ &\mathbf{n}' = \mathbf{n} + a\bar{\mathbf{m}} + \bar{a}\mathbf{m} + a\bar{a}\,\mathbf{l}, \end{aligned} \tag{A2}$$

where a is an arbitrary complex scalar. Then $\mathbf{l}'$ is a repeated principal null vector if a can be found such that both Ψ_4' and Ψ_3' are zero, i.e.,

$$\begin{aligned} \Psi_4' &= 0 = \Psi_4 + 4\bar{a}\Psi_3 + 6\bar{a}^2\Psi_2 - 4\bar{a}^3\bar{\Psi}_3 + \bar{a}^4\bar{\Psi}_4, \\ \Psi_3' &= 0 = \Psi_3 + 3\bar{a}\Psi_2 - 3\bar{a}^2\bar{\Psi}_3 + \bar{a}^3\bar{\Psi}_4. \end{aligned} \tag{A3}$$

It follows after some straightforward algebra that the Ψ_A in this case must satisfy

$$\begin{aligned} &(1 - a\bar{a})\Psi_4 + 2\bar{a}\Psi_3 = 0, \\ &6\bar{a}^2\Psi_2 - (1 - 4a\bar{a} + a^2\bar{a}^2)\Psi_4 = 0, \end{aligned} \tag{A4}$$

from which it follows that

$$a^2\Psi_4 = \bar{a}^2\bar{\Psi}_4, \quad \Psi_4\bar{\Psi}_3^2 = \bar{\Psi}_4\Psi_3^2. \tag{A5}$$

Now consider the similar null rotation which leaves $\mathbf{n}'$ fixed:

$$\begin{aligned} &\mathbf{n}'' = \mathbf{n}', \quad \mathbf{m}'' = \mathbf{m}' + b\mathbf{n}', \\ &\mathbf{l}'' = \mathbf{l}' + b\bar{\mathbf{m}}' + \bar{b}\mathbf{m}' + b\bar{b}\mathbf{n}', \end{aligned} \tag{A6}$$

where b is an arbitrary complex scalar. Then $\Psi_3'' = \Psi_4'' = 0$. With a satisfying (A4) and with

$$b = -a/(1 + a\bar{a}), \tag{A7}$$

it follows that $\Psi_0'' = \Psi_1'' = 0$ and thus the only zero Ψ_A'' is Ψ_2''. In this case the Petrov type is D. If no such a exists the type is I.

Another way of obtaining this result is to start with the Ψ_A in canonical form, one Petrov type at a time, and show that there are no rotations (A2) and (A6) such that the Ψ_A'' end up satisfying the double prime version of (A.1), except for the Petrov types D and, of course, I.

This last argument can be used in Sec. 3(c) when $\mathbf{v}$ is spacelike and

$$\Psi_0 = \bar{\Psi}_4, \quad \Psi_1 = \bar{\Psi}_3, \quad \Psi_2 = \bar{\Psi}_2. \tag{A8}$$

In this case the canonical forms of all the Petrov types can be taken one at a time and the Ψ_A put into this form by means of the null rotations (A2) and (A6). Thus the Weyl tensor of such a spacetime can be of any single Petrov type.

[1]C. B. G. McIntosh, Gen. Rel. Grav. **7**, 199 (1976).
[2]J. Ehlers and W. Kundt, *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), Chap. 2.
[3]G. C. Debney, J. Math. Phys. **12**, 1088 (1971).
[4]C. D. Collinson and J. Fugère, J. Phys. A **10**, 1877 (1977).
[5]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[6]R. Debever, R. G. McLenaghan, and N. Tariq, Gen. Rel. Grav. **10**, 853 (1979).
[7]C. B. G. McIntosh and W. D. Halford, J. Math. Phys. **23**, 436 (1982).
[8]G. S. Hall and C. B. G. McIntosh, "On the determination of the metric from the curvature in general relativity" (preprint).
[9]C. D. Collinson, J. Math. Phys. **11**, 818 (1970).
[10]J. Plebański, Acta Phys. Polon. **26**, 963 (1964).
[11]C. B. G. McIntosh, J. Foyster, and A. W.-C. Lun, J. Math. Phys. **22**, 2620 (1981).

The metric dependence of four-dimensional formulations of electromagnetism

R. R. Birss
Physics Department, University of Salford, Salford M5 4WT, United Kingdom

(Received 10 August 1981; accepted for publication 7 January 1982)

Three-dimensional and four-dimensional formulations of electromagnetism are compared with particular regard to whether it is necessary to postulate a metric locally (i.e., to define an inner product in the tangent space). In three dimensions, a metric-independent formulation is possible but in four dimensions the required set of fundamental differential forms can not be defined by experiment on a differential manifold without a metric.

PACS numbers: 04.40. + c

A recent pair of papers[1,2] in this journal contained a description of, and comments upon, an interpretation of flux, charge and angular momentum quantization in terms of a set of naturally independent concepts relating to the periods of one-, two- and three-dimensional cyclic integrals in space–time. In making[1] these suggestions, Kiehn acknowledged the importance of prior work by Post[3,4] (relating flux and charge quanta to one- and two-dimensional periods) and studied the topology of fields built on a four-dimensional space–time in terms of a set of fundamental differential forms which are assumed to exist for all physical systems that are usually described in terms of a field theory.

An important feature of this work is the metric invariance[2,4] and the (much stronger) topological invariance[2] of the periods of these integrals. Correspondingly, the space–time manifold itself is neither equipped[4] with a particular topology nor, indeed, endowed[4] with a metric (which would, of course, automatically induce a topology). However, the purpose of the present paper is to question whether the required set of fundamental differential forms can, in fact, be defined on a differentiable manifold without a metric, when the physical theory is the four-dimensional formulation of electromagnetism.

The pair two-form F involving the fields **E** and **B** is expressed by Post[5] in a form equivalent to

$$F = F_{\lambda_1\lambda_2}\, d_1 q^{[\lambda_1} d_2 q^{\lambda_2]} \quad (\lambda_1,\lambda_2 = 0,1,2,3), \tag{1}$$

where the index under the symbol d indicates a restriction, namely that $d_1 q^{\lambda_1}$, for example, must correspond to an increment along the q^{λ_1} coordinate line. More generally, however, no such restriction obtains, since an exterior differential p-form can be defined[6] either as (a) an element of order p of the exterior algebra $T_n^{*\Lambda(p)}$ constructed on the vector space of Pfaffian forms (i.e., the space T_n^* of linear functionals on the tangent space T_n) or, equivalently, as (b) an element of $(T_n^{\Lambda(p)})^*$, that is a linear functional defined on the space of order p of the exterior algebra $T_n^{\Lambda(p)}$ constructed on T_n. Alternative (b) leads to an (unrestricted) form such as

$$F = F_{\lambda_1\lambda_2}\, du^{[\lambda_1} dv^{\lambda_2]}, \tag{2}$$

where **du** and **dv** are *independent* vectors in the tangent space T_4, and whether Eq. (1) or (2) is considered to be the more appropriate depends upon an examination of the experimental procedures used in making field measurements.

The restricted form (1) corresponds to a belief that a field measurement is in actual fact a measurement of an integral of a field in such a way that the field in the domain of integration may be assumed to be constant. On this view, any specification of the experimental arrangement is irrelevant to the specification of F in (1) but defines, instead, the domain over which the integration of F must be evaluated. An unrestricted form such as (2) corresponds to the belief that the experimental process of measurement defines[7] the corresponding differential form directly, notwithstanding the fact that the experimental procedures involve dimensions that are not truly infinitesimal but only small enough to allow spatial variations of fields to be ignored. At this point, however, it is important to make a clear distinction between the three- and four-dimensional cases.

In three dimensions, experiment defines unambiguously multilinear antisymmetrical maps[7] $\sigma : T_3 \to R$, which in turn define the linear maps $\tau : T_3^{\Lambda(p)} \to R$ that constitute differential forms of type (2), such as $\epsilon = \tau_1(\mathbf{du}) = E_{\mu_1}\, \mathbf{du}^{\mu_1}$ and $\Phi = \tau_2(\mathbf{dv}_\Lambda \mathbf{dw}) = B_{\mu_1\mu_2} dv^{[\mu_1} dw^{\mu_2]}$, where $\mu_1,\mu_2 = 1,2,3$ and the generalized fields E_{μ_1} and $B_{\mu_1\mu_2}$ correspond to a particular choice of basis $\mathbf{e}_{\mu_1}$ for T_3. Such maps can be established experimentally not only *before* any coordinates are introduced but also *without* any need to postulate a metric, even locally (i.e., it is not necessary to define an inner product in the tangent space). In three dimensions, therefore, a metric-independent formulation such as that involving the F of Eq. (2) is entirely appropriate and the same conclusion may be drawn about the differential form (G) involving the other two fields (**H** and **D**), except that, as correctly emphasized by Post,[5] G is (unlike F) an *impair* two-form. In four dimensions, however, the situation is different.

Two inertial observers, O and $\bar{O}$, in relative motion can both perform experimentation to determine maps of the form τ but, clearly, these maps must alter in character on passing from O to $\bar{O}$, because they reflect how a particular observer partitions the electromagnetic field into electrical and magnetic parts. What remains invariant, however, is the composite map $\phi : T_4 \to R$ with values

$$\begin{aligned}\phi &= \tau_1(dU^{\mu_1}\mathbf{e}_{\mu_1})dV^0 - \tau_1(dV^{\mu_2}\mathbf{e}_{\mu_2})dU^0 \\ &\quad + \tau_2(dU^{\mu_1}\mathbf{e}_{\mu_1\Lambda} dV^{\mu_2}\mathbf{e}_{\mu_2}) \\ &= E_{\mu_1} dU^{\mu_1} dV^0 - E_{\mu_2} dV^{\mu_2} dU^0 \\ &\quad + B_{\mu_1\mu_2} dU^{[\mu_1} dV^{\mu_2]},\end{aligned} \tag{3}$$

 0022-2488/82/061153-02$02.50

where **dU** and **dV** correspond to events, i.e., $dU^{\lambda_1}, dV^{\lambda_2} \in T_4$, with $\lambda_1, \lambda_2 = 0,1,2,3$. This is the required differential two-form (i.e., F) and the reason that Eq. (3) is written in a different form from Eq. (2) is to emphasize that the experimental definition of ϕ involves two different sets of vectors, namely those involved in the definition of the generalized fields and those involved in the specification of two arbitrary events. That these two can not coincide can be readily seen by considering measurements of magnetic flux: a selection by O of fixed vectors **du** and **dv** in his reference frame allows him to measure $\Phi = B_{\mu_1\mu_2} du^{[\mu_1} dv^{\mu_2]}$ but the corresponding quantity $\bar{\Phi} = \bar{B}_{\mu_1\mu_2} d\bar{u}^{[\mu_1} d\bar{v}^{\mu_2]}$, obtained from the first by a Lorentz transformation involves vectors that are not fixed in $\bar{O}$'s frame and (despite the algebraic similarity) it does not, therefore, correspond to a measurement of flux in $\bar{O}$'s reference frame.

In fact, ϕ is a composite mapping the invariance of which rests on two sets of experimental input: first, measurements of τ_1 and τ_2 performed by individual inertial observers and, secondly, the coordination by these observers of events that must be chosen arbitrarily to justify the existence of a linear functional defined on the space of order 2 of the exterior algebra $T_4^{\Lambda(2)}$ constructed on T_4. Although the former procedure does not require that a metric be defined, the latter cannot be carried out without a metric, because the coordinatization of events involves not the raw data received by an observer but his logging of events after allowing for the time taken for signals to reach him. It may therefore be concluded that the required set of fundamental differential forms (e.g., F and G) can *not* be defined by experiment on a differential manifold without a metric. The corollary that space–time must be regarded from the outset as being endowed with a semidefinite metric makes it easier to accommodate the fact that all known residues of G appear to relate to two-dimensional cyclic domains of a purely spatial nature, because, as pointed out by Post,[8] the semidefinite nature of the space–time metric retains a distinction between space and time through Sylvester's theorem of inertia.

[1] R. M. Kiehn, J. Math. Phys. **18**, 614 (1977).
[2] E. J. Post, J. Math. Phys. **18**, 2084 (1977).
[3] E. J. Post, "Current Status of Geometrodynamics," McGill University Progress Report to the "Foundations and Philosophy of Science Unit," 1972.
[4] E. J. Post, Phys. Rev. D **9**, 3379 (1974).
[5] E. J. Post, Found. Phys. **7**, 255 (1977).
[6] A. Lichnerowicz, *Linear Algebra and Analysis* (Holden-Day, San Francisco, 1967).
[7] R. R. Birss, Phys. Lett. A **78**, 223 (1980).
[8] E. J. Post, Found. Phys. **9**, 831 (1979).

On a class of exact solutions to the Fokker–Planck equations

L. Garrido and J. Masoliver
Departamento de Física Teórica, Universidad de Barcelona, Diagonal, 647, Barcelona-28, Spain

(Received 9 December 1981; accepted for publication 19 February 1982)

In this paper we study under which circumstances there exists a general change of gross variables that transforms any Fokker–Planck equation into another of the Ornstein–Uhlenbeck class that, therefore, has an exact solution. We find that any Fokker–Planck equation will be exactly solvable by means of a change of gross variables if and only if the curvature tensor and the torsion tensor associated with the diffusion is zero and the transformed drift is linear. We apply our criteria to the Kubo and Gompertz models.

PACS numbers: 05.30. – d

I. INTRODUCTION

The Fokker–Planck equation (FPE) describes a generalized diffusion process for the probability density $P(q,t) = P(q^1, \ldots q^n; t)$ in an n-dimensional physical space spanned by the n gross variables $\{q^\nu, \nu = 1, \ldots, n\}$. If $D^{\mu\nu}(q)$ and $f^\mu(q)$ are the diffusion matrix and the drift, the FPE reads

$$\dot{P}(q,t) = -\partial_\mu [f^\mu(q) - \tfrac{1}{2}\partial_\nu D^{\mu\nu}(q)] P(q,t) \tag{1.1}$$

when ∂_μ means $\partial/\partial q^\mu$.

It has been possible to find an exact solution to the FPE when the diffusion is constant and when the drift is linear.[1] There exists also a class of FPE's that have exact solution and whose model is an FPE with linear drift and diffusion $\delta^{\mu\nu}$.[2] Any FPE of this class can be reduced to its model by means of a change in the gross variables. We may ask ourselves immediately the following question: Under which conditions does there exist a change of gross variables that transforms any FPE into another FPE with linear drift and diffusion $\delta^{\mu\nu}$, i.e., that has an exact solution? This last question has a complete answer and constitutes the central result of this paper, in which we intend to give the necessary and sufficient conditions to determine if any FPE has an exact solution related with the exact solution of the Ornstein–Uhlenbeck process by means of a general change of gross variables.

In Sec. II we show how any FPE can be transformed into another one with exact solution. We use here the covariant formulation of FPE[3,4] by means of which we obtain a clear and rigorous method for such a transformation. In Sec. III we give the necessary and sufficient conditions to be satisfied by our original FPE so that such a transformation of gross variables exists: the curvature and torsion tensors have to be zero. These criteria characterized with precision a class of FPE's that have exact solutions.[2] In Sec. IV we study some subclasses of FPE that have physical importance, among which we find the Kubo[5] and Gompertz[6] models.

In this paper we suppose the sum over repeated indices except those that are within a parenthesis.

II. TRANSFORMATION OF AN FPE INTO ANOTHER ONE WHOSE DIFFUSION IS $\delta^{\mu\nu}$

Let M be the manifold formed by the physical states of the system.[7] This manifold is characterized by two sets of gross variables, $\{q^\mu\}$ and $\{q'^\mu\}$, related among themselves by means of continuous and differentiable functions and which conserve the number of gross variables. The diffusion tensor $D^{\mu\nu}(q)$ is the metric tensor of the manifold.

In this section we are going to transform a general diffusion process (1.1) into an FPE whose diffusion matrix is $\delta^{\mu\nu}$. This last equation has exact solution if the transformed drift is linear (Ornstein–Uhlenbeck process). We write Eq. (1.1) covariantly by[4]

$$\dot{S}(q,t) = -\nabla_\mu(h^\mu S - \tfrac{1}{2}D^{\mu\nu}\nabla_\mu S) - \tfrac{1}{2}D^{\sigma\tau}\nabla_\nu D_{\tau\sigma}(h^\nu S - \tfrac{1}{2}D^{\mu\nu}\nabla_\mu S), \tag{2.1}$$

where ∇_μ is the covariant derivative of the Riemann connection of $D^{\mu\nu}$[8]

$$\nabla_\mu h^\nu = \partial_\mu h^\nu + \Gamma^\nu_{\rho\mu} h^\rho \tag{2.2}$$

where $\Gamma^\nu_{\rho\mu}$ are the Christoffel symbols

$$\Gamma^\nu_{\rho\mu} = \tfrac{1}{2} D^{\nu\sigma}(\partial_\mu D_{\sigma\rho} + \partial_\rho D_{\sigma\mu} - \partial_\sigma D_{\rho\mu}). \tag{2.3}$$

The function $S(q,t)$ is the scalar probability density

$$S(q,t) = P(q,t) D^{1/2} \quad (D = \det D^{\mu\nu}); \tag{2.4}$$

the vector $h(q)$ is the covariant drift[4]

$$h^\nu(q) = f^\nu(q) - \tfrac{1}{2} D^{1/2}\partial_\mu(D^{\mu\nu}/D^{1/2}), \tag{2.5}$$

$$D_{\mu\nu} = (D^{-1})^{\mu\nu}. \tag{2.6}$$

Let $\{q'^\nu\}$ be the new "gross variables" related to the old ones by means of

$$J^\sigma_\rho = \frac{\partial q'^\sigma}{\partial q^\rho}, \quad J^{-1\rho}_\sigma = \frac{\partial q^\rho}{\partial q'^\sigma}, \tag{2.7}$$

which obviously satisfies

$$\partial_\sigma J^\nu{}_\mu = \partial_\mu J^\nu_\sigma. \tag{2.8}$$

Let us suppose that the new variables $\{q'^\nu\}$ are such that

$$D'^{\mu\nu} = \delta^{\mu\nu}. \tag{2.9}$$

The corresponding FPE is

$$\dot{P}'(q',t) = -\partial'_\mu [(f'^\mu(q') - \tfrac{1}{2}\partial'_\mu] P'(q',t) \tag{2.10}$$

since for the Euclidean metric $\delta^{\mu\nu}$ the Christoffel symbols are zero and besides

$$S'(q',t) = P'(q',t) \equiv P(q(q'),t),$$
$$h'^\mu(q') = f'^\mu(q') \equiv f^\mu(q(q')). \tag{2.11}$$

Remembering that $f'^\mu = h'^\mu = J_\nu{}^\mu h^\nu$, the transformed drift

 0022-2488/82/061155-04$02.50

is

$$f'^{\mu} = J_{\nu}{}^{\mu}(f^{\nu} + \tfrac{1}{2}D^{\alpha\beta}\Gamma^{\nu}_{\alpha\beta}). \tag{2.12}$$

The transformed FPE (2.10) is exactly solvable if it represents an Ornstein–Uhlenbeck process. The solution of (2.10) with the initial condition

$$P'(q', 0) = \delta^{n}(q' - q_0) \tag{2.13}$$

is[1]

$$P'(q', t) = [\Pi^{n}\det \sigma_{\mu\nu}(t)]^{-1/2} \times \exp\{-\sigma^{-1}_{\mu\nu}(t)[q'^{\mu} - q'^{\mu}(t)][q'^{\nu} - q'^{\nu}(t)]\} \tag{2.14}$$

if from the transformed linear drift

$$f'^{\mu}(q') = A^{\mu}{}_{\nu}q'^{\nu} + B^{\mu} \tag{2.15}$$

we define

$$q'^{\mu}(t) \equiv \gamma^{\mu}_{\rho}(t)[q'^{\rho}_0 + A_{z}^{-1\rho}B^{z}] - A_{\rho}^{-1\mu}B^{\rho}, \tag{2.16}$$

where $\gamma^{\mu}_{\rho}(t)$ is defined by

$$\dot{\gamma}^{\mu}_{\rho}(t) = A^{\mu}_{\alpha}\gamma^{\alpha}_{\rho}(t), \tag{2.17}$$

$$\gamma^{\mu}_{\rho}(0) = \delta^{\mu}_{\rho}, \tag{2.18}$$

and

$$\sigma_{\mu\nu}(t) = \{\gamma^{\rho}_{\mu}(0)\gamma^{\tau}_{\nu}(0) - \gamma^{\rho}_{\mu}(t)\gamma^{\tau}_{\nu}(t)\}\Omega_{\rho\tau} \tag{2.19}$$

$$A^{\alpha}_{\mu}\Omega^{\nu}{}_{\alpha} + \Omega^{\alpha}{}_{\mu}A^{\nu}_{\alpha} = -2\delta^{\nu}{}_{\mu} \tag{2.20}$$

Therefore, if there exists a system of "gross variables" $\{q'^{\nu}\}$ that satisfies Eqs. (2.9) and (2.15), the solution of the original FPE is (2.14) when we substitute, for q', $q'(q)$ and, for $P(q'(q),t)$, $P(q,t)D^{1/2}$. *In this case the integrations of the FPE is simply reduced to the integration of the change of variables (2.7).*

III. NECESSARY AND SUFFICIENT CONDITIONS TO SOLVE EXACTLY AN FPE BY MEANS OF A CHANGE OF VARIABLES

There will not always exist a change of variables $q = q(q')$ such that $D'^{\mu\nu} = \delta^{\mu\nu}$. In the Appendix we show that the necessary and sufficient conditions, in order that $D'^{\mu\nu} = \delta^{\mu\nu}$, are that the curvature tensor $R_{\mu\nu\alpha\beta}$ and torsion tensor $T^{\mu}_{\nu\alpha}$[8] associated with the diffusion matrix be zero.

Therefore the conditions

$$R_{\mu\nu\alpha\beta} = \tfrac{1}{2}(\partial^{2}_{\nu\alpha}D_{\mu\beta} + \partial^{2}_{\mu\beta}D_{\nu\alpha} - \partial^{2}_{\nu\beta}D_{\mu\alpha} - \partial^{2}_{\mu\alpha}D_{\nu\beta}) + D_{\sigma\rho}(\Gamma^{\sigma}_{\nu\alpha}\Gamma^{\rho}_{\mu\beta} - \Gamma^{\sigma}_{\nu\beta}\Gamma^{\rho}_{\mu\alpha}) = 0, \tag{3.1}$$

$$T_{\nu\sigma}{}^{\mu} = \Gamma^{\mu}_{\nu\alpha} - \Gamma^{\mu}_{\alpha\nu} = 0 \tag{3.2}$$

guarantee the existence of a change of variables determined by the matrix $J_{\lambda}{}^{\mu}$ such that in the new variables the diffusion matrix $D'^{\mu\nu}$ becomes $\delta^{\mu\nu}$. If we work with Riemann's connection $D^{\mu\nu}$, the Christoffel symbols are given by Eq. (2.3), and condition (3.2) is satisfied identically.

In the same appendix [Eq. (A3)] we show that the matrix $J_{\lambda}{}^{\mu}$ for the variables transformations must satisfy the relation

$$\partial_{\mu}J_{\nu}{}^{\lambda} = \Gamma^{\alpha}_{\mu\nu}J_{\alpha}{}^{\lambda}. \tag{3.3}$$

For this equation the *formal solution* is

$$J_{\nu}{}^{\lambda}(q) = S^{\alpha}{}_{\nu}(q, q_0)J_{\alpha}{}^{\lambda}(q_0), \tag{3.4}$$

where

$$S^{\alpha}{}_{\nu}(q, q_0) \equiv \exp\left[\int_{q_0}^{q} \Gamma^{\alpha}_{\mu\nu}(q'')\, dq''^{\mu}\right]. \tag{3.5}$$

On this formal solution we should impose "initial conditions" that are not given here.

The diffusion matrix in the variables $\{q'^{\nu}\}$ becomes a constant matrix; therefore, the general solution of (3.4) should satisfy:

$$J_{\mu}^{-1\sigma}J_{\nu}^{-1\rho}D^{\mu\nu}(q) = A^{\sigma\rho} \tag{3.6}$$

or, equivalently,

$$J^{\sigma}_{\mu}J^{\rho}_{\nu}A_{\sigma\rho} = D_{\mu\nu}(q), \tag{3.7}$$

where $A_{\sigma\rho}$ is a constant matrix. Concretely, we could take $A_{\sigma\rho} = \delta_{\sigma\rho}$; in this case the transformation matrix satisfies

$$D_{\mu\nu}(q) = J^{\sigma}_{\mu}J^{\rho}_{\nu}\delta_{\sigma\rho}. \tag{3.8}$$

Relation (3.8) is used in Ref. 2 to characterize a whole class of FPE that have exact solutions.

In order that the transformed FPE represent an Ornstein–Uhlenbeck process, which has an exact solution, it is also necessary that the transformed drift $f'^{\mu}(q')$ be linear, i.e.,

$$\partial^{2}_{\sigma\lambda}f'^{\mu}(q') = 0. \tag{3.9}$$

In the coordinates q' the covariant derivation ∇' coincides with the ordinary derivatives and besides the drift f'^{μ} is the covariant drift h'^{μ}. Therefore, the covariant expression for (3.9) is

$$\nabla'_{\gamma}\nabla'_{\gamma}h'^{\mu} = 0, \tag{3.10}$$

which if written in the original variables becomes

$$\nabla_{\gamma}\nabla_{\lambda}h^{\mu} = 0. \tag{3.11}$$

Let us observe that the condition (3.11) does not require a knowledge of the change of gross variables: it is sufficient to know the Christoffel symbols $\Gamma^{\mu}_{\nu\alpha}$ that can be evaluated differentiating the diffusion $D^{\mu\nu}$.

From what has been said above, we can check whether or not any FPE (1.1) represents, in a certain set of gross variables, an Ornstein–Uhlenbeck process that has an exact solution. We could test if the diffusion matrix $D^{\mu\nu}$ satisfies the condition (3.1) and if the drift satisfies the condition (3.11). If the answer is affirmative, we can integrate the change of variables by means of a matrix J^{σ}_{μ} of the type (3.7). The solution of the FPE is the one transformed from Eq. (2.14).

IV. APPLICATIONS

An important case are those processes whose diffusion is *diagonal*:

$$D^{\mu\nu}(q) = D_{(\mu)}(q)\delta^{\mu\nu}. \tag{4.1}$$

The only elements of curvature tensor that are not zero are $R_{\mu\nu\mu\rho}$ and $R_{\mu\nu\mu\nu}$. As it is easy to check a diffusion matrix of the form

$$D^{\mu\nu}(q) = \varphi_{(\mu)}(q_{\mu})\delta^{\mu\nu} \tag{4.2}$$

with $\varphi_{(\mu)}(q_{\mu}) > 0$ satisfies simultaneously the conditions

$$R_{\mu\nu\mu\rho} = 0, \tag{4.3}$$

$$R_{\mu\nu\mu\nu} = 0. \tag{4.4}$$

The corresponding matrix of the change of gross variables is any solution of the equations

$$\partial_\mu J^\rho_\nu = 0 \quad (\mu \neq \nu), \tag{4.5}$$

$$\partial_\nu J^\rho_\nu = \tfrac{1}{2}\partial_\nu \ln \varphi_{(\nu)} J^\rho_\nu. \tag{4.6}$$

The general solution of (4.5) and (4.6) is

$$J^\lambda_\nu = M^\lambda_\beta (\varphi_{(\beta)}(q_\beta))^{1/2} \delta^\beta{}_\nu, \tag{4.7}$$

where M^λ_β is a nonsingular constant matrix.

With the metric (4.2) the covariant drift in its original variables is

$$h^\mu = f^\mu + \tfrac{1}{4}\varphi^{-1}_{(\mu)} \partial_\mu \varphi_{(\mu)}, \tag{4.8}$$

and the condition that the original drift $f^\mu(q)$ should satisfy in order that the transformed drift $f'^\nu(q')$ be linear is

$$\nabla_\lambda \nabla_\gamma (f^\mu + \tfrac{1}{4}\varphi^{-2}_{(\mu)} \partial_\mu \varphi_{(\mu)}) = 0. \tag{4.9}$$

To the subclass of models represented by Eq. (4.2) belong the Kubo and Gompertz models. Let the function $\varphi_{(\mu)}(q_\mu)$ be of the form

$$\varphi_{(\mu)}(q_\mu) = a^2_{(\mu)} q^2_{(\mu)}; \tag{4.10}$$

if the transformed drift is constant, we have n monodimensional Kubo models.

Similarly, if we have

$$\varphi_{(\mu)}(q_\mu) = a^2_{(\mu)} q^2_{(\mu)} |\log(b_{(\mu)} q_{(\mu)})|^2 \tag{4.11}$$

when the transform drift is constant, we obtain n monodimensional Gompertz models.[2,6]

Another example is the case when the diffusion matrix is *conformally flat*:

$$D^{\mu\nu}(q) = D(q)\delta^{\mu\nu} \tag{4.12}$$

with $D(q) > 0$. In such a case there will exist a change of variables such that $D'^{\mu\nu} = \delta^{\mu\nu}$ if the function $D(q)$ satisfies the equations

$$3D^{-1}\partial_\nu D \partial_\rho D - 2\partial^2_{\rho\nu} D = 0 \quad (\rho \neq \nu), \tag{4.13}$$

$$D^{-1}\left[(\partial_{(\mu)} D)^2 + (\partial_{(\nu)} D)^2 - \tfrac{1}{2}\sum_{\sigma \neq \mu,\nu} (\partial_\sigma D)^2\right] - (\partial^2_{(\mu)(\mu)} D + \partial_{(\nu)(\nu)} D) = 0 \quad (\mu \neq \nu). \tag{4.14}$$

A solution of these equations is

$$D(q) = A \exp[B(q_\mu + q_\nu) + c] \tag{4.15}$$

with an associated transformation matrix

$$J_\sigma{}^\rho = A^{1/2} \exp\{\tfrac{1}{2}B(q_\mu + q_\nu) + c\} M^\rho_\sigma. \tag{4.16}$$

V. CONCLUSIONS

Using the powerful methods of the differential geometry we have found a class of FPE that are exactly solvable by means of change of variables.

Choosing as the metric tensor of the physical space the diffusion matrix, the corresponding FPE will be exactly solvable by means of a change of gross variables if and only if the curvature tensor and the torsion tensor associated with the diffusion is zero, and the transformed drift is linear.

To write the solution of the FPE in the original variables, we have to know the functions $q' = q'(q)$ of the change of variables and substitute them in the solution (2.14). Therefore, with the method presented in this paper the solution of an FPE is reduced to integrating a change of gross variables. Anyway, with the method presented above we can test any FPE to see whether or not it has an exact solution by means of a change of coordinates: The diffusion should satisfy Eqs. (3.1) and (3.2) and the drift equation (3.11).

And so to write the solution of the FPE in the original variables, we should know the functions of such a change of variables $q' = q'(q)$ and enter into the solution (2.14). Therefore, the integration of an FPE equation is reduced with the method presented here to the integration of a change of variables.

ACKNOWLEDGMENTS

One of us (J. M.) would like to thank Dr. Llosa for his help and his useful suggestions.

APPENDIX

We are going to show that the necessary and sufficient conditions which makes that the metric $D^{\mu\nu}(q)$ (diffusion matrix) become Euclidean in some coordinate system are that the associated curvature tensors and the torsion be zero.

Indeed, let us suppose that there exists a coordinate system $q'^\alpha = q'^\alpha(q)$ such that $D'^{\mu\nu}(q') = \delta^{\mu\nu}$; in this case the Christoffel symbols of the metric D' are zero:

$$\Gamma'^\alpha_{\mu\nu}(q') = 0.$$

Since the relation beween the Christoffel symbol for different coordinates system is given by[8]

$$\Gamma^\alpha_{\mu\nu}(q) = \Gamma'^\lambda_{\rho\sigma}(q') \frac{\partial q^\alpha}{\partial q'^\lambda} \frac{\partial q'^\rho}{\partial q^\mu} \frac{\partial q'^\sigma}{\partial q^\nu} + \frac{\partial^2 q'^\lambda}{\partial q^\mu \partial q^\nu} \frac{\partial q^\alpha}{\partial q'^\lambda}, \tag{A1}$$

we will have

$$\frac{\partial q'^\lambda}{\partial q^\alpha} \Gamma^\alpha_{\mu\nu}(q) = \frac{\partial^2 q'^\lambda}{\partial q^\mu \partial q^\nu}. \tag{A2}$$

By means of Eq. (2.7), the relation (A2) may be written as

$$J^\lambda_\alpha \Gamma^\alpha_{\mu\nu}(q) = \partial_\mu J^\lambda_\nu. \tag{A3}$$

If there exists a system of coordinates $\{q'^\nu\}$ such that $D'^{\mu\nu}(q') = \delta^{\mu\nu}$, or, equivalently, such that $\Gamma'^\alpha_{\mu\nu}(q') = 0$, there will exist a transformation matrix solution of Eq. (A3). Therefore, (A3) should satisfy the integrability equations

$$\partial^2_{\rho\mu} J^\lambda_\mu - \partial^2_{\mu\rho} J^\lambda_\nu = 0. \tag{A4}$$

Deriving (A3) and entering in (A4), we conclude that J exists if

$$J^\lambda_\alpha [\partial_\rho \Gamma^\alpha_{\mu\nu}(q) - \partial_\mu \Gamma^\alpha_{\rho\nu}(q)] + J^\lambda_\beta [\Gamma^\beta_{\rho\alpha}(q)\Gamma^\alpha_{\mu\nu}(q) - \Gamma^\beta_{\mu\nu}(q)\Gamma^\alpha_{\rho\nu}(q)] = 0;$$

this condition is satisfied if

$$R^\beta{}_{\mu\nu\rho} \equiv \partial_\rho \Gamma^\beta_{\mu\nu} - \partial_\mu \Gamma^\beta_{\rho\nu} + \Gamma^\beta_{\rho\alpha}\Gamma^\alpha_{\mu\nu} - \Gamma^\beta_{\mu\alpha}\Gamma^\alpha_{\rho\nu} = 0, \tag{A5}$$

where $R^\beta{}_{\mu\nu\rho}$ is the curvature tensor of the manifold.[8] And,

since

$$R_{\alpha\mu\nu\rho} \equiv D_{\alpha\beta} R^{\beta}{}_{\mu\nu\rho} = 0, \tag{A6}$$

the matrix $J_{\alpha}{}^{\lambda}$ exists if *all components* of the curvature tensor are zero. But we could ask ourselves whether or not this matrix corresponds to a continuous and differentiable change of gross variables, i.e., if we can write $J_{\alpha}{}^{\lambda} = \partial_{\alpha} q'^{\lambda}$. The well-known condition

$$\partial_{\mu} J^{\lambda}_{\alpha} = \partial_{\alpha} J^{\lambda}_{\mu} \tag{A7}$$

should then be satisfied, which is equivalent to

$$J_{\beta}{}^{\lambda} (\Gamma^{\beta}_{\mu\alpha} - \Gamma^{\beta}_{\alpha\mu}) = 0, \tag{A8}$$

as can be seen making use of (A3). This implies that the torsion tensor is zero

$$T_{\mu\alpha}{}^{\beta} \equiv \Gamma^{\beta}_{\mu\alpha} - \Gamma^{\beta}_{\alpha\mu} = 0. \tag{A9}$$

[1] H. Haken, Rev. Mod. Phys. **47**, 67 (1975).
[2] M. San Miguel, Z. Phys. B **33**, 307 (1979).
[3] R. Graham, Z. Phys. B **26**, 397 (1977).
[4] L. Garrido, Physica A **100**, 140 (1980).
[5] W. Horsthemke and A. Bach, Z. Phys. B **22**, 189 (1975).
[6] H. Haken, Z. Phys. B **24**, 321 (1976).
[7] L. Garrido and J. Llosa, "The intrinsic Fokker–Planck equation," in *Systems Far from Equilibrium*, edited by L. Garrido (Springer-Verlag, New York, 1980).
[8] L. D. Landau and E. M. Lifshitz, *Teoría clasica de campos* (Reverté, Barcelona, 1966); S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972); N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, London, 1974).

On boson condensation into an infinite number of low-lying levels

M. van den Berg
Institute for Theoretical Physics, State University of Groningen, P. O. Box 800, Groningen, The Netherlands

(Received 20 November 1981; accepted for publication 25 January 1982)

The complicated structure of the condensate of a free boson gas in two dimensions with mixed boundary conditions is examined.

PACS numbers: 05.30.Jp

I. INTRODUCTION AND THEOREM

It has been shown in recent papers[1,2] that there exist three types of condensation in the free boson gas:

(I) Macroscopic occupation of a finite number of single-particle levels.

(II) Macroscopic occupation of an infinite number of single-particle levels.

(III) Nonextensive condensation (no levels macroscopically occupied).

In this paper we will present an example of type II. Let B_L be a finite region in R^d with volume L^d and let $E_1 < E_2^L \leqslant E_3^L \cdots$ be the spectrum of the single-particle Hamiltonian H_L; then [if we assume that $\exp(-\beta H_L)$ is trace-class] the mean number $\langle n_k \rangle_L$ of particles per volume in the k th state for a free boson gas is given by (in the grand canonical ensemble[3])

$$\langle n_k \rangle_L = \frac{1}{L^d} \frac{\zeta(L)}{e^{\eta_k^L} - \zeta(L)}, \tag{1}$$

where $\eta_k^L = E_k^L - E_1^L$ and $\zeta(L)$ is the positive solution of

$$\sum_k \langle n_k \rangle_L = \rho; \tag{2}$$

ρ is the mean number of particles per unit volume. The thermodynamic limit is the limit in which $L \to \infty$ and the number density ρ is kept fixed. If the spectrum of H_L is such that the critical density ρ_c, defined by

$$\rho_c = \lim_{\zeta \uparrow 1} \lim_{L \to \infty} \frac{1}{L^d} \sum_k (e^{\eta_k^L} - \zeta)^{-1}, \tag{3}$$

is finite, then condensation of type I, II, or III takes place if $\rho > \rho_c$. In the case when $H_L = -\Delta/2$ (with Dirichlet or Neuman conditions on the boundary ∂B_L of B_L) the critical density ρ_c is finite only for $d > 2$. However, if we impose attractive boundary on the eigenfunctions, then ρ_c is finite for all dimensions $d = 1,2,\cdots$. This was noticed by Robinson[4] and discussed in detail by Landau and Wilde.[5] The reason that ρ_c is finite in this case is that the attractive boundary conditions cause a gap in the single-particle spectrum:

$$E_j^L - E_1^L > g, \quad \text{for some } j > 1 \text{ and all } L > L_0. \tag{4}$$

It was noticed before by Bijl, De Boer, and Michels[6] that a gap in the spectrum will change the thermodynamical properties.

In this paper we will take $H_L = -\Delta/2$ and impose position-dependent boundary conditions. We will see that they lead to condensation of type II for $d = 2$ and of type I for $d = 1$, $d = 3, \cdots$. If the boundary conditions are position-independent, then there is condensation for $\rho > \rho_c$ of type I only (as in Ref. 4). Let us take for B_L the cuboid

$$B_L = \{x \in R^d : 0 < x_1 < L, \ldots, 0 < x_d < L\} \tag{5}$$

and let the boundary conditions be

$$\frac{\partial \phi}{\partial x_1} = -\sigma \phi \quad \text{for } x_1 = L, 0 < x_2 < L, \ldots, 0 < x_d < L,$$
$$\phi = 0 \quad \text{for any other point of the boundary.} \tag{6}$$

The boundary conditions are said to be attractive if $\sigma < 0$. We have expressed all lengths in units $(\beta\hbar^2/m)^{1/2}$, where β is the inverse of the temperature times Boltzmann's constant. Let $\{\Phi_k^L\}$ be the eigenfunction of the Laplacian in B_L with boundary conditions (6) so that

$$-\tfrac{1}{2}\Delta \Phi_k^L = E_k^L \Phi_k^L. \tag{7}$$

Then

$$\Phi_k^L = \left(\frac{2}{L}\right)^{d/2} \left(1 - \frac{\sin 2L(2\epsilon_{k_1})^{1/2}}{2L(2\epsilon_{k_1})^{1/2}}\right)^{-1/2} \times \sin x_1 (2\epsilon_{k_1})^{1/2} \prod_{i=2}^{d} \sin \frac{\pi k_i x_i}{L},$$
$$E_k^L = \epsilon_{k_1} + \sum_{i=2}^{d} \frac{\pi^2 k_i^2}{2L^2}$$
$$k = (k_1, \ldots, k_d),$$
$$k_i = 1,2,3,\cdots \quad (i = 1, \ldots, d), \tag{8}$$

and $\{\epsilon_{k_1}\}$ are the positive solutions for ϵ of

$$\tan L(2\epsilon)^{1/2} = -(2\epsilon)^{1/2}/\sigma, \tag{9}$$

ordered so that $\epsilon_1 < \epsilon_2 < \epsilon_3 \cdots$. For $\sigma < -1/L$ and $L \gg 1$, the spectrum is approximately given by

$$\epsilon_{k_1} \sim (\pi^2/2L^2)(k_1 - 1)^2, \quad k_1 = 2,3,\cdots, \tag{10}$$

while there is a bound state with energy

$$\epsilon_1 \sim -\sigma^2/2. \tag{11}$$

Our main result is the following:

Theorem: For $\rho < \rho_c$, $\zeta(L)$ tends to the positive solution of

$$\rho = \sum_{n=1}^{\infty} \frac{\zeta^n}{(2\pi n)^{d/2}} e^{-n\sigma^2/2}. \tag{12}$$

For $\rho > \rho_c$, $\zeta(L)$ tends to 1 as follows:

$$\zeta(L) \sim 1 - 1/AL^2, \quad d = 2,$$
$$\zeta(L) \sim 1 - 1/(\rho - \rho_c)L^d, \quad d = 1,3,\cdots,$$
$$\rho_c = \sum_{n=1}^{\infty} \frac{e^{-n\sigma^2/2}}{(2\pi n)^{d/2}}, \tag{13}$$

 0022-2488/82/061159-03$02.50

and A is the unique solution of

$$\sum_{k_2=1}^{\infty}\left(\frac{\pi^2}{2}(k_2^2-1)+\frac{1}{A}\right)^{-1}=\rho-\rho_c. \tag{14}$$

We see that due to (13) and (1)

$$\lim_{L\to\infty}\langle n_k\rangle_L=\begin{cases}[(\pi^2/2)(k_2^2-1)+1/A\,]^{-1}\\ 0,\end{cases}$$

$$\left.\begin{aligned}k_1=1,\ k_2=1,2,3,\cdots\\ k_1=2,3,\cdots,\ k_2=1,2,\cdots\end{aligned}\right\}d=2, \tag{15}$$

$$\lim_{L\to\infty}\langle n_k\rangle_L=\left.\begin{cases}\rho-\rho_c, & k=(1,\ldots,1)\\ 0, & k\neq(1,\ldots,1)\end{cases}\right\}d\neq 2.$$

It follows that in two dimensions there is macroscopic occupation of an infinite set of low-lying single-particle states (condensation of type II) whereas in 1,3,4,··· dimensions only the ground state is macroscopically occupied (type I).

This can also be seen by looking at the scaled spatial particle density ν_L defined by

$$\nu_L(u)=\sum_k\frac{\zeta(L)}{e^{\eta_k^L}-\zeta(L)}\,[\Phi_k^L(Lu)]^2. \tag{16}$$

One finds that

$$\lim_{L\to\infty}\nu_L(u)=\begin{cases}(\rho-\rho_c)\delta(1-u_1)+\rho_c, & d=1,\ \rho>\rho_c,\\ \delta(1-u_1)\displaystyle\sum_{j=1}^{\infty}\frac{2\sin^2\pi u_2}{(j^2-1)\pi^2/2+1/A}+\rho_c, & d=2,\ \rho>\rho_c,\\ (\rho-\rho_c)\delta(1-u_1)\displaystyle\prod_{m=2}^{d}2\sin^2\pi u_m+\rho_c, & d\geqslant 3,\ \rho>\rho_c,\\ \rho, & d\geqslant 1,\rho<\rho_c,\end{cases} \tag{17}$$

where $u\in B_1$ and $\delta(1-u_1)$ is the Dirac delta function supported on the hyperplane $u_1=1$. We see that the expressions [apart from (12) and (13)] are identical to the corresponding expressions for the free boson gas in the presence of an external field of power form in one direction (see Ref. 1). This is for the following reason: The attractive boundary condition causes a gap in the spectrum and forces the wavefunction to have a maximum near the attractive boundary. The same is true for one-particle Hamiltonian with an external potential with an absolute minimum at the boundary (and Dirichlet boundary conditions). We will give a sketch of the proof using the ideas of Lewis and Pulè.[7]

II. SKETCH OF THE PROOF

Let $f_L(z)$ be defined by

$$f_L(z)=\frac{1}{L^d}\sum_{n=1}^{\infty}z^n\sum_{\{k:k_1=1\}}e^{-n\eta_k^L}. \tag{18}$$

The first step is to show that for $z\in[0,1]$ we have

$$\lim_{L\to\infty}f_L(z)=\sum_{n=1}^{\infty}\frac{z^n}{(2\pi n)^{d/2}}e^{-n\sigma^2/2}. \tag{19}$$

This can be done using the asymptotic behavior of the E_k^L for large L given by (8),...,(11). Since it follows from (1) and (2) that for each $L>0$ and $\rho>0$ there is always a unique $\zeta(L)\in[0,1)$ we find that if $\zeta(L)$ tends to $\zeta\in[0,1)$ for $L\to\infty$, then

$$\rho=\sum_{n=1}^{\infty}\frac{\zeta^n}{(2\pi n)^{d/2}}e^{-n\sigma^2/2}+\lim_{L\to\infty}\frac{1}{L^d}\sum_{n=1}^{\infty}\zeta^n\sum_{\{k:k_1=1\}}e^{-n\eta_k^L}. \tag{20}$$

Since

$$\lim_{L\to\infty}\frac{1}{L^d}\sum_{n=1}^{\infty}\zeta^n\left[\sum_{k=1}^{\infty}\exp\frac{-n\pi^2}{2L^2}(k^2-1)\right]^{d-1}\leqslant\lim_{L\to\infty}\frac{1}{L^d}\sum_{n=1}^{\infty}\zeta^n\left(1+\frac{L}{(2\pi n)^{1/2}}\right)^{d-1}=0,$$

We arrive at Eq. (12). Suppose now that $\zeta(L)\to 1$; then

$$\frac{1}{L^d}\sum_{n=1}^{\infty}[\zeta(L)]^n\sum_{\{k:k_1=1\}}e^{-n\eta_k^L}\to\rho-\rho_c. \tag{21}$$

Equation (21) implies that for $d=1$ we have macroscopic occupation of the ground state. So for $d=2$ we have

$$\frac{1}{L^d}\sum_{k=1}^{\infty}\frac{\zeta(L)}{e^{\pi^2(k^2-1)/2}-\zeta(L)}\to\rho-\rho_c, \tag{22}$$

from which it follows that $\zeta(L)\sim 1-1/AL^2$, where A is the solution of (14). For $d\geqslant 3$ we have the following estimate:

$$\begin{aligned}&\frac{1}{L^d}\sum_{\{k:k_1=1\}}\frac{\zeta(L)}{e^{\eta_k^L}-\zeta(L)}-\frac{\zeta(L)}{1-\zeta(L)}\\ &=\frac{1}{L^d}\sum_{\left\{\begin{smallmatrix}k:k_1=1\\ k\neq(1,\ldots,1)\end{smallmatrix}\right\}}\frac{\zeta(L)}{e^{\eta_k^L}-\zeta(L)}\\ &\leqslant\frac{1}{L^d}\sum_{\left\{\begin{smallmatrix}k:k_1=1\\ k\neq(1,\ldots,1)\end{smallmatrix}\right\}}\frac{1}{e^{\eta_k^L}-1}\\ &\leqslant\frac{d}{L^d}\sum_{n=1}^{\infty}\sum_{k=2}^{\infty}\exp\left(-\frac{n\pi^2}{2L^2}(k^2-1)\right)\times\left[\sum_{k=1}^{\infty}\exp\left(-\frac{n\pi^2}{2L^2}(k^2-1)\right)\right]^{d-2}\\ &\leqslant\frac{d}{L^d}\sum_{n=1}^{\infty}e^{-n\pi^2/L^2}\frac{L}{(2\pi n)^{1/2}}\left(1+\frac{L}{(2\pi n)^{1/2}}\right)^{d-2}\\ &\leqslant\frac{d\,2^d}{L^d}\sum_{n=1}^{\infty}e^{-n\pi^2/L^2}\frac{L}{(2\pi n)^{1/2}}\left(1+\frac{L^{d-2}}{(2\pi n)^{(d-2)/2}}\right),\end{aligned} \tag{23}$$

which goes to zero as $L\to\infty$. So for $\rho>\rho_c$

$$\frac{1}{L^d}\frac{\zeta(L)}{1-\zeta(L)}\to\rho-\rho_c, \tag{24}$$

and we have macroscopic occupation of the ground state only (condensation of type I).

ACKNOWLEDGMENTS

The author wishes to thank Professor Dr. J. T. Lewis for his helpful discussions. The work in this paper is part of the research program of the Foundation of Fundamental Research on Matter, which is financially supported by the Netherlands Organization for Pure Research.

[1]M. van den Berg and J. T. Lewis, "On the free boson gas in a weak external potential," Commun. Math. Phys. **81**, 475 (1981).
[2]M. van den Berg and J. T. Lewis, "On generalized condensation in the free boson gas," Physica A (to be published).
[3]R. M. Ziff, G. M. Uhlenbeck, and M. Kac, Phys. Lett. **32**, 171 (1977).
[4]D. W. Robinson, Commun. Math. Phys. **50**, 53 (1976).
[5]L. J. Landau and I. F. Wilde, Commun. Math. Phys. **70**, 43 (1979).
[6]A. Bijl, J. de Boer, and A. Michels, Physica **8**, 655 (1941).
[7]J. T. Lewis and J. V. Pulè, Commun. Math. Phys. **36**, 1 (1974).

Higher spin one-dimensional Ising lattice in arbitrary external field

J. K. Percus [a)]
Departamento de Matematica, Pontificia Universidade Catolica do Rio de Janeiro, Brazil and Courant Institute of Mathematical Sciences and Physics Department, New York University, [b)] New York, New York 10012

(Received 4 December 1981; accepted for publication 25 January 1982)

We consider an integer lattice in one dimension whose site variables take on the values $\nu = 0,1,...,D$ with a fixed nearest neighbor interaction but an arbitrary site-dependent external potential. By first eliminating the external potential in favor of the site probability density, an expression is found in principle for the potential as a functional of the density. This relation is worked out in detail for basic spin $\frac{1}{2}$ model, Z_3 lattice, random walk ensemble, and a special continuous spin model. The direct correlation function in all cases has only nearest neighbor support, and the thermodynamic potential as a functional of the density couples only nearest neighbor sites.

PACS numbers: 05.50. + q

1. INTRODUCTION

One-dimensional systems, despite their restricted physical applicability, nonetheless illustrate a number of the phenomena associated with the three-dimensional world. They have the great advantage of being explicitly solvable in a number of different contexts, allowing them to serve as benchmarks against which to assess approximations that are required in the more physical domains. And of course, in some situations, such as flow in very narrow channels, they may be very good models of reality. Perhaps the simplest context worth investigating is that of classical thermal equilibrium, where such models have been used both to check quite general physical assertions—such as the signature of lattice vertex functions[1]—as well as to generate effective approximate models of real fluids.[2] The key to using such correspondences is that of having maximum control over the models, which has generally turned out to imply solving them in the presence of an arbitrary external field, such as that which mimics inertial forces in dynamical considerations.

In this paper, we shall continue previous investigations in this vein, and solve the equilibrium statistical mechanics of an Ising lattice with arbitrary external field, in which the "spin" at a given site is not a dichotomic variable but can assume $D + 1$ values, say $\nu = 0,1,...,D$. We shall consider only the case of position-independent nearest neighbor interaction $\phi(\nu_x,\nu_{x-1})$ between integer sites x and $x - 1$; these functions, assumed symmetric, occupy a linear space of dimension $\frac{1}{2}(D + 1)(D + 2)$, $D + 1$ dimensions of which can be incorporated into an external potential. The applied external potential $w_x(\nu_x)$ at site x is a member of a linear space of dimension $D + 1$, one dimension of which sets the local reference potential and does not, e.g., affect distribution functions. Indeed, the primary information we seek will be the dependence of the single-site distribution $n_x(\nu)$—the probability that $\nu_x = \nu$—on the applied field $\{w_x(\nu)\}$. All distributions can, of course, be generated by differentiation of this relation. We will find, as in the previously solved spin $\frac{1}{2}$ case, that whereas the problem expressed in this form does not lead to a simple closed form solution, the inverse problem—the dependence of $w_x(\nu)$ on $\{n_x(\nu)\}$—does. We will solve this problem in several special cases, including infinite spin and continuous spin limits, and discuss some consequences of the solutions.

2. BASIC SOLUTIONS

Our system is represented by the partition function

$$\Xi = \sum_{\{\nu_x\}} e^{-\beta \Sigma_x w_x(\nu_x)} e^{-\beta \Sigma_x \phi(\nu_x,\nu_{x-1})}. \tag{2.1}$$

In order to take the domain of the integer index x as $-\infty$ to ∞ we may, for example, distinguish one state $\nu = 0$ (say, the empty state) and assume that $\exp[-\beta w_x(\nu)] \to \delta_{\nu,0}$ as $|x| \to \infty$. We then take advantage of the one-dimensional nearest neighbor character of the system by considering instead

$$\Xi_L(\nu) \equiv \sum_{\{\nu_x\}} \exp\left\{-\beta \sum_{-\infty}^{L-1} w_x(\nu_x)\right\} \exp\{-\beta w_L(\nu)\} \times \exp\left\{-\beta \sum_{-\infty}^{L-1} \phi(\nu_x,\nu_{x-1})\right\} \exp\{-\beta\phi(\nu,\nu_{L-1})\}, \tag{2.2}$$

$$\hat{\Xi}_L(\nu) \equiv \sum_{\{\nu_x\}} \exp\left\{-\beta \sum_{L+1}^{\infty} w_x(\nu_x)\right\} \exp\{-\beta w_L(\nu)\} \times \exp\left\{-\beta \sum_{L+2}^{\infty} \phi(\nu_x,\nu_{x-1})\right\} \exp\{-\beta\phi(\nu_{L+1},\nu)\}, \tag{2.3}$$

the left and right truncations of Ξ at site L, spin ν. It is clear that the probability of spin ν at site L is given by

$$n_L(\nu) \equiv \Pr(\nu_L = \nu) = \exp\{\beta w_L(\nu)\} \Xi_L(\nu) \hat{\Xi}_L(\nu)/\Xi. \tag{2.4}$$

It is also clear that the recursion relations

$$\Xi_L(\nu) = \exp\{-\beta w_L(\nu)\} \sum_{\nu'} \exp\{-\beta\phi(\nu,\nu')\} \Xi_{L-1}(\nu'),$$

[a)]Supported in part by NSF Grant CHE-80011285, DOE Contract DE-AC02-76ER03077, and CNPq 40.0438/79 (Brazil).
[b)]Permanent address.

 0022-2488/82/061162-06$02.50

$$\hat{\Xi}_L(\nu) = \exp\{-\beta w_L(\nu)\} \sum_{\nu'} \exp\{-\beta\phi(\nu',\nu)\}\Xi_{L+1}(\nu') \tag{2.5}$$

are satisfied.

For notational convenience, let us define the vector and matrices

$$W_x(\nu) \equiv \exp\{-\beta w_x(\nu)\}, \quad e(\nu,\nu') \equiv \exp\{-\beta\phi(\nu,\nu')\}, \tag{2.6}$$

$$W_x(\nu,\nu') \equiv W_x(\nu)\delta_{\nu,\nu'},$$

so that (2.5) and (2.4) become

$$\Xi_L = W_L e \Xi_{L-1}, \quad \hat{\Xi}_L = W_L e \hat{\Xi}_{L+1}, \tag{2.7}$$

$$W_L(\nu) = \Xi_L(\nu)\hat{\Xi}_L(\nu)/n_L(\nu)\Xi. \tag{2.8}$$

Our strategy now calls for the elimination of $W_L(\nu)$ between (2.8) and each of (2.7), resulting in

$$\hat{\Xi}_L e \Xi_{L-1} = n_L \Xi, \quad \Xi_L e \hat{\Xi}_{L+1} = n_L \Xi, \tag{2.9}$$

which may also be written as

$$\Xi_L/\Xi = e^{-1}(n_{L+1}/\hat{\Xi}_{L+1}), \quad \hat{\Xi}_L/\Xi = e^{-1}(n_{L-1}/\Xi_{L-1}). \tag{2.10}$$

Thus Ξ and $\hat{\Xi}$ are determined in principle as functionals of $\{n_x\}$.

There are at least three convenient forms in which (2.9) and (2.10) can be recast. First, we eliminate Ξ_L from the second of (2.9) and the first of (2.10) to read (after $L \to L-1$)

$$\sum_{\nu',\nu''} e(\nu,\nu')e^{-1}(\nu,\nu'')n_L(\nu'')\{\hat{\Xi}_L(\nu')/\hat{\Xi}_L(\nu'')\} = n_{L-1}(\nu). \tag{2.11}$$

These are $D+1$ nonlinear relations for any D independent ratios of the form $\hat{\Xi}_L(\nu')/\hat{\Xi}_L(\nu'')$, but one relation is superfluous since summing (2.11) over yields the identity $1 = \Sigma_\nu n_{L-1}(\nu)$. In the same fashion, the first of (2.9) and second of (2.10) combine to

$$\sum_{\nu',\nu''} e(\nu,\nu')e^{-1}(\nu,\nu'')n_L(\nu'')\{\Xi_L(\nu')/\Xi_L(\nu'')\} = n_{L+1}(\nu), \tag{2.12}$$

to be solved for the $\Xi_L(\nu')/\Xi_L(\nu'')$. The solution to the inverse problem is then given, following (2.8), by

$$\frac{W_L(\nu')}{W_L(\nu'')} = \frac{\Xi_L(\nu')}{\Xi_L(\nu'')}\frac{\hat{\Xi}_L(\nu')}{\hat{\Xi}_L(\nu'')}\frac{n_L(\nu'')}{n_L(\nu')}. \tag{2.13}$$

If a local reference—say $W_L(0)$—is specified, the $W_L(\nu)$ are thereby determined as functionals of $\{n_x(\nu)\}$.

A second formulation is obtained by eliminating $\hat{\Xi}$ between the two equations of (2.9): $n_L \Xi = \Xi_L e \hat{\Xi}_{L+1}$ $= \Xi_L e(n_{L+1}\Xi/e\Xi_L)$, and similarly eliminating Ξ. Thus, appending (2.8), we have

$$n_L(\nu) = \Xi_L(\nu)e\left(\frac{n_{L+1}(\nu)}{e\Xi_L(\nu)}\right), \quad n_L(\nu) = \hat{\Xi}_L(\nu)e\left(\frac{n_{L-1}(\nu)}{e\hat{\Xi}_L(\nu)}\right), \tag{2.14a}$$

$$W_L(\nu) = \frac{1}{\Xi}\frac{\Xi_L(\nu)\hat{\Xi}_L(\nu)}{n_L(\nu)}. \tag{2.14b}$$

Here, at given L, $\Xi_L(\nu)$ and $\hat{\Xi}_L(\nu)$ are each determined only to within a multiplicative constant, and Ξ is likewise undetermined, but the required ratio in (2.14b) is again fixed by inspection of a local reference such as $W_L(0)$.

If e^{-1}, rather than e, has an especially simple form, the same procedure can be applied to the two equations of (2.10). For this purpose, let us first define

$$Y_L(\nu) \equiv n_L(\nu)/\Xi_L(\nu), \quad \hat{Y}_L(\nu) \equiv n_L(\nu)/\hat{\Xi}_L(\nu), \tag{2.15}$$

so that (2.10) becomes

$$n_L(\nu)/\Xi = Y_L(\nu)e^{-1}\hat{Y}_{L+1}(\nu), \; n_L(\nu)/\Xi = \hat{Y}_L(\nu)e^{-1}Y_{L-1}(\nu). \tag{2.16}$$

The analog of Eqs. (2.14) is seen to be

$$n_L(\nu) = Y_L(\nu)e^{-1}\left(\frac{n_{L+1}(\nu)}{e^{-1}Y_L(\nu)}\right),$$

$$n_L(\nu) = \hat{Y}_L(\nu)e^{-1}\left(\frac{n_{L-1}(\nu)}{e^{-1}\hat{Y}_L(\nu)}\right), \tag{2.17a}$$

$$W_L(\nu) = \frac{1}{\Xi}\frac{n_L(\nu)}{Y_L(\nu)\hat{Y}_L(\nu)}, \tag{2.17b}$$

and similar comments as to normalization are appropriate: the unknown constant in $\Xi Y_L(\nu)\hat{Y}_L(\nu)$ is fixed by imposing a local reference for $W_L(\nu)$.

3. THE SPIN ½ ISING MODEL

To make contact with previous work, let us consider the case $D = 1$, i.e., dichotomic spin variable, but use the lattice gas notation $\nu = 0,1$. Without loss of generality we can choose $\phi(\nu,\nu') = J\nu\nu'$, $w_x(\nu) = \nu u_x$ [so that $W_L(0) = 1$], and then

$$e = \begin{pmatrix} 1 & 1 \\ 1 & e^{-\beta J} \end{pmatrix}, \quad e^{-1} = \begin{pmatrix} e^{-\beta J} & -1 \\ -1 & 1 \end{pmatrix} / (e^{-\beta J} - 1),$$

$$W_x = \begin{pmatrix} 1 & \\ & e^{-\beta u_x} \end{pmatrix}. \tag{3.1}$$

The formulation (2.11)–(2.13) is convenient. We first define $F(\nu',\nu'';n,n')$ as that solution, in the form

$$F(\nu',\nu'',n,n') = G(\nu';n,n')/G(\nu'';n,n'), \tag{3.2}$$

of

$$\sum_{\nu,\nu''} e(\nu,\nu')e^{-1}(\nu,\nu'')n(\nu'')F(\nu',\nu'';n,n') = n'(\nu). \tag{3.3}$$

Then (2.13) reads

$$\frac{W_L(\nu')}{W_L(\nu'')} = F(\nu',\nu'';n_L,n_{L-1}) \times F(\nu',\nu'';n_L,n_{L+1})\frac{n_L(\nu'')}{n_L(\nu')} \tag{3.4}$$

or, for that matter,

$$W_L(\nu) = \alpha_L G(\nu;n_L,n_{L-1})G(\nu;n_L,n_{L+1})/n_L(\nu) \tag{3.5}$$

with undetermined α_L.

In the present case, we need only the single ratio $F(n,n') = F(1,0;n,n')$ and the single equation corresponding to $\nu = 0$. According to (3.3), then,

$$e(0,0)e^{-1}(0,0)n(0) + e(0,0)e^{-1}(0,1)n(1)/F(n,n') + e(0,1)e^{-1}(0,0)n(0)F(n,n') + e(0,1)e^{-1}(0,1)n(1) = n'(0) \tag{3.6}$$

or, in terms of (3.1),

$$2F(n,n') = -\left(1 - \frac{f}{e}\frac{n'(0)}{n(0)} - \frac{1}{e}\frac{n(1)}{n(0)}\right) + \left[\left(1 - \frac{f}{e}\frac{n'(0)}{n(0)} - \frac{1}{e}\frac{n(1)}{n(0)}\right)^2 + \frac{4}{e}\frac{n(1)}{n(0)}\right]^{1/2}, \tag{3.7}$$

where $e \equiv e^{-\beta J}$, $f \equiv e^{-\beta J} - 1$, the sign of the square root being set by the positivity of F. Equation (3.4) now becomes

$$e^{-\beta u_x} = F(n_x, n_{x-1})F(n_x, n_{x+1})\frac{n_x(1)}{n_x(0)}, \tag{3.8}$$

and the pair (3.7) and (3.8), with $n_x(1) = \rho(x) = 1 - n_x(0)$, is identical with the solution given in Ref. 1.

4. A HIGHER SPIN MODEL

The technical job of solving Eqs. (3.2) and (3.3) depends very much on the nature of the interaction given by $e(\nu,\nu')$. Even the next simplest case of spin 1 can lead to substantial complications. Let us consider one model which is quite readily solved, the spin 1 or Z_3 lattice, with $\nu = 0,1,2$. Here ν is associated with the two-dimensional vector $a(\nu) = \exp\{2\pi i\nu/3\}$, and the interaction is given by $\phi(\nu,\nu') = -Ja(\nu)\cdot a(\nu')$, i.e.,

$$(\phi(\nu,\nu')) = -J\begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix}. \tag{4.1}$$

Hence

$$(e(\nu,\nu')) = \begin{pmatrix} e & 1/e & 1/e \\ 1/e & e^2 & 1/e \\ 1/e & 1/e & e^2 \end{pmatrix}, \quad e \equiv e^{(1/2)\beta J}, \tag{4.2}$$

and

$$(e^{-1}(\nu,\nu')) = \begin{pmatrix} e^3+1 & -1 & -1 \\ -1 & e^3+1 & -1 \\ -1 & -1 & e^3+1 \end{pmatrix}\frac{e}{(e^3-1)(e^3+2)}. \tag{4.3}$$

This in fact is a member of the general class of interactions

$$e(\nu,\nu') = a\delta_{\nu,\nu'} - b,$$
$$e^{-1}(\nu,\nu') = \frac{1}{a}\left(\delta_{\nu,\nu'} + \frac{b}{a-(D+1)b}\right) \tag{4.4}$$

for arbitrary dimension $D+1$. The whole class is readily solved because its limitation to two distinct eigenvalues renders it only slightly more difficult than the spin $\frac{1}{2}$ prototype.

We assume the "ferromagnetic" case $a > (D+1)b$. For the class (4.4), the pair (3.2), (3.3) becomes simply

$$n_0(\nu) - \frac{b}{a}\frac{n_0(\nu)}{G(\nu)}\sum G(\nu') + \frac{b}{a-(D+1)b}G(\nu)\sum\frac{n_0(\nu')}{G(\nu')} - \frac{b^2}{a(a-(D+1)b)}\sum G(\nu')\sum\frac{n_0(\nu'')}{G(\nu'')} = n'(\nu),$$

so that

$$\frac{G(\nu)}{\Sigma G(\nu')} = \frac{-(n_0(\nu) - n'(\nu) - (b/a)H) + [(n_0(\nu) - n'(\nu) - (b/a)H)^2 + 4(b/a)n_0(\nu)H]^{1/2}}{2H}, \tag{4.5}$$

where

$$H = \frac{b}{a-(D+1)b}\sum G(\nu')\sum\frac{n_0(\nu'')}{G(\nu'')},$$

the sign of the square root occasioned again by the positivity of $G(\nu)$. Summing (4.5) over ν and using the fact that $\Sigma n_0(\nu) = \Sigma n'(\nu) = 1$, we have the algebraic equation for H,

$$\frac{2a-(D+1)b}{a}H = \sum_\nu\left[\left(n_0(\nu) - n'(\nu) - \frac{b}{a}H\right)^2 + 4\frac{b}{a}n_0(\nu)H\right]^{1/2}. \tag{4.6}$$

In other words, we conclude from (3.5) on redefining α_L and setting $K = (b/a)H$, that

$$e^{-\beta w_L(\nu)} = \alpha_L G(\nu;n_L,n_{L+1})G(\nu;n_L,n_{L-1})/n_L(\nu),$$

where

$$G(\nu;n,n') = -[n(\nu) - n'(\nu) - K(n,n')] + [(n(\nu) - n'(\nu) - K(n,n'))^2 + 4n(\nu)K(n,n')]^{1/2} \tag{4.7}$$

and

$$\frac{2a-(D+1)b}{b}K(n,n') = \sum_\nu[(n(\nu) - n'(\nu) - K(n,n'))^2 + 4n(\nu)K(n,n')]^{1/2}.$$

It goes without saying that higher spin can also be used to simulate more complicated spin $\frac{1}{2}$ Ising models, such as those with non-nearest neighbor interaction, or with a finite number of interacting rows. In these cases, it is convenient to employ a nonsymmetric $e(\nu,\nu')$, a generalization that is trivially made, starting with (2.7) modified to read

$$\Xi_L = W_L e\Xi_{L-1}, \quad \hat{\Xi}_L = W_L e^T\hat{\Xi}_{L+1}. \tag{4.8}$$

5. INFINITE SPIN MODELS

The state specifying parameter ν can denote spatial location, in which case the one-dimensional order can loosely be regarded as temporal. This interpretation must be used with caution, however, since the weight of a chain chosen by the partition function depends upon the full history and need not correspond to a reasonable dynamical process. Indeed, the most direct realization is in terms of an ensemble of polymers, sequences of links between lattice points. Let us consider the simplest model of this type, in which ν denotes

location on a one-dimensional integer lattice. The elementary interaction will be a step to the left or to the right, so that

$$e(\nu,\nu') = \begin{cases} 1, & \nu' = \nu \pm 1 \\ 0, & \nu' \neq \nu \pm 1 \end{cases} . \tag{5.1}$$

The formulation (2.14) is now appropriate: if $n_L \to n$, $n_{L+1} \to n'$, Ξ_L, $\hat{\Xi}_L \to z$, then we have

$$n(\nu) = z(\nu)\frac{n'(\nu+1)}{z(\nu)+z(\nu+2)} + z(\nu)\frac{n'(\nu-1)}{z(\nu)+z(\nu-2)} . \tag{5.2}$$

Rewriting (5.2) as $n(\nu) - n'(\nu+1) = n'(\nu-1)z(\nu)/[z(\nu)+(z-2)] - n'(\nu+1)z(\nu+2)/[z(\nu+2)-z(\nu)]$, it is solved at once as $n'(\nu+1)z(\nu+2)/[z(\nu+2)-z(\nu)]$ $= \Sigma_0^\infty [n'(\nu+1-2p) - n(\nu-2p)]$, or

$$z(\nu+2)/z(\nu) = \sum_0^\infty [n'(\nu+1-2p) - n(\nu-2p)] / \sum_0^\infty [n'(\nu-1-2p) - n(\nu-2p)] . \tag{5.3}$$

Hence

$$z(\nu) = \alpha \prod_1^\infty \sum_0^\infty [n'(\nu+1-2p-2q) - n(\nu-2p-2q)] / \sum_0^\infty [n'(\nu-1-2p) - n(\nu-2p-2q)] , \tag{5.4}$$

on substituting in (2.14),

$$n_L(\nu)e^{-\beta w_L(\nu)} = \alpha_L \prod_{q=1}^\infty \frac{\Sigma_{p=0}^\infty [n_{L+1}(\nu+1-2p-2q) - n_L(\nu-2p-2q)]\Sigma_{p=0}^\infty [n_{L-1}(\nu+1-2p-2q) - n_L(\nu-2p-2q)]}{\Sigma_{p=0}^\infty [n_{L+1}(\nu-1-2p-2q) - n_L(\nu-2p-2q)]\Sigma_{p=0}^\infty [n_{L-1}(\nu-1-2p-2q) - n_L(\nu-2p-2q)]} . \tag{5.5}$$

The alternating structure of (5.5) is not an accident, but rather a consequence of the fact that (5.1) connects odd ν only to even ν and vice versa. Indeed, generalization to polymer ensembles on higher dimensional ν-space lattices is quite direct. We again work in the formulation (2.14), and suppose a division of ν space into sublattices Λ_α such that two jumps always return the chain to the same sublattice:

$$\text{if } e(\nu,\nu'')e(\nu'',\nu') \neq 0 \text{ and } \nu \in \Lambda_\alpha \tag{5.6}$$

$$\text{then } \nu' \in \Lambda_\alpha .$$

Under these circumstances (2.14), simplified as in (5.2), now reads

$$n_\alpha(\nu) = z_\alpha(\nu) \sum_{\gamma,\nu''} e(\nu,\nu'') \frac{n'_\gamma(\nu'')}{\Sigma_{\nu'} e(\nu'',\nu')z_\alpha(\nu')} , \tag{5.7}$$

where the relevant sublattice is given as index. Thus the equation decomposes into sublattice equations for the $z_\alpha(\nu)$.

6. CONTINUOUS SPIN MODELS

As a final example, we let ν represent a continuous amplitude at the site in question. The developments of Sec. 2 go through with minor obvious modifications. The one-dimensional nearest neighbor interaction (not Coulomb since only nearest neighbor)

$$\phi(\nu,\nu') = J|\nu - \nu'| \tag{6.1}$$

is one that occurs in several applications, e.g., self-sorting of biological cells,[3] solid-on-solid model,[4].... Now

$$e(\nu,\nu') = e^{-\beta J|\nu-\nu'|} , \tag{6.2}$$

$$e^{-1} = \frac{1}{2}\left(-\frac{1}{\beta J}\frac{d^2}{d\nu^2} + \beta J\right),$$

and so the form (2.15), (2.17) is suggested. Then with $n_L \to a$, $n_{L\pm1} \to b$, Y_L, $\hat{Y}_L \to y$, we have

$$a(\nu) = y(\nu)\left(\frac{d^2}{d\nu^2} - \beta^2 J^2\right)\frac{b(\nu)}{(d^2/d\nu^2 - \beta^2 J^2)y(\nu)} \tag{6.3}$$

which simplifies, via the substitution

$$z = ye^{\beta J\nu} , \quad x = e^{2\beta J\nu} , \tag{6.4}$$

$$A = ae^{-2\beta J\nu} , \quad B = be^{-2\beta J\nu} ,$$

to

$$A(x) = z(x)\frac{d^2}{dx^2}\frac{B(x)}{z''(x)} . \tag{6.5}$$

Equation (6.5) can be reduced in differential order by rewriting it as

$$A(x) = \frac{d^2}{dx^2}\frac{B(x)z(x)}{z''(x)} - 2\frac{d}{dx}\frac{B(x)z'(x)}{z''(x)} + B(x) , \tag{6.6}$$

so that

$$\frac{d}{dx}\frac{B(x)z(x)}{z''(x)} - 2\frac{B(x)z'(x)}{z''(x)} = \int [A(x) - B(x)]dx , \tag{6.7}$$

and further reduced by writing

$$\zeta(x) = z'(x)/z(x) , \tag{6.8}$$

yielding

$$\frac{d}{dx}\frac{B(x)}{\zeta'(x) + \zeta(x)^2} - 2\frac{B(x)\zeta(x)}{\zeta'(x) + \zeta(x)^2} = \int [A(x) - B(x)]dx . \tag{6.9}$$

While the general properties of the second order differential equation (6.9) can be assessed, explicit solutions are not easy to come by. It is therefore politic to examine a modification of (6.2), in which the broken slope at $\nu = 0$ is softened:

$$e(\nu,\nu') = \tfrac{1}{2}\,\mathrm{sech}\beta J(\nu - \nu') . \tag{6.10}$$

On a sufficiently restricted space, we then have $e^{-1} = (2\beta J/\pi)\cos(\pi/2\beta J\, d/d\nu)$, which supplies the formal motivation for the following solution of (2.17): Define as well

$$\sigma(\nu,\nu') = \tfrac{1}{2}\tanh\beta J(\nu - \nu') , \tag{6.11}$$

and introduce operators, diagonal on Fourier space, by

$$C(k) = \cosh \pi k/2\beta J\,, \quad S(k) = -i \sinh \pi k/2\beta J\,. \tag{6.12}$$

Then indeed,

$$e = \pi/2\beta J C^{-1}\,, \quad \sigma = \pi/2\beta J \mathscr{P} S^{-1}\,, \tag{6.13}$$

where $\mathscr{P}$ denotes principal part. Further, with $n_L \to a$, $Y_L, \hat{Y}_L \to y$, $n_{L\pm 1} \to b$, (2.17) now takes the form

$$a = yC(b/Cy)\,. \tag{6.14}$$

The most important property of (6.12) is that

$$E_\pm \equiv C \pm iS \text{ are automorphisms}\,. \tag{6.15}$$

One consequence is that if F is a linear operator and we adopt the convention

$$\begin{aligned} Ff\cdot g &= F(fg)\,, \quad F_1 f\cdot g = F(f)g\,, \quad F_2 f\cdot g \\ &= fF(g)\,, \end{aligned} \tag{6.16}$$

then $C \pm iS = (C_1 \pm iS_1)(C_2 \pm iS_2)$, so that

$$C = C_1C_2 - S_1S_2\,, \quad S = C_1s_2 + C_2S_1\,. \tag{6.17}$$

It follows that

$$C_2 = SS_1 + CC_1 \tag{6.18}$$

and hence from (6.14) that $a = C_2 y\cdot b/Cy$ $= (SS_1 + CC_1)y\cdot b/Cy$, or

$$a = S(bSy/Cy) + Cb\,. \tag{6.19}$$

The null space of S is that of constant functions, and indeed $a - Cb$ has no constant component. Thus, (6.19) can be solved in the form

$$\frac{Sy}{cy} = \frac{1}{b} S^{-1}(a - Cb) + \frac{\alpha}{b}\,, \tag{6.20}$$

where S^{-1} and $S^{-1}C$ can be taken as the generalized inverses of (6.12),

$$\begin{aligned} S^{-1}(\nu,\nu') &= (\beta J/\pi)\tanh \beta J(\nu - \nu')\,, \\ S^{-1}C(\nu,\nu') &= (\beta J/\pi)\cosh \beta J(\nu - \nu')\,, \end{aligned} \tag{6.21}$$

and α is as yet an undetermined constant.

Another consequence of (6.12) and (6.15) is that

$$\begin{aligned} Sy/Cy &= -i(E_+ - E_-)y/(E_+ + E_-)y \\ &= -iE_+(y - E_-^2 y)/(y + E_-^2 y) \\ &= iE_+(E_-^2 y/y - 1)/(E_-^2 y/y + 1) \\ &= E_+ \tan(-\tfrac{1}{2} i \ln E_-^2 y/y), \text{ or} \\ Sy/Cy &= \tan(S \ln y)\,. \end{aligned} \tag{6.22}$$

Hence α in (6.20) is determined by

$$\int \arctan\left(\frac{1}{b} S^{-1}(a - Cb) + \frac{\alpha}{b}\right) d\nu = 0 \tag{6.23}$$

and (6.20) has as its solution

$$\ln y = S^{-1} \arctan\left(\frac{1}{b} S^{-1}(a - Cb) + \frac{\alpha}{b}\right) + \gamma \tag{6.24}$$

for some constant γ. We therefore conclude from (2.17), (6.21), and (6.24) that

$$\begin{aligned} &\ln n_L(\nu) + \beta w_L(\nu) \\ &= \frac{\beta J}{\pi} \int \tanh \beta \surd(\nu - \nu') \\ &\quad \times \Big\{ \arctan \frac{\beta J/\pi}{n_{L+1}(\nu')} \Big[\alpha_L^+ + \int (\tanh \beta J(\nu' - \nu'') n_L(\nu'') \\ &\quad - \operatorname{csch} \beta J(\nu' - \nu'') n_{L+1}(\nu'')) d\nu'' \Big] \\ &\quad + \arctan \frac{\beta J/\pi}{n_{L-1}(\nu')} \Big[\alpha_L^- + \int (\tanh \beta J(\nu' - \nu'') n_L(\nu'') \\ &\quad - \operatorname{csch} \beta J(\nu' - \nu'') n_{L-1}(\nu'')) d\nu'' \Big] \Big\} d\nu'\,, \end{aligned} \tag{6.25}$$

which is the desired result. The constant in $Y_L \hat{Y}_L$ has been absorbed as a local reference in $w_L(\nu)$.

7. DISCUSSION

We have examined a number of nearest neighbor spin models. The possibility of solving them completely in the presence of an arbitrary external potential depends very much on the details of the interaction. But there are structural properties common to all such models. For example, consider the modified direct correlation or linear response

$$\hat{C}_2(x\nu, y\nu') \equiv \frac{\partial - \beta w_x(\nu)}{\partial n_y(\nu')}\,, \tag{7.1}$$

whose short range is the key to so many approximations in the theory of fluids. Here $\hat{C}_2$ is in fact not defined because the $\{n_y(\nu')\}$ are constrained by the condition

$$\sum_{\nu'} n_y(\nu') = 1\,. \tag{7.2}$$

This difficulty may be overcome either by eliminating one component, e.g., $n_y(0) = 1 - \Sigma_{\nu \neq 0} n_y(\nu)$, or more simply by restricting derivatives to the surface tangent to (7.2). Since $(\partial/\partial n_y(\nu') - \partial/\partial n_y(\nu''))\Sigma n_y(\nu) = 0$, this amounts to considering only

$$\hat{C}(x\nu, y\nu'\nu'') \equiv C_2(x\nu, y\nu') - \hat{C}_2(x\nu, y\nu'')\,. \tag{7.3}$$

From (2.17), it is seen that

$$\hat{C}_2(x\nu, y\nu'\nu'') = 0 \quad \text{if} \quad |x - y| \neq 0, 1, \tag{7.4}$$

and, as the short range property, is indeed common to all models we have considered. It is clear that this property extends to the higher derivatives as well.

The potential

$$\Omega = -\frac{1}{\beta} \ln \Xi \tag{7.5}$$

is a more basic construct, since all distributions can be obtained from it by differentiation. Although Ω appears prominently in our formulation, the formulation is not particularly appropriate for its explicit construction. Let us, however, see how this can be done, without actually carrying out the messy calculations. The key is the relation

$$n_x(\nu) = \partial\Omega/\partial w_x(\nu)\,, \tag{7.6}$$

so that under any variation, $\delta\Omega = \Sigma \partial\Omega/\partial w_x(\nu) \delta w_x(\nu)$ $= \Sigma n_x(\nu)\delta w_x(\nu) = \delta \Sigma n_x(\nu) w_x(\nu) - \Sigma w_x(\nu) \delta n_x(\nu)$. Thus, we have the two equivalent forms

$$\delta\Omega = \sum n_x(\nu)\delta w_x(\nu)\,, \tag{7.7}$$

$$\delta\Big(\sum n_x(\nu)w_x(\nu) - \Omega\Big) = \sum w_x(\nu)\delta n_x(\nu)\,. \tag{7.8}$$

A convenient way to generate Ω is to start at an empty lattice and then turn on the density according to

$$n_x(\nu;\lambda) = \lambda n_x(\nu) \quad \text{for} \quad \nu \neq 0\,,$$
$$n_x(0;\lambda) = 1 - \lambda \sum_{i\neq 1} n_x(\nu) \tag{7.9}$$

as λ goes from 0 to 1. If the potential reference is chosen so that $w_x(0) = 0$, (7.7) and (7.8) then imply

$$\begin{aligned}\Omega - \Omega_0 &= \sum_{x,\nu} n_x(\nu) \int_0^1 \lambda \frac{\partial}{\partial\lambda} w_x(\nu;\lambda)\,d\lambda \\ &= \sum_{x,\nu} n_x(\nu)w_x(\nu) - \sum_{x,\nu} n_x(\nu) \int_0^1 w_x(\nu;\lambda)\,d\lambda\,,\end{aligned} \tag{7.10}$$

where the λ dependence of w_x and n_x has been set off by a semicolon. One thing is clear, e.g., from $w_x(\nu) = -(1/\beta)\ln W_x(\nu)$ and (2.17), and that is that the nearest neighbor form of $\hat{C}_2$ is mirrored in Ω, which necessarily takes the form

$$\Omega = \sum_{x,\nu} n_x(\nu)[f_x(\nu;n_x,n_{x+1}) + f_x(\nu;n_x,n_{x-1})]\,. \tag{7.11}$$

ACKNOWLEDGMENT

I would like to thank Professor P. J. Paes-Leme for a number of discussions that led to the writing of this paper.

[1] J. K. Percus, J. Stat. Phys. **16**, 299 (1976).
[2] J. K. Percus, J. Chem. Phys. **75**, 1316 (1981).
[3] S. Childress and J. K. Percus, in *Lectures on Mathematics in the Life Sciences*, **14**, edited by S. Childress (American Mathematical Society, Providence, R. I., 1981), p. 59.
[4] J. D. Weeks, J. Chem. Phys. **67**, 3106 (1977).

Quasicontinual representation of a field given on a finite lattice. Interpolation properties

Włodzimierz Garczyński and Jerzy Stelmach
Institute of Theoretical Physics, University of Wrocław, ul. Cybulskiego 36, 50-205 Wrocław, Poland

(Received 17 December 1980; accepted for publication 10 April 1981)

A general theory is systematically presented for handling sets of discrete quantities distributed over finite lattice–lattice functions. This is a more general case than we treated before and in the limit when a lattice becomes infinite our previous results are recovered. We find a one-to-one correspondence between lattice functions which interpolate between discrete lattice values. The uniqueness of interpolations of lattice functions with some entire analytic functions, forming the quasicontinua QC(a,b) and QC(b,a)*, is a key result used.

PACS numbers: 05.50. + q, 02.30. + g

1. INTRODUCTION

In our previous papers[1,2] a general concept of a quasicontinuum QC(a,0) was discussed. It is composed of entire analytic functions of degree not greater than one and of type $\Lambda = \pi/a$ (a—lattice constant),[3–5]

$$\phi(x + iy) = \frac{1}{\sqrt{2\pi}} \int_{-\Lambda}^{\Lambda} dp\, \tilde{\phi}(p) \exp[ip(x + iy)] \in \mathrm{QC}(a,0) \tag{1.1}$$

(where $\tilde{\phi}$ is the Fourier transform of ϕ),

$$|\phi(x + iy)| \leqslant C \exp(\Lambda\, |y|). \tag{1.2}$$

It is a remarkable property of the quasicontinuum QC(a,0) that the interpolation problem possesses a unique solution within this set of functions.[6] This means that for a given set of lattice values

$$\{\phi(na) \equiv \phi(n);\ n \in \mathbf{Z}\}, \tag{1.3}$$

such that

$$\phi(n) \to 0, \quad |n| \to \infty,$$

we may find a unique function $\phi(x) \in \mathrm{QC}(a,0)$ such as its Fourier transform $\tilde{\phi}(p) \in L^2(-\Lambda,\Lambda)$, and assuming prescribed values $\phi(n)$ at lattice sites $x = na, n \in \mathbf{Z}$. The function $\phi(x)$ has the following expansion:

$$\phi(x) = a \sum_{n=-\infty}^{\infty} \phi(n) \delta_a(x - na), \tag{1.4}$$

where

$$\delta_a(x) = \sin(\Lambda x)/\pi x. \tag{1.5}$$

Conversely, given $\phi(x)$ one can find the lattice values as follows:

$$\phi(n) = \int_{-\infty}^{\infty} dx\, \phi(x) \delta_a(x - na). \tag{1.6}$$

The one-to-one correspondence between the set of lattice values[3] and the functions from QC(a,0) permits the use of techniques of continuous analysis when handling discrete quantities (like a field on a lattice,[7] for example). It also permits definition of a derivative of $\phi(n)$ as a value of the $\phi'(x)$ at $x = na$ (cf., e.g., Ref. 8),

$$\phi'(n) = \sum_{\substack{m=-\infty \\ m \neq n}}^{\infty} \frac{(-1)^{n-m}}{(n-m)a} \phi(m). \tag{1.7}$$

In this paper we would like to present a generalization of the formalism to the case of a finite lattice composed of the points $\{na; n \in \mathbf{Z}_N\}$, where $\mathbf{Z}_N$ is the set of integers ranging from $-N$ to N (N—some natural number),

$$\mathbf{Z}_N = \{n; n = 0, \pm 1, \ldots, \pm N\}. \tag{1.8}$$

In this case the function $\phi(x)$ will be defined on some finite interval $[-L/2, L/2]$ only, and the formula[1] cannot be true anymore. Vanishing outside of the mentioned interval would require vanishing everywhere, by analyticity. Thus we fall into a realm of Fourier series rather than integral, and the periodic boundary conditions naturally emerge, as will be evident in the following section.

2. CONCEPTS OF QUASICONTINUA QC(a,b) AND QC(b,a)*

A. The interpolation problem

Let us consider two sets of lattice values,

$$\{\phi(n) \equiv \phi(na); n \in \mathbf{Z}_N\} \tag{2.1}$$

and

$$\{\tilde{\phi}(m) \equiv \tilde{\phi}(mb); m \in \mathbf{Z}_N\}, \tag{2.2}$$

such that for $m,n \in \mathbf{Z}_N$ the following relations hold:

$$\tilde{\phi}(m) = \frac{a}{\sqrt{2\pi}} \sum_{n=-N}^{N} \phi(n) \exp(-imnab) \tag{2.3}$$

and

$$\phi(n) = \frac{b}{\sqrt{2\pi}} \sum_{m=-N}^{N} \tilde{\phi}(m) \exp(imnab). \tag{2.4}$$

Here a and b are some positive numbers—lattice constants,

$$a = \pi/\Lambda, \quad b = 2\pi/L,$$
$$L = (2N + 1)a, \quad 2\Lambda = (2N + 1)b. \tag{2.5}$$

Formulas (2.3) and (2.4) define the discrete Fourier transforms from $\phi(n)$ to $\tilde{\phi}(m)$ and conversely.

The interpolating functions for $\phi(n)$ and $\tilde{\phi}(m)$ may be

 0022-2488/82/061168-07$02.50

obtained from (2.3) and (2.4) upon the replacements $mb \to p, na \to x$,

$$\phi(x) \equiv \frac{b}{\sqrt{2\pi}} \sum_{m=-N}^{N} \tilde{\phi}(m) \exp(imbx), \quad x \in \left[-\frac{L}{2}, \frac{L}{2}\right] \equiv \mathfrak{B}, \tag{2.6}$$

$$\tilde{\phi}(p) \equiv \frac{a}{\sqrt{2\pi}} \sum_{n=-N}^{N} \phi(n) \exp(-ina p), \quad p \in [-\Lambda, \Lambda] \equiv \mathfrak{A}. \tag{2.7}$$

Substituting formulas (2.3) and (2.4) into (2.6) and (2.7) we get the expansions

$$\phi(x) = a \sum_{n=-N}^{N} \phi(n) \delta_{a,b}(x - na), \quad x \in \mathfrak{B}, \tag{2.8}$$

$$\tilde{\phi}(p) = b \sum_{m=-N}^{N} \tilde{\phi}(m) \delta_{b,a}(p - mb), \quad p \in \mathfrak{A}, \tag{2.9}$$

where the function $\delta_{a,b}(x)$ is given by the formula

$$\delta_{a,b}(x) = \frac{b}{2\pi} \sum_{m=-N}^{N} \exp(imbx) = \frac{b \sin(\Lambda x)}{2\pi \sin((b/2)x)} = \frac{\delta_a(x)}{L\delta_L(x)}. \tag{2.10}$$

The function $\delta_{b,a}(p)$ may be obtained from the previous formula upon the replacements $a \longleftrightarrow b$ and $x \longleftrightarrow p$,

$$\delta_{b,a}(p) = \delta_b(p)/2\Lambda\delta_{2\Lambda}(p). \tag{2.11}$$

The following properties of $\delta_{a,b}(x)$ can be deduced from the definition:

$$\delta_{a,b}(x) = \delta_{a,b}(-x) = \delta^*_{a,b}(x) = \delta_{a,b}(x + L), \tag{2.12a}$$

(where * = complex conjugate)

$$\delta_{a,b}(na) = a^{-1}\delta_{n0}, \quad n \in \mathbf{Z}_N, \tag{2.12b}$$

$$\int_{\mathfrak{B}} dx\, \delta_{a,b}(x - na) = 1, \quad n \in \mathbf{Z}_N, \tag{2.12c}$$

$$a \sum_{n=-N}^{N} \delta_{a,b}(x - na) = 1, \quad n \in \mathbf{Z}_N, \tag{2.12d}$$

$$a \int_{-L/2}^{L/2} dx\, \delta_{a,b}(x - na)\delta_{a,b}(x - ma) = \delta_{mn}, \tag{2.12e}$$

$$a \sum_{n=-N}^{N} \delta_{a,b}(x - na)\delta_{a,b}(y - na) = \delta_{a,b}(x - y), \tag{2.12f}$$

$$\int_{\mathfrak{B}} dx\, \delta_{a,b}(x - y)\phi(x) = \phi(y)$$

for a function $\phi(x)$ of the form (2.6), (2.12g)

$$\lim_{a \to 0, b\text{—fix}} \delta_{a,b}(x) = \frac{b}{2\pi} \sum_{n=-\infty}^{\infty} \exp(inbx) = \sum_{n=-\infty}^{\infty} \delta(x - nL), \tag{2.12h}$$

$$\lim_{b \to 0, a\text{—fix}} \delta_{a,b}(x) = \delta_{a,0}(x) \equiv \delta_a(x) = \frac{\sin(\Lambda x)}{\pi x}. \tag{2.12i}$$

Corresponding properties of the function $\delta_{b,a}(p)$ may be obtained from the above upon the following operations:

(i) $a \longleftrightarrow b$,

(ii) $x \longleftrightarrow p$.

It is not difficult to verify the following Parseval relations:

$$\int_{\mathfrak{B}} dx\, \phi^*_1(x)\phi_2(x) = a \sum_{n=-N}^{N} \phi^*_1(n)\phi_2(n) = b \sum_{m=-N}^{N} \tilde{\phi}^*_1(m)\tilde{\phi}_2(m) = \int_{\mathfrak{A}} dp\, \tilde{\phi}^*_1(p)\tilde{\phi}_2(p) \tag{2.13}$$

for any $\phi_k(x), \tilde{\phi}_l(p)$ of the form (2.6) and (2.7), $k,l = 1,2$.

B. Analytic properties of the interpolations

The function $\phi(x)$ defined by (2.6) permits an analytic continuation to an entire and real, periodic function

$$\phi(z) = \frac{b}{2\pi} \sum_{m=-N}^{N} \tilde{\phi}(m) \exp(imbz) = \phi(z + L), \quad z = x + iy, \tag{2.14}$$

satisfying the estimate

$$|\phi(z)| \leqslant C_b \exp(\Lambda\, |\mathrm{Im}\, z|), \tag{2.15}$$

with

$$C_b = \frac{b}{\sqrt{2\pi}} \sum_{m=-N}^{N} |\tilde{\phi}(m)|.$$

Similarly, the function $\tilde{\phi}(p)$ defined by (2.7) can be extended to an entire analytic, real, periodic function

$$\tilde{\phi}(r) = \frac{a}{\sqrt{2\pi}} \sum_{n=-N}^{N} \phi(n) \exp(-inar) = \tilde{\phi}(r + 2\Lambda), \tag{2.16}$$

$$r = p + iq,$$

satisfying the estimate

$$|\tilde{\phi}(r)| \leqslant C_a \exp[(L/2)|\mathrm{Im}\, r|], \tag{2.17}$$

$$C_a = \frac{a}{\sqrt{2\pi}} \sum_{n=-N}^{N} |\phi(n)|.$$

Introducing the variables

$$\zeta = \exp(ibz), \quad \rho = \exp(-iar), \tag{2.18}$$

one may express the functions $\phi(z)$ and $\tilde{\phi}(r)$ as follows:

$$\phi(z) = \zeta^{-N} \mathscr{P}_{2N}(\zeta), \tag{2.19}$$

$$\tilde{\phi}(r) = \rho^{-N} \tilde{\mathscr{P}}_{2N}(\rho), \tag{2.20}$$

where $\mathscr{P}_{2N}$ and $\tilde{\mathscr{P}}_{2N}$ are polynomials of degree $2N$,

$$\mathscr{P}_{2N}(\zeta) = \frac{b}{\sqrt{2\pi}} \sum_{k=0}^{2N} \tilde{\phi}(k - N)\zeta^k, \tag{2.21}$$

$$\tilde{\mathscr{P}}_{2N}(\rho) = \frac{a}{\sqrt{2\pi}} \sum_{k=0}^{2N} \phi(k - N)\rho^k. \tag{2.22}$$

The entire analytic functions $\phi(z)$ periodic with the period L, satisfying the estimate (2.15), and being of the form (2.19), compose a quasicontinuum QC(a,b). Correspondingly, the functions $\tilde{\phi}(r)$, periodic with a period 2Λ, entire analytic functions satisfying the estimate (2.17), and being of the form (2.20), compose the quasicontinuum QC$(b,a)^*$.

C. Main formulas of the quasicontinual approach

The following formulas interrelating lattice values and their interpolations are valid:

$$\phi(n) = \frac{b}{\sqrt{2\pi}} \sum_{m=-N}^{N} \tilde{\phi}(m) \exp(imnab) \in \mathcal{N}(\mathbf{Z}_N) \tag{2.23a}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\mathfrak{A}} dp\, \tilde{\phi}(p) \exp(inap) \tag{2.23b}$$

$$= \int_{\mathfrak{B}} dx \delta_{a,b}(x - na)\phi(x), \quad n \in \mathbf{Z}_N, \tag{2.23c}$$

$$\tilde{\phi}(m) = \frac{a}{\sqrt{2\pi}} \sum_{n=-N}^{N} \phi(n) \exp(-imnab) \in \mathcal{M}(\mathbf{Z}_N) \tag{2.23d}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\mathfrak{B}} dx\, \phi(x) \exp(-imbx) \tag{2.23e}$$

$$= \int_{\mathfrak{A}} dp\, \delta_{b,a}(p - mb)\tilde{\phi}(p), \quad m \in \mathbf{Z}_N, \tag{2.23f}$$

$$\phi(x) = \frac{b}{\sqrt{2\pi}} \sum_{m=-N}^{N} \tilde{\phi}(m) \exp(imbx) \in X(\mathfrak{B}) \tag{2.23g}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\mathfrak{A}} dp\, \tilde{\phi}(p) \left\{ a \sum_{n=-N}^{N} \delta_{a,b}(x - na) \exp(inap) \right\} \tag{2.23h}$$

$$= a \sum_{n=-N}^{N} \phi(n)\delta_{a,b}(x - na), \quad x \in \mathfrak{B}, \tag{2.23i}$$

$$\tilde{\phi}(p) = \frac{a}{\sqrt{2\pi}} \sum_{n=-N}^{N} \phi(n) \exp(-inap) \in P(\mathfrak{A}) \tag{2.23j}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\mathfrak{B}} dx\, \phi(x) \left\{ b \sum_{m=-N}^{N} \delta_{b,a}(p - mb) \times \exp(imbx) \right\} \tag{2.23k}$$

$$= b \sum_{m=-N}^{N} \tilde{\phi}(m)\delta_{b,a}(p - mb), \quad p \in \mathfrak{A}. \tag{2.23l}$$

These formulas, together with the Parseval relations, establish isometric correspondences between the sets $\mathcal{N}(\mathbf{Z}_N)$, $\mathcal{M}(\mathbf{Z}_N)$, $X(\mathfrak{B})$, and $P(\mathfrak{A})$. (See Fig. 1.)

In practical applications one is interested in the limit when $N \to \infty$ in such a way that a is fixed and $b \to 0$ (or the other way round). In this case the formulas (2.23) go over into the formulas of our previous papers.[1,2]

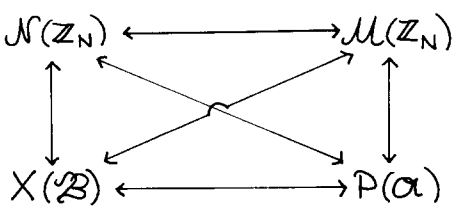


FIG. 1. Isometric relationships between the sets of functions.

3. UNIQUENESS OF INTERPOLATIONS

The two interpolations $\phi_1(x),\phi_2(x)$ of the same lattice function $\phi(n)$ differ by a function vanishing at every lattice site. Therefore, using (2.18) and (2.19), we have

$$\phi_1(x) - \phi_2(x) = \zeta^{-N}[\mathcal{P}^{(1)}_{2N}(\zeta) - \mathcal{P}^{(2)}_{2N}(\zeta)] = \zeta^{-N}\mathcal{P}_{2N}(\zeta), \tag{3.1}$$

where $\mathcal{P}_{2N}$ is again a polynomial of order not greater than $2N$. Since ζ^{-N} never vanishes, this polynomial has to vanish for every ζ_n corresponding to $x = na, n \in \mathbf{Z}_N$,

$$\zeta_n = \exp(in\theta), \quad \theta = 2\pi/(2N+1). \tag{3.2}$$

It is only possible when $\mathcal{P}_{2N}$ vanishes identically. Hence the interpolation problem within the class of functions QC(a,b) has a unique solution. The same is true for the interpolation problem within the class QC$(b,a)^*$. The function $\tilde{\phi}(p)$ is uniquely determined by its lattice values $\tilde{\phi}(m), m \in \mathbf{Z}_N$.

The uniqueness of interpolation is significant since it permits us to replace the lattice functions $\phi(n),\tilde{\phi}(m)$, $m,n \in \mathbf{Z}_N$ by their interpolations $\phi(x)$ and $\tilde{\phi}(p), x \in \mathfrak{B}, p \in \mathfrak{A}$ without losing any information. The inverse is also true. One can recover lattice values from $\phi(x)$ and $\tilde{\phi}(p)$ using formulas (2.23). We would like to maintain the uniqueness of interpolation also in the limits $b \to 0, a$ fixed and $a \to 0, b$ fixed. One knows that this requires a compactification of the set $\mathfrak{A}$ in the first case and the set $\mathfrak{B}$ in the case of second limit if one admits functions $\phi(x),\tilde{\phi}(p)$ rising at infinity not faster than some power of $|x|$ and, correspondingly, $|p|$ (cf., e.g., Ref. 1). For this reason we shall consider functions $\phi(n),\tilde{\phi}(m),\phi(x),\tilde{\phi}(p)$ on compactified sets $\mathbf{Z}_N^*$, $\mathfrak{A}^*$, and $\mathfrak{B}^*$ even before the above-mentioned limits are taken. Compactification means identification of end points in relevant sets and thus introducing the equivalence relations. For instance, the set $\mathbf{Z}_N^*$ consisting of equivalence classes,

$$\mathbf{Z}_N^* = \{n; n \in \mathbf{Z}, n \sim n + 2N + 1\}, \tag{3.3}$$

may be identified with the set of roots $r(n)$ of the order $2N + 1$ from unity,

$$r(n) = \exp(in\theta), \quad n \in \mathbf{Z}_N^*, \tag{3.4}$$

where θ is given by formula (3.2) and the natural identification is made,

$$r(n) = r(n|_{\mathrm{mod}(2N+1)}). \tag{3.5}$$

The roots $r(n)$ lie on a unit circle which now replaces the set $\mathbf{Z}_N$ (see Fig. 2).

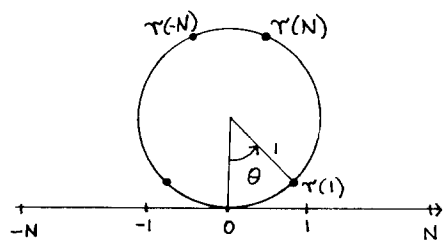


FIG. 2. Compactification of the set $\mathbf{Z}_N$ to the set $\mathbf{Z}_N^*$ of roots of degree $2N + 1$ from unity.

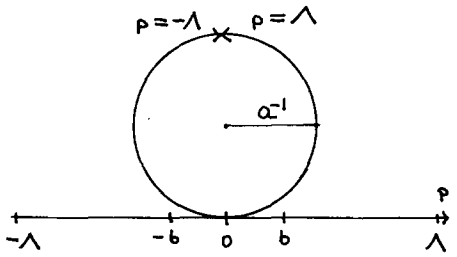


FIG. 3. Compactification of the interval $[-\Lambda,\Lambda]$ to the circle $C_{a^{-1}} = \mathfrak{A}^*$.

The set $\mathfrak{A}$, after introducing the equivalence of points $p \sim p + 2\Lambda$, goes over into circle $C_{a^{-1}}$ of the radius a^{-1} (see Fig. 3).

Similarly, identifying points x and $x + L$ we obtain a circle $C_{b^{-1}} = \mathfrak{B}^*$ containing all admissible positions (see Fig. 4).

Formulas (2.23) remain valid for compactified sets $\mathbf{Z}_N^*$, $\mathfrak{A}^*$, and $\mathfrak{B}^*$, due to the periodicity properties of the functions involved.

4. MULTIPLICATION OF INTERPOLATING FUNCTIONS

Let $\phi_1(n)$ and $\phi_2(n)$ be two lattice functions from the set $\mathcal{N}(\mathbf{Z}_N^*)$. One may readily compose their product and get another lattice function

$$\phi_{12}(n) \equiv \phi_1(n)\phi_2(n) \in \mathcal{N}(\mathbf{Z}_N^*). \tag{4.1}$$

The interpolating function $\phi_{12}(x)$ has the form

$$\phi_{12}(x) = a \sum_{n=-N}^{N} \phi_1(n)\phi_2(n)\delta_{a,b}(x - na) \equiv (\phi_1 \cdot \phi_2)(x) \tag{4.2}$$

and will be called the dot product of $\phi_1(x)$ and $\phi_2(x)$, which interpolates $\phi_1(n)$ and $\phi_2(n)$, correspondingly. From (2.23d) we have

$$\phi_1 \cdot \phi_2(m) = \frac{a}{\sqrt{2\pi}} \sum_{n=-N}^{N} \phi_1(n)\phi_2(n) \exp(-imnab) \in \mathcal{M}(\mathbf{Z}_N^*). \tag{4.3}$$

Calculating the convolutions denoted by $*$, of the functions

$$f_m(n) = \exp(imnab) \in \mathcal{N}(\mathbf{Z}_N^*), \tag{4.4}$$

$$\tilde{f}_n(m) = \exp(-imnab) \in \mathcal{M}(\mathbf{Z}_N^*) \tag{4.5}$$

we obtain

$$f_{m'} * f_{m''}(n) \equiv a \sum_{n'=-N}^{N} f_{m'}(n - n'|_{\mathrm{mod}(2N+1)}) f_{m''}(n') = (2\pi/b)\delta_{m'm''} \exp(im'nab), \tag{4.6}$$

$$\tilde{f}_{n'} * \tilde{f}_{n''}(m) \equiv b \sum_{m'=-N}^{N} \tilde{f}_{n'}(m - m'|_{\mathrm{mod}(2N+1)}) \tilde{f}_{n''}(m') = (2\pi/a)\delta_{n'n''} \exp(-imn'ab). \tag{4.7}$$

Therefore, formula (4.3) may be written in the form

$$\phi_1 \cdot \phi_2(m) = \frac{1}{\sqrt{2\pi}} \left[\frac{a}{\sqrt{2\pi}} \sum_{n'=-N}^{N} \phi_1(n')\tilde{f}_{n'} \right] * \left[\frac{a}{\sqrt{2\pi}} \sum_{n''=-N}^{N} \phi_2(n'')\tilde{f}_{n''} \right](m) = \frac{1}{\sqrt{2\pi}} \tilde{\phi}_1 * \tilde{\phi}_2(m) = \frac{b}{\sqrt{2\pi}} \sum_{m'=-N}^{N} \tilde{\phi}_1(m - m'|_{\mathrm{mod}(2N+1)}) \tilde{\phi}_2(m'). \tag{4.8}$$

Using (2.23g) we obtain from it

$$\phi_1 \cdot \phi_2(x) = \int_{\mathfrak{B}^*} dx' \int_{\mathfrak{B}^*} dx'' M_{a,b}(x,x',x'') \phi_1(x') \phi_2(x''), \tag{4.9}$$

where the form factor $M_{a,b}(x,x',x'')$ is given by the formula

$$M_{a,b}(x,x',x'') = \left(\frac{b}{2\pi} \right)^2 \sum_{m,m'=-N}^{N} \exp\{ib\,[mx - (m - m'|_{\mathrm{mod}(2N+1)})x' - m'x'']\}. \tag{4.10}$$

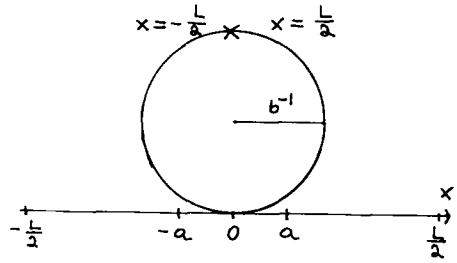


FIG. 4. Compactification of the interval $[-L/2, L/2]$ to the circle $C_{b^{-1}} = \mathfrak{B}^*$.

Performing the above rather tedious calculations we get the result

$$M_{a,b}(x,x',x'') = L^{-2} \left\{ \frac{\sin(\Lambda x)\sin(\Lambda x')}{\sin[(b/2)(x'' - x')]\sin[(b/2)(x'' - x)]} + \frac{\sin(\Lambda x)\sin(\Lambda x'')}{\sin[(b/2)(x' - x'')]\sin[(b/2)(x' - x)]} + \frac{\sin(\Lambda x')\sin(\Lambda x'')}{\sin[(b/2)(x - x')]\sin[(b/2)(x - x'')]} \right\}. \tag{4.11}$$

It is easy to see that in the limit $b \to 0, a$ fixed, we recover from this the form factor $M_a(x,x',x'')$, from our previous paper.[1]

From the Parseval relations (2.13) it follows that

$$\int_{\mathfrak{A}} dp\, \tilde{\phi}(p) = b \sum_{m=-N}^{N} \tilde{\phi}(m) \tag{4.12}$$

and

$$\int_{\mathfrak{B}} dx\, \phi(x) = a \sum_{n=-N}^{N} \phi(n). \tag{4.13}$$

Hence, formula (4.8) may be rewritten as follows:

$$\widetilde{\phi_1 \cdot \phi_2}(m) = \frac{1}{\sqrt{2\pi}} \int_{\mathfrak{A}} dp'\, \tilde{\phi}_1(mb - p'|_{\mathrm{mod}\, 2A})\tilde{\phi}_2(p'). \tag{4.14}$$

Passing to the interpolation $\phi_1 \cdot \phi_2(p)$ we get from it

$$\widetilde{\phi_1 \cdot \phi_2}(p) = \frac{1}{\sqrt{2\pi}} \int_{\mathfrak{A}} dp'\, \tilde{\phi}_1(p - p'|_{\mathrm{mod}\, 2A})\tilde{\phi}_2(p'). \tag{4.15}$$

The whole construction may be carried over into a product of two elements $\tilde{\phi}_1(m), \tilde{\phi}_2(m) \in \mathcal{M}(\mathbf{Z}_N^*)$. Corresponding formulas may be obtained by the familiar replacements $a \longleftrightarrow b, x \longleftrightarrow p$ and a complex conjugation.

The dot products and the convolutions are commutative and associative operations. The dot products of more than two interpolating functions lead to form factors of higher order. For example,

$$(\phi_1 \cdot \cdots \cdot \phi_n)(x) = \int_{\mathfrak{B}} dx_1 \cdots \int_{\mathfrak{B}} dx_n\, M_{a,b}(x, x_1, \ldots, x_n)\phi_1(x_1) \ldots \phi_n(x_n), \tag{4.16}$$

where the form factor of order $n+1$ can be expressed through form factors of the lower order by means of the formula

$$M_{a,b}(x, x_1, \ldots, x_n) = \int_{\mathfrak{B}^*} dy\, M_{a,b}(x, x_1, \ldots, x_{n-2}, y) M_{a,b}(y, x_{n-1}, x_n). \tag{4.17}$$

Applying this rule to the case $n = 3$ we get, after rather lengthy calculations, the result

$$\begin{aligned} M_{a,b}(x, x_1, x_2, x_3) = L^{-3} \Bigg\{ & \frac{\cos(\Lambda x)\sin(\Lambda x_1)\sin(\Lambda x_2)\sin(\Lambda x_3)}{\sin[(b/2)(x_1 - x)]\sin[(b/2)(x_2 - x_1)]\sin[(b/2)(x_3 - x)]} \\ & + \frac{\sin(\Lambda x)\cos(\Lambda x_1)\sin(\Lambda x_2)\sin(\Lambda x_3)}{\sin[(b/2)(x_1 - x_3)]\sin[(b/2)(x_2 - x_1)]\sin[(b/2)(x_3 - x_1)]} \\ & + \frac{\sin(\Lambda x)\sin(\Lambda x_1)\cos(\Lambda x_2)\sin(\Lambda x_3)}{\sin[(b/2)(x_1 - x_2)]\sin[(b/2)(x_3 - x_2)]\sin[(b/2)(x - x_2)]} \\ & + \frac{\sin(\Lambda x)\sin(\Lambda x_1)\sin(\Lambda x_2)\cos(\Lambda x_3)}{\sin[(b/2)(x_1 - x_3)]\sin[(b/2)(x_2 - x_3)]\sin[(b/2)(x - x_3)]} \Bigg\}. \end{aligned} \tag{4.18}$$

From formula (4.17) and from the explicit form of the third order form factor one may easily deduce the following properties of the higher order form factors:

1. $M_{a,b}(x_1, \ldots, x_n)$ is a real and symmetric function with respect to the permutations of its arguments,
2. $M_{a,b}(x_1, \ldots, x_n) = M_{a,b}(x_1 + ka, \ldots, x_n + ka), \quad k \in \mathbf{Z},$
3. $M_{a,b}(x_1 + L, x_2, \ldots, x_n) = M_{a,b}(x_1, x_2, \ldots, x_n),$
4. $M_{a,b}(-x_1, \ldots, -x_n) = M_{a,b}(x_1, \ldots, x_n),$
5. $\int_{\mathfrak{B}^*} dx_n M_{a,b}(x_1, \ldots, x_{n-1}, x_n) = M_{a,b}(x_1, \ldots, x_{n-1}), \quad n > 3,$
6. $\int_{\mathfrak{B}^*} dx_3 M_{a,b}(x_1, x_2, x_3) = \delta_{a,b}(x_1 - x_2),$
7. $\lim_{\substack{b \to 0 \\ a \text{ fixed}}} M_{a,b}(x_1, x_2, x_3) = M_{a,0}(x_1, x_2, x_3) = M_a(x_1, x_2, x_3),$
8. $\int_{\mathfrak{B}^*} dx\, \phi_1 \cdot \phi_2(x) = \int_{\mathfrak{B}^*} dx\, \phi_1(x)\phi_2(x) = a \sum_{n=-N}^{N} \phi_1(n)\phi_2(n) = b \sum_{m=-N}^{N} \tilde{\phi}_1(m)\tilde{\phi}_2(m) = \int_{\mathfrak{A}^*} dp\, \tilde{\phi}_1(p)\tilde{\phi}_2(p) = \int_{\mathfrak{A}^*} dp\, \tilde{\phi}_1 \cdot \tilde{\phi}_2(p).$

Corresponding properties of the form factors $M_{b,a}(p_1, \ldots, p_n)$ may be obtained from the list above upon the replacements $a \longleftrightarrow b, x \longleftrightarrow p$, and a complex conjugation.

The last relation demonstrates that the dot product of any two interpolating functions coincides with the usual product, when integrated. The equality may also happen without the integrations. For instance, when supports of $\tilde{\phi}_1(m)$ and $\tilde{\phi}_2(m)$ are confined to the domains $|m| \leqslant N_1, |m| \leqslant N_2$ with $N_1 + N_2 \leqslant N$, then the convolution (4.8) coincides with the ordinary one and its support does not exceed the interval $[-N, N]$. As the result of Fourier transformation one then gets the equality

$$\phi_1 \cdot \phi_2(x) = \phi_1(x)\phi_2(x). \tag{4.19}$$

5. DIFFERENTIATION

Owing the uniqueness of interpolations one can calculate unique derivatives of $\phi(x)$ and $\tilde{\phi}(p)$ at every point including lattice sites. For instance, from formula (2.23i) we have

$$\phi'(x) = a \sum_{n=-N}^{N} \phi(n)\delta'_{a,b}(x - na), \quad x \in \mathfrak{B}^*. \tag{5.1}$$

Now, using once more (2.23i), we find the expansion

$$\delta'_{a,b}(x - na) = a \sum_{m=-N}^{N} \delta'_{a,b}(ma - na)\delta_{a,b}(x - ma), \tag{5.2}$$

where

$$\delta'_{a,b}(ma-na)=\begin{cases}0, & m=n,\\ \dfrac{b(-1)^{m-n}}{2a\sin[(ab/2)(m-n)]}, & m\neq n.\end{cases} \tag{5.3}$$

Therefore, the function $\phi'(x)$ may be written as the interpolation of the lattice function $\phi'(n)$,

$$\phi'(x)=a\sum_{n=-N}^{N}\phi'(n)\delta_{a,b}(x-na), \tag{5.4}$$

with

$$\phi'(n)=\frac{b}{2}\sum_{\substack{m=-N\\ m\neq n}}^{N}\frac{(-1)^{m-n}}{\sin[(ab/2)(m-n)]}\phi(m), \quad n\in\mathbf{Z}_N^*. \tag{5.5}$$

The lattice function $\phi'(n)$ may be understood as a derivative of the lattice function $\phi(n)$. All higher derivatives may be found in the same way.

From formula (2.23e) we now get

$$\tilde{\phi}'(m)=imb\tilde{\phi}(m)\in\mathcal{M}(\mathbf{Z}_N^*), \tag{5.6}$$

and further, from (2.23l) it follows that

$$\tilde{\phi}'(p)=ip\tilde{\phi}(p)\in P(\mathfrak{A}). \tag{5.7}$$

Let us now calculate the derivative of the dot product of two interpolating functions. Using (2.23g) and (4.8) we find

$$(\phi_1\cdot\phi_2)'(x)=\frac{b^2}{2\pi}\sum_{m,m'=-N}^{N}\tilde{\phi}_1(m-m'|_{\mathrm{mod}(2N+1)})\tilde{\phi}_2(m)imb\exp(imbx). \tag{5.8}$$

Taking into account the identity

$$m=m-m'+m'=m-m'|_{\mathrm{mod}(2N+1)}+m'+\Delta(m,m'), \tag{5.9}$$

where

$$\Delta(m,m')=(2N+1)[\theta(m-m'-N-\tfrac{1}{2})-\theta(m'-m-N-\tfrac{1}{2})] \tag{5.10}$$

and θ is the usual step function, we may write formula (5.8) as follows:

$$(\phi_1\cdot\phi_2)'(x)=\phi'_1\cdot\phi_2(x)+\phi_1\cdot\phi'_2(x)+\Delta_{a,b}(x;\phi_1,\phi_2). \tag{5.11}$$

Here $\Delta_{a,b}(x;\phi_1,\phi_2)$ corresponds to the third term in formula (5.9) and is called the "defect of the Leibnitz rule,"

$$\Delta_{a,b}(x;\phi_1,\phi_2)=\int_{\mathfrak{B}^*}dx'\int_{\mathfrak{B}^*}dx''\Delta_{a,b}(x,x',x'')\phi_1(x')\phi_2(x''), \tag{5.12}$$

where

$$\Delta_{a,b}(x,x',x'')=\frac{ib^3}{2\pi}\sum_{m,n=-N}^{N}\Delta(m,n)\times\exp\{ib[mx-(m-n|_{\mathrm{mod}(2N+1)})x'-mx'']\}. \tag{5.13}$$

On the other hand, using expression (4.9) for the dot product, we obtain upon partial integrations with use of the periodicity property of the form factor the formula

$$\Delta_{a,b}(x,x',x'')=(\partial+\partial'+\partial'')M_{a,b}(x,x',x''). \tag{5.14}$$

Therefore, using formula (4.11) for the form factor we get

$$\Delta_{a,b}(x,x',x'')=\Lambda L^{-2}\left\{\frac{\sin[\Lambda(x+x')]}{\sin[(b/2)(x''-x')]\sin[(b/2)(x''-x)]}+\frac{\sin[\Lambda(x+x'')]}{\sin[(b/2)(x'-x'')]\sin[(b/2)(x'-x)]}+\frac{\sin[\Lambda(x'+x'')]}{\sin[(b/2)(x-x')]\sin[(b/2)(x-x'')]}\right\}. \tag{5.15}$$

This expression coincides with $\Delta_a(x,x',x'')$ calculated in Ref. 1, in the limit $b\to 0$, a fixed.

All the relevant formulas for the differentiation of $\tilde{\phi}(p)$ may be obtained from the present ones by the replacements $a\longleftrightarrow b, x\longleftrightarrow p$, and complex conjugation.

6. GENERALIZATIONS TO HIGHER DIMENSIONS

The whole construction presented so far permits an immediate generalization to higher—let us say D—dimensions. In order to minimize the required changes in the above formulas we shall use the following compact notations:

$$x=(x^1,\ldots,x^D), \quad x^\mu\in\mathfrak{B}_\mu, \ \mu=1,\ldots,D, \tag{6.1}$$

$$p=(p_1,\ldots,p_D), \quad p_\nu\in\mathfrak{A}_\nu, \nu=1,\ldots,D, \tag{6.2}$$

$$n=(n^1,\ldots,n^D), \quad n^\mu\in\mathbf{Z}_{N_\mu}, \tag{6.3}$$

$$m=(m_1,\ldots,m_D), \quad m_\nu\in\mathbf{Z}_{N_\nu}, \tag{6.4}$$

$$\mathfrak{B}_\mu=\{x^\mu;-L_\mu/2\leq x^\mu\leq L_\mu/2\}, \tag{6.5}$$

$$\mathfrak{A}_\nu=\{p_\nu;-\Lambda_\nu\leq p_\nu\leq\Lambda_\nu\}, \tag{6.6}$$

$$\mathbf{Z}_{N_\mu}=\{n^\mu;-N_\mu\leq n^\mu\leq N_\mu\}, \tag{6.7}$$

$$L_\mu=(2N_\mu+1)a_\mu, 2\Lambda_\nu=(2N_\nu+1)b_\nu, \tag{6.8}$$

$$L_\mu=2\pi/b_\mu, \quad 2\Lambda_\nu=2\pi/a_\nu. \tag{6.9}$$

Therefore, one may have different lattice constants on different axes and the number of lattice sites may also vary. We find it convenient to use the following shorthand:

$$na=(n^1a_1,\ldots,n^Da_D), \tag{6.10}$$

$$mb=(m_1b_1,\ldots,m_Db_D), \tag{6.11}$$

$$\delta_{a,b}(x)=\prod_{\mu=1}^{D}\delta_{a_\mu,b_\mu}(x^\mu), \tag{6.12}$$

$$\delta_{b,a}(p)=\prod_{\nu=1}^{D}\delta_{b_\nu,a_\nu}(p_\nu), \tag{6.13}$$

$$\kappa(a)=\prod_{\mu=1}^{D}a_\mu, \quad \kappa(b)=\prod_{\nu=1}^{D}b_\nu, \tag{6.14}$$

$$mb\cdot na=\sum_{\mu=1}^{D}m_\mu n^\mu a_\mu b_\mu, \tag{6.15}$$

$$\mathfrak{A}=\mathfrak{A}_1\times\cdots\times\mathfrak{A}_D, \tag{6.16}$$

$$\mathfrak{B}=\mathfrak{B}_1\times\cdots\times\mathfrak{B}_D, \tag{6.17}$$

$$\mathbf{Z}_N=\mathbf{Z}_{N_1}\times\cdots\times\mathbf{Z}_{N_D}. \tag{6.18}$$

Let us consider two sets of lattice values,

$$\{\phi(n)\equiv\phi(na), \quad n\in\mathbf{Z}_N\}, \tag{6.19}$$

$$\{\tilde{\phi}(m)\equiv\tilde{\phi}(mb), \quad m\in\mathbf{Z}_N\}, \tag{6.20}$$

such that

$$\tilde{\phi}(m) = \frac{\kappa(a)}{(2\pi)^{D/2}} \sum_{n\in\mathbb{Z}_N} \phi(n) \exp(-imb\cdot na), \tag{6.21}$$

$$\phi(n) = \frac{\kappa(b)}{(2\pi)^{D/2}} \sum_{m\in\mathbb{Z}_N} \tilde{\phi}(m) \exp(imb\cdot na). \tag{6.22}$$

The interpolating fields $\phi(x)$ and $\tilde{\phi}(p)$ are obtained from the above lattice functions upon the replacements $mb \to p$ and $na \to x$. In the same way as before one gets

$$\phi(x) = \kappa(a) \sum_{n\in\mathbb{Z}_N} \phi(n)\delta_{a,b}(x - na), \quad x\in\mathfrak{B}, \tag{6.23}$$

$$\tilde{\phi}(p) = \kappa(b) \sum_{m\in\mathbb{Z}_N} \tilde{\phi}(m)\delta_{b,a}(p - mb), \quad p\in\mathfrak{A}. \tag{6.24}$$

The whole bulk of formulas (2.23) possesses obvious generalization to higher dimensions and will not be reproduced here. The dot product of interpolations is now equal to

$$\phi_1\cdot\phi_2(x) = \int_{\mathfrak{B}^*} dx' \int_{\mathfrak{B}^*} dx''\, M_{a,b}(x,x',x'')\phi_1(x')\phi_2(x''), \tag{6.25}$$

where the form factor is given by the formula

$$M_{a,b}(x,x',x'') = \prod_{\mu=1}^{D} M_{a_\mu,b_\mu}(x^\mu,x'^\mu,x''^\mu). \tag{6.26}$$

Differentiation of the dot product of two interpolations yields

$$\partial_\mu(\phi_1\cdot\phi_2)(x) = \partial_\mu\phi_1\cdot\phi_2(x) + \phi_1\cdot\partial_\mu\phi_2(x) + \Delta^{(\mu)}_{a,b}(x;\phi_1,\phi_2), \tag{6.27}$$

where the defect of the Leibnitz rule is of the form

$$\Delta^{(\mu)}_{a,b}(x,x',x'') = (\partial_\mu + \partial'_\mu + \partial''_\mu)M_{a,b}(x,x',x'') = \frac{M_{a,b}(x,x',x'')}{M_{a_\mu,b_\mu}(x^\mu,x'^\mu,x''^\mu)} \Delta_{a_\mu,b_\mu}(x^\mu,x'^\mu,x''^\mu). \tag{6.28}$$

and its kernel might be expressed through the form factor

$$\Delta^{(\mu)}_{a,b}(x,x',x'') = (\partial_\mu + \partial'_\mu + \partial''_\mu)M_{a,b}(x,x',x'') = \frac{M_{a,b}(x,x',x'')}{M_{a_\mu,b_\mu}(x^\mu,x'^\mu,x''^\mu)} \Delta_{a_\mu,b_\mu}(x^\mu,x'^\mu,x''^\mu). \tag{6.29}$$

[1] W. Garczyński and J. Stelmach, "Quasicontinual Approach to a Field Theory on a Lattice. General Theory," J. Math. Phys. **22**, 1106 (1981).
[2] W. Garczyński and J. Stelmach, "A Lattice Field Theory in Quasicontinual Representation," Bull. Acad. Polon. Sci. (in press).
[3] D. Rogula, Bull. Acad. Polon. Sci., série Sci. Techn. **13**, 337 (1965). See also "Quasicontinual Theory of Crystals," Arch. Mech. **28**, 563 (1976).
[4] J. A. Krumhansl, "Generalized Continuum Field Representation for Lattice Vibrations," in *Lattice Dynamics*, edited by R. F. Wallis (Pergamon, New York, 1965), p. 627.
[5] I. A. Kunin, Prikl. Mat. Mech. **30**, 542 (1966). See also *Teoria uprugikh sred s mikrostrukturoj* (Nauka, Moscow, 1975) (in Russian).
[6] Ja. I. Khurgin, W. P. Jakowlew, *Finitnyje funkcji v fizikie i technikie* (Nauka, Moscow 1971) (in Russian).
[7] K. Wilson, Phys. Rev. D **10**, 2445 (1974).
[8] S. D. Drell, M. Weinstein, and S. Yankielowicz, Phys. Rev. D **13**, 3342 (1976).

Soliton solutions for self-dual SU(2) gauge fields on Euclidean space

Patricio S. Letelier
Departamento de Física, Universidade de Brasília, 70910 Brasília D. F., Brazil

(Received 22 October 1981; accepted for publication 15 January 1982)

Axially symmetric soliton solutions for self-dual SU (2) gauge fields on Euclidean four-dimensional flat space are found using an extension of the Belinsky–Zakharov solution generating technique.

PACS numbers: 11.10.Np

I. INTRODUCTION

Recently Yang[1] reduced the problem of finding self-dual SU (2) gauge fields on Euclidean four-dimensional flat space to the problem of solving the system of equations

$$\phi(\phi_{\xi\bar{\xi}}+\phi_{\eta\bar{\eta}})-\phi_{\xi}\phi_{\bar{\xi}}-\phi_{\eta}\phi_{\bar{\eta}}+\rho_{\xi}\bar{\rho}_{\bar{\xi}}+\rho_{\eta}\bar{\rho}_{\bar{\eta}}=0, \tag{1.1a}$$

$$\phi(\rho_{\xi\bar{\xi}}+\rho_{\eta\bar{\eta}})-2\rho_{\xi}\phi_{\bar{\xi}}-2\rho_{\eta}\phi_{\bar{\eta}}=0, \tag{1.1b}$$

$$\phi(\bar{\rho}_{\xi\bar{\xi}}+\bar{\rho}_{\eta\bar{\eta}})-2\bar{\rho}_{\bar{\xi}}\phi_{\xi}-2\bar{\rho}_{\bar{\eta}}\phi_{\eta}=0, \tag{1.1c}$$

where the subscripts denote partial differentiation and

$$2^{1/2}\xi=x+iy,\quad 2^{1/2}\bar{\xi}=x-iy, \tag{1.2a}$$

$$2^{1/2}\eta=z-ix_4,\quad 2^{1/2}\bar{\eta}=z+ix_4. \tag{1.2b}$$

ϕ and ρ are a real and a complex functions, respectively. If we restrict ϕ and ρ to be functions only of $r=(2\xi\bar{\xi})^{1/2}$ and $z=(\eta+\bar{\eta})/2^{1/2}$, we find that (1.1) reduces to

$$\phi(\phi_{rr}+\phi_r/r+\phi_{zz})-\phi_r^2-\phi_z^2+\rho_r\bar{\rho}_r+\rho_z\bar{\rho}_z=0, \tag{1.3a}$$

$$\phi(\rho_{rr}+\rho_r/r+\rho_{zz})-2(\rho_r\phi_r+\rho_z\phi_z)=0, \tag{1.3b}$$

$$\phi(\bar{\rho}_{rr}+\bar{\rho}_r/r+\bar{\rho}_{zz})-2(\bar{\rho}_r\phi_r+\bar{\rho}_z\phi_z)=0. \tag{1.3c}$$

The system of Eq. (1.3) has been studied by many authors using Bäcklund transformations.[2–6] One can find particular solutions to this system of equations noticing that for a ρ of the form $\rho=\sigma e^{i\alpha}$, where α is a real constant and σ a real function, the system of Eqs. (1.3) is equivalent to the Ernst equation for axially symmetric gravitational fields.[7] Solutions to this last equation are well known.

The purpose of this paper is to find explicit pure soliton solutions to Eqs. (1.3). The method used is a simple extension of the Belinsky–Zakharov solution generating technique[8,9] that we present in Sec. II. In Sec. III we study the equation for the "wavefunction" ψ_0. The knowledge of this field allows us to find the different solutions. In Sec. IV we exhibit a one-soliton solution to (1.3). And finally, in Sec. V we study a two-soliton solution, and we show how to construct n-soliton solutions to (1.3).

II. THE SOLUTION GENERATING ALGORITHM

The system of equations (1.3) can be written in a complete equivalent form as

$$\partial_r(rg_rg^{-1})+\partial_z(rg_zg^{-1})=0, \tag{2.1}$$

where

$$g=\frac{r}{\phi}\begin{pmatrix} 1 & -\bar{\rho} \\ -\rho & \rho\bar{\rho}+\phi^2 \end{pmatrix}. \tag{2.2}$$

Note that $g=g^+$ and $\det g=r^2$. The BZ method of Refs. 8 and 9 cannot be used directly to solve (2.1) since g is not a symmetric matrix and its determinant is positive. Thus we need to extend the BZ method to the present case; the extension is straightforward, so we shall only present the essential formulas.

The BZ method for solving Eq. (2.1) is based on the fact that the condition of integrability for the system of equations

$$D_r\psi=\frac{rU+\lambda V}{\lambda^2+r^2}\psi, \tag{2.3a}$$

$$D_z\psi=\frac{rV-\lambda U}{\lambda^2+r^2}\psi, \tag{2.3b}$$

where

$$D_r\equiv\partial_r+\frac{2\lambda r}{\lambda^2+r^2}\partial_\lambda, \tag{2.4a}$$

$$D_z\equiv\partial_z-\frac{2\lambda^2}{\lambda^2+r^2}\partial_\lambda, \tag{2.4b}$$

$$U\equiv rg_rg^{-1},\quad V\equiv rg_zg^{-1}, \tag{2.5}$$

is just the same Eq. (2.1). ψ is a 2×2 complex matrix function of r, z, and the expectral parameter λ. Putting $\lambda=0$ in (2.3), we have that $\psi(\lambda=0)=g$. Solutions with pure soliton character are associated with solutions of Eqs. (2.3) of the form

$$\psi=\chi\psi_0, \tag{2.6}$$

$$\chi\equiv I+\sum_{k=1}^{n}\frac{R_k}{\lambda-\mu_k}, \tag{2.7}$$

where ψ_0 is a solution to Eqs. (2.3) for a known g, say g_0. R_k are complex matrices functions of r and z only, and μ_k are scalar complex functions of r and z only. The pure soliton character of the solution is associated with the particular form of χ given by (2.7), i.e., with the existence of simple poles in the matrix χ. The number of poles will tell us the number of solitons appearing in the solution. Note that letting $\lambda=0$ in (2.6), we get

$$g=(\chi|_{\lambda=0})g_0. \tag{2.8}$$

Now we look for a condition that guarantees the fact that $g=g^+$. For real g, this condition is

$$g=\chi(-r^2/\lambda,r,z)g_0\,\tilde{\chi}(\lambda,r,z), \tag{2.9}$$

where the tilde indicates transposition. The most natural generalization of (2.9) for the present case is

$$g=\chi(-r^2/\bar{\lambda},r,z)g_0[\chi(\lambda,r,z)]^+. \tag{2.10}$$

 0022-2488/82/061175-04$02.50

The condition (2.10) tells us that if $g_0 = g_0^+$, then $g = g^+$, when g is given by (2.6). The proof of this statement is a straightforward generalization of the one given in Ref. 8.

From (2.3)–(2.10) we find

$$g_{ab} = (g_0)_{ab} - \sum_{kl} \frac{\bar{N}_a^{(l)}(\Gamma^{-1})_{lk} N_b^{(k)}}{\mu_k \bar{\mu}_l}, \tag{2.11}$$

$$\Gamma_{kl} \equiv \frac{m^{(k)} \cdot \bar{m}^{(l)}}{r^2 + \mu_k \bar{\mu}_l} = \bar{\Gamma}_{lk}, \tag{2.12}$$

$$m^{(k)} \cdot \bar{m}^{(l)} \equiv m_a^{(k)} (g_0)_{ab} \bar{m}_b^{(l)}, \tag{2.13}$$

$$N_a^{(k)} \equiv m_b^{(k)} (g_0)_{ba}, \tag{2.14}$$

$$m_a^{(k)} \equiv m_{0b}^{(k)} M_{ba}^{(k)}, \tag{2.15}$$

$$M^{(k)} = \psi_0^{-1}|_{\lambda = \mu_k}, \tag{2.16}$$

$$\mu_k = \alpha_k - z \pm [(\alpha_k - z)^2 + r^2]^{1/2}. \tag{2.17}$$

The sum convention on the indices a and b is assumed. a and b take the values 1 and 2. $m_{0b}^{(k)}$ and α_k are sets of arbitrary complex constants. Note that the solution (2.11) is completely determined by g_0, ψ_0, and these sets of constants. Regardless of the fact that the matrix whose elements are (2.11) is Hermitian, in general, it is not possible to cast it in the form (2.2), since the matrix (2.2) has determinant equal to r^2. To remedy this problem, we can define a new matrix,

$$g^{\mathrm{Ph}} \equiv rg/(\det g)^{1/2}, \tag{2.18}$$

that satisfies the two conditions $(g^{\mathrm{Ph}})^+ = g^{\mathrm{Ph}}$ and $\det g^{\mathrm{Ph}} = r^2$. Taking the trace of Eq. (2.1), one can prove that the new g^{Ph} is also a solution to (2.1) whenever g is a solution. The determinant of (2.11) can be explicitly computed:

$$\det g_{(n)} = (-1)^n r^{2n} \prod_{l=1}^{n} |\mu^l|^{-2} \det g_0. \tag{2.19}$$

Since $\det g_0 = r^2$ for a g_0 of the form (2.2) and $\det g$ needs to be positive in order to normalize its value to r^2, from (2.19) we conclude that we can only have an even number of solitons associated with a g_0 given by (2.2). We can have an odd number of solitons defining a new matrix

$$g_0' = \begin{pmatrix} r/\phi & 0 \\ 0 & -r\phi \end{pmatrix}, \tag{2.20}$$

with $\det g_0' = -r^2$. Equation (2.1) for g_0' is equivalent to (1.3) with $\rho = 0$. Thus, for an odd n the new g' constructed using Eq. (2.11) will be Hermitian with positive determinant, and in consequence we can define a new g'^{Ph} like (2.18) that will be a solution to (2.1) (by construction) and can be cast as (2.2).

III. THE FUNCTION ψ_0

The function ψ_0 obeys the differential equations[9] (2.3) with g replaced by g_0, i.e.,

$$\left(\partial_r + \frac{2\lambda r}{\lambda^2 + r^2}\partial_\lambda\right)\psi_0 = \frac{rU_0 + \lambda V_0}{\lambda^2 + r^2}\psi_0, \tag{3.1a}$$

$$\left(\partial_z - \frac{2\lambda^2}{\lambda^2 + r^2}\partial_\lambda\right)\psi_0 = \frac{rV_0 - \lambda U_0}{\lambda^2 + r^2}\psi_0, \tag{3.1b}$$

where $U_0 \equiv r(g_0)_r g_0^{-1}$ and $V_0 = r(g_0)_z g_0^{-1}$. Furthermore, ψ_0 must satisfy the initial condition

$$\psi_0|_{\lambda=0} = g_0. \tag{3.2}$$

From (3.1) we get

$$\lambda \partial_r \psi_0 + r\partial_z \psi_0 = V_0 \psi_0. \tag{3.3}$$

In this section we study the system of equations (3.1a), (3.3) with the boundary condition (3.2) when g_0 is a diagonal matrix, i.e., $(g_0)_{12} = (g_0)_{21} = 0$. If g_0 is diagonal, one may assume that ψ_0 is also a diagonal matrix with these assumptions; (3.1a), (3.3), and (3.2) yield

$$\left(\partial_r + \frac{2\lambda r}{\lambda^2 + r^2}\partial_\lambda\right)\ln(\det\psi_0) = \frac{2r}{\lambda^2 + r^2}, \tag{3.4a}$$

$$(\lambda\partial_r + r\partial_z)\ln(\det\psi_0) = 0, \tag{3.4b}$$

$$\det\psi_0|_{\lambda=0} = r^2. \tag{3.5}$$

A solution to Eq. (3.4) with the condition (3.5) is

$$\det\psi_0 = r^2 - \lambda^2 - 2\lambda z. \tag{3.6}$$

Note that this result does not depend on the explicit form of the matrix g_0.

We shall take as g_0 the particular solution to (2.1) given by

$$(g_0)_{11} = r^{1-b}\exp\{-[az + c(\tfrac{1}{2}r^2 - z^2)]\}, \tag{3.7a}$$

$$(g_0)_{22} = r^2 (g_0)_{11}^{-1}, \tag{3.7b}$$

where a, b, and c are arbitrary real constants. Multiplying the rhs of (3.7a) by a constant, we can obtain a new solution to (2.1), but, since this constant can be absorbed in the arbitrary constants $m_0^{(k)}$, the final solution will not be generalized by the inclusion of the above-mentioned constant.

For the g_0 given by (3.7), Eqs. (3.1a), (3.3), and (3.2) are equivalent to

$$\left(\partial_r + \frac{2\lambda r}{\lambda^2 + r^2}\partial_\lambda\right)\ln(\psi_0)_{11} = \frac{(1-b)r - \lambda r(a - 2cz) - cr3}{\lambda^2 + r^2}, \tag{3.8a}$$

$$(\lambda\partial_r + r\partial_z)\ln(\psi_0)_{11} = -r(a - 2cz), \tag{3.8b}$$

$$\ln(\psi_0)_{11}|_{\lambda=0} = (1-b)\ln r - az + c(\tfrac{1}{2}r^2 - z^2), \tag{3.8c}$$

$$\ln(\psi_0)_{22} = \ln(r^2 - \lambda^2 - 2\lambda z) - \ln(\psi_0)_{11}. \tag{3.6'}$$

A direct computation shows that

$$(\psi_0)_{11} = (r^2 - \lambda^2 - 2\lambda z)^{(1-b)/2} \times \exp(-\{a(z + \tfrac{1}{2}\lambda) + c[\tfrac{1}{2}r^2 - (z + \tfrac{1}{2}\lambda)^2]\}), \tag{3.9a}$$

is a solution to (3.8a) and (3.8b), and that satisfies the condition (3.8c). From (3.9a) and (3.6'), we have

$$(\psi_0)_{22} = (r^2 - \lambda^2 - 2\lambda z)^{(1+b)/2} \times \exp\{a(z + \tfrac{1}{2}\lambda) + c[\tfrac{1}{2}r^2 - (z + \tfrac{1}{2}\lambda)^2]\}. \tag{3.9b}$$

Note that the solution ψ_0' for the case of a g_0' with negative determinant can be obtained doing the following change of signal in (3.7) and (3.9), respectively:

$$(g_0')_{11} = (g_0)_{11}, \quad (g_0')_{22} = -(g_0)_{22}, \tag{3.10}$$

$$(\psi_0')_{11} = (\psi_0)_{11}, \quad (\psi_0')_{22} = -(\psi_0)_{22}. \tag{3.11}$$

The key stone of the BZ method is the function ψ_0; one needs to have it in an explicit form, since, alas, one can solve (3.1) explicitly only for very special cases. The case presented

here is a generalization of the one that appears in Ref. 9 which can be obtained letting $b = 1$ and $a = c = 0$ in (3.9).

IV. ONE-SOLITON SOLUTION

Following the indications for odd number solitons given at the end of Secs. II and III, we find that the solution (3.7) can be "vested" in the following way:

$$g_{11} = -r(r^2|m_{01}\psi_2/\mu|^2 + |m_{02}\psi_1|^2\phi^2)/(\phi\Delta), \quad (4.1a)$$

$$g_{12} = \bar{g}_{21} = 2m_{01}m_{02}(r^2 + |\mu|^2)\bar{\alpha}\bar{\mu}r\phi/(|\mu|^2\Delta), \quad (4.1b)$$

$$g_{22} = -r(|m_{01}\psi_2|^2 + r^2|m_{02}\psi_1/\mu|^2\phi^2)\phi/\Delta, \quad (4.1c)$$

where

$$\Delta = |m_{01}\psi_2|^2 - |m_{02}\psi_1|^2\phi^2, \quad (4.2a)$$

$$\psi_1 = (-2\alpha\mu)^{(1-b)/2}\exp\{a(z + \tfrac{1}{2}\mu) + c[\tfrac{1}{2}r^2 - (z + \tfrac{1}{2}\mu)^2]\}, \quad (4.2b)$$

$$\psi_2 = 2\alpha\mu/\psi_1, \quad (4.2c)$$

$$\phi = r^b\exp[az + c(\tfrac{1}{2}r^2 - z^2)], \quad (4.2d)$$

$$\mu = z - \alpha \pm [(z-\alpha)^2 + r^2]^{1/2}. \quad (4.2e)$$

In (4.1a)–(4.2e) we have omitted an index (1) because it is not necessary in this case. The "physical" g, i.e., g^{Ph}, is obtained multiplying the expressions (4.1) by $|\mu|/r$. The soliton character of the solution can be perceived by noticing the special form of the function Δ appearing in (4.1).

V. TWO-SOLITON AND n-SOLITON SOLUTIONS

When the matrix χ has two poles $\lambda = \mu_1$ and $\lambda = \mu_2$, the expression (2.11) can be written as

$$g = g_0 - \hat{g}, \quad (5.1a)$$

$$\begin{aligned}\hat{g}_{ab}\Delta = {} & \frac{m^{(2)}\cdot\bar{m}^{(2)}}{(r^2 + |\mu_2|^2)|\mu_1|^2}\bar{N}_a^{(1)}N_b^{(1)} \\ & - \frac{m^{(2)}\cdot\bar{m}^{(1)}}{(r^2 + \bar{\mu}_1\mu_2)\mu_1\bar{\mu}_2}\bar{N}_a^{(2)}N_b^{(1)} \\ & - \frac{m^{(1)}\cdot\bar{m}^{(2)}}{(r^2 + \mu_1\bar{\mu}_2)\bar{\mu}_1\mu_2}\bar{N}_a^{(1)}N_b^{(2)} \\ & + \frac{m^{(1)}\cdot m^{(1)}}{(r^2 + (\mu_1)^2|\mu_2|^2}\bar{N}_a^{(2)}N_b^{(2)}, \end{aligned} \quad (5.1b)$$

where

$$\Delta = \bar{\Delta} \equiv \frac{m^{(1)}\cdot\bar{m}^{(1)}}{r^2 + |\mu_1|^2}\,\frac{m^{(2)}\cdot\bar{m}^{(2)}}{r^2 + |\mu_2|^2} - \left|\frac{m^1\cdot\bar{m}^2}{r^2 + \mu_1\bar{\mu}_2}\right|^2, \quad (5.2a)$$

$$m^{(k)}\cdot\bar{m}^{(l)} \equiv m_a^{(k)}(g_0)_{ab}m_b^{(l)}. \quad (5.2b)$$

The indices k, l, a, and b take the values 1 and 2. From (2.14), (3.7), (3.9), (2.16), and (2.17), we get

$$N_1^{(k)} = m_{01}^{(k)}r^{1-b}(-2\alpha_k\mu_k)^{(b-1)/2} \times\exp[-c(\mu_k z + \tfrac{1}{4}\mu_k^2) + \tfrac{1}{2}a\mu_k], \quad (5.3a)$$

$$N_2^{(k)} = m_{02}^{(k)}r^{1+b}(-2\alpha_k\mu_k)^{-(b+1)/2} \times\exp[c(\mu_k z + \tfrac{1}{4}\mu_k^2) - \tfrac{1}{2}a\mu_k]. \quad (5.3b)$$

Let us first consider the general case, i.e., all $m_{0\alpha}^{(k)} \neq 0$. In this case we can cast the different functions appearing in (5.1) as

$$\Delta = \frac{r|B|^2}{4|\alpha_1\alpha_2\mu_1\mu_2|} \times\left(\frac{\cosh(y_1 + \delta_1)\cosh(y_2 + \delta_2)}{(r^2 + |\mu_1|^2)(r^2 + |\mu_2|^2)} - \left|\frac{\cosh(x + \epsilon)}{r^2 + \mu_1\bar{\mu}_2}\right|^2\right),$$

$$m^{(k)}\cdot\bar{m}^{(k)} = \frac{rA_k}{2|\alpha_k\mu_k|}\cosh(y_k + \delta_k), \quad (5.5)$$

$$m^1\cdot\bar{m}^2 = (m^2\cdot\bar{m}^1)^* = \frac{rB}{2(\alpha_1\bar{\alpha}_2\mu_1\bar{\mu}_2)^{1/2}}\cosh(x + \epsilon), \quad (5.6)$$

where the new constants A_k, δ_k, B, and ϵ are related to the old constants by

$$A_k = 2|m_{01}^{(k)}m_{02}^{(k)}|, \quad (5.7)$$

$$\tanh\delta_k = \frac{|m_{01}^{(k)}|^2 - |m_{02}^{(k)}|^2}{|m_{01}^{(k)}|^2 + |m_{02}^{(k)}|^2}, \quad (5.8)$$

$$B^2 = 4m_{01}^{(1)}\bar{m}_{01}^{(2)}m_{02}^{(1)}\bar{m}_{02}^{(2)}, \quad (5.9)$$

$$\tanh\epsilon = \frac{m_{01}^{(1)}\bar{m}_{01}^{(2)} - m_{02}^{(1)}\bar{m}_{02}^{(2)}}{m_{01}^{(1)}\bar{m}_{01}^{(2)} + m_{02}^{(1)}\bar{m}_{02}^{(2)}}. \quad (5.10)$$

The functions y_k and x are related to r and z by

$$y_k = a(z + \mathrm{Re}\mu_k) + c[\tfrac{1}{2}r^2 + z^2 - 2\mathrm{Re}(z + \tfrac{1}{2}\mu_k)^2] + b\ln|2\alpha_k\mu_k/r|,$$

$$\begin{aligned} x = {} & a[z + \tfrac{1}{2}(\mu_1 + \bar{\mu}_2)] \\ & + c[\tfrac{1}{2}r^2 - z^2 - z(\mu_1 + \bar{\mu}_2) - \tfrac{1}{4}(\mu_1^2 + \bar{\mu}_2^2)] \\ & + \tfrac{1}{2}b\ln(4\alpha_1\alpha_2\mu_1\bar{\mu}_2/r^2). \end{aligned} \quad (5.11)\text{–}(5.12)$$

Note that A_k and δ_k are real and B and ϵ are complex constants and also that y_1 and y_2 are real and x is a complex function.

The physical g is obtained from (5.1) as follows:

$$g^{\mathrm{Ph}} = (|\mu_1\mu_2|/r^2)(g_0 - \hat{g}). \quad (5.13)$$

Now we shall study particular cases of (5.1). Taking $m_{02}^{(1)} = m_{01}^{(2)} = 0$, we get

$$g_{12}^{\mathrm{Ph}} = -|\mu_2/\mu_1|(g_0)_{11}, \quad (5.14a)$$

$$g_{22}^{\mathrm{Ph}} = r^2/g_{11}^{\mathrm{Ph}}, \quad g_{12}^{\mathrm{Ph}} = 0. \quad (5.14b)$$

Putting $m_{01}^{(1)} = m_{02}^{(2)} = 0$, we have

$$g_{11}^{\mathrm{Ph}} = -|\mu_1/\mu_2|(g_0)_{11}; \quad (5.15)$$

g_{22}^{Ph} and g_{12}^{Ph} are given by (5.14b). Choosing either $m_{02}^{(1)} = m_{02}^{(2)} = 0$ or $m_{01}^{(1)} = m_{01}^{(2)} = 0$, after some algebra, we get

$$g_{11}^{\mathrm{Ph}} = (r^2/|\mu_1|\ |\mu_2|)(g_0)_{11}; \quad (5.16)$$

g_{22}^{Ph} and g_{12}^{Ph} are given by (5.14b).

The particular form of the solutions (5.14)–(5.16) suggest that *any diagonal* solution to (2.1), say g_0, give rise to different two-soliton solutions given by

$$g_{11}^{\mathrm{Ph}} = \pm|\mu_1/r|^{\epsilon_1}|\mu_2/r|^{\epsilon_2}(g_0)_{11}; \quad (5.17)$$

g_{22}^{Ph} and g_{12}^{Ph} are given as before, and ϵ_1 and ϵ_2 are constants that can take the values ± 1.

That (5.17) is a solution to (2.1) for any diagonal g_0 can be proved directly. Also, if ϵ_1 and ϵ_2 are arbitrary constants, (5.17) is a solution to (2.1), but in this case we cannot say that the new solution is a two-soliton solution since the soliton

character is given by the existence of single poles in the "scattering" matrix χ and for single poles we have $\epsilon_{1,2} = \pm 1$.

The solution (5.17) can be also obtained by construction of a one-soliton solution. Taking either $m_{01} = 0$ or $m_{02} = 0$ in (4.1), and multiplying the result by $|\mu|/r$, we get

$$g_{11}^{\mathrm{Ph}} = |\mu/r|^{\epsilon}(g_0)_{11}; \tag{5.18}$$

$\epsilon = 1$ is associated with $m_{01} = 0$ and $\epsilon = -1$ with $m_{02} = 0$. We can repeat the procedure one more time taking (5.18) as $(g_0)_{11}$; in this form we end up with Eq. (5.17), mod ± 1. The g_{22}^{Ph} and g_{12}^{Ph} are obtained as before. This procedure can be repeated n times to give

$$g_{11}^{\mathrm{Ph}} = \left(\prod_{l=1}^{n} |\mu_l/r|^{\epsilon_l}\right)(g_0)_{11}; \tag{5.19}$$

g_{22}^{Ph} and g_{12}^{Ph} are obtained as in (5.14b). For an odd n we consider g_0 to be given by Eq. (2.20). This shows how to construct an n-soliton solution from any known diagonal solution to (2.1). As a final remark, we want to add that the solutions (4.1) and (5.1), in the general case, are not solutions to the Ernst equation since $\rho = -g_{21}^{\mathrm{Ph}}/g_{11}^{\mathrm{Ph}}$ and for these solutions ρ cannot be cast in the form $\rho = \sigma e^{i\alpha}$, where α is a real constant and σ a real function. Nevertheless, the particular cases (5.17), (5.18), and the n-soliton solution (5.10) are solutions to the Ernst equation, because $g_{12}^{\mathrm{Ph}} = 0$ for these solutions; then we have $\sigma = 0$.

[1]C. N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[2]M. A. Lohe, Nucl. Phys. B **142**, 236 (1978), and references therein.
[3]D. Ray, Phys. Lett. B **97**, 113 (1981).
[4]G. Neugebaur, J. Phys. A: Math Gen. **12**, 67 (1979).
[5]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978).
[6]E. Kyriakoupoulos, Gen. Relativ. Gravit. **12**, 989 (1980).
[7]L. Witten, Phys. Rev. D. **19**, 718 (1979).
[8]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)].
[9]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **77**, 3 (1979) [Sov. Phys. JETP **50**, 1 (1979)].

An SL(2,C)-invariant representation of the Dirac equation

Levere C. Hostler
Physics Department, Wilkes College, Wilkes-Barre, Pennsylvania 18766

(Received 17 March 1981; accepted for publication 30 July 1981)

The relativistic Pauli-type equation $\{\Pi_\mu\Pi_\mu + m^2 + ie\boldsymbol{\sigma}\cdot(\mathbf{E} + i\mathbf{B})\}\phi = 0$ (one of a pair of second order equations that follow from the Dirac equation) is investigated. Here ϕ is a 2×1 Pauli spinor, and σ_a, a = 1, 2, 3, are the usual Pauli spin matrices. We investigate the correspondence between a description of a Dirac particle based on this second order equation and a single 2×1 spinor wave function ϕ, and a description employing the conventional Dirac equation. To this end, a representation is derived that expresses the exact interacting Dirac propagator in an arbitrary external electromagnetic field in terms of the 2×2 interacting propagator of the Pauli-type equation. Techniques for discussing scattering of a Dirac particle using the above wave equation are presented. The Feynman rules, at least for the external c-number field problems considered here, are basically those of scalar electrodynamics. A notable difference is the replacement of the factor $-ie(p_{f\mu} + p_{i\mu})$ for the one-photon vertex of scalar electrodynamics by the new factor $-ie(p_{f\nu}(\delta_{\nu\mu} + i\sigma_{\nu\mu}) + (\delta_{\mu\nu} + i\sigma_{\mu\nu})p_{i\nu})$ for our one-photon vertex. $\sigma_{\mu\nu}$ is a self-dual spin tensor whose components are the usual Pauli matrices, times ± 1.

PACS numbers: 11.10Qn

I. INTRODUCTION

We will here investigate a relativistic Pauli-type wave equation,

$$\{\Pi_\mu\Pi_\mu + m^2 + ie\boldsymbol{\sigma}\cdot(\mathbf{E} + i\mathbf{B})\}\phi = 0, \qquad \Pi_\mu \equiv \partial/i\partial x^\mu - eA_\mu, \tag{1.1}$$

for a Dirac particle moving in an arbitrary c-number electromagnetic field, $A_\mu \equiv (A_1, A_2, A_3; A_4 \equiv iV)$. In this equation ϕ is a 2×1 Pauli spinor and σ_a, a = 1, 2, 3, are the usual Pauli spin matrices. Equation(1.1) and the analogous equation,

$$\{\Pi_\mu\Pi_\mu + m^2 - ie\boldsymbol{\sigma}\cdot(\mathbf{E} - i\mathbf{B})\}\eta = 0, \tag{1.2}$$

arise naturally in the theory of the Dirac equation if one works in a representation in which γ_5 is diagonal. In a work of Laporte and Uhlenbeck[1] Eqs. (1.1) and (1.2) are derived from a spinor form of Dirac's equation due to van der Waerden.[2] Also, with the help of the spinor-analysis formalism of van der Waerden,[3] it is possible to motivate Eqs. (1.1) and (1.2) without recourse to the Dirac equation.[4]

It is evident that a single one of the second-order equations (1.1) or (1.2) alone should already contain all the physics of a Dirac particle. Accordingly, it should be possible, for example, to fully describe a Dirac particle using a single 2×1 spinor ϕ obeying the second order equation (1.1). Because of the small dimension of the matrices involved, Eq. (1.1) for a Dirac particle may prove to be quite useful for applications, for example to quantum electrodynamics. Also, the equation involves the mass only quadratically. This feature may make Eq. (1.1) quite suitable for application to the " indefinite mass theory" of Hostler[5]; or to the similar "mass Hamiltonian" method of Feynman, Kislinger, and Raundal,[6] and analogous formalisms of Hiroshi Enatsu and others.[7-9]

We will here investigate the relationship between the description of a Dirac particle using Eq. (1.1) and a single 2×1 spinor wave function ϕ, and a description employing the conventional Dirac equation. The material in Sec. IIA, although not especially new, is included for completeness, since this background material is needed to do the physics of a Dirac particle using Eq. (1.1). The propagator of Eq. (1.1) is investigated in Sec. IIB. A representation, Eq. (2.41), is derived expressing the exact interacting Dirac propagator S_F in an arbirtrary external electromagnetic field in terms of the 2×2 interacting prpagator, g_+, of Eq. (1.1). The equivalence of Eq. (1.1) with the original form of Dirac's equation is made manifest through this representation. For example, the equivalence of the two bound state spectra, if one exists, follows, since two functions related as in Eq. (2.41) must have the same pole structure in the complex energy plane.

In Sec. IIC amplitudes for scattering, pair production, and pair annihilation are compared for the new formalism and for the original form of the Dirac equation, The representation (2.41) of the Dirac propagator plays a central role in this comparison. Feynman rules for Eq. (1.1), at least for the external c-number field problems considered here, are basically those of scalar electrodynamics. A notable differnce is the replacement of the factor $-ie(p_{f\mu} + p_{i\mu})$ for the one-photon vertex of scalar electrodynamics with the new factor $-ie(p_{f\nu}(\delta_{\nu\mu} + i\sigma_{\nu\mu}) + (\delta_{\mu\nu} + i\sigma_{\mu\nu})p_{i\nu})$ for our one-photon vertex. The spin tensor $\sigma_{\mu\nu}$ occurring here is a self-dual tensor made up of the Pauli matrices [see Eq. (2.12)].

II. PHYSICS OF THE SL(2, C)-INVARIANT WAVE EQUATION

A. General formalism

1. Relation to the Dirac equation

Our Dirac equation is written in the form

$$(\Pi - im)\psi = 0, \quad \Pi \equiv \gamma_\mu\Pi_\mu, \tag{2.1}$$

$$\gamma \equiv \begin{pmatrix} 0 & -i\boldsymbol{\sigma} \\ i\boldsymbol{\sigma} & 0 \end{pmatrix}, \quad \gamma_4 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.2}$$

Equation (2.1) is equivalent to the two equations[2]

$$m\phi = \bar{\tau}_\mu\Pi_\mu\eta \tag{2.3}$$

 0022-2488/82/061179-06$02.50

and

$$m\eta = -\tau_\mu \Pi_\mu \phi, \tag{2.4}$$

for a pair of 2×1 Pauli spinors ϕ and η. The matrices τ_μ are defined through the equations $\tau_{1,2,3}\equiv\sigma_{1,2,3}$ and $\tau_4\equiv i$

$$\tau_1=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\qquad \tau_2=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},$$
$$\tau_3=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\qquad \tau_4=\begin{pmatrix}i & 0\\ 0 & i\end{pmatrix}, \tag{2.5}$$

while the matrices $\bar\tau_\mu$ are defined as

$$\bar\tau_\mu\equiv(\tau_\mu)^\dagger. \tag{2.6}$$

Equations (2.3) and (2.4) can be derived by writing the 4×1 Dirac spinor ψ in terms of ϕ and η

$$\psi\equiv\begin{pmatrix}\frac{1}{2}(\phi+\eta)\\ \frac{1}{2}(\phi-\eta)\end{pmatrix}. \tag{2.7}$$

When the representation (2.7) is substituted in the original Dirac equation (2.1), the Dirac equation splits into the pair of equation (2.3) and (2.4). The second-order equations (1.1) and (1.2) now follow by eliminating η and ϕ, respectively, from Eqs. (2.3) and (2.4).

For future reference we prefer to express the second order equations (1.1) and (1.2) in the manifestly covariant forms[1]

$$(\Pi_\mu\Pi_\mu + m^2 - e\tfrac{1}{2}\sigma_{\mu\nu}F_{\mu\nu})\phi = 0 \tag{2.8}$$

and

$$(\Pi_\mu\Pi_\mu + m^2 - e\tfrac{1}{2}\bar\sigma_{\mu\nu}F_{\mu\nu})\eta = 0, \tag{2.9}$$

respectively. We here require the identities

$$\bar\tau_\mu\tau_\nu = \delta_{\mu\nu} + i\sigma_{\mu\nu}, \tag{2.10}$$

and

$$\tau_\mu\bar\tau_\nu = \delta_{\mu\nu} + i\bar\sigma_{\mu\nu}. \tag{2.11}$$

In these identities, $\sigma_{\mu\nu}$ is the self-dual spin tensor

$$\sigma_{\mu\nu}\equiv\begin{pmatrix}0 & \sigma_3 & -\sigma_2 & \sigma_1\\ -\sigma_3 & 0 & \sigma_1 & \sigma_2\\ \sigma_2 & -\sigma_1 & 0 & \sigma_3\\ -\sigma_1 & -\sigma_2 & -\sigma_3 & 0\end{pmatrix}; \tag{2.12}$$

and $\bar\sigma_{\mu\nu}$, whose dual is equal to minus $\bar\sigma_{\mu\nu}$, is the spin tensor

$$\bar\sigma_{\mu\nu}\equiv\begin{pmatrix}0 & \sigma_3 & -\sigma_2 & -\sigma_1\\ -\sigma_3 & 0 & \sigma_1 & -\sigma_2\\ \sigma_2 & -\sigma_1 & 0 & -\sigma_3\\ \sigma_1 & \sigma_2 & \sigma_3 & 0\end{pmatrix}. \tag{2.13}$$

Lorentz invariance of Eqs. (2.3), (2.4), (2.8), and (2.9) is best discussed in terms of the spinor analysis formalism of van der Waerden mentioned above. Relevant transformation properties are discussed in Appendix A, but the reader is referred to the sources cited in Ref. 3 for derivations. We record here for future reference the algebra of the matrices[10] $\sigma_{\mu\nu}$:

$$\sigma_{\mu\nu}\sigma_{\lambda\rho} = \delta_{\mu\lambda}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\lambda} + \epsilon_{\mu\nu\lambda\rho} + i(\delta_{\mu\lambda}\sigma_{\nu\rho} - \delta_{\mu\rho}\sigma_{\nu\lambda} + \delta_{\nu\rho}\sigma_{\mu\lambda} - \delta_{\nu\lambda}\sigma_{\mu\rho}). \tag{2.14}$$

From Eq. (2.14) we learn that the spin tensor $\sigma_{\mu\nu}/2$ obeys the Lie algebra of the homogeneous Lorentz group

$$[\sigma_{\mu\nu};\sigma_{\lambda\rho}] = 2i(\delta_{\mu\lambda}\sigma_{\nu\rho} - \delta_{\mu\rho}\sigma_{\nu\lambda} + \delta_{\nu\rho}\sigma_{\mu\lambda} - \delta_{\nu\lambda}\sigma_{\mu\rho}). \tag{2.15}$$

2. Unitary dot product

We here take the point of view suggested in the introduction of treating just one of the second-order equations, chosen to be Eq. (2.8), as providing a full description of a Dirac particle. In order to ensure the physical equivalence of the original Dirac equation and our relativistic "Pauli equation" (2.8), we require a generalized inner product $(\phi_b;\phi_a)$, in the space of the states ϕ which agrees with the Dirac inner product of the corresponding four-spinors, ψ,

$$(\phi_b;\phi_a)\equiv\int d^3r\,\psi_b^\dagger\psi_a. \tag{2.16}$$

From Eqs.(2.7) and (2.4) we obtain

$$(\phi_b;\phi_a) = \int d^3r\,\phi_b^\dagger\,\frac{m^2+\overleftarrow{B}\,\overrightarrow{B}}{2m^2}\,\phi_a, \tag{2.17}$$

where

$$B\equiv\tau_\mu\Pi_\mu. \tag{2.18}$$

For future reference, we record here an analogous definition

$$A\equiv\bar\tau_\mu\Pi_\mu. \tag{2.19}$$

Note that in Eq. (2.17) and throughout this paper we adopt the convention that a derivative acting to the left shall mean *minus* differentiation of the objects on the left; e.g., $\phi^\dagger\overleftarrow{\partial}_\mu\equiv-\partial\phi^\dagger/\partial x^\mu$. Equations (2.8) and (2.17) enable us to discuss the physics of a Dirac particle without further reference to the original Dirac equation.

Equation (2.8) admits a conserved transition current

$$(J_\mu)_{ba} = -\bar\phi_b\,\frac{\overleftrightarrow{\Pi}_\mu}{2m^2}\,\phi_a - \partial_\nu\left(\bar\phi_b\,\frac{\sigma_{\mu\nu}}{2m^2}\,\phi_a\right), \tag{2.20}$$

$$\bar\phi\equiv\phi^\dagger\overleftarrow{B}. \tag{2.21}$$

We define $(J_4)_{ba}\equiv i\rho_{ba}$. It is straightforward to verify that $\int d^3r\,\rho_{ba}$ differs from the original inner product (2.17) only in that the present expresion for $\int d^3r\,\rho_{ba}$ incorporates the equation of motion.

The object $\bar\phi$ defined in Eq. (2.21) plays a role for the relativistic Pauli equation (2.8) which is quite analogous to the role played by $\bar\psi\equiv\psi^\dagger\gamma_4$ in conventional Dirac theory. In this connection it is worth pointing out the SL(2,C) transformation laws

$$\phi' = U^{-1}\phi \tag{2.22}$$

and

$$\bar\phi' = \bar\phi U \tag{2.23}$$

of ϕ and $\bar\phi$. According to these transformation laws the combination $\bar\phi_b\phi_a$ is Lorentz invariant. For future reference we note the expression for the charge conjugate $C[\phi]$ of the wave function of Eq. (2.8):

$$C[\phi]\equiv\frac{\sigma_y}{m}\,\bar{\bar\phi}. \tag{2.24}$$

B. Propagators

Two related propagators $g_\pm(2,1)$ are defined as follows:

$$\{(\Pi_\mu)_2(\Pi_\mu)_2 + m^2 + ie\boldsymbol{\sigma}\cdot(\mathbf{E}(2) + i\mathbf{B}(2))\}g_+(2,1) = -\delta(t_2 - t_1)\delta^3(r_2 - r_1), \tag{2.25}$$

$$\{(\Pi_\mu)_2(\Pi_\mu)_2 + m^2 - ie\boldsymbol{\sigma}\cdot(\mathbf{E}(2) - i\mathbf{B}(2))\}g_-(2,1) = -\delta(t_2 - t_1)\delta^3(r_2 - r_1). \tag{2.26}$$

In accordance with hole-theory concepts, we assume Feynman boundary conditions. As is known, the Feynman boundary conditions can be incorporated by giving the mass parameter m a small negative imaginary part. This prescription will be implicit in the following.

To proceed, it is convenient to go over to an abstract operator notation. We introduce space-time coordinate eigenkets, $|1\rangle \equiv |r_1,t_1\rangle$ defined through the equations

$$x_\mu|1\rangle = (x_\mu)_1|1\rangle, \tag{2.27}$$

and

$$\langle 2|1\rangle = \delta(t_2 - t_1)\delta^3(r_2 - r_1). \tag{2.28}$$

The time component of the four-vector x_μ is thus treated as an operator on the same footing as x, y, z.[11] From Eq. (2.28) the completeness relation

$$1 = \int dt_1 \int d^3 r_1 |1\rangle\langle 1| \tag{2.29}$$

follows. The propagators $g_\pm(2,1)$ in Eqs. (2.25) and (2.26) are visualized as coordinate space representatives of abstract operators $g_\pm : g_\pm(2,1) \equiv \langle 2|g_\pm|1\rangle$. In this notation Eq. (2.25) reads

$$\{\Pi_\mu\Pi_\mu + m^2 + ie\boldsymbol{\sigma}\cdot(\mathbf{E} + ie\mathbf{B})\}g_+ = -1,$$

or, in terms of the operators B and A of Eqs. (2.18) and (2.19),

$$(AB + m^2)g_+ = -1. \tag{2.30}$$

The analogous equation for g_- is

$$(BA + m^2)g_- = -1. \tag{2.31}$$

The Green's function g_- can be expressed in terms of g_+ as follows. Multiplying Eq. (2.31) on the left by A we find

$$(AB + m^2)Ag_- = -A,$$

$$Ag_- = -(AB + m^2)^{-1}A = g_+A.$$

Now multiply $g_+A = Ag_-$ on the left by B and use Eq. (2.31) again: $Bg_+A = (BA + m^2 - m^2)g_- = -1 - m^2g_-$. Thus,

$$g_- = -\frac{Bg_+A + 1}{m^2}. \tag{2.32}$$

Next we explore the relationship between the propagator g_+ of the relativistic Pauli equation and the propagator S_F of the usual Dirac equation. The latter is defined (in the same abstract notation, and with the same m having $\mathrm{Im}(m) > 0$ by

$$S_F = \frac{-1}{\Pi - im}. \tag{2.33}$$

The propagator S_F is first expressed in terms of the propagator G_F of the second-order Dirac equation

$$S_F = (\Pi + im)G_F, \tag{2.34}$$

and the equation

$$\{\Pi_\mu\Pi_\mu + m^2 + ie(\boldsymbol{\alpha}\cdot\mathbf{E} + i\boldsymbol{\sigma}\cdot\mathbf{B})\}G_F = -1 \tag{2.35}$$

for G_F is integrated in terms of g_+. Here $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ are the usual 4×4 Dirac matrices. The same notation $\boldsymbol{\sigma}$ will be used to denote either the 4×4 Dirac spin matrices or the 2×2 Pauli matrices, depending on the context. Thus we write

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} \text{ and } \boldsymbol{\sigma} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}.$$

Equation (2.35) can be integrated in terms of g_+ and g_- by projecting onto subspaces defined using the perpendicular projectors $\frac{1}{2}(1 \pm \gamma_5)$, where $\gamma_5 = \gamma_4\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. By inserting a resolution of the identity in Eq. (2.35), we find

$$\{O\tfrac{1}{2}(1 + \gamma_5) + O\tfrac{1}{2}(1 - \gamma_5)\}G_F = -1, \tag{2.36}$$

$$O \equiv \Pi_\mu\Pi_\mu + m^2 + ie(\boldsymbol{\alpha}\cdot\mathbf{E} + i\boldsymbol{\sigma}\cdot\mathbf{B}). \tag{2.37}$$

Next, it is convenient to write the 4×4 matrices $O\frac{1}{2}(1 \pm \gamma_5)$ as Kronecker products of two 2×2 matrices. The explicit prescription for this is

$$(A \otimes B)_{\alpha\beta,\mu\nu} \equiv A_{\beta\nu}B_{\alpha\mu}. \tag{2.38}$$

Equation (2.36) then goes over to the form

$$\{O_+ \otimes \tfrac{1}{2}(1 + \sigma_x) + O_- \otimes \tfrac{1}{2}(1 - \sigma_x)\}G_F = -1, \tag{2.39}$$

where

$$O_+ \equiv \Pi_\mu\Pi_\mu + m^2 + ie\boldsymbol{\sigma}\cdot(\mathbf{E} + i\mathbf{B}),$$

and

$$O_- \equiv \Pi_\mu\Pi_\mu + m^2 - ie\boldsymbol{\sigma}\cdot(\mathbf{E} - i\mathbf{B}).$$

Since the 2×2 operators $\frac{1}{2}(1 + \sigma_x)$ are a set of perpendicular projectors, we can write the solution of (2.39) in the form $G_F = -(O_+)^{-1} \otimes \frac{1}{2}(1 + \sigma_x) - (O_-)^{-1} \otimes \frac{1}{2}(1 - \sigma_x)$. The relations $-(O_\pm)^{-1} = g_\pm$ then lead to $G_F = g_+ \otimes \frac{1}{2}(1 + \sigma_x) + g_- \otimes \frac{1}{2}(1 - \sigma_x)$, or more explicitly,

$$G_F = \frac{1}{2}\begin{pmatrix} g_+ + g_- & g_+ - g_- \\ g_+ - g_- & g_+ + g_- \end{pmatrix}. \tag{2.40}$$

Next we substitute the representation (2.32) for g_-, and apply the operator $(\Pi + im)$ in accordance with Eq. (2.34). The result is a representation expressing the exact 4×4 Dirac propagator with arbitary external electromagnetic field in terms of the 2×2 propagator g_+ of the relativistic Pauli equation (2.8). This representation is

$$iS_F = \begin{pmatrix} \dfrac{(B - m)g_+(A + m) + 1}{2m} & \dfrac{(B - m)g_+(-A + m) + 1}{2m} \\ -\dfrac{(B + m)g_+(A + m) + 1}{2m} & -\dfrac{(B + m)g_+(-A + m) + 1}{2m} \end{pmatrix}. \tag{2.41}$$

To obtain this identity repeated use of the propagator equation (2.30) is required. The representation (2.41) makes the equivalence of the new and the old formalism quite manifest. For example, it is clear from (2.41) that a bound-state spec-

trum, if it exists, must be the same for g_+ and for S_F, since two functions related by Eq. (2.41) will have the same pole structure in the complex energy plane.

C. Scattering amplitudes

To discuss scattering it is convenient to rewrite Eq. (2.8) in the form

$$(p^2 + m^2 + V)\phi = 0,$$
$$V = -ep_\mu A_\mu - eA_\mu p_\mu + e^2 A_\mu A_\mu - e\tfrac{1}{2}\sigma_{\mu\nu}F_{\mu\nu}. \tag{2.42}$$

The spin dependent part of V can be transformed as follows:

$$\begin{aligned} &-e\tfrac{1}{2}\sigma_{\mu\nu}F_{\mu\nu} \\ &\quad = -e\tfrac{1}{2}\sigma_{\mu\nu}(\partial_\mu A_\nu - \partial_\nu A_\mu) \\ &\quad = -ie\sigma_{\mu\nu} i^{-1}\partial A_\nu/\partial x^\mu \\ &\quad = -ie\sigma_{\mu\nu}[p_\mu; A_\nu] \\ &\quad = -iep_\mu \sigma_{\mu\nu}A_\nu - ieA_\nu \sigma_{\nu\mu}p_\mu. \end{aligned}$$

Through the use of this transformation, the spin term of V merges with the orbital terms to give

$$\begin{aligned} V = &-ep_\mu(\delta_{\mu\nu} + i\sigma_{\mu\nu})A_\nu \\ &-eA_\nu(\delta_{\nu\mu} + i\sigma_{\nu\mu})p_\mu + e^2 A_\mu A_\mu. \end{aligned} \tag{2.43}$$

The mathematics of integrating Eq.(2.42) with interaction potential (2.43) is basically not different from the familiar mathematics of the Klein–Gordon equation. Indeed, Feynman rules developed for Eqs. (2.42) and (2.43) will be basically the same as for scalar electrodynamics, at least for the external c-number field problems considered here. The principal difference between the two sets of rules will be the replacement of the one-photon vertex $-ie(p_{f\mu} + p_{i\mu})$ of scalar electrodynamics by the new factor $-ie(p_{f\nu}(\delta_{\nu\mu} + i\sigma_{\nu\mu}) + (\delta_{\mu\nu} + i\sigma_{\mu\nu})p_{i\nu})$ for our one-photon vertex.

The integration of Eq. (2.42) involves the standard lore of scattering theory. Therefore the following results on scattering can be reported briefly. We denote by $\phi_i{}^+(\mathbf{x})$ the solution of Eq. (2.42) that describes the action of the potential on a positive or negative frequency free particle, described by a plane wave state, $\phi_i(\mathbf{x})$. We go over to the abstract notation of Sec. IIB and write Eq. (2.42) in integral form as follows:

$$|\phi_i{}^+\rangle = |\phi_i\rangle + g_0 V|\phi_i{}^+\rangle, \tag{2.44}$$

where

$$\phi_i{}^+(\mathbf{x}) \equiv \langle \mathbf{x}|\phi_i{}^+\rangle, \quad \phi_i(\mathbf{x}) \equiv \langle \mathbf{x}|\phi_i\rangle,$$

and $g_0 \equiv -1/(p^2 + m^2)$ is the free particle propagator. Feynman boundary conditions are assumed and the same small negative imaginary part of m is implicitly assumed. Equation (2.44) can be solved by iteration in a Born series $|\phi_i{}^+\rangle = (1 + g_0 V + g_0 V g_0 V + \cdots)|\phi_i\rangle$; equivalently,

$$|\phi_i{}^+\rangle = g_+(-(p^2 + m^2))|\phi_i\rangle, \tag{2.45}$$

g_+ being the propagator (2.30). Although the representation (2.45) would have the indeterminant form 0/0 for real m, the representation is well defined here because m has a small negative imaginary part. In the following the representation (2.45) will be convenient because it avoids any expansion in the potential.

Scattering amplitudes can now be computed as

$$S_{fi} \equiv (\phi_f; \phi_i{}^+) = \int d^3 r\, \bar{\phi}_f(i\overleftrightarrow{\Pi}_4/2m^2)\phi_i{}^+,$$

where ϕ_f is the final plane wave state. Using standard reduction techniques[12] this amplitude for $f \neq i$ can be transformed into

$$S_{fi} = \epsilon_f \langle \bar{\phi}_f|(-i(p^2 + m^2)/2m^2)|\phi_i{}^+\rangle,$$

or in view of Eq. (2.45), into

$$\begin{aligned} S_{fi} = (i\epsilon_f/2m^2)\, &\overline{\langle \phi_f|}(-(p^2 + m^2)) \\ &\times g_+(-(p^2 + m^2))|\phi_i\rangle, \quad f \neq i. \end{aligned} \tag{2.46}$$

Here

$$\overline{\langle \phi\,|} \equiv \langle \phi\,|B \tag{2.47}$$

and $\epsilon_f = +1$ for electrons, $\epsilon_f = -1$ for positrons.

Next we substitute in Eq. (2.46) explicit expressions for initial and final plane wave states, $\phi_{i,f}$. These expressions are here required in the forms[13]:

$$|\phi_i\rangle = (2E_i(E_i + m))^{-1/2}(A_0 + m)u^{\rho'}(0)|\mathbf{p}_i\rangle \quad \text{(electrons)}, \tag{2.48}$$

$$|\phi_i\rangle = (2E_i(E_i + m))^{-1/2}(-A_0 + m)u^{\rho'}(0)|-\mathbf{p}_i\rangle \quad \text{(positrons)}, \tag{2.49}$$

$$\overline{\langle \phi_f|} = \langle \mathbf{p}_f|u^{\rho'}(0)^\dagger(B_0 - m)m(2E_f(E_f + m))^{-1/2} \quad \text{(electrons)}, \tag{2.50}$$

and

$$\overline{\langle \phi_f|} = \langle -\mathbf{p}_f|u^{\rho'}(0)^\dagger(B_0 + m)m(2E_f(E_f + m))^{-1/2} \quad \text{(positrons)}. \tag{2.51}$$

In these equations $A_0 \equiv -ip_4 + \boldsymbol{\sigma}\cdot\mathbf{p}$ denotes the free particle form of the operator A of Eq. (2.19), and $B_0 \equiv ip_4 + \boldsymbol{\sigma}\cdot\mathbf{p}$ is the free-particle from of B defined in Eq. (2.18). Also, the subscripts i and f for "initial" and "final" are to be interpreted as "unperturbed" and "perturbation" respectively, and do not necessarily refer to the actual chronological ordering of events. When the representations (2.48)–(2.51) are substituted in Eq. (2.46), we obtain the following expressions for S_{fi}:

$$C\langle p_f|u^{\rho'}(0)^\dagger(-(p^2 + m^2))(B_0 - m)g_+(A_0 + m)(-(p^2 + m^2))u^{\rho'}(0)|\mathbf{p}_i\rangle \quad \text{(electron-electron scattering, } f \neq i); \tag{2.52}$$

$$-C\langle -p_f|u^{\rho'}(0)^\dagger(-(p^2 + m^2))(B_0 + m)g_+(A_0 + m)(-(p^2 + m^2))u^{\rho'}(0)|\mathbf{p}_i\rangle \quad \text{(pair annihilation)}; \tag{2.53}$$

$$C\langle p_f|u^{\rho'}(0)^\dagger(-(p^2 + m^2))(B_0 - m)g_+(-A_0 + m)(-(p^2 + m^2))u^{\rho'}(0)|-\mathbf{p}_i\rangle \quad \text{(pair production)}; \tag{2.54}$$

and

$$-C\langle -p_f|u^{\rho'}(0)^\dagger(-(p^2+m^2))(B_0+m)g_+(-A_0+m)(-(p^2+m^2))u^{\rho'}(0)|-\mathbf{p}_i\rangle$$

(positron-positron scattering, $f\neq i$), (2.55)

$$C\equiv(i/2m)(2E_i(E_i+m))^{-1/2}(2E_f(E_f+m))^{-1/2}.$$

We are now in a position to compare our scattering amplitudes with those of the original Dirac theory. The scattering amplitudes for the latter can be written in the form [analogous to Eq.(2.46)]

$$S_{fi}=\epsilon_f\,\overline{\langle\psi_f|}\,(p-im)S_F(p-im)|\psi_i\rangle,\qquad f\neq i,\quad \overline{\langle\psi_f|}\equiv\langle\psi_f|\gamma_4. \tag{2.56}$$

When explicit expressions for the Dirac plane wave states are substituted in Eq. (2.56) the factors $(p-im)$ are converted into factors (p^2+m^2), and we find simply

$$S_{fi}=-\epsilon_f(2E_i(E_i+m))^{-1/2}(2E_f(E_f+m))^{-1/2}\times\langle\epsilon_f\mathbf{p}_f|\bar{u}^{\rho'}(0)(-(p^2+m^2))S_F(-(p^2+m^2))u^{\rho'}(0)|\epsilon_i\mathbf{p}_i\rangle, \tag{2.57}$$

$f\neq i$; and $\epsilon_{i,f}=+1$ for electrons, $\epsilon_{i,f}=-1$ for positrons. In Eq. (2.57) we substitute the representation (2.41) of S_F and then select spinors $u^{\rho'}(0)$ and $\bar{u}^{\rho'}(0)$ as appropriate to describe one of the processes (2.52)–(2.55). Doing this will automatically select only one sector of the 2×2 matrix of matrices (2.41), and we obtain structures almost identical to Eqs. (2.52) through (2.55). For example, if we select $u^{\rho'}(0)$ and $\bar{u}^{\rho'}(0)$ to be electron states in Eq. (2.57); Then when (2.41) is substituted in Eq. (2.57) we find

$$S_{fi}=-(2E_i(E_i+m))^{-1/2}(2E_f(E_f+m))^{-1/2}(-i/2m)\times\langle\mathbf{p}_f|u^{\rho'}(0)^\dagger(-(p^2+m^2))\times[(B-m)g_+(A+m)+1](-(p^2+m^2))u^{\rho'}(0)|\mathbf{p}_i\rangle,\qquad f\neq i. \tag{2.58}$$

Now $u^{\rho\rho'}(0)$ denote 2×1 column matrices. Because of the condition $f\neq i$, orthogonality permits dropping the term involving the unit matrix in the bracket of (2.58). Equivalence of (2.58) and (2.52) now follows if we are permitted to make the replacements $A\to A_0$ and $B\to B_0$ in Eq. (2.58). But this is permitted because in the limit $\mathrm{Im}(m)\to0^-$ the factors $(-(p^2+m^2))$ in (2.58) will act on the initial and final states and annul terms in A and B that involve the potential.[14] The amplitudes for the other processes (2.53)–(2.55) can be similarly seen to be equal to the corresponding amplitudes calculated with the original Dirac theory.

APPENDIX

Lorentz invariance will be reviewed here briefly. As mentioned in the introduction, the reader is referred to the sources cited in Ref. 3 for derivations.

To each Lorentz transformation $x'_\mu=x_\nu R_{\nu\mu}$ we associate a linear transformation U of SL(2, C) according to the rule

$$U\bar{\tau}_\mu U^\dagger=\bar{\tau}_\nu R_{\nu\mu}, \tag{A1}$$

equivalently,

$$(U^{-1})^\dagger\tau_\mu U^{-1}=\tau_\nu R_{\nu\mu}. \tag{A2}$$

Equations (A1) or (A2) establish a homomorphism between the homogeneous Lorentz group and the group SL(2, C), and show that $\bar{\tau}_\mu$ and τ_μ in Eqs. (2.3) and (2.4) will behave as four-vectors, thereby assuring the Lorentz invariance of these equations. Note that ϕ is transformed according to Eq. (2.22), and η according to $\eta'\equiv U^\dagger\eta$. From Eqs. (2.10) and (2.11) and the transformation laws (A1) and (A2), it is straightforward to show that the spin matrices $\sigma_{\mu\nu}$ and $\bar{\sigma}_{\mu\nu}$ have the expected tensor transformation properties

$$U\sigma_{\mu\nu}U^{-1}=\sigma_{\lambda\rho}R_{\lambda\mu}R_{\rho\nu}, \tag{A3}$$

and

$$(U^{-1})^\dagger\bar{\sigma}_{\mu\nu}U^\dagger=\bar{\sigma}_{\lambda\rho}R_{\lambda\mu}R_{\rho\nu}. \tag{A4}$$

Equations (A3) and (A4) assure the Lorentz invariance of Eqs. (2.8) and (2.9).

The homomorphism between the homogeneous Lorentz group and SL(2, C) can be exhibited more explicitly by writing both matrices R and U in canonical form with the same canonical parameters $\omega_{\mu\nu}$: $R=\exp(-\frac{1}{2}i\omega_{\mu\nu}S_{\mu\nu})$ and $U=\exp(-\frac{1}{4}i\omega_{\mu\nu}\sigma_{\mu\nu})$. The canonical parameters are defined as:

$$\omega_{\mu\nu}=\begin{array}{c}\nu\to\\ \mu\downarrow\end{array}\left(\begin{array}{ccc|c}0&\theta_3&-\theta_2&iv_1\\-\theta_3&0&\theta_1&iv_2\\\theta_2&-\theta_1&0&iv_3\\\hline-iv_1&-iv_2&-iv_3&0\end{array}\right), \tag{A5}$$

in which the real angles $\theta_{1,2,3}$ describe spacial rotations and the pure imaginary angles $iv_{1,2,3}$ describe Lorentz boots. The matrix $U(\omega_{\mu\nu})$ has the property $U(\omega_{\mu\nu}{}^*)^{-1}=U(\omega_{\mu\nu})^\dagger$. The generators, $S_{\mu\nu}$, of Lorentz transformations in coordinate space are self-adjoint matrices with matrix elements

$$(S_{\mu\nu})_{\alpha\beta}\equiv i^{-1}(\delta_{\mu\alpha}\delta_{\nu\beta}-\delta_{\mu\beta}\delta_{\nu\alpha}).$$

[1] O. Laporte and G. E. Uhlenbeck, Phys. Rev. **37**, 1380 (1931).
[2] B. L. van der Waerden, *Group Theory and Quantum Mechanics* (Springer, New York, 1974), p. 137.
[3] See Ref. 2 and the review article cited in Ref. 1. Other references on spinor analysis are: W. L. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1952); H. Umezawa, *Quantum Field Theory* (North-Holland, New York, 1956); and E. M. Corson, *Introduction to Tensors, Spinors, and Relativistic Wave-equations* (Hafner, New York, 1953).
[4] H. M. Pilkuhn, *Relativistic Particle Physics* (Springer, New York, 1979), p. 18.
[5] L. Hostler, J. Math. Phys. **21**, 2461 (1980).
[6] R. P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971).
[7] H. Enatsu, A. Takenaka, and M. Okazaki, Nuovo Cimemto **43**, 575 (1978).
[8] K. Fujimura, T. Kobayashi, and M. Namiki, Prog. Theor. Phys. (Kyoto) **44**, 193 (1970).
[9] In the work of W. C. Davidon [Phys. Rev. **97**, 1131 (1955)] a form of Eq. (1.1) has already found application in such a theory.
[10] The analogous algebra for $\bar{\sigma}_{\mu\nu}$ is

$$\bar{\sigma}_{\mu\nu}\bar{\sigma}_{\lambda\rho} = \delta_{\mu\lambda}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\lambda} - \epsilon_{\mu\nu\lambda\rho} + i(\delta_{\mu\lambda}\bar{\sigma}_{\nu\rho} - \delta_{\mu\rho}\bar{\sigma}_{\nu\lambda} + \delta_{\nu\rho}\bar{\sigma}_{\mu\lambda} - \delta_{\nu\lambda}\bar{\sigma}_{\mu\rho}.$$

Products involving one factor τ_μ and one factor of $\sigma_{\mu\nu}$ or $\bar{\sigma}_{\mu\nu}$ can be evaluated using

$$\tau_\mu\sigma_{\nu\rho} = -i\epsilon_{\mu\nu\rho\pi}\tau_\pi - i(\delta_{\mu\nu}\tau_\rho - \delta_{\mu\rho}\tau_\nu)$$

and

$$\bar{\sigma}_{\mu\nu}\tau_\rho = -i\epsilon_{\mu\nu\rho\pi}\tau_\pi + i(\delta_{\rho\mu}\tau_\nu - \delta_{\rho\nu}\tau_\mu).$$

By taking the adjoint, similar identities for $\sigma_{\nu\rho}\bar{\tau}_\mu$ and $\bar{\tau}_\rho\bar{\sigma}_{\mu\nu}$ can be found. The identities noted here and the identities (2.10), (2.11), and (2.14) of Sec. II A1 are the only binary products of the matrices τ_μ, $\bar{\tau}_\mu$, $\sigma_{\mu\nu}$, and $\bar{\sigma}_{\mu\nu}$ that will have a tensor structure in their Lorentz indices. The reason is that only the above mentioned matrix products express the spinor operations of outer product and contraction of spinor indices of the same (dotted or undotted) type.

[11]The notation is that of Julian Schwinger, Proc. Nat. Acad. Sci. **37**, 455 (1951).

[12]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[13]In order to obtain positive frequency free-particle wave functions which correspond to the usual Dirac plane wave states, we start with the wave function for a spin up (down) electron in the comoving frame, and then perform a Lorentz transformation taking us back to a general frame. Positron wave functions $\phi^{3,4}(\mathbf{p},\mathbf{x})$ are obtained from the electron wave functions $\phi^{1,2}(\mathbf{p},\mathbf{x})$ by application of the charge conjugation operation (2.24) $\phi^{3,4}(\mathbf{p},\mathbf{x}) \equiv C[\phi^{1,2}(\mathbf{p},\mathbf{x})]$.
The normalization is

$$(\phi^{\rho_2}(\mathbf{p}_2,\mathbf{x});\phi^{\rho_1}(\mathbf{p}_1,\mathbf{x})) = (2\pi)^3\delta^3(\mathbf{p}_2 - \mathbf{p}_1)\delta_{\rho_2\rho_1}, \qquad \rho_1,\rho_2 = 1, 2, 3, 4;$$

where the unitary dot product is that of Eq. (2.17).

[14]To clarify this point let us examine the effect of substituting the Born expansion for g_+. A typical term involves the factor

$$\langle \mathbf{p}_f|u^{\rho_f}(0)^\dagger(-(p^2+m^2))(B-m)g_0V\cdots Vg_0(A+m)(-(p^2+m^2))u^{\rho_i}(0)|\mathbf{p}_i\rangle$$
$$= \langle \mathbf{p}_f|u^{\rho_f}(0)^\dagger(-(p^2+m^2))[(B_0-m) - e(iA_4 + \boldsymbol{\sigma}\cdot\mathbf{A})]g_0V\cdots Vg_0$$
$$\times[(A_0+m) - e(-iA_4 + \boldsymbol{\sigma}\cdot\mathbf{A})](-(p^2+m^2))u^{\rho_i}(0)|\mathbf{p}_i\rangle.$$

Now we expand each expression in the square brackets into two separate terms, generating four terms in all. In the term

$$\langle \mathbf{p}_f|u^{\rho_f}(0)^\dagger(-(p^2+m^2))(B_0-m)g_0V\cdots Vg_0(A_0+m)$$
$$\times(-(p^2+m^2))u^{\rho_i}(0)|\mathbf{p}_i\rangle$$

the factors $(-(p^2+m^2))$ cancel against corresponding factors of g_0 and we obtain a matrix element

$$u^{\rho_f}(0)^\dagger(B_{0f}-m)\langle \mathbf{p}_f|V\cdots V|\mathbf{p}_i\rangle(A_{oi}+m)u^{\rho_i}(0)$$

that remains finite as $\mathrm{Im}(m)\to 0^-$. The analogous structure

$$\langle \mathbf{p}_f|(iA^4 + \boldsymbol{\sigma}\cdot\mathbf{A})g_0V\cdots V|\mathbf{p}_i\rangle$$

will likewise remain finite as $\mathrm{Im}(m)\to 0^-$.
Consequently the term

$$u^{\rho_f}(0)^\dagger(-(p_f^2+m^2))(-e)\langle \mathbf{p}_f|(iA_4 + \boldsymbol{\sigma}\cdot\mathbf{A})g_0V\cdots V|\mathbf{p}_i\rangle(A_{0i}+m)u^{\rho_i}(0)$$

from the expansion of the bracket is finite as $\mathrm{Im}(m)\to 0^-$ before the factor $(-(p^2+m^2))$ is applied. Because of this finiteness, and because $(-(p^2+m^2))\to 0$ as $\mathrm{Im}(m)\to 0^-$, the whole expression is annulled by the factor $(-(p^2+m^2))$ in the limit $\mathrm{Im}(m)\to 0^-$. Treating the other potential terms from the expansion of the bracket similarly completes the justification of the desired replacements $(B-m)\to(B_0-m)$ and $(A+m)\to(A_0+m)$ in the brackets of Eq. (2.58).

High-energy asymptotic expansion and Regge trajectories for the general even power potential in the Klein–Gordon equation with applications

L. K. Sharma[a)]
Department of Physics, University of Kaiserslautern, 6750 Kaiserslautern, West Germany

V. P. Iyer
Department of Applied Physics, Government Engineering College, Jabalpur-482011, India

(Received 21 July 1981; accepted for publication 21 August 1981)

Large energy asymptotic expansions, wavefunctions, and Regge trajectories of the generalized even power potential $V(r) = -g^2\Sigma_{j=0}^{\infty} N_{2j} r^{2j}$ in the Klein–Gordon equation are obtained. These general expansions are then used to obtain Regge trajectories for the anharmonic oscillator and Gauss potentials. Finally, the charmonium spectroscopy for an harmonic oscillator is studied.

PACS numbers: 11.60. + c

1. INTRODUCTION

With the recent discovery of several new particles[1] interest in hadron spectroscopy has been revived. Many attempts have been made to study the hadron spectrum using the nonrelativistic Schrödinger equation with a linear potential.[2,3] A linear potential is suggested by the gauge theory of quark confinement[4] and by exactly soluble two-dimensional QED.[5] The resulting spectroscopy seems to be quite satisfactory and even has significant advantages over earlier schemes in the case of ordinary mesons and baryons.[3] Because of this success it is important to go beyond the nonrelativistic approach, which must be regarded as a first approximation for a complicated hadronic system.

A complete treatment should incorporate both relativistic and quantum effects, and in addition, requires a full understanding of the underlying dynamics of the quarks. The well-known complexity encountered in the bound state problem in relativistic field theory suggests that the complete solution of this problem is rather remote at the present time.

In a paper by Kang and Schnitzer[6] meson spectra have been calculated using a linear potential function as the fourth component of a four-vector in the Klein–Gordon equation. Gunion and Li[7] have studied the same potential as a Lorentz scalar in both Klein–Gordon and Dirac equations.

A phenomenological view point was adopted by Ram and Halasa[8] to investigate the meson spectra using a harmonic oscillator potential function. A semirelativistic quark model for mesons has also been studied by Horwitz[9] using a square-well potential model.

The main result of the present investigation is the derivation of the high eigenenergy and Regge trajectory expansions for a general even power potential in the Klein–Gordon equation. The general even power potential is of particular interest in potential theory. Since such well-known potentials as the harmonic oscillator, the Gauss potential and the anharmonic oscillator potentials (with even anharmonicities) may be derived from this as particular cases. The general even power potential has already[10] been studied in the Schrödinger frame work using large coupling constants. In the present investigations we use the perturbation technique explained in Ref. 11. In Sec. 2, we derive the large energy expansion and Regge trajectories for the general even power potential. In Sec. 3, we give applications of the general eigenenergy expansion to the Gauss and the anharmonic oscillator potentials. Meson spectra of the charmed quarks is also investigated using the harmonic oscillator potential. Finally in Sec. 4, we give a brief discussion of our results.

2. EIGENENERGIES FOR THE GENERAL EVEN POWER POTENTIAL

We consider the radial part of the Klein–Gordon equation given by

$$\frac{d^2\psi(r)}{dr^2} + \left((E - V)^2 - m^2 - \frac{l(l+1)}{r^2} \right)\psi(r) = 0. \qquad (2.1)$$

($\hbar = c = 1$)

Setting $E^2 = m^2 + k^2$, $K = ik$,

$$Z = -2ikr, \quad \dot{X} = \frac{dX}{dZ} \qquad (2.2)$$

and

$$\psi(r) = e^{ikr} r^{l+1} X(r) \qquad (2.3)$$

we obtain

$$Z\ddot{X} + (2l + 2 - Z)\dot{X} - \{(l+1) - (Z/4K^2)(V^2 - 2EV)\}X = 0. \qquad (2.4)$$

We now assume a potential of the form

$$V(r) = -g^2 \sum_{j=0}^{\infty} N_{2j} r^{2j}, \qquad (2.5)$$

where the coefficients N_{2j} can be positive or negative.

Substituting (2.5) in (2.4) we obtain an equation which may be put in the form

$$Z\ddot{X} + (2l + 2 - z)\dot{X} - (l+1)X = -\sum_{j=0}^{\infty} (g^4 M_j + 2Eg^2 N_{2j}) \left(\frac{z^{2j+1}}{(4K^2)^{j+1}} \right), \qquad (2.6)$$

where

$$M_j = \sum_{i=0}^{j} N_{2(j-i)} N_{2i}. \qquad (2.7)$$

[a)] On leave from Department of Applied Physics, Government-Engineering College, Jabalpur, India.

 0022-2488/82/061185-05$02.50

The right-hand side of Eq. (2.6) is of order $1/K^2$. To a first approximation it may be neglected for $|K^2|\to\infty$. Hence the equation reduces to

$$Z\ddot{X}^{(1)} + (2l+2-z)\dot{X}^{(1)} - (l+1)X^{(1)} = 0. \tag{2.8}$$

A solution of (2.8) is

$$X^{(1)} = \phi(a,b;z), \tag{2.9}$$

where ϕ is a confluent hypergeometric function with

$$a = l+1 \tag{2.10}$$

and

$$b = 2l+2.$$

This solution will be a normalized bound state wavefunction, if

$$a = -n \text{ for } n = 0,\ 1,\ 2,.... \tag{2.11}$$

Hence, in our original problem we may write

$$l+1 = a + \Delta(K^2)/4K^2 = -n + \Delta(K^2)/4K^2, \tag{2.12}$$

where $\Delta(K^2)$ is an (as yet) undetermined expansion in powers of $1/K^2$.

Inserting (2.12) in (2.6) we have

$$D_n X = \left[\Delta(K^2)h - \sum_{j=0}^{\infty}(g^4M_j + 2Eg^2N_{2j})Z^{2j+1}h^{j+1}\right]X, \tag{2.13}$$

where

$$D_n = Z\frac{d^2}{dz^2} + (b-z)\frac{d}{dz} + n \tag{2.14}$$

and

$$h = 1/4K^2. \tag{2.15}$$

As a first approximation to X we have (apart from an overall normalization constant)

$$X = X^{(1)} = \phi_n(z). \tag{2.16}$$

This approximation obviously leaves uncompensated terms on the right-hand side of (2.13) amounting to

$$R^{(1)} = \left[\Delta h - \sum_{j=0}^{\infty}(g^4M_j + 2Eg^2N_{2j})z^{2j+1}h^{j+1}\right]\phi_n(z). \tag{2.17}$$

For convenience we set $\phi_n(z) = \phi(a,b;z) = \phi(a)$ and write the recurrence relation for $\phi(a)$ in the form

$$\begin{aligned} Z\phi(a) = {} & (a,a+1)\phi(a+1) + (a,a)\phi(a) \\ & + (a,a-1)\phi(a-1), \end{aligned} \tag{2.18}$$

where

$$\begin{aligned} &(a,a+1) = a = -n = l+1-\Delta h, \\ &(a,a) = b - 2a = 2(l+n+1) = 2\Delta h, \\ &(a,a-1) = a - b = -(2l+2+n) = n - 2\Delta h. \end{aligned} \tag{2.19}$$

By repeated application of (2.18) we obtain the following general relation:

$$Z^m\phi(a) = \sum_{j=-m}^{m} S_m(a,a+j)\phi(a+j), \tag{2.20}$$

where the coefficients $S_m(a,a+r)$ satisfy the recurrence relation

$$\begin{aligned} S_m(a,a+r) = {} & S_{m-1}(a,a+r-1)(a+r-1,a+r) \\ & + S_{m-1}(a,a+r)(a+r,a+r) \\ & + S_{m-1}(a,a+r+1)(a+r+1,a+r), \end{aligned} \tag{2.21}$$

with

$$S_0(a,a) = 1,$$

all

$$S_0(a,a+i) = 0 \text{ for } i \neq 0 \tag{2.22}$$

and

$$S_m(a,a+r) = 0 \text{ for } |r| > m.$$

The expansion $R^{(1)}$ may now be written

$$R^{(1)} = \sum_{j=0}^{\infty} h^{j+1} \sum_{K=-(2j+1)}^{(2j+1)} [a,a+K]_{j+1}\phi(a+K), \tag{2.23}$$

where

$$[a,a]_1 = \Delta - (g^4M_0 + 2Eg^2N_0)S_1(a,a)$$

and

$$[a,a+K]_{j+1} = -(g^4M_j + 2Eg^2N_{2j})S_{2j+1}(a,a+K), \tag{2.24}$$

for j and K not simultaneously equal to zero.

We now observe that $D_n\phi(a+K) = K\phi(a+K)$ so that a term $\mu\phi(a+K)$ may be removed by adding to $X^{(1)}$ the contribution $(\mu/K)\phi(a+K)$ except of course when $K=0$. Hence the next contribution to $X^{(1)}$ becomes

$$X^{(2)} = \sum_{j=0}^{\infty} h^{j+1} \sum_{K=-(2j+1)}^{2j+1} \frac{[a,a+K]_{j+1}}{K}\phi(a+K). \tag{2.25}$$

This contribution leaves uncompensated a sum of terms $R^{(2)}$ which again lead to $X^{(3)}$. Repeating this process successively and adding these contributions to $X^{(1)}$ we obtain.

$$X = X^{(1)} + X^{(2)} + X^{(3)} + \cdots. \tag{2.26}$$

However, (2.26) will be a solution of our equation only if the sum of all terms containing $\phi(a)$ in $R^{(1)}$, $R^{(2)}$,..., etc., left uncompensated so far is set equal to zero. Thus

$$\begin{aligned} 0 = {} & h[a,a]_1 + h^2\left\{[a,a]_2 + \frac{[a,a+1]_1[a+1,a]_1}{1} - \frac{[a,a-1]_1[a-1,a]_1}{1}\right\} \\ & + h^3\left\{[a,a]_3 + \frac{[a,a+1]_2[a+1,a]_1}{1} + \frac{[a,a-1]_2[a-1,a]_1}{-1}\right. \\ & + \frac{[a,a+1]_1[a+1,a]_2}{1} + \frac{[a,a-1]_1[a-1,a]_2}{-1} + \frac{[a,a+1]_1[a+1,a+1]_1[a+1,a]_1}{1} \\ & \left. + \frac{[a,a-1]_1[a-1,a-1]_1[a-1,a]_1}{1}\right\} + O(h^4). \end{aligned} \tag{2.27}$$

The expansion (2.26) is then an eigensolution and (2.27), the appropriate secular equation which enables us to calculate Δ and hence the eigenenergy. Explicit calculation of terms up to $O(h^3)$ yields the following expression for the eigenenergies for large K^2 or small h

$$
\begin{aligned}
n+l+1=\Delta h=\ & \frac{1}{(2n+1)(2P_0^2-12P_1)+h\{P_2(288n^3+360n^2-232n-200)\}}\,[-1/h^2+2P_0/h \\
& -4P_1(3n^2+3n-1)+2P_1^2+2P_0^2(n^2+n)+h\{P_2(60n^4+120n^3-140n^2-200n+48) \\
& +24nP_1P_0(n^2-2)-4P_0^3(n^2+n-1)\} \\
& +(\Delta h)^2\{8P_1+h[P_2(80n^2+80n+280)+48P_1P_0]\} \\
& +(\Delta h)^3[160P_2h(2n+1)+(\Delta h)^4\cdot 32P_2h]+O(h^2),
\end{aligned}
\tag{2.28}
$$

where

$$P_i=g^4M_i+2Eg^2N_{2i} \quad \text{for } i=0,1,2,3,\cdots,\text{etc.}$$

Expanding the denominator of (2.28) in powers of h and then iterating for Δh we finally obtain

$$
\begin{aligned}
l=\ & -n-1+(1/h^7)[32P_2B^5]-(1/h^6)[32P_2B^5(8P_0+5A)] \\
& +(1/h^5)[32P_2B^5(15A^2+40P_0A+24P_0^2-4x)-B^4D] \\
& +(1/h^4)[32P_2B^5(-35A^3-120P_0A^2-120P_0^2A+20Ax+24P_0x-4a-32P_0^3)+B^4D(6P_0+4A)+8B^3P_1] \\
& +(1/h^3)[32P_2B^5(70A^4+280A^3P_0+360P_0^2A^2-60A^2x-120P_0Ax+20Aa+160P_0^3A \\
& +16P_0^4+6x^2-48P_0^2x+24P_0A)+B^4D(3x-12P_0^2-24AP_0-10A^2)+B^3(-32P_1P_0+c-16AP_1)] \\
& +(1/h^2)[32P_2B^5(-126A^5-560A^4P_0-840A^3P_0^2+140A^3x+360A^2P_0x-60A^2a-480A^2P_0^3-80P_0^4A \\
& -30x^2A+240P_0^2xA-120AP_0a+12ax-24P_0x^2-32P_0^2a+32P_0^3x-18P_0^2a) \\
& +B^4D(8P_0^3+3a-12P_0x-12Ax+48P_0^2A+60P_0A^2+20A^3) \\
& +B^3(4P_0^2b-2bx-4P_0c+8P_0Ab+3A^2b-2Ac)-B] \\
& +(1/h)[32P_2B^5\{210A^6+1008A^5P_0+1680A^4P_0^3-280A^4x-840A^3P_0x \\
& +140A^3a+1120A^3P_0^3+240A^2P_0^4+90A^2x^2-720A^2P_0^2x+360A^2P_0a \\
& -60Ax+120AP_0x^2+160AP_0^2a-160AP_0^3x+90AP_0^2a+(4P_0x-2a)^2+2a^2-16P_0xa \\
& -8P_0a^2+2(4P_0^2-2x)(x^2+4P_0a)\} \\
& +B^4D(120P_0^2x-3x^2-12P_0a-32P_0^3A-12Aa+48P_0Ax+30A^2x-120A^2P_0^2-120A^3P_0-35A^4) \\
& +B^3\{3A^2c-4A^3b-12P_0A^2b+8P_0Ac+(4P_0^2-2x)(c-2Ab)+4P_0x\}+B(2P_0+A)] \\
& +[-330A^7-1680A^6P_0-3024A^5P_0^2+504A^5x+1680A^4P_0x-280A^4a-2240A^4P_0^3 \\
& -560A^3P_0^4-210A^4x^2+1680A^3P_0^2x-840A^3P_0a+60A^2xa-120A^2P_0x^2-480A^2P_0^2a \\
& +30A^2(4P_0^2-2x)(4P_0x-2a)-5A(4P_0x-2a)^2-10Aa^2+80AP_0xa+40P_0Aa^2 \\
& -10A(4P_0^2-2x)(x^2+4P_0a)+4ax(4P_0^2-2x)+2(4P_0x-2a)(x^2+4P_0a) \\
& +B^4D(12P_0^2a+6x^2P_0-6xa-48P_0^2Ax+12Ax^2+48AP_0a+80A^2P_0^3+30A^2a-120A^2P_0x \\
& -60A^3x+240A^3P_0^2+150A^4P_0+56A^5)+B^3\{5A^4b-4A^3c-12A^2P_0c+16A^3P_0b \\
& +(4P_0^2-2x)(3A^2b-2Ac)+4P_0x(c-2Ab)+b(x^2-2a+4P_0a)\} \\
& +B(x-2P_0A-A^2)]+h[495A^8+2640A^7P_0+5040A^6P_0^2-840A^6x \\
& +3024A^4P_0x+504A^5a+4032A^5P_0^3+1120A^4P_0^4+420A^4x^2-3360A^4P_0^2x+1680A^4P_0a-140A^3xa \\
& +280A^3P_0x^2+1120A^3P_0^2a+30A^2(4P_0^2-2x)(x^2+4P_0a)-20Axa(4P_0^2-2x) \\
& -10A(4P_0x-2a)(x^2+4P_0a)+B^4D\{x^3-a^2+12P_0xa-48AP_0^2a-24Ax^2P_0 \\
& +24Axa+120A^2P_0^2x-30A^2x^2-120A^2P_0a-160A^3P_0^3-60A^3a+240A^3P_0x+105A^4x \\
& -420A^4P_0^2-336A^5P_0-84A^6\}-B^3\{5A^4c-6A^5b-20P_0A^4b+16P_0A^3c \\
& +(4P_0^2-2x)(3A^2c-4A^3b)+12P_0A^2bx-8P_0Acx \\
& +2bxa+x^2-2a+4P_0a(c-2Ab)\}+B\{a-Ax-2A^2P_0+A^3\}]+O(h^2),
\end{aligned}
\tag{2.29}
$$

where

$$
\begin{aligned}
& A=[P_2/(2n+1)(2P_0^2-12P_1)][288n^3+360n^2-232n-200-(2n+1)(72P_1P_0+4P_0^3)], \\
& B=[1/(2n+1)(2P_0^2-12P_1)], \quad x=-4P_1(3n^2+3n-1)+2P_1^2+2P_0^2(n^2+n), \\
& c=P_2(80n^2+80n+280)+48P_1P_0, \\
& a=P_2(60n^4+120n^3-140n^2-200n+48)+24nP_1P_0(n^2-2)-4P_0^3(n^2+n-1), \\
& b=8P_1, \\
& D=160P_2(2n+1).
\end{aligned}
\tag{2.30}
$$

Although terms up to $O(h^3)$ have been calculated but for brevity we have mentioned terms only up to $O(h)$ in our expansion (2.29). However, for calculating Regge trajectories and for studying the charmonium spectroscopy for the harmonic oscillator, eigenenergy expansion up to $O(h^3)$ have been used.

3. APPLICATIONS OF THE GENERAL EIGENENERGY EXPANSION

We now apply the eigenenergy expansion (2.28) to the following cases.

A. Gauss potential

The Gauss potential is given by

$$V(r) = -g^2 e^{-\alpha^2 r^2} \tag{3.1}$$

so that

$$N_{2j} = (-1)^j(\alpha^{2j}/j!). \tag{3.2}$$

Hence the eigenenergy expansion is obtained by making the following substitution in Eq. (2.28).

$$\begin{aligned} P_0 &= g^4 + 2Eg^2, \\ P_1 &= 2\alpha^2 g^4 - 2Eg^2\alpha^2, \\ P_2 &= 2\alpha^4 g^4 + Eg^2\alpha^4. \end{aligned} \tag{3.3}$$

The ground state Regge trajectories are shown in Fig. 1.

B. Anharmonic oscillator

We consider the potential

$$V(r) = -g^2(N_0 + N_2 r^2 + N_4 r^4). \tag{3.4}$$

Hence in (2.28) we make the substitutions

$$\begin{aligned} P_0 &= g^4 N_0^2 + 2Eg^2 N_0, \\ P_1 &= 2g^4 N_2 N_0 + 2Eg^2 N_2, \\ P_2 &= g^4(N_2^2 + 2N_0 N_4) + 2Eg^2 N_4. \end{aligned} \tag{3.5}$$

The Regge trajectories are shown in Fig. 2.

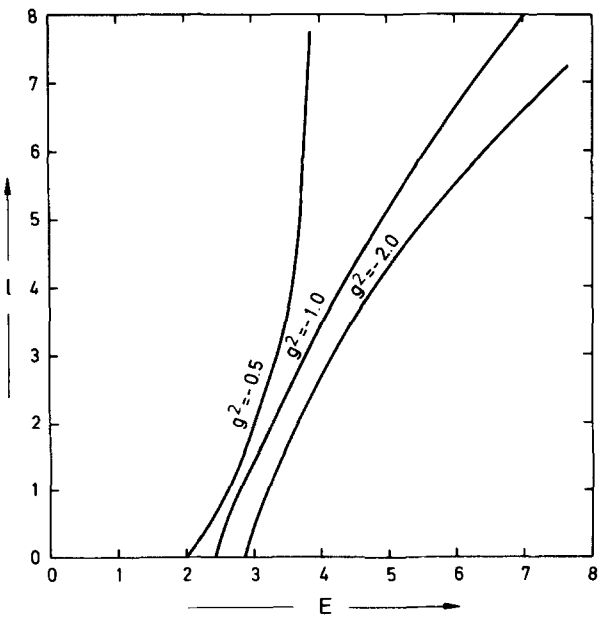


FIG. 1. Ground state Regge trajectories for the Gauss potential with $m = 1$, $\alpha^2 = 1$ and different values of the coupling parameter g^2.

C. Charmonium spectroscopy with a harmonic oscillator

We study the potential

$$V(r) = -g^2(N_0 + N_2 r^2). \tag{3.6}$$

The total energy E in the center of mass of a classical system of a quark and an antiquark interacting by means of this potential is given by[6]

$$E - V = (\mathbf{p}^2 + m_1^2)^{1/2} + (p^2 + m_2^2)^{1/2}, \tag{3.7}$$

where $\mathbf{p}$ is the three momentum of the quark. For equal masses,

$$m_1 = m_2 = m, \tag{3.8}$$

$$\tfrac{1}{4}(E - V)^2 = \mathbf{p}^2 + m^2. \tag{3.9}$$

Making the usual quantum identification we arrive at the Klein–Gordon equation given by

$$[\nabla^2 + \tfrac{1}{4}(E - V)^2 - m^2]\,\psi(r) = 0. \tag{3.10}$$

Hence in our original Eq. (2.2) we make the following alteration

$$E^2/4 = m^2 + k^2. \tag{3.11}$$

So that Eq. (2.15) is now transformed to

$$h = 1/16K^2. \tag{3.12}$$

The eigenenergy expansion is obtained from (2.28) by making the substitutions

$$\begin{aligned} P_0 &= g^4 N_0^2 + 2Eg^2 N_0, \\ P_1 &= 2N_2 N_0 g^4 + 2Eg^2 N_2, \\ P_2 &= g^4 N_2^2. \end{aligned} \tag{3.13}$$

In fact for obtaining the Regge trajectories for the above potentials and for calculating the meson masses it has been found convenient to solve Eq. (2.28) (which is a fourth power

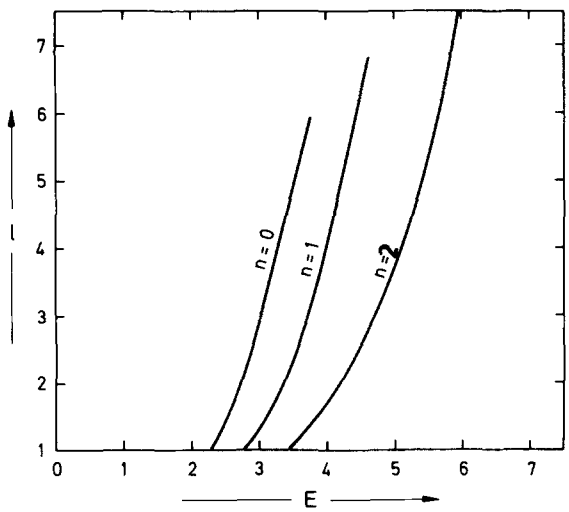


FIG. 2. Regge trajectories for the even power potential ($N_{2j} = 1$) for different values of n. The coupling parameter g^2 has been taken to be -1.

TABLE I. Predicted masses in GeV for mesons with charmed quark pairs with the oscillator potential of parameters $N_0 = -0.002$ GeV, $N_2 = 0.65$ GeV^3, $g = 1$, and $m = 1.5$ GeV.

n \ l	0	1	2	3	4
0	3.095	3.45	3.62	3.94	4.12
1	3.69	3.95	4.21	4.44	4.68
2	4.10	4.31	4.52	4.78	4.83
3	4.4	4.6	4.83	5.01	5.2

equation in Δh) rather than Eq. (2.29).

The meson masses are given in Table I and Regge trajectories for the same are shown in Fig. 3. The predicted masses for mesons with charmed quarks are in agreement with those obtained earlier by Ram *et al.*[8] with the oscillator potential in the Klein–Gordon equation using a numerical method in the WKB approximation.

4. DISCUSSION

It is common knowledge that the nonrelativistic harmonic oscillator leads to rising trajectories. The same view point is established here for a relativistic case also. The Regge trajectories shown in Fig. 3 also appear to be nearly parallel. The meson masses with charmed quarks–antiquarks have also been calculated using this potential. The results are very much in agreement with those obtained by Ram *et al.*[8] who have however used a numerical method for solving the harmonic oscillator potential in the K–G equation using WKB approximation. Although numerical methods have been found to be quite useful in such type of study, it has been thought worthwhile here to obtain theoretical expression for the eigenenergies and wave function using a perturbation theory.

The Regge trajectories for the anharmonic oscillator (with r^4 anharmonicity) and the Gauss potential shown in Figs. 1 and 2, respectively, are also linear in the range of E considered here. The slope of the ground state trajectories decreases as the coupling constant is decreased for a Gauss potential. A similar effect is observed for the Regge trajectories of a general even power potential with increasing n.

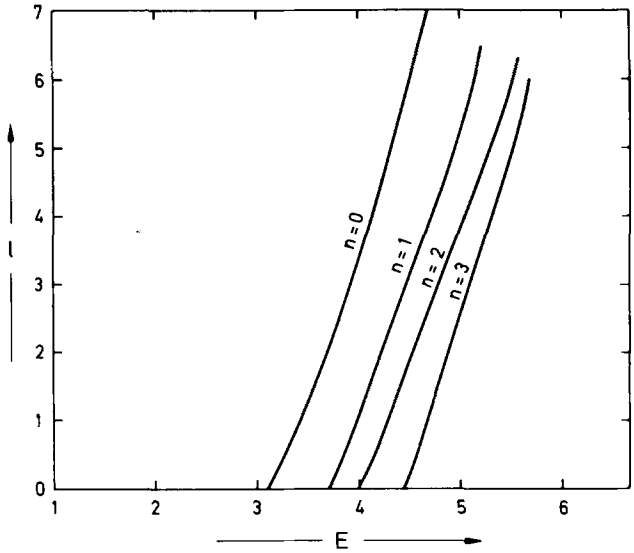


FIG. 3. Regge trajectories for the harmonic oscillator potential in the Klein–Gordon equation with parameters $N_0 = -0.002$ GeV, $N_2 = 0.65$ GeV^3, $g = 1$ and $m = 1$, and $m = 1.5$ GeV.

ACKNOWLEDGMENTS

The authors are indebted to Professor H. J. W. Müller-Kirsten for many helpful suggestions. This work was supported in part by a DAAD visiting fellowship awarded to the first author.

[1]J. J. Aubert *et al.*, Phys. Rev. Lett. **33**, 1404 (1974); J. E. Augustin *et al.*, *ibid.* **33**, 1406 (1974); C. Bacci *et al.*, *ibid.* **33**, 1408 (1974); G. S. Abrams *et al.*, *ibid.* **33**, 1453 (1974).
[2]E. Eichten *et al.*, Phys. Rev. Lett. **34**, 369 (1975); B. J. Harrington, S. Y. Park, and A. Yildiz, *ibid.* **34**, 168 (1975); K. Jhung, K. Cheng, and R. S. Willey, Phys. Rev. D **12**, 1999 (1975).
[3]J. F. Gunion and R. S. Willey, Phys. Rev. D **12**, 174 (1975).
[4]S. Weinberg, Phys. Rev. Lett. **31**, 494 (1973); D. Gross and F. Wilczek, Phys. Rev. D **8**, 3633 (1973); K. Wilson, *ibid.* **10**, 2445 (1974).
[5]J. Kogut and L. Susskind, Phys. Rev. D **9**, 3501 (1974).
[6]J. S. Kang and H. J. Schnitzer, Phys. Rev. D **12**, 841 (1975).
[7]J. F. Gunion and L. F. Li, Phys. Rev. D **12**, 3583 (1975).
[8]B. Ram and R. Halasa, Phys. Rev. D **19**, 3467 (1979).
[9]P. Horwitz, Phys. Rev. **161**, 1415 (1967).
[10]L. K. Sharma, J. Choubey, and H. J. W. Müller-Kirsten, J. Math. Phys. **21** (6), 1533 (1980).
[11]H. J. W. Müller-Kirsten, Phys. Rev. D **22**, 1952 (1980), and other publications by the same author given in the references in that paper.

On the determination of πN phase shifts from isospin constraints and fixed t analyticity

I. Sabba Stefanescu
Institut für Theoretische Kernphysik der Universität Karlsruhe, 75 Karlsruhe 1, West Germany

(Received 30 July 1980; accepted for publication 17 April 1981)

This paper treats the problem of the unique and constructive determination of πN scattering amplitudes by means of fixed t dispersion relations and isospin invariance from data on the differential cross section and polarization of the elastic and charge exchange reactions. The conclusion is that the amplitudes can be uniquely constructed in a neighborhood $(-\alpha, a)$ of $t=0$ if they are continuous functions of both real s and t for s on the cuts and $4\mu^2 > t > -\alpha_0$, $\alpha_0 > \alpha$ and there exists ν_0 such that, for $|\nu| \geqslant \nu_0$, both $|A^+(\nu,t)|/|A^+(\nu,t) + \nu B^+(\nu,t)| < \text{const}$, $t \in (-\alpha_0, 4\mu^2)$ (as indicated by spin rotation measurements) and $|\mathrm{Re}(A^+ + \nu B^+)|/\mathrm{Im}(A^+ + \nu B^+) < \text{const}$, $t \in (-\alpha_0, 0)$. This result is specific for πN scattering, as it depends on a small number of experimental facts concerning this process, which are enumerated in the text.

PACS numbers: 13.75.Gx, 13.85.Dz, 11.20.Fm

I. INTRODUCTION

Over the last years, considerable technical progress has been achieved in phase shift analysis. As a result, one is now able to produce amplitudes which, in contrast to the fixed energy analyses,[1,2] fulfil exactly analyticity constraints in energy, either at fixed momentum transfer[3] (t) or along curves in the s–t space.[4] More precisely, in the special case of the πN phase shift analysis based on fixed t analyticity, which is the object of this paper, one can claim nowadays that one knows to a good approximation a pair of functions $A(s,t)$, $B(s,t)$ with the following properties.

(P1) At fixed t, in some interval $0 \geqslant t \geqslant -t_0$, $A(s,t)$, $B(s,t)$ are real holomorphic in s in the complex s plane cut for $s \geqslant (m+\mu)^2$, $u \geqslant (m+\mu)^2$ (s,t,u are the Mandelstam variables, m = mass of the nucleon, μ = mass of the pion);

(P2) at fixed $s = s_0 + i\epsilon$ (or $u = u_0 + i\epsilon$), the functions $A(s,t)$, $B(s,t)$ are holomorphic in t in an ellipse with foci at the end points of the physical region and extending to $t = 4\mu^2$;

(P3) the partial waves projected out of $A(s,t)$, $B(s,t)$ satisfy the unitarity constraints;

(P4) the observables constructed from these amplitudes agree with their measured values, within experimental errors.

In practice, one has just a large, but finite, set of holomorphic functions of s, $A, B(s,t_j)$, for t_j in an interval $(-t_0, 0)$, $t_0 = 1.5\ \mathrm{GeV}^2$, and a finite set of holomorphic functions of t, $A, B(s_j,t)$ for $s_j \geqslant (m+\mu)^2$ [and $u_j \geqslant (m+\mu)^2$].

A normal question to ask in the face of this achievement is whether the pair of functions satisfying properties (P1)–(P4) is unique, that is, whether $A, B(s,t)$ represent indeed the true amplitudes.

This question has received much attention in connection with the phase shift analyses at fixed energy, where analyticity in s is ignored. The conclusion of these studies was negative; it was, namely, shown that there exists a discrete and a continuum ambiguity in the determination of the phase, even if analyticity in t is taken into account.[5–12] The continuum ambiguity is related to inelastic unitarity; the discrete one can appear in both the elastic and inelastic domains.[13] A detailed practical investigation of these ambiguities has been performed for $\pi^+ p$ scattering.[6,12] Their extent increases, (very) roughly speaking, with the energy.

With few exceptions,[14,15] however, there exist no detailed studies concerning the ambiguities of phase shift analyses, which include analyticity constraints in the energy. The reason for this lies, in the opinion of the author, partly in the fact that it is more difficult to formulate a clear and relevant mathematical problem related to this question than in the fixed-energy case.

Indeed, consider for simplicity a spinless reaction described by an amplitude $A(s,t)$ and assume that $|A(s,t)|$ is known, at fixed $t < 0$ along the whole s and u cuts $(s_{\mathrm{th}}, \infty)$, $(u_{\mathrm{th}}, \infty)$ (including the unphysical region). If one ignores information coming from unitarity, then analyticity in energy, at fixed t, cannot determine the phase of the amplitude $A(s,t)$ from its modulus. There exists, namely a large ambiguity, partly given by products of the type

$$B(z,z_1,\ldots,z_n) = \prod_{i=1}^{n} \frac{z - z_i}{1 - z z_i^*}, \tag{1.1}$$

$$z = \frac{(s_{\mathrm{th}} - s)^{1/2} - (u_{\mathrm{th}} - u)^{1/2}}{(s_{\mathrm{th}} - s)^{1/2} + (u_{\mathrm{th}} - u)^{1/2}}. \tag{1.2}$$

Each factor in (1.1) has unit modulus along the cuts and carries a zero in the cut s plane at the image of z_i through the inverse of (1.2). Even if one invokes vague arguments to fix the number n of such factors, there remains, in principle, a large uncertainty concerning the actual values of the parameters z_i in (1.1).

On the other hand, if one uses unitarity, one concludes that the amplitude can be obtained exactly in a sufficiently small neighborhood of threshold[7,8,16] and therefore that it is known in the whole cut s plane by the uniqueness of analytic continuation. The value of this argument is limited in practice by the fact that the errors of the data render the continuation of the phase to higher energies impossible. So, one would like to have a proof of uniqueness (or a study of ambiguities) which would avoid any steps of analytic continuation.

 0022-2488/82/061190-20$02.50

This requirement is however, too strong, since in order to draw conclusions from fixed t dispersion relations, one must have information on the amplitude in the unphysical region present at $t<0, s\geqslant s_{\rm th}$. One can obtain it by continuing $|A(s,t)|^2$ at fixed energy from its measured values in the physical region to unphysical points by means of

$$M_s(t)\equiv|A(s,t)|^2 = A(s,t)A^*(s,t^*), \tag{1.3}$$

which is holomorphic in t in the same domain as the amplitude. We shall adopt, then, the rules of game of Ref. 14, according to which one takes the modulus of the spinless amplitude as being known (with finite errors) in a domain $D=\{4\mu^2>t>-t_0,\ s,u\geqslant(m+\mu)^2\}$, by means of (1.3) and tries to avoid in the discussion of uniqueness any steps of analytic continuation in energy. We can also modify slightly this assumption[15] in that we can replace the modulus of the amplitude on the unphysical part of the cut, for $t<0$, by its imaginary part, as extrapolated from a supposedly unique phase shift analysis on some interval $(m+\mu)^2<s<s_0$. The larger analyticity domain of the imaginary part makes the extrapolation in the angle practically more reliable.

Studies of ambiguities at fixed energy show that the constraining power of unitarity for the phase decreases with increasing energy. To get a manageable problem for a large range of energies, it appears natural to ignore the exact expression of unitarity, except for its simple consequence, valid at all energies: the positivity of the imaginary part of the partial waves.

Further, in πN scattering, there exist many high precision data not only on the $\pi^{\pm}p$ elastic processes, but also on the charge exchange one $\pi^-p\to\pi^0n$. The amplitudes for the latter processes can be related linearly to those of the elastic ones if we assume invariance of the T matrix elements under rotations in isospin space. For our purpose it is relevant that this linear relation turns out to produce non-negligible constraints on the phase of the amplitudes, if data on all three reactions are taken into account. In fact, we take in Sec. III as a first-order problem that of describing the ambiguities allowed by fixed t analyticity and isospin invariance, with a constraint derived from fixed s analyticity. We show in Secs. IV and V that, under some smoothness assumptions concerning the phase, all the solutions of this problem can be found. We then use the positivity property of the partial waves, experimental information, and some assumptions, which seem acceptable, to show that the remaining ambiguity is actually nonexistent for all t in some neighborhood $(-\alpha,\alpha)$ of $t=0$, $\alpha,a>0$.

So, we achieve "uniqueness", and the next question concerns the stability against errors of the method which is required, according to this study, for the construction of amplitudes from data (differential cross sections, polarizations, and their extensions at fixed s; see Sec. II). In Secs. IV–VII we show that the determination of the amplitudes is equivalent (under the assumptions to which we referred above) to the construction of a meromorphic function in some unbounded domain (the ratio of two polynomially bounded holomorphic functions) from its values (modulus and phase) along the boundary.

This process might still appear to be unstable against small errors, unless one can control the number of zeros and poles of the meromorphic functions to be determined. We shall show, however, in Sec. VIII that a (slight) strengthening of the assumptions we needed to establish uniqueness allows us also to settle the question of stability. This is discussed in detail in Sec. VIII.

To give an idea of the assumptions and experimental facts that are involved, we shall next explain roughly the argument for uniqueness in the simple (and artificial) case of a spinless amplitude $a(s,t)$ with normal thresholds in the s and u channels. We assume $a(s,t)$ satisfies (twice-subtracted) dispersion relations in some interval $-t_0\leqslant t\leqslant t_1$, $t_0,t_1>0$ and wish to reconstruct it from knowledge of

$$|a(\tilde\nu,t)|\equiv f(\tilde\nu,t), \tag{1.4}$$

$$|a^-(\tilde\nu,t)|\equiv|a(\tilde\nu,t)-a(-\tilde\nu,t)|\equiv f_0(\tilde\nu,t), \tag{1.5}$$

where $\tilde\nu=s-u$. Equations (1.4) and (1.5) hold on the cuts $|\tilde\nu|\geqslant\tilde\nu_{\rm th}$, for $-t_0\leqslant t\leqslant t_1$. Equation (1.5) represents the isospin relation: for $\tilde\nu\geqslant\tilde\nu_{\rm th}$ we know the modulus of a linear combination of the amplitudes $a(\tilde\nu,t)$, $a(-\tilde\nu,t)$ (for the s and u channel processes). Clearly, in (1.5), $f_0(\tilde\nu,t)=f_0(-\tilde\nu,t)$. At each $\tilde\nu\geqslant\tilde\nu_{\rm th}$ and t, the function $f_0(\tilde\nu,t)$ is restricted by the inequalities

$$|f(\tilde\nu,t)-f(-\tilde\nu,t)|\leqslant f_0(\tilde\nu,t)\leqslant f(\tilde\nu,t)+f(-\tilde\nu,t), \tag{1.6}$$

i.e., $f_0(\tilde\nu,t), f(\tilde\nu,t), f(-\tilde\nu,t)$ should make up an "isospin triangle." We further assume that the combination $a^+(\tilde\nu,t)=a(\tilde\nu,t)+a(-\tilde\nu,t)$ has positivity properties in $0\leqslant t\leqslant t_1$, and we shall confine ourselves to the construction of a in this interval of t values.

We proceed as follows. From (1.4) and (1.5), we construct along the cuts

$$|a^+(\tilde\nu,t)|^2 = 2(f^2(\tilde\nu,t)+f^2(-\tilde\nu,t))-f_0^2(\tilde\nu,t). \tag{1.7}$$

We can then determine the ratio $R(\tilde\nu,t)=a^-(\tilde\nu,t)/a^+(\tilde\nu,t)$ at each $\tilde\nu$ and t up to a twofold ambiguity. Indeed, dividing (1.5) by (1.4) and squaring, we find

$$\mathrm{Re}\,\frac{a(\tilde\nu,t)}{a(-\tilde\nu,t)}=\frac{1}{2}\,\frac{f^2(\tilde\nu,t)+f^2(-\tilde\nu,t)-f_0^2(\tilde\nu,t)}{f^2(-\tilde\nu,t)} \tag{1.8}$$

and thus we can determine $\mathrm{Im}(a(\tilde\nu,t)/a(-\tilde\nu,t))$ from (1.5) and (1.8) up to a sign ambiguity. A simple calculation shows that this ambiguity is precisely that of the area of the isospin triangle constructed from $f_0(\tilde\nu,t), f(\tilde\nu,t), f(-\tilde\nu,t)$. The area of this triangle is a holomorphic function of t at fixed s, in the same domain as the amplitude. Consequently, its orientation is known by analyticity, once it is known at some t value. At $t=0$, however, the optical theorem (assumed to be valid for this artificial example)

$$\mathrm{Im}a(\tilde\nu,0)\sim\sigma_{\rm tot}(\tilde\nu) \tag{1.9}$$

allows one to fix the amplitude completely,[14] in conjunction with (1.4), and in particular the sign of the isospin triangle. Consequently, we can assume we know $a(\tilde\nu,t)/a(-\tilde\nu,t)$ and thus $R(\tilde\nu,t)=a^-(\tilde\nu,t)/a^+(\tilde\nu,t)$ at all points of the domain $D=\{-t_0\leqslant t\leqslant t_1, |\tilde\nu|\geqslant\tilde\nu_{\rm th}\}$. The poles of $R(\tilde\nu,t)$ coincide with the zeros of $a^+(\tilde\nu,t)$ unless $a^-(\tilde\nu,t)$ happens to have a zero at the same place. Unless such coincident zeros occur, we can determine $a^+(\tilde\nu,t)$ completely from knowledge of the poles of $R(\tilde\nu,t)$.

Indeed, the supposed positivity of the imaginary part of $a^+(\tilde{\nu},t)$ in $t_1 \geqslant t \geqslant 0$ implies that the latter has only a finite number of zeros ($\leqslant 2$) in the complex $\tilde{\nu}$ plane[14] and that it is representable as

$$a^+(\tilde{\nu},t) = B^+(\tilde{\nu},t)E^+(\tilde{\nu},t), \tag{1.10}$$

where $B^+(\tilde{\nu},t)$ is a function of the form (1.1) carrying the zeros of $a^+(\tilde{\nu},t)$ and $E^+(\tilde{\nu},t)$ is a function having no zeros in the cut $\tilde{\nu}$ plane and the modulus of a^+ on the cuts.[17] If we know $a^+(\tilde{\nu},t)$, the other amplitudes can be obtained algebraically.

We conclude that the construction of $a(\tilde{\nu},t)$ is equivalent to the determination of $R(\tilde{\nu},t)$ from its boundary values, provided we can show that no common zeros of a^+, a^- occur. It is possible to show that this occurence is forbidden if the following (virtual) "experimental facts" are verified:

(f1) There exists a domain $D_1 = \{\tilde{\nu}_{\text{th}} \leqslant \tilde{\nu} \leqslant \tilde{\nu}_1, -t_3 \leqslant t \leqslant t_2\}$ so that, for no point $(\tilde{\nu},t)$ in D_1 do $f(\tilde{\nu},t), f(-\tilde{\nu},t), f_0(\tilde{\nu},t)$ vanish simultaneously;

(f2) there exists a constant c, so that $\text{Im}a^+(\tilde{\nu},0) > c\tilde{\nu}$, for $\tilde{\nu} \geqslant \tilde{\nu}_1$ (i.e., $\sigma_{\text{tot}}(\tilde{\nu}) + \sigma_{\text{tot}}(-\tilde{\nu}) > c, \tilde{\nu} \geqslant \tilde{\nu}_1$);

(f3) the amplitude $a(\tilde{\nu},0)$ has no symmetrical zeros in the $\tilde{\nu}$ plane and if it it true that

(a1) the amplitude $a(\tilde{\nu},t)$ is continuous in $\tilde{\nu}$ and t at all points of the region D.

We shall only sketch the reason for this, since the proofs are done for the πN case, from similar facts and assumptions, in the body of the paper. From (f3), one sees that $a^+(\tilde{\nu},0), a^-(\tilde{\nu},0)$ do not have common zeros. So, if there exist any such zeros in the interval $0 \leqslant t \leqslant t_1$, they must disappear throught the cut or move to infinity, as $t \to 0$. However, from (f2) and the positivity of a^+, it is possible to show (see Sec. VI) that $a^+(\tilde{\nu},t)$ cannot vanish for $|\tilde{\nu}|$ sufficiently large in the complex plane. So, the latter possibility is discarded. But, for $|\tilde{\nu}| \geqslant \tilde{\nu}_1, t \geqslant 0$, $\text{Im}a^+(\tilde{\nu},0) \geqslant c\tilde{\nu} > 0$ and so a^+ cannot vanish. For $\tilde{\nu} \leqslant \tilde{\nu}_1$, we can show that (a1) implies (see Sec. VI) that if two coincident zeros reach the cut, then the isospin triangle vanishes at that point. This is in turn forbidden by (f1).

So the poles of $R(\tilde{\nu},t)$ determine *all* the zeros of $a^+(\tilde{\nu},t)$ in $0 \leqslant t \leqslant t_1$ and thus $a^+(\tilde{\nu},t)$ can be constructed from Eq. (1.10) Let us notice that, since in this simple case the number of zeros of $a^+(\tilde{\nu},t)$ cannot be larger than two, no stability problem occurs in the determination of $R(\tilde{\nu},t)$ from its boundary values (see Sec. VIII).

The problem of πN scattering is more complicated than this example. There are two amplitudes to be determined and the extrapolated modulus of the relevant amplitudes does not coincide with the modulus of the extrapolated amplitudes. Further, positivity is known to occur on both s and u cuts for one special combination of amplitudes only.

However, it seems to the author that the same conclusions can be drawn as for this simplified example, with some supplementary assumptions. In Sec. II we introduce the necessary definitions and state the correct mathematical problem for πN scattering. The assumptions and experimental facts [analogous to (a1) and (f1)–(f3)] that are needed to achieve uniqueness will be introduced in the course of the paper and are listed again in the conclusions (Sec. X).

II. NOTATIONS AND STATEMENT OF THE PROBLEM

(a). The notation used in the following coincides with that of Ref. 18.

Consider pion nucleon elastic scattering. Let $p,p',u_\alpha(p),u_\beta(p')$ be the momenta and Dirac spinors for the initial and final states of the nucleon and q,q' be the initial and final pion momenta: $p+q=p'+q'$. The T matrix element between these states can be expressed in terms of two complex amplitudes: $A(s,t), B(s,t)$:

$$T_{\alpha\beta}(p,p') = \bar{u}_\beta(p')\left(A(s,t) + \frac{(q+q')_\mu}{2}\gamma^\mu B(s,t)\right)u_\alpha(p),$$
$$s=(p+q)^2,\ t=(p'-p)^2. \tag{2.1}$$

If exact isospin symmetry holds, the amplitudes for the charge-exchange process are expressed in terms of those for elastic scattering by

$$A_0(s,t) = 2^{-1/2}[A_+(s,t) - A_-(s,t)], \tag{2.2}$$

and similarly for $B_0(s,t)$. The indices $+, -, 0$ stand for π^+p elastic, π^-p elastic, and charge-exchange scattering, respectively. It is useful to define

$$C(s,t) = A + \frac{\nu B}{1 - t/4m^2}, \tag{2.3}$$

with $\nu = (s-u)/4m$.

We need the following analyticity properties of $A(s,t)$, $B(s,t)$: for all t in $|t| < 4\mu^2$, $A(s,t), \nu B(s,t)$ are holomorphic in the s plane cut from $s=(m+\mu)^2$ to the right and $u=(m+\mu)^2$ to the left, apart from a pole at $u=m^2$ in $\nu B(s,t)$ (the nucleon pole), and satisfy twice-subtracted dispersion relations. Further, for each $s \geqslant (m+\mu)^2$ above the cut, the amplitudes are holomorphic in an ellipse (which we call the small Lehmann–Martin ellipse), with foci at $Z_s = \pm 1$ and large semiaxes extending to $Z_{\max} = (1+\mu^2/q_s^2)^{1/2}$ with

$$q_s^2 = [s-(m-\mu)^2][s-(m+\mu)^2]/4s, \quad Z_s = 1 + t/2q_s^2. \tag{2.4}$$

The amplitudes of Ref. 3 are even analytic in an ellipse extending to $Z_{\max} = 1 + 2\mu^2/q_s^2$.

This is also true for $u \geqslant (m+\mu)^2$ above the cut.

$A_\pm, B_\pm$ are related by crossing:

$$A_+(-\tilde{\nu},t) = A_-(\tilde{\nu},t),\ B_+(-\nu,t) = -B_-(\nu,t). \tag{2.5}$$

These properties can be deduced, as is well known, from field theory (Refs. 16 and 19–21). Thus, if analyticity, crossing, and isospin symmetry are taken into account, πN elastic and charge-exchange scattering is completely described by two complex functions, which we can choose to be $A_+(s,t), B_+(s,t)$.

(b). Information about A_+, B_+ comes from measurements of the differential cross sections $(d\sigma/d\Omega)_{\pm,0}$ and polarizations $P_{\pm,0}$. We regard these as "the data" and ignore (until Sec. VII) any information coming from the few spin rotation measurements. The relation of observables to amplitudes is most easily expressed by means of the functions $h_\pm, g_\pm$ of Ref. 3, defined by

$$h_\pm = (4m^2 - t)^{1/2}C_\pm + [t((m^2-\mu^2)^2 - su)]^{1/2}/(4m^2-t)^{1/2}B_\pm, \tag{2.6a}$$

$$g_{\pm} = (4m^2 - t)^{1/2} C_{\pm} - [t((m^2 - \mu^2)^2 - su)]^{1/2}/(4m^2 - t)^{1/2} B_{\pm}. \quad (2.6b)$$

The square root in (2.6) is defined so that, for $t \leqslant 0$, it has a cut for $|\nu| > \bar{\nu}(t)$, where

$$\bar{\nu}(t) = ((\mu^2 - t/4)(1 - t/4m^2))^{1/2}, \quad (2.7)$$

and is positive for $|\nu| < \bar{\nu}(t)$. The physical region (in which measurements can be performed) extends for $|\nu| \geqslant \bar{\nu}(t)$, at $t \leqslant 0$. The relation of $h_{\pm}, g_{\pm}$ to the data is given[3] for $t \leqslant 0$ and $\nu \geqslant \bar{\nu}(t)$ by

$$|h_{\pm}(\nu,t)|^2 = 64\pi^2 s (1 + P_{\pm}) d\sigma_{\pm}/d\Omega, \quad (2.8a)$$

$$|g_{\pm}(\nu,t)|^2 = 64\pi^2 s (1 - P_{\pm}) d\sigma_{\pm}/d\Omega. \quad (2.8b)$$

Because of Eq. (2.5) it is true that

$$h_{\pm}(-\nu,t) = g_{\mp}(\nu,t). \quad (2.9a)$$

Therefore, for $t \leqslant 0$, we can take $h_{\pm}(\nu,t)$ as the two amplitudes to be determined and define

$$|h_{+}(\nu,t)| = f_{+}(\nu,t), \quad (2.10a)$$

$$|h_{-}(\nu,t)| = f_{-}(\nu,t), \quad (2.10b)$$

for all ν, with $|\nu| \geqslant \bar{\nu}(t)$. The function f_{+}, f_{-} are known by (2.8). We define, further

$$h_0(\nu,t) = h_{+}(\nu,t) - h_{-}(\nu,t), \quad (2.11a)$$

$$g_0(\nu,t) = g_{+}(\nu,t) - g_{-}(\nu,t). \quad (2.11b)$$

It follows that

$$h_0(-\nu,t) = -g_0(\nu,t). \quad (2.12)$$

It is also true that

$$|h_0(\nu,t)|^2 = 128\pi^2 s(1 + P_0) d\sigma_0/d\Omega \equiv f_0^2(\nu,t), \quad (2.13a)$$

$$|h_0(-\nu,t)|^2 = 128\pi^2 s(1 - P_0) d\sigma_0/d\Omega \equiv f_0^2(-\nu,t), \quad (2.13b)$$

$\nu \geqslant \nu(t)$. We assume that data are available for all $\nu \geqslant \nu(t)$, and for all t in an interval $0 \geqslant t \geqslant -t_0$, for some t_0.

(c). We next discuss the information that can be obtained on the unphysical part of the cut, for $t < 0$, by extrapolation at fixed energy from the data available in the physical region.

The definition we chose in the previous paragraph for the root in Eq. (2.6), namely with cuts in s, whose location depends on t, is awkward for extrapolation at fixed s to complex or unphysical t values. We choose another definition for real $s > (m + \mu)^2$, by letting the cut lie across the physical region $|Z_s| \leqslant 1$ and the root be negative imaginary above it. [We can extend by continuity this definition of the root to other real (and complex s), by letting a cut run in the t plane between $t = 0$ and $t = (2s(m^2 + \mu^2) - s^2 - (m^2 - \mu^2)^2)/s$. The surfaces of discontinuity of the root in Eq. (2.6) that are thus obtained do not coincide with those one gets by uniformizing the two-sheeted function of s at each point in the complex t plane. Hence the different notations h vs $\tilde{h}$.] With this new choice for the root, we define functions $\tilde{h}_{\pm,0}(s,t), \tilde{g}_{\pm,0}(s,t)$ by the same formulas as h and g in (2.6). These functions have two sheets in t at each fixed s and have the property that, for t in the physical region of the s channel,

$$\tilde{h}_{\pm}(s,t + i0) = h_{\pm}(\nu + i0,t) = \tilde{g}_{\pm}(s,t - i0), \quad (2.14a)$$

$$\tilde{g}_{\pm}(s,t + i0) = g_{\pm}(\nu + i0,t) = \tilde{h}_{\pm}(s,t - i0), \quad (2.14a)$$

For $t < 0$ and s in the unphysical region, $\nu_{\text{th}}(t) \leqslant \nu \leqslant \bar{\nu}(t)$, one gets

$$\tilde{h}_{\pm}(s,t) = h_{\pm}(\nu,t), \quad (2.15a)$$

$$\tilde{g}_{\pm}(s,t) = g_{\pm}(\nu,t). \quad (2.15b)$$

Apart from the cut, for $|Z_s| \leqslant 1, \tilde{h}_{\pm,0}\tilde{g}_{\pm,0}$ are holomorphic in the Martin ellipse. The two-sheeted Z_s plane can be uniformized by means of the variable

$$w_s = Z_s + (Z_s^2 - 1)^{1/2} \quad (2.16)$$

so that the upper and lower lips of the cut on the first sheet come onto the upper and lower halves of the unit circle in the w_s plane. We call $U_{+}(w_s)$ the unique function representing $\tilde{h}_{+}, \tilde{g}_{+}$ in the w_s plane. The sheets I and II are indexed so that $U_{+}(t_{\text{I}}) = \tilde{h}_{+}$, $U_{+}(t_{\text{II}}) = \tilde{g}_{+}$. From (2.10) and (2.14) we know the modulus of $U_{+}(w_s)$ for $|w_s| = 1$

$$|U_{+}(w_s)|^2 = U_{+}(w_s)U_{+}^{*}(w_s) = U_{+}(w_s)U_{+}^{*}\left(\frac{1}{w_s^*}\right) = \begin{cases} |h_{+}|^2, & 0 \leqslant \arg w_s \leqslant \pi \\ |g_{+}|^2, & \pi \leqslant \arg w_s \leqslant 2\pi \end{cases} \quad (2.17)$$

The function $U(w_s)$ is holomorphic in a ring:

$$1/r_s \leqslant |w_s| \leqslant r_s, \quad r_s = \mu/q_s + (1 + \mu^2/q_s^2)^{1/2}, \quad (2.18)$$

which is the image of the small Martin ellipse lying on the two sheets of the Z_s plane. Clearly, $U_{+}^{*}(1/w_s^*)$ is holomorphic in the same domain, so that (2.17) can be extended analytically to $1/r_s \leqslant |w_s| \leqslant r_s$. We are interested in the points with w_s real, outside the physical region. We do not obtain by analytic continuation of $|U_{+}(w_s)|^2$ the modulus of the amplitude at these points, but rather the complex values $K_{+}(\nu,t)$ of the combination (for $|w_s| > 1$):

$$K_{+}(\nu,t) \equiv K_{+}(s,w_s) = U_{+}(w_s)U_{+}^{*}\left(\frac{1}{w_s}\right) = \tilde{h}_{+}(s,Z_s)\tilde{g}_{+}^{*}(s,Z_s) = h_{+}(\nu,t)g_{+}^{*}(\nu,t). \quad (2.19)$$

Equation (2.19) follows from the fact that the points w_s, and $1/w_s$ on the real axis correspond, by (2.16), to the same value of Z_s, $|Z_s| > 1$, but on different sheets of (2.16). If we had chosen $w_s < 1$ in (2.19), we would have obtained $K_{+}^{*}(\nu,t)$.

We conclude that in the unphysical region of the cuts at $t < 0, \mu + t/4m \leqslant \nu \leqslant \bar{\nu}(t)$, we have the following information concerning h_{+}, h_{-}:

$$h_{+}(\nu,t)g_{+}^{*}(\nu,t) = h_{+}(\nu,t)h_{-}^{*}(-\nu,t) = K_{+}(\nu,t), \quad (2.20a)$$

$$h_{-}(\nu,t)g_{-}^{*}(\nu,t) = h_{-}(\nu,t)h_{+}^{*}(-\nu,t) = K_{-}(\nu,t), \quad (2.20b)$$

$$h_0(\nu,t)g_0^{*}(\nu,t) = -h_0(\nu,t)h_0^{*}(-\nu,t) = K_0(\nu,t), \quad (2.21)$$

where use has been made of the crossing relations (2.9) and (2.12). Using the isospin relation (2.11) we can write (2.21) as

$$h_{+}(\nu,t)h_{+}^{*}(-\nu,t) + h_{-}(\nu,t)h_{-}^{*}(-\nu,t) = K_{-}(\nu,t) + K_{+}(\nu,t) - K_0(\nu,t) \equiv K(\nu,t). \quad (2.22)$$

Equations (2.20) and (2.22) represent the information on $h_{+}(\nu,t), h_{-}(\nu,t)$ on the unphysical part of the cut, for $t < 0$.

(d). Following Ref. 14 and the discussion of the Introduction, we now formulate the problem of phase shift analy-

sis for $0 < t \leqslant 4\mu^2$.

Clearly, information on the amplitude in this interval is obtained by extrapolating the moduli of $h_{\pm,0}, g_{\pm,0}$, known in the physical region [Eq. (2.10–(2.13)]. We define $h(\nu,t), g(\nu,t)$ for $t \geqslant 0$ by the same formula (2.6) where we now choose the cut of the square root to lie along $|\nu| \leqslant \bar{\nu}(t)$ and be negative for $\nu \geqslant \bar{\nu}(t)$. This is not what one would obtain by continuing the definition for $t < 0$, through a circuit of complex t values around the origin but has the merit that it respects Eq. (2.15). (We do not introduce new symbols for h,g in $t > 0$). With this new choice of the cut, the crossing relations read

$$h_+(-\nu,t) = h_-(\nu,t), \tag{2.23a}$$

$$g_+(-\nu,t) = g_-(\nu,t). \tag{2.23b}$$

We call $H(\nu,t)$ the two-sheeted function which is equal to $h_+(\nu,t)$ on the first sheet of the ν plane and to $g_+(\nu,t)$ on the second sheet. Clearly, apart from the cut along $(-\bar{\nu}(t),\bar{\nu}(t))$, $H(\nu,t)$ has cuts on both sheets, running at $|\nu| \geqslant \mu + t/4m$. By extrapolating as in the preceding paragraph the moduli of $h_{\pm,0}, g_{\pm,0}$, we conclude that our knowledge of the function $H(\nu,t)$ for $\nu \geqslant \mu + t/4m$ above the cut is summarized by

$$\begin{aligned} h_+(\nu,t)g^*_+(\nu,t) &= H(\nu_{\rm I},t)H^*(\nu_{\rm II},t) \\ &= \bar{K}_+(\nu,t), \end{aligned} \tag{2.24a}$$

$$\begin{aligned} h_-(\nu,t)g^*_-(\nu,t) &= H(-\nu_{\rm I},t)H^*(-\nu_{\rm II},t) \\ &= \bar{K}_-(\nu,t), \end{aligned} \tag{2.24b}$$

$$h_0(\nu,t)g^*_0(\nu,t) = \bar{K}_0(\nu,t). \tag{2.25}$$

By using (2.11) we reformulate (2.25) as

$$\begin{aligned} &h_-(\nu,t)g^*_+(\nu,t) + h_+(\nu,t)g^*_-(\nu,t) \\ &\quad \equiv H(-\nu_{\rm I},t)H^*(\nu_{\rm II},t) + H(\nu_{\rm I},t)H^*(-\nu_{\rm II},t) \\ &\quad = \bar{K}_+(\nu,t) + \bar{K}_-(\nu,t) - \bar{K}_0(\nu,t) \\ &\quad = \bar{K}(\nu,t). \end{aligned} \tag{2.26}$$

Equations (2.24) and (2.26) summarize our knowledge about $H(\nu,t)$ for $t \geqslant 0$.

(e). Independent information on the amplitude at $t = 0$ comes from measurements of $\sigma_{\rm tot}(\pi^\pm p)$, by the optical theorem

$$\mathrm{Im}C_\pm(s) = (q_s s^{1/2}/m)\sigma_{\rm tot}(\pi^\pm p). \tag{2.27}$$

We assume $\sigma_{\rm tot}$ is known at all energies. As a consequence, it can be shown that, if (2.10) is used at $t = 0$, $C_\pm(\nu, t = 0)$ is a known complex function of ν. A proof of this is given in Ref. 14, Sec. 4.

(f). We are now in a position to state the first-order problem mentioned in the Introduction, concerning the ambiguities allowed by fixed t analyticity and the isospin constraint. The statement of the problem is different, according to whether $t > 0$ or $t \leqslant 0$.

Problem I. $t > 0$; determine all functions $H(\nu,t)$, holomorphic in the two-sheeted ν plane [cut along $-\bar{\nu}(t),\bar{\nu}(t)$], except for cuts extending on the two sheets for $|\nu| \geqslant \mu + t/4m$, and poles at $u = m^2$ on both sheets, polynomially bounded as $|\nu| \to \infty$ (on both sheets), and satisfying Eqs. (2.24) and (2.26) for $\nu > \mu + t/4m$.

Problem II. $t < 0$; determine all pairs of functions $(h_+(\nu,t), h_-(\nu,t))$, holomorphic in the ν plane cut for $|\nu| \geqslant \mu + t/4m$, with a pole at $u = m^2$, polynomially bounded as $|\nu| \to \infty$, and which satisfy Eqs. (2.10) and (2.13) for $|\nu| \geqslant \bar{\nu}(t)$ and Eqs. (2.20) and (2.22) for $\mu + t/4m \leqslant |\nu| \leqslant \bar{\nu}(t)$.

In formulating these problems, we have ignored the constraint of fixed s analyticity. In Sec. III, we show how we can take it partly into account, by altering slightly the formulation of (I) and (II).

(g). Before proceeding to the solutions of these problems, we make an assumption right now, which will be constantly used in the following

(A1): the amplitudes $A_+(s,t), B_+(s,t)$ describing the scattering are continuous functions of both variables s,t in the region $D = \{4\mu^2 > t > -t_0, s, u \geqslant (m+\mu)^2\}$ (this is the region in which we assume information on the amplitudes is available). Further, we assume there exist constants K,N, so that $\mathrm{Im}|A_+(s,t)|, \mathrm{Im}|B_+(s,t)| < K|s|^N$, for s,t in D.

Clearly, this assumption cannot be tested experimentally. One can only state that phase-shift analysis would not be practically feasible if continuity in s and t were not assumed for the observables, since the latter must currently be displaced by small amounts in energy and angle to form bins, on which statistical averages can be performed. This assumption does not imply (A1) but makes it reasonable.

As a first consequence of (A1), the functions $K_{\pm,0}, K$ in Eqs. (2.20)–(2.22) are continuous functions of ν and t, defined on $\mu + t/4m \leqslant \nu \leqslant \bar{\nu}(t)$, $-t_0 \leqslant t \leqslant 0$. Also, the functions $\bar{K}_{\pm,0}, \bar{K}$ in (2.24)–(2.26) are continuous in both ν and t on their domain of definition.

III. THE ISOSPIN MAP

It is well known that the isospin constraint (2.11) imposes restrictions on the possible values of $(d\sigma/d\Omega)_{\pm,0}, P_{\pm,0}$ at fixed energy and angle. These are usually[22–24] expressed by the requirement that the square of the areas of the triangles constructed with $(1+P)(d\sigma/d\Omega)_{\pm,0}, (1-P)(d\sigma/d\Omega)_{\pm,0}$ be positive for all ν,t. With the notation of (2.10) and (2.13), this constraint on the area can be expressed for all ν real, $|\nu| \geqslant \bar{\nu}(t), t \leqslant 0$ by

$$\begin{aligned} 16\bar{F}^2(\nu,t) \equiv{}& 4f^2_+(\nu,t)f^2_-(\nu,t) \\ &- (f^2_0(\nu,t) - f^2_+(\nu,t) - f^2_-(\nu,t))^2 \geqslant 0. \end{aligned} \tag{3.1}$$

The function $\bar{F}(\nu,t)$ is the area of the triangle constructed with $(1+P)d\sigma/d\Omega$ for $\nu \geqslant \bar{\nu}(t)$, and the area of the triangle constructed with $(1-P)d\sigma/d\Omega$ for $\nu \leqslant -\bar{\nu}(t)$.

Equivalently, we may consider the same function at fixed s, using Eq. (2.14):

$$\begin{aligned} F(s,t) &= \tfrac{1}{2}\,\mathrm{Im}(\tilde{h}_+\tilde{h}^*_-) = \tfrac{1}{2}|\tilde{h}_+||\tilde{h}_-|\sin\theta_{h_+h_-} \\ &= (1/4i)(\tilde{h}_+(s,t)\tilde{h}^*_-(s,t) - \tilde{h}_-(s,t)\tilde{h}^*_+(s,t)), \end{aligned} \tag{3.2}$$

where $\theta_{h_+h_-}$ is the angle between the complex numbers $\tilde{h}_+, \tilde{h}_-$. From the analyticity properties of $\tilde{h}(s,t)$ at fixed s, we see that $F(s,t)$ can be extended analytically to the small Lehmann ellipse, with a square root cut for $|Z_s| \leqslant 1$. The continuation of $F(s,t)$ through the cut is obtained by replacing h by g in (3.2) and is again analytically extendable to the Lehmann–Martin ellipse. Using the variable w_s, Eq. (2.14), we can write for all w_s in a ring $1/r_s < |w_s| < r_s$ with r_s given by (2.16):

$$F(s,w_s) = (1/4i)[U_+(s,w_s)U^*_-(s,1/w_s^*) - U_-(s,w_s)U^*_+(s,1/w_s^*)], \tag{3.3}$$

where $U(s,w_s)$ is the function introduced in (2.15). Notice $F(s,w_s)$ is real along $|w_s| = 1$.

Clearly, the values of $F(s,w_s)$ are known along $|w_s| = 1$ from the measured quantities f_+, f_-, f_0 up to a sign, by Eq. (3.1) [recall Eq. (2.14)]. Further, if the sign of $F(s,w_s)$ is known along some angular interval $\theta_1 \leqslant \arg w_s \leqslant \theta_2$, it is known everywhere on $|w_s| = 1$ by analytic continuation. We point out that the determination of $F(s,w_s$ from its values on the interval (θ_1,θ_2) is not a problem of numerical analytic continuation, but one of finding the unique function that is holomorphic in the ring $1/r_s < |w_s| < r_s$ in the finite set of functions obtained by assigning signs in all possible ways to $|F(s,w_s)|$ outside the interval (θ_1,θ_2). [Since $F(s,w_s)$ is holomorphic in $1/r_s < |w_s| < r_s$, it can change sign only a finite number of times, at fixed s.] This selection can be easily done in practice (if data are available on the whole circle $|w_s| = 1$) and we show an example of this in Sec. IX.

So, we can assume that at each $s > (m+\mu)^2$ (or at least where data exist over the whole angular range), we know $F(s,w_s)$ up to an overall sign ambiguity. But, according to (A1), $F(s,w_s)$ is a continuous function of s, so that its sign cannot change arbitrarily as we move in s, for $|w_s| = 1$.

In fact, maps of the lines where $F(s,w_s) = 0$ have been drawn for a long time[25] in connection with the saturation of the isospin bounds. Their up-to-date shape can be found in Ref. 18. From these maps, we can take the following to be true.

(F0′). There is no energy, so that the isospin triangles are simultaneously degenerate (with zero area), along the whole angular interval. Because of (F0′), we can determine the sign of $F(s,w_s)$ by analyticity at fixed s, at all points where the whole angular region is covered by data, from knowledge of the sign at just one point. We can choose the latter to lie in the forward direction, since $C_+(\nu,t=0)$ is completely known. It is indeed true that:

(F0″). There exists ν_1, so that the phase of $C_+(\nu_1,t=0)$ is different (mod π) from the phase of $C_+(-\nu_1,t=0)$. This means that there exist energies where the isospin triangle constructed from the forward amplitudes is not degenerate. We conclude that, with (A1), (F0′), and (F0″), we can assume the function $F(s,w_s)$ to be known at all energies, where data exist on the whole angular region. It turns out that, at higher energies, where data are limited to an interval of low t, one can nevertheless fix the sign of the triangles by continuity from low s (see the figures of Ref. 18). We shall take it from now on as a fact that:

(F0). The function $F(s,w_s)$ is known over the whole domain $s \geqslant (m+\mu)^2, 4\mu^2 > t \geqslant -t_0$. We now alter slightly the formulations of Problems I and II in that we ask only for that subset of their solutions for which the area $\bar{F}(\nu,t)$ is also given. More precisely, this means that, for problem I, the solutions should satisfy

$$(1/4i)[H(\nu_1,t)H^*(-\nu_{II},t) - H(-\nu_1,t)H^*(\nu_{II},t)] \equiv \bar{F}(\nu,t) = F(s,w_s), \tag{3.4}$$

[where we have used the crossing relations (2.23)].

For Problem II, knowledge of $F(s,w_s)$ means knowledge of the combination

$$(1/4i)[h_+(\nu,t)h^*_+(-\nu,t) - h_-(\nu,t)h_-^*(-\nu,t)] \equiv \bar{F}(\nu,t) = F(s,w_s), \tag{3.5}$$

for $\bar{\nu}(t) \geqslant |\nu| \geqslant \mu + t/4m$ [we have used the crossing relation (2.9)], and of

$$(1/4i)[h_+(\nu,t)h^*_-(-\nu,t) - h_-(\nu,t)h^*_+(\nu,t)] = \bar{F}(\nu,t), \tag{3.6}$$

for $|\nu| \geqslant \bar{\nu}(t)$.

Equations (3.4)–(3.6) are the only way we make use at this stage of analyticity in t at fixed s.

IV. THE INTERVAL $0 \leqslant t \leqslant 4\mu^2$

We now turn to the explicit solution of Problem I, with the restriction (3.4). It is convenient to introduce the uniformizing variable

$$\omega_t(\nu) = \nu/\bar{\nu}(t) + (\nu^2/\bar{\nu}^2(t) - 1)^{1/2}, \tag{4.1}$$

which maps the two sheeted ν plane onto the whole ω plane in such a way that the upper lip of the cut $|\nu| \leqslant \bar{\nu}(t)$ comes onto the semicircle $|\omega_t| = 1$ lying in the upper half-plane. The second sheet lies inside the circle. Points that lie at the same position on the two sheets of the ν plane have their images in the relation

$$\omega_I(\nu)\omega_{II}(\nu) = 1. \tag{4.2}$$

There are two nucleon poles lying at

$$\omega_{B\,1,2} = \nu_B(t)/\bar{\nu}(t) \pm i(1 - \nu_B^2(t)/\bar{\nu}^2(t))^{1/2}, \tag{4.3}$$

where $\nu_B = (2\mu^2 - t)/4m$ ($|\omega_{B\,1,2}| = 1$).

We wish to determine all functions $H(\omega)$, holomorphic in the ω plane cut for $|\omega| \geqslant \omega(\nu_{th}) \equiv \omega_{th}(\nu_{th} \equiv \mu + t/4m)$ and $|\omega| \leqslant 1/\omega_{th}$, with simple poles at $\omega_{B\,1,2}$, polynomially bounded in ω, behaving like $1/\omega^n$ near $\omega = 0$, for some $n > 0$, and obeying for $\omega \geqslant \omega_{th}$ (above the cut).

$$H(\omega)H^*(1/\omega) = \bar{K}_+(\omega), \tag{4.4}$$

$$H(-\omega)H^*(-1/\omega) = \bar{K}_-(\omega), \tag{4.5}$$

$$H(-\omega)H^*(1/\omega) + H(\omega)H^*(-1/\omega) = \bar{K}(\omega). \tag{4.6}$$

We divide (4.4) and (4.6) by (4.5) and get

$$\frac{H(\omega)}{H(-\omega)}\frac{H^*(1/\omega)}{H^*(-1/\omega)} = \frac{\bar{K}_+(\omega)}{\bar{K}_-(\omega)}, \tag{4.7}$$

$$\frac{H(\omega)}{H(-\omega)} + \frac{H^*(1/\omega)}{H^*(-1/\omega)} = \frac{\bar{K}(\omega)}{\bar{K}_-(\omega)}. \tag{4.8}$$

From Eqs. (4.7) and (4.8) we can determine at each $\omega \geqslant \omega_{th}$ the values of $x_1 = H(\omega)/H(-\omega), x_2 = H^*(1/\omega)/H^*(-1/\omega)$ up to a twofold ambiguity (namely the assignment of the roots) by solving the equation

$$x^2\bar{K}_-(\omega) - x\bar{K}(\omega) + \bar{K}_+(\omega) = 0, \tag{4.9}$$

$$x_{1,2} = \frac{\bar{K}_- + \bar{K}_+ - \bar{K}_0}{2\bar{K}_-} \pm [(\bar{K}_- + \bar{K}_+ - \bar{K}_0)^2 - 4\bar{K}_+\bar{K}_-]^{1/2}/2\bar{K}_-. \tag{4.10}$$

Expressing (3.4) in the variable ω, one can check that the following identities, inspired by (4.10) and (3.1) hold:

$$2\frac{H(\omega)}{H(-\omega)}\bar{K}_-(\omega)-\bar{K}(\omega)=4i\bar{F}(\omega,t), \tag{4.11}$$

$$2\frac{H^*(1/\omega)}{H^*(-1/\omega)}\bar{K}_-(\omega)-\bar{K}(\omega)=-4i\bar{F}(\omega,t). \tag{4.12}$$

Expressions (4.11) and (4.12) [which we could have written without resorting to (4.9) and (4.10)] show that $r(\omega)\equiv H(\omega)/H(-\omega)$ is fixed for all $\omega\in(0,1/\omega_{\rm th})\cup(\omega_{\rm th},\infty)$. We next define

$$S(\omega)=H(\omega)+H(-\omega)=S(\omega^2), \tag{4.13a}$$

$$A(\omega)=H(\omega)-H(-\omega). \tag{4.13b}$$

The ratio $A(\omega)/S(\omega)$ is known on the cuts at the same time with $r(\omega)=H(\omega)/H(-\omega)$.

$$\frac{A(\omega)}{S(\omega)}=R(\omega)=\frac{r(\omega)-1}{r(\omega)+1}. \tag{4.14}$$

Clearly, $R(-\omega)=-R(\omega)$. All the zeros of $S(\omega)$, which do not have the same position as the zeros of $A(\omega)$, are to be found among the poles of $R(\omega)$. We next use Eq. (4.4) to determine $S(\omega)$,

$$\begin{aligned}\bar{K}_+(\omega)&=H(\omega)H^*(1/\omega)\\&=S(\omega)S^*(1/\omega)(1+R(\omega))(1+R^*(1/\omega)).\end{aligned} \tag{4.15}$$

Defining

$$\begin{aligned}\kappa(\omega^2)&=\frac{\bar{K}_+(\omega)}{(1+R(\omega))(1+R^*(1/\omega))}\\&=\bar{K}_+(\omega)+\bar{K}_-(\omega)+\bar{K}(\omega)\end{aligned} \tag{4.16}$$

[which we can also obtain by addition of (4.4), (4.5), and (4.6)], for $\omega\geqslant\omega_{\rm th}$, we have reduced the problem to the following:

Find all functions $S(\omega^2)$, real holomorphic in the ω^2 plane cut along $(0,1/\omega_{\rm th}^2)\cup(\omega_{\rm th}^2,\infty)$ with two simple poles at $\omega_{B\,1,2}^2$, polynomially bounded in ω^2, behaving like $1/\omega^n$, $n>0$, at $\omega^2=0$ and satisfying for $\omega^2\geqslant\omega_{\rm th}^2$ on the cut

$$S(\omega^2)S^*(1/\omega^2)=\kappa(\omega^2). \tag{4.17}$$

We are interested in the subset of solutions $S(\omega^2)$ which vanish at the positions of the poles of $R(\omega)$. From now on we shall make the following assumption

(A2) The phase $\phi(\nu,t)$ of $S(\nu,t)$ can be defined at fixed t along the cuts, on both sheets of the ν plane, by continuity from threshold, except for a finite number of points, where it has finite discontinuities. The magnitude of the latter is uniquely determined by small excursions in the complex ν plane (we assume no accumulation of zeros occurs at points of the cut). Moreover, we assume for simplicity that $\phi(\nu,t)$ approaches a finite limit $\phi_0(t)$, as $\nu\to\infty$, in such a way that $\nu^\beta(\phi(\nu,t)-\phi_0(t))\to0$, for some $\beta>0$, and that it is Hölder continuous on its intervals of continuity.

It is easy to see that any solution of (4.17) leads to a solution of our problem, provided it has zeros at the poles of $R(\omega)$. Indeed, from (4.14) we can then find $A(\omega)$ and we know that it is holomorphic in the correct domain. We then construct $H(\omega)$, $H(-\omega)$ and verify that all Eqs. (4.4)–(4.6) are satisfied.

It is convenient to move over to the variables

$$\zeta=\lambda\frac{\omega^2-1}{\omega^2+1}, \tag{4.18}$$

which maps the intervals $(0,1/\omega_{\rm th}^2),(\omega_{\rm th}^2,\infty)$ of the ω^2 plane onto the lines $(-\lambda,-\zeta_{\rm th}),(\zeta_{\rm th},\lambda)$, respectively $\zeta_{\rm th}=\zeta(\omega_{\rm th}^2)$. In terms of this variable, Eq. (4.17) turns into

$$S(\zeta)S^*(-\zeta)=\kappa(\zeta),\ \zeta\in(\zeta_{\rm th},\lambda). \tag{4.19}$$

The behavior of S at $\pm\lambda$ should be proportional to $1/(\zeta-\lambda)^p, 1/(\zeta+\lambda)^q$, for some $p,q>0$ and $S(\zeta)\to$const as $\zeta\to\infty$ (the constant could be zero). To simplify the argument, we shall make the assumption that $\kappa(\zeta)$ has no zeros on $(\zeta_{\rm th},\lambda)$. The general case is treated in Appendix A. It follows from (A2) that the phase $\varphi_K(\zeta)$ of $\kappa(\zeta)$ in (4.19) is Hölder continuous on $[\zeta_{\rm th},\lambda]$. So one can define[26]

$$\Omega_K(\zeta)=\exp\left[\frac{1}{\pi}\int_{\zeta_{\rm th}}^{\lambda}\frac{\psi_K(\zeta')}{\zeta'-\zeta}d\zeta'\right], \tag{4.20}$$

with $\psi_K(\zeta)=\varphi_K(\zeta)$ or $\varphi_K(\zeta)-\pi$, according to whether $\varphi_K(\zeta_{\rm th})=0$ or π. We define further

$$S_1(\zeta)\equiv S(\zeta)/\Omega_K(\zeta),\kappa_1(\zeta)\equiv\kappa(\zeta)/(\Omega_K(\zeta)\Omega_K(-\zeta)). \tag{4.21}$$

It is clear that, as $\zeta\to\infty$, $\Omega_K(\zeta)\to1$; at $\zeta=\lambda,\Omega_K(\zeta)\sim(\zeta-\lambda)^{\psi_K(\lambda)/\pi}$; $\arg\Omega_K(\zeta)=\arg\kappa(\zeta)(\mathrm{mod}\,\pi)$ for all $\zeta\in[\zeta_{\rm th},\lambda]$. We conclude $\kappa_1(\zeta)$ is a real nonvanishing continuous function of ζ on $[\zeta_{\rm th},\lambda]$ and

$$S_1(\zeta)S_1^*(-\zeta)=\kappa_1(\zeta). \tag{4.22}$$

Clearly, $S_1(\zeta)$ behaves at λ like $1/(\zeta-\lambda)^{p'}$, with $p'=p-\psi_K(\lambda)/\pi$. From (4.22) and the real analyticity of $S_1(\zeta)$ we conclude that $\arg S_1(\zeta)=\arg S_1(-\zeta)(\mathrm{mod}\,\pi)$, for $\zeta\in(\zeta_{\rm th},\lambda)$. Let then $S_1^0(\zeta)$ be a solution of (4.22), and let its phase on $(\zeta_{\rm th},\lambda)$ be $\varphi_1^0(\zeta)$. We are free to choose $\varphi_1^0(\zeta_{\rm th})=0$ since, according to (4.22), if $S_1^0(\zeta)$ is a solution, $-S_1^0(\zeta)$ is also a solution. We consider then the function

$$\Omega_0(\zeta)=\exp\left[\frac{2}{\pi}\int_{\zeta_{\rm th}}^{\lambda}\frac{\varphi_1^0(\zeta')\zeta'\,d\zeta'}{\zeta'^2-\zeta^2}\right]. \tag{4.23}$$

This function is even in ζ, and has the same phase (mod π) as $S_1^0(\zeta)$ on the cuts $(\zeta_{\rm th},\lambda),(-\lambda,-\zeta_{\rm th})$. Consider then $\bar{S}_1(\zeta)\equiv S_1^0(\zeta)/\Omega_0(\zeta)$. This function has no cuts, but just the zeros of $S_1^0(\zeta)$ [it follows from (A2) that there are a finite number of such zeros [27]] and its poles, and possible poles or zeros at $\zeta=\pm\lambda$. There cannot be essential singularities at these last two points because of the known endpoint behavior[26] of (4.23) and the boundedness condition at $\pm\lambda$ on $S_1^0(\zeta)$.

We conclude,

$$\bar{S}_1(\zeta)=\left(\prod_{i=1}^{n}(\zeta-\zeta_i)\right)/(\zeta-\lambda)^{p_1}/(\zeta+\lambda)^{q_1}/(\zeta^2-\zeta_B^2), \tag{4.24}$$

where p_1, q_1 are integers of any sign. Let $p_1-q_1=p'-q=m$. We see that, if we know the positions of the zeros and the number m, we can define

$$S_2(\zeta)=S_1(\zeta)(\zeta-\lambda)^m\Big/\prod_{i=1}^{n}(\zeta-\zeta_i), \tag{4.25}$$

and we know that $S_2(\zeta)$ is an even function of ζ. Then let

$$\kappa_2(\zeta)\equiv\kappa_1(\zeta)(\lambda^2-\zeta^2)^m/\prod_{i=1}^{n}(\zeta_i^2-\zeta^2). \tag{4.26}$$

It is easy to verify that $\kappa_2(\zeta)$ is positive for $\zeta\in(\zeta_{\rm th},\lambda)$. It follows that the even function of $\zeta, S_2(\zeta)$, satisfies

$$|S_2(\zeta)|^2=\kappa_2(\zeta^2),\quad \zeta^2\in(\zeta_{\rm th}^2,\lambda^2). \tag{4.27}$$

Further, it has no zeros in the cut ζ^2 plane, it has one nucleon pole (on the negative real axis) at ζ_B^2, and falls off more rapidly than $(1/\zeta)^{2r}$ at infinity, where $r=[(n-m+1)/2]$, and $[x]$ means the greatest integer smaller than x. Such a function is easily constructed by a series of mappings, as follows: let $\sigma=\zeta^2, \sigma'=2\cdot(\sigma-(\lambda^2+\zeta_{\rm th}^2)/2)/(\lambda^2+\zeta_{\rm th}^2)$ [which brings the cut $(\zeta_{\rm th}^2,\lambda^2)$ onto $(-1,1)$], and finally $\eta=\sigma'-(\sigma'^2-1)^{1/2}$ (which brings the cut onto the unit disk, so that infinity comes to the center). In terms of η, the solution to our problem is

$$S_2(\eta)=\eta^{r_1}\frac{1-\eta\eta_B}{\eta-\eta_B}\times\exp\left[\frac{1}{2\pi}\int_0^{\pi}\frac{1-\eta^2}{1-2\eta\cos\theta+\eta^2}\ln\kappa_2(\theta)d\theta\right], \tag{4.28}$$

where $\eta_B=\eta(\zeta_B^2)$ and r_1 is an integer larger than r. In (4.28) we have applied Poisson's formula. The existence of the integral in (4.28) is ensured if $|\ln|S(\eta)||$ is integrable over $[0,\pi]$. The latter follows from (A1) and the fact that $S(\nu)$ satisfies dispersion relations in our interval of t values.[17] Equation (4.28) solves Problem I completely. The zeros of $S(\omega)$ in (4.25) are free parameters and can be chosen to agree with the poles of $R(\omega)$. There could be more zeros, which are then coincident with zeros of $A(\omega)$. Knowledge of $R(\omega)$ also furnishes the minimal possible value of r_1 in (4.28). Higher values of r_1 are possible if $S(\omega), A(\omega)$ have common zeros at $\omega_\infty=\omega(\zeta=\infty)$. We conclude that the ambiguity left in the solution of Problem I consists of possible common zeros of $S(\omega)$ and $A(\omega)$, and in the difference m between the behaviors of $S_1(\zeta)$ at λ and $-\lambda$. The overall ambiguity $H(\omega)\to -H(\omega)$ can be solved by continuity from the foward direction.

There is also the problem of the practical construction of $R(\omega)$ from its values on the cut. We turn to this in Secs. VIII and IX.

V. THE INTERVAL $t<0$

We now turn to Problem II, Sec. II, with the restriction on the area given by (3.5) and (3.6). We show first that Eqs. (2.10), (2.13),(2.20), and (2.22), together with (3.5) and (3.6), determine the ratio $h_+(\nu)/h_-(\nu)$ on the cuts $|\nu|\geqslant\mu+t/4m$. Indeed, from (2.10) and (2.13) we get, for $|\nu|\geqslant\bar\nu(t)$,

$$|(h_+(\nu)/h_-(\nu))-1|^2=(f_0^2(\nu,t)/f_-^2(\nu,t)). \tag{5.1}$$

Using (2.10) again, we get from (5.1)

$$\mathrm{Re}\,(h_+(\nu)/h_-(\nu))=\tfrac{1}{2}((f_+^2+f_-^2-f_0^2)/f_-^2). \tag{5.2}$$

Further, using (3.6),

$$\mathrm{Im}\,(h_+(\nu)/h_-(\nu))=\mathrm{Im}\,(h_+(\nu)h_-^*(\nu))/|h_-|^2=(\bar F(\nu,t)/f_-^2), \tag{5.3}$$

which shows that $h_+(\nu)/h_-(\nu)$ is completely known for $|\nu|\geqslant\bar\nu(t)$. For the interval $\bar\nu(t)\geqslant|\nu|\geqslant\mu+t/4m$, we follow Sec. IV and construct

$$2\frac{h_+(\nu,t)}{h_-(\nu,t)}K_-(\nu)-K(\nu)=4i\bar F(\nu,t), \tag{5.4}$$

$$2\frac{h_-^*(-\nu,t)}{h_+^*(-\nu,t)}K_-(\nu)-K(\nu)=-4i\bar F(\nu,t), \tag{5.5}$$

where we have used Eqs. (2.20), (2.22), and (3.5). So Eqs. (5.2)–(5.5) fix h_+/h_- on the whole cut $|\nu|\geqslant\mu+t/4m$.

It is convenient for the following to move over to $h^+=h_+(\nu,t)+h_-(\nu,t), h^-(\nu,t)=h_+(\nu,t)-h_-(\nu,t)=h^0$; notice, h^+, h^- do not have any symmetry in ν at fixed t. According to the above, $\bar R(\nu,t)=h^-(\nu,t)/h^+(\nu,t)$ is known for $|\nu|\geqslant\nu_{\rm th}$. From (2.10) and (2.13), we get, for $|\nu|\geqslant\bar\nu(t)$,

$$|h^+|^2=2f_+^2(\nu,t)+2f_-^2(\nu,t)-f_0^2(\nu,t)\equiv f^{+2}(\nu,t). \tag{5.6a}$$

and from (2.20) and (2.21), for $\nu_{th}\leqslant|\nu|\leqslant\bar\nu(t)$,

$$h^+(\nu,t)h^{+*}(-\nu,t)=2K_+(\nu,t)+2K_-(\nu,t)-K_0(\nu,t)\equiv K^+(\nu,t). \tag{5.6b}$$

Equations (5.6) represent our knowledge about $h^+(\nu,t)$ on $\nu_{\rm th}\leqslant|\nu|$. The zeros of h^+ are determined from the poles of $\bar R(\nu,t)$, in so far as $h^+(\nu,t), h^-(\nu,t)$ do not have common zeros: any solution of (5.6) with the correct zeros leads then to a solution of Problem II.

To simplify the treatment, from now on we make the assumption (A2'): The phase of $h^+(\nu,t)$ has the same properties as the phase of $S(\nu,t)$ in (A2), on both cuts $|\nu|\geqslant\nu_{\rm th}$.

It follows then from the polynomial boundedness of $h^+(\nu,t)$ and from (A2') that $h^+(\nu,t)$ has only a finite number of zeros $\nu_1,\nu_2,\ldots,\nu_n$ in the interior of the ν plane.[27]

To find all solutions $h^+(\nu,t)$ of (5.6), we define

$$E^+(\nu)=\exp\left[\frac{1}{2\pi}\int_0^{\pi}\frac{1-\xi^2(\nu)}{1-2\xi(\nu)\cos\theta+\xi^2(\nu)}\times\ln|h^+(\theta)|^2d\theta\right], \tag{5.7}$$

where $\xi(\nu)$ is the mapping of the ν plane cut for $|\nu|\geqslant\bar\nu(t)$ onto the unit disk, in such a way that the points $\pm\infty$ come to $\pm i$, respectively. We also define

$$B^+(\nu)=\frac{1-\xi_B\xi}{\xi-\xi_B}\prod_{i=1}^{n}\frac{\xi(\nu)-\xi(\nu_i)}{1-\xi(\nu)\xi^*(\nu_i)}, \tag{5.8}$$

with $\xi_B=\xi(\nu_B)$. Then let

$$\bar h^+(\nu)=h^+(\nu)/(B^+(\nu)E^+(\nu)). \tag{5.9}$$

The function $\bar h^+(\nu)$ has no zeros in the ν plane, cut for $|\nu|\geqslant\nu_{\rm th}$, and modulus one for $|\nu|\geqslant\bar\nu(t)$. We can then define [possibly changing the sign of $\bar h^+(\nu)$]

$$\bar L^+(\nu)=\ln\bar h^+(\nu) \tag{5.10}$$

which is real analytic in ν, in the ν plane cut for $|\nu|\geqslant\nu_{\rm th}$, satisfies

$$\bar L^+(\nu)+\bar L^{+*}(-\nu)=\ln\frac{K^+(\nu)}{B^+(\nu)B^+(-\nu)E^+(\nu)E^+(-\nu)}\equiv\bar K_1(\nu) \tag{5.11}$$

for $\nu_{\rm th}\leqslant|\nu|\leqslant\bar\nu$, is purely imaginary for $|\nu|\geqslant\bar\nu(t)$ and, from (A2'), is bounded by a constant in the whole ν plane. [The discontinuity of the phase of K^+ is positive if it occurs at a point where $\bar h^+(\nu)$ has a zero and is negative at a point where

$\bar{h}^+(-\nu)=0$.] We define further

$$L^+(\nu)=\bar{L}^+(\nu)/(\bar{\nu}^2-\nu^2)^{1/2}, \tag{5.12}$$

with the root cut for $|\nu|\geqslant\bar{\nu}$ and positive for $|\nu|\leqslant\bar{\nu}$.

The function $L^+(\nu)$ goes to zero as $\nu\to\infty$, and satisfies on $\nu_{\rm th}\leqslant|\nu|\leqslant\bar{\nu}(t)$ a condition similar to (5.11) for $\bar{L}^+(\nu)$, with $\bar{K}_1(\nu)$ replaced by $K_1(\nu)\equiv\bar{K}_1(\nu)/(\bar{\nu}^2-\nu^2)^{1/2}$. Now let

$$l^+(\nu)=\frac{1}{\pi}\int_{\nu_{th}}^{\bar{\nu}}\frac{\mathrm{Im}K_1(\nu')}{\nu'-\nu}d\nu' \tag{5.13}$$

and define

$$\tilde{L}^+(\nu)=L^+(\nu)-l^+(\nu). \tag{5.14}$$

The existence of the integral in (5.13) is guaranteed by (A2'). Since $l^+(\nu)$ is real for $\nu\leqslant\mu+t/4m$, we conclude that $\tilde{L}^+(\nu)$ satisfies for $\mu+t/4m\leqslant|\nu|\leqslant\bar{\nu}(t)$ the equation

$$\tilde{L}^+(\nu)+\tilde{L}^{+*}(-\nu)=K_2(\nu)\equiv K_1(\nu)-\mathrm{Im}\,l^+(\nu), \tag{5.15}$$

where $K_2(\nu)$ is a real function. We conclude from (5.15) that $\mathrm{Im}\tilde{L}^+(\nu)=\mathrm{Im}\tilde{L}^+(-\nu)$. Let then $\tilde{L}_0^+(\nu)$ be a solution of (5.15) and let $\tilde{L}_1^+(\nu)$ be the function defined by

$$\tilde{L}_1^+(\nu)=\frac{2}{\pi}\int_{\nu_{th}^2}^{\bar{\nu}^2}\frac{\mathrm{Im}\tilde{L}_0^+(\nu')\nu'd\nu'}{\nu'^2-\nu^2}. \tag{5.16}$$

Consider $\Delta L(\nu)\equiv\tilde{L}_0^+-\tilde{L}_1^+$. This function has no cuts, goes to zero at infinity, and has no singularities in the whole complex plane. So, $\Delta L(\nu)\equiv0$ and it follows that $\tilde{L}_0^+(\nu)$ is an even function of ν. Using the variable $\eta=\nu^2$, we then reduce our problem to that of finding a holomorphic function of η in the η plane cut along $((\mu+t/4m)^2,\bar{\nu}^2(t))$ and having a given real part $K_2(\nu)/2$ along this cut. The solution of this problem is easily obtained by

$$\tilde{L}^+(\eta)=(1/\pi)[(\eta-(\mu+t/4m)^2)(\eta-\bar{\nu}^2(t))]^{1/2}\times\int_{(\mu+t/4m)^2}^{\bar{\nu}^2(t)}\frac{K_2(\eta')d\eta'}{(\eta'-\eta)[(\eta'-(\mu+t/4m)^2)(\eta'-\bar{\nu}^2(t))]^{1/2}}. \tag{5.17}$$

In this way, we have completely solved Problem II. We see that, analogously to Problem I, the ambiguity of the solution is that of common zeros of h^+ and h^-. There is no ambiguity concerning the behavior at infinity [$\tilde{L}^+(\eta)$ defined by Eq. (5.17) will not, in general, vanish for $\eta\to\infty$ unless $K_2(\eta')$ satisfies a sum rule. A nonvanishing $\tilde{L}^+$ leads to a phase which increases indefinitely as $\nu\to\infty$.]; the overall sign of the solution is fixed by continuity from $t=0$. The practical question still to be solved is the one of finding the zeros and poles of $h^-(\nu)/h^+(\nu)$ from its values along the cut. We discuss this in Sec. VIII and IX.

One could wonder whether a qualitative change appears if, instead of assuming the combinations (2.20) and (2.22) to be given on the unphysical part of the cut, for $t<0$, one were to take as input the imaginary part obtained from the presumably unique πN phase shift analysis at fixed low energies. This question is treated in Appendix B. The conclusion is that this problem always has at most one solution, but its construction requires unstable steps of analytic continuation in energy at fixed t, unless the two amplitudes $h_+(\nu,t),h_-(\nu,t)$ have no common zeros in the complex ν plane. So there is no practically significant change from the solution of Problem II. It might be surprising, however, that the CDD ambiguity, which is usual in this type of problem, disappears because of the isospin constraint.

VI. REMOVAL OF AMBIGUITIES FOR $t\geqslant0$

In this section we wish to show that, if we use some experimental facts concerning πN scattering amplitudes together with some plausible assumptions, we can remove completely the ambiguities left in the determination of $S(\nu),A(\nu)$ in some interval $(0,a)$, $a>0$, of t values.

We shall take the following facts to be experimentally established.

(F1) There exists no point (s,t) in the physical region, restricted to $t\geqslant-t_0,t_0>0$, where any of the isospin triangles collapses to a point. To be sure, there are many points where their area vanishes, but none where *all* sides of one of them simultaneously vanish.

(F2) There exists a constant c_σ and an energy s_σ so that, for $s>s_\sigma,\sigma_{\rm tot}(\pi^+p)+\sigma_{\rm tot}(\pi^-p)>c_\sigma$; in fact, s_σ can be taken arbitrarily close to threshold, provided c_σ is sufficiently small.

(F3) The forward amplitude of elastic scattering $C_+(\nu,t=0)$ [see Eq. (2.3)] has no symmetrical zeros, i.e., there exists no ν in the complex plane, so that $C_+(\nu,t=0)=C_+(-\nu,t=0)=0$.

Notice the analogy of (F1–3) to (f1–3) in the Introduction. Apart form (A1) and (A2), we shall make the assumption

(A3) There exist constants c,ν_0 and an interval $I_t=(-\alpha',a')$ of t values such that for $t\in I_t$ and $\nu\geqslant\nu_0$,

$$|A^+(\nu,t)|/|A^+(\nu,t)+\nu B^+(\nu,t)|\leqslant c. \tag{6.1}$$

This inequality is verified, for $t\leqslant0$, by spin rotation measurements and the value of the constant is small for large enough ν (see Sec. IX). Its validity for $t>0$ can in principle be checked by analytic continuation in angle at fixed energy.

Before proceeding to the discussion of ambiguities, we recall that, as a consequence of positivity,[20,21]

(a) the combination

$$D^+(s,t)=A^+(s,t)+\nu B^+(s,t) \tag{6.2}$$

has, at fixed s, a monotonically increasing imaginary part in $0\leqslant t\leqslant4\mu^2$;

(b) the amplitudes $A^+,\nu B^+$ satisfy a twice-subtracted, uniformly convergent dispersion relation in $|t|\leqslant r_0<4\mu^2$.

In Appendix C we show that assumption (A1) concerning the continuity of the amplitudes in two variables in the physical region, and statement (b) above, have as a consequence that the amplitudes are continuous even in three variables Res, Ims, and Ret at points of the physical region. This means that, for any sequence of points (s_n,t_n) (s_n complex, t_n real) converging to a point (s,t) of the physical region, $f(s_n,t_n)\to f(s,t)$ where f is any one of the amplitudes. As a consequence, we know that any complex zero of the amplitudes in the cut s plane, which approaches a point of the cut, gives rise to a zero of the measurable modulus of the amplitudes.

We next discuss the ambiguity in the solution of Problem I concerning the possible common zeros of of $S_t(s),A_t(s)$

in the complex (two sheeted) s plane. (The values of t is used as a subscript) Clearly, if common zeros occur only at isolated values of t in $(0,4\mu^2)$, then the solution is uniquely fixed for all t in this interval. It can be obtained at those exceptional points by analytic continuation in t at fixed s. Assume therefore that $S_t(s)$, $A_t(s)$ have common zeros $\mathscr{S}_s(t)$, $\mathscr{S}_A(t)$ in the open complex cut s plane on some interval $(t',t'')\subset(0,4\mu^2)$. We can now follow $\mathscr{S}_s(t)\equiv\mathscr{S}_A(t)$ analytically as t goes towards $t=0$. The function $\mathscr{S}_s(t)$ can have at most branch points of finite order, according to Weierstrass's preparation theorem[28]; the identity $\mathscr{S}_s(t)\equiv\mathscr{S}_A(t)$ holds for all branches and we follow the function analytically along all of them, in case we meet a branch point. In this process, one of the following three situations can occur:

(i) $\mathscr{S}_s(t)$ can be analytically continued to $t=0$. In this situation, the functions $S_{t=0}(\nu)$, $A_{t=0}(\nu)$ have common zeros at $t=0$. This means that the forward amplitude has symmetrical zeros, which contradicts F3).

(ii) $\mathscr{S}_s(t)$ reaches the cut at $t=t_1$ and according to Appendix C, gives rise to a zero of the modulus of $S_t(\nu)$. Since $\mathscr{S}_A(t)\equiv\mathscr{S}_S(t)$, $A_t(\nu)$ will have a zero at the same position. We shall show that, as a consequence of (A1), (A3), (F1), and (F2), this is impossible if t_1 is less than a certain constant a_0.

(iii) $\mathscr{S}_s(t)$ gets unbounded as we continue to lower values of t. This means that, for any N, there exists a t_N such that for any $t<t_N$, if $\mathscr{S}_s(t)$ can be continued down to t then $|\mathscr{S}_s(t)|\geqslant N$. We shall show, however, that (A3) and (F2) imply the existence of a certain interval $[0,a_0']$ of t values where the positions of the zeros cannot acquire an arbitrarily high modulus.

We next explicitly discard possibility (ii). To this end, we write the t channel isospin even combinations of h,g, Eq. (2.6), as functions of the amplitudes D^+ and B^+

$$h^+(\nu,t)=(4m^2-t)^{1/2}\times\left\{D^+ + \frac{(t)^{1/2}\nu B^+}{4m^2-t}\left[(t)^{1/2}+\frac{2m(\nu^2-\bar{\nu}^2(t))^{1/2}}{\nu}\right]\right\}, \tag{6.3a}$$

$$g^+(\nu,t)=(4m^2-t)^{1/2}\times\left\{D^+ + \frac{(t)^{1/2}\nu B^+}{4m^2-t}\left[(t)^{1/2}-\frac{2m(\nu^2-\bar{\nu}^2(t))^{1/2}}{\nu}\right]\right\}. \tag{6.3b}$$

Clearly, h^+ and g^+ are the values on the two sheets of the ν plane of the function $S_t(\nu)$. It follows from (A3) that there exists c_1 so that, for $\nu\geqslant\nu_0$, $|\nu B^+|/|D^+|\leqslant c_1$, for all $t\in(-\alpha',a')$. Let $\nu_1=\max(\nu_0,\nu_\sigma)$, $\nu_\sigma=\nu_t(s_\sigma)$ of (F2). For $\nu\geqslant\nu_1$, $\operatorname{Im}D^+\geqslant\bar{c}_\sigma\nu$, because of consequence (a) of positivity, for $t\in(0,4\mu^2(\bar{c}_\sigma$ is a positive constant). Further, (A3) implies for $\nu\geqslant\nu_1$

$$|h^+(\nu,t)|\geqslant(4m^2-t)c_\sigma\nu\times\left\{1-\frac{c_1(t)^{1/2}}{4m^2-t}\left[(t)^{1/2}+\frac{2m(\nu^2-\bar{\nu}^2(t))^{1/2}}{\nu}\right]\right\}, \tag{6.4}$$

where we have used $|D^+|\geqslant\operatorname{Im}D^+\geqslant\bar{c}_\sigma\nu$. So $|h^+|$ cannot vanish for ν high enough and $t\leqslant(m-\mu)^2/c_1^2$; in particular, it cannot vanish simultaneously with h^-. The same reasoning can be used for g^+,g^- and we again call $[0,a_1]$ the resulting interval of t values.

If a common zero of $S_t(\nu),A_t(\nu)$ occurs for $\nu<\nu_1$, at some $t>0$, it follows that $H(\nu),H(-\nu)$ vanish there simultaneously, by (4.13), so that the complex quantities $\bar{K}_+,\bar{K}_-,\bar{K}$, Eqs. (4.4)–(4.6), all vanish at that point. But $\mathscr{K}(\nu,t)=|\bar{K}_+|^2+|\bar{K}_-|^2+|\bar{K}|^2$ is a continuous function of ν and t on the compact domain $\mu+t/4m\leqslant\nu\leqslant\nu_1, 0\leqslant t\leqslant r_0<4\mu^2$ and is therefore uniformly continuous there. According to (F1), $\mathscr{K}(\nu,t)$ is different from zero at $t=0$, and attains there its minimal value $K_{\min}$. It follows that there exists a whole interval of t values $(0,a_2)$, $a_2=a_2(K_{\min})$, so that $\mathscr{K}(\nu,t)$ is different from zero there, for $\nu\leqslant\nu_1$. So, if we take $a_0=\min(a_1,a_2)$ we have discarded possibility (ii) for $t<a_0$.

We can obviously avoid this argument if we dare state as a "fact" that $\bar{K}_+,\bar{K}_-,\bar{K}_0$, as obtained by extrapolation at fixed energy to $t>0$, do not vanish simultaneously at any point in a region $0\leqslant t\leqslant\alpha,\nu_{\text{th}}\leqslant\nu<\infty$ [as we did in the Introduction in (f1)]. This we shall assume in the next section.

To discuss possibility (iii), we notice first the following: Fact (F2) and the positivity of D^+ in $[0,4\mu^2)$ imply that there exists $N(c_\sigma)$ [c_σ in (F2)] such that, for $|\nu|\geqslant N$, D^+ does not vanish in the complex ν plane.

To show this we write a once-subtracted dispersion relation for D^+ in the ν^2 plane. We denote $x=\nu^2=|\nu|^2(\cos\theta+i\sin\theta)$,

$$D^+(x,t)=a_0(t)+\frac{2x}{\pi}\times\int_{\nu_{\text{th}}^2}^{\infty}\frac{\operatorname{Im}D^+(x',t)}{x'(x'-x)}\,dx'-\frac{g^2}{m}\frac{1}{x-\nu_B^2}. \tag{6.5}$$

Let $\bar{a}=\max a_0(t)$ for $t\in[0,a']$ and $\bar{d}=\max\operatorname{Im}D^+(x',t)$ for $(x',t)\in[0,a']\times[\nu_{\text{th}},\nu_\sigma]$. Both $\bar{a}$ and $\bar{d}$ are clearly finite since the domains are compact and the functions $a_0(t)$, $\operatorname{Im}D^+$ continuous. We next estimate a lower bound for D^+ on a circle of radius $|x|$.

Consider to this end first the angular interval $0<\theta\leqslant3\pi/4$:

$$\begin{aligned}|\tilde{D}^+|&\equiv\left|\frac{2x}{\pi}\int_{\nu_1^2}^{\infty}\frac{\operatorname{Im}D^+(x',t)\,dx'}{x'(x'-x)}\right|\\
&\geqslant\frac{2|x|}{\pi}\left|\operatorname{Im}\int_{\nu_1^2}^{\infty}\frac{\operatorname{Im}D^+(x',t)\,dx'}{x'(x'-x)}\right|=\frac{2|x|}{\pi}\int_{\nu_1^2}^{\infty}\frac{\operatorname{Im}D^+(x',t)x\sin\theta}{(x'-|x|\cos\theta)^2+|x|^2\sin^2\theta}\frac{dx'}{x'}\\
&\geqslant\frac{2|x|}{\pi}\bar{c}_\sigma\int_{\nu_1^2}^{\infty}\frac{dx'\,x\sin\theta}{(x')^{1/2}[(x'-|x|\cos\theta)^2+|x|^2\sin^2\theta]}=\frac{2|x|}{\pi}c_\sigma\operatorname{Im}\int_{\nu_1^2}^{\omega}\frac{1}{(x')^{1/2}}\frac{1}{x'-x}\,dx'\\
&\geqslant2|x|c_\sigma\left(\operatorname{Im}\frac{1}{(\sqrt{}-x)}-\frac{A}{|x|}\right),\end{aligned} \tag{6.6}$$

where $A/|x|$ is an upper bound for

$|\mathrm{Im}\int_0^{\nu_1^2} dx'/((\sqrt{x'})(x'-x))|$ for large $|x|$. In (6.6) we have used the positivity property (a) of $\mathrm{Im}D^+$. Using $|x|\,\mathrm{Im}[1/(-x)]^{1/2} = |\nu|\cos(\theta/2)$, we conclude that, for $|\nu|$ sufficiently large, $0 \leqslant \theta \leqslant (3\pi/4)$,

$$|D^+| \geqslant 2\bar{c}_\sigma|\nu|\cos(\theta/2) - 2\bar{c}_\sigma A - \bar{a} - \bar{d}A_1 - A_2, \tag{6.7}$$

with A_1, A_2 upper bounds, for large $|x|$, of

$$\frac{2|x|}{\pi}\int_{\nu_{\mathrm{th}}^2}^{\nu_1^2} \frac{dx'}{x'-x}, \frac{g^2}{m}\frac{1}{x-\nu_B^2}.$$

We now turn to the interval $\pi \geqslant \theta \geqslant 3\pi/4$. There

$$|\bar{D}^+(\nu,t)| \geqslant \frac{2|x|}{\pi}\left|\mathrm{Re}\int_{\nu_1^2}^{\infty}\frac{\mathrm{Im}D^+(x',t)\,dx'}{x'(x'-x)}\right|$$

$$\approx \frac{2|x|}{\pi}\int_{\nu_1^2}^{\infty}\frac{\mathrm{Im}D^+(x',t)(x'-|x|\cos\theta)}{(x'-|x|\cos\theta)^2+|x|^2\sin^2\theta}\frac{dx'}{x'}$$

$$\geqslant \frac{2|x|}{\pi}\bar{c}_\sigma\,\mathrm{Re}\int_{\nu_1^2}^{\infty}\frac{dx'}{(x')^{1/2}(x'-x)}dx' \geqslant 2\bar{c}_\sigma(|x|^{1/2}\sin\frac{\theta}{2} - A'). \tag{6.8}$$

for a certain A'. Consequently, for $3\pi/4 \leqslant \theta \leqslant \pi$,

$$|D^+(\nu,t)| \geqslant 2\bar{c}_\sigma|\nu|\sin(\theta/2) - 2\bar{c}_\sigma A' - \bar{a} - \bar{d}A_1 - A_2. \tag{6.9}$$

Inequalities (6.7) and (6.9) show that, if $|\nu|$ is larger than a certain $N(\bar{c}_\sigma)$, $|D^+|$ cannot vanish any more. Further, $N(\bar{c}_\sigma)$ is independent of $t, t \in [0,a']$.

Now consider the ratio $\nu B^+/D^+$. This ratio is holomorphic for $|\nu| \geqslant N$, since D^+ has no zeros there. Further, we have seen that along the cuts $|\nu B^+|/|D^+| \leqslant c_1$ for $|\nu| \geqslant \nu_0$ [from (A3)]. But $1/D^+$ is polynomially bounded for large $|\nu|$ as one sees from (6.7) and (6.9).

Since νB^+ is also polynomially bounded, it follows from the Phragmen–Lindelöf theorem that $|\nu B^+|/|D^+| \leqslant c_1'$ for all $|\nu| \geqslant N$, where $c_1' = \max(c_1, \sup_{t\in[0,a'],|\nu|=N}(|\nu B^+|/|D^+|))$. But then it is easy to see that, for $|\nu| \geqslant N$,

$$|h^+(\nu,t)| \geqslant (4m^2-t)^{1/2}|D^+|(1 - \frac{(t)^{1/2}c_1'}{4m^2-t}c_2), \tag{6.10}$$

where c_2 is an upper bound for $|\nu| \geqslant N, 0 \leqslant t \leqslant a'$ of $|t^{1/2} + 2m(\nu^2 - \bar{\nu}^2(t))^{1/2}/\nu|$. Since $|D^+|$ does not vanish for $|\nu| \geqslant N$, we conclude that, for $0 \leqslant t \leqslant \min(a', (4m^2-t)^2/(c_1'^2, c_2^2)) = a$, h^+ cannot vanish for $|\nu_I| \geqslant N$ in the complex ν plane.

The same reasoning is clearly true for $g^+(\nu,t) = S_t(\nu_{\mathrm{II}})$. Therefore, there cannot be common zeros of S_t and A_t in the ν plane for large $|\nu|$.

This way we have removed the ambiguity of the common zeros in Problem I. We now turn to the one concerning the number m, [Eq. (4.25)]. To this end, we recall that we have just seen in Eqs. (6.7)–(6.9) that there exist a constant K_1 such that ($t \in [0,a]$):

$$|S_t(\nu_{\mathrm{I}})| > K_1|\nu_{\mathrm{I}}|, \quad |S_t(\nu_{\mathrm{II}})| > K_1|\nu_{\mathrm{II}}| \tag{6.11}$$

for large $|\nu_{\mathrm{I}}|, |\nu_{\mathrm{II}}|$, on the two sheets of the complex ν plane. Equation (6.11) can be reconciled with the requirement of twice subtracted dispersion relations for $S_t(\nu)$ at most for one value m_0 of m: If $|m - m_0| > 0$, the behavior of $S_t(\nu)$ on one of the sheets is changed by a power of $|\nu|^2$ as $\nu \to \infty$, which is not admissible.

We have shown in this way that, for each t in an interval $[0,a]$, we can construct at most one pair of πN amplitudes consistent with fixed t analyticity, isospin invariance, fixed s analyticity, positivity, assumptions (A1)–(A3), and the experimental facts (F0)–(F3). Its explicit construction is given by formulas (4.25), (4.28), (4.20), (4.21), (4.11), and (4.12).

VII. REMOVAL OF AMBIGUITIES FOR $t \leqslant 0$

We can clearly obtain the amplitudes for $t < 0$ by analytic continuation from $t > 0$. Few people would believe this is feasible practically. We now investigate supplementary conditions which allow a unique and stable construction of the amplitudes at fixed $t < 0$.

For this interval of t values, we replace the fact (F2) concerning the forward amplitude by

(F'2) There exists a constant $c_+ > 0$, such that $|h^+(\nu,t)| \geqslant c_+|\nu|$, for $|\nu| \geqslant \nu_+$, in some interval $(-\alpha',0)$ of t values and ν_+ sufficiently large. Evidence for this is discussed in Sec. IX.

Since there exist measurements which directly verify (A3), we may take it as a fact and not as an assumption for $t \leqslant 0$.

We now show that a sufficient hypothesis for a unique determination of the phase in some interval $(-\alpha,0)$ of t values is that

(A4) There exists a constant $c_R > 0$ such that, for $t \in (-\alpha_R, 0), \nu \geqslant \nu_R$,

$$|\mathrm{Re}C^+(\nu,t)| \leqslant c_R\,\mathrm{Im}\,C^+(\nu,t). \tag{7.1}$$

Clearly, (A4) is true in a model where the amplitude C^+ has Regge asymptotic behavior (Pomeron exchange) and is plausible by continuity from the forward direction, where $|\mathrm{Re}C^+(\nu,0)|/\mathrm{Im}C^+(\nu,0) \to 0$ as $\nu \to \infty$. Most physicists are accustomed to (A4).

To show that (A4) is sufficient, we recall the discussion of Sec. V, according to which the only ambiguity in the deter-

mination of the amplitudes $h_+(\nu,t)$, $h_-(\nu,t)$ consists of possible common zeros lying in the complex ν plane. We then follow the reasoning of the previous section: Since h_+, h_- have no such common zeros in the forward direction and, for some interval $-\alpha'' \leqslant t \leqslant 0$, there is no value of ν on the cut $|\nu| \geqslant \nu_{\text{th}}$ where the isospin triangle collapses to a point (by invoking, as before, the continuity of the observables in s and t for the unphysical region at low energies), it follows that common zeros may appear in the ν plane only from infinity. So, we remove all ambiguities of the phase for $-\alpha \leqslant t \leqslant 0$ if we show that there exists an N_1 such that $h_+(\nu,t)$, $h_-(\nu,t)$ do not have common zeros for $|\nu| \geqslant N_1$, $-\alpha \leqslant t \leqslant 0$. A sufficient condition for this to happen is that $h^+(\nu,t)$ does not vanish for $|\nu| \geqslant N_1$, $-\alpha \leqslant t \leqslant 0$. Writing (A3) as $|\nu B^+|/|C^+| \leqslant c_4$ for $\nu \geqslant \nu_0$, $-\alpha' \leqslant t \leqslant 0$ and a certain constant c_4, we get from (2.6)

$$c_+|\nu| \leqslant |h^+(\nu,t)| \leqslant |C^+(\nu,t)|(4m^2 + t_0)^{1/2}(1 + (c_4\sqrt{-t})/2m) \leqslant c_5|C^+|, \tag{7.2}$$

where c_5 is an upper bound of the brackets for $-\alpha' \leqslant t \leqslant 0$. From (7.2) and (7.1) we get that, for $\nu \geqslant \nu_0$, $\operatorname{Im} C^+(\nu,t) \geqslant c_+\nu/(c_5(1 + c_R^2))$, It follows from Sec. VI that there exists an N_1, such that $C^+(\nu,t)$ does not vanish for $|\nu| \geqslant N_1$ (N_1 is independent of t for $-\alpha' \leqslant t \leqslant 0$). We deduce as before from the Phragmen–Lindelöf theorem that $|\nu B^+|/|C^+| \leqslant c_4'$ for all $\nu, |\nu| \geqslant N_1$, $t \in [-\alpha', 0]$ and therefore that

$$|h^+(\nu,t)| \geqslant |C^+|\left((4m^2 - t)^{1/2} - \frac{2m(\sqrt{|t|})|(\sqrt{(\nu^2 - \bar{\nu}^2)}|}{|\nu|(4m^2 - t)^{1/2}}c_4'\right)$$

cannot vanish for $|\nu| \geqslant N_1$ and t sufficiently small. This concludes the argument. Here, as before $c_4' = \max(c_4, \sup_{|\nu| = N_1, t \in [-\alpha', 0]}|\nu B^+|/|C^+|)$.

So, we have seen that under assumptions (A1), (A2'), and (A4) and because of the facts (A3), (F0), (F1), (F'2), and (F3) we can uniquely construct the amplitudes $h^+(\nu,t)$, $h^-(\nu,t)$ in some interval $(-\alpha, 0)$ of t values by means of formulas (5.2)–(5.5),(5.11),(5.13), (5.14), and (5.17) of Sec. V.

VIII. THE MAGNITUDE OF THE INTERVAL $(-\alpha,a)$ AND STABILITY QUESTIONS

So far, we have argued that if the area is known as a two sheeted function for each $s \geqslant (m + \mu)^2$, in some interval $-t_0 \leqslant t \leqslant 4\mu^2$, if the functions $R(\nu,t) = A_t(\nu)/S_t(\nu), \bar{R}(\nu,t) = h^-(\nu,t)/h^+(\nu,t)$ can be determined in the whole ν plane, and if assumptions (A1)–(A4) in Sections VI and VII are correct, then there exists an interval $-\alpha \leqslant t \leqslant a$ where we can construct the amplitudes from data by explicit formulas, given in the preceding sections. We have, namely, shown that *all* their zeros—which appear as parameters in these formulas–are determined for $t \in (-\alpha,a)$ by the meromorphic functions $R(\nu,t)$, $\bar{R}(\nu,t)$.

There still persist the following reasons for dissatisfaction:

(i) These assumptions do not permit a determination of the interval $(-\alpha,a)$, where the construction can be done; (ii) We do not yet know how difficult it is to construct accurately the meromorphic functions, $R(\nu,t), \bar{R}(\nu,t)$.

In this paragraph, we show that essentially by strengthening assumption (A4) and assuming that the analysis of Ref. 3 performs a correct extrapolation (with *finite* errors) of the observables to the unphysical region, we can give an answer to these questions.

Namely, instead of (A4), let us assume that

(A4') the phase of the amplitude $C^+(\nu,t)$ has a limit $\phi_C(t)$ as $\nu \to \infty$, for each $-t_0 < t < 0$, and this limit is uniform with respect to t in an interval $0 \geqslant t \geqslant -\alpha_0$.

We assume for simplicity that the values α' in (F'2) and (A3) also equal α_0. In fact, we can choose $\alpha_0 = 0.5\ \text{GeV}^2$ (see Sec. IX).

Clearly, (A4') implies (A4), since (i) $\phi_C(0) = \pi/2$ and (ii) $|C^+(\nu,t)| > 0$ for ν large enough and $0 \geqslant t \geqslant -\alpha_0$ by (F'2) and (A3). According to (A4'), $\phi_C(t)$ is then the uniform limit of continuous functions, is itself continuous, and so there exists an interval $(-\alpha_R, 0)$ where it does not vanish.

From Ref. 3, we shall take the following to be true

(AF1): The functions $K_+, K_-, K_0(\nu,t)$ [Eqs. (2.18)–(2.19)] do not vanish simultaneously at any point in $0 \geqslant t \geqslant -\alpha_0, \bar{\nu}(t) \geqslant \nu \geqslant \nu_{\text{th}}$.

We emphasize that, strictly speaking, the errors of analytic continuation off the physical region are infinite.

We consider in detail in the following only negative values of t. The situation for positive t is simpler and can be treated by analogy.

With (A4') and (AF1) the determination of the value α [so that uniqueness is achieved in $(-\alpha,a)$], as was done in the previous section, still appears difficult because it contains *a priori* unknown quantities like α_R in (A4), N_1 (the radius of the circle such that C^+ does not have zeros in $0 \geqslant t \geqslant -\alpha_R, |\nu| \geqslant N_1$), and

$$c_4^0 = \max_{\substack{|\nu| = N_1 \\ 0 \geqslant t \geqslant -\alpha_R}} |\nu B^+(\nu,t)/C^+(\nu,t)|.$$

We can clearly apply the construction of amplitudes of the previous section for any t, but the problem is that, below a certain $t = -\bar{\alpha}$, we might no longer obtain the true amplitudes by this method.

One can, however, show the following:

Assume we know functions $\bar{h}^+, \bar{h}^-$ (and $\bar{C}, \bar{B}$), for $0 \geqslant t \geqslant -\alpha_1$, constructed according to the assumption that they do not have common zeros (in the s plane at fixed t). Assume further, that these functions are such that

(a) $\bar{C}^+(\nu,t)$ does not vanish outside a circle of radius $|\nu| = N$ in $0 \geqslant t \geqslant -\alpha_1$;

(b) if $\bar{c}_4 = \max_{|\nu| = N, -\alpha_1 < t < 0}|\nu\bar{B}|/|\bar{C}^+|$, then $\bar{t}_0(\bar{c}_4) = (2m^2/\bar{c}_4)^2 \geqslant \alpha_1$;

(c) $\pi - \phi_0 \geqslant \bar{\phi}_C(t) \geqslant \phi_0 > 0$, where $\phi_C(t)$ is the limiting phase of $\bar{C}^+$. Then if h^+, h^- are the true amplitudes, fulfill assumptions (A1)–(A4') for $0 \geqslant t \geqslant -\alpha_1$ [and the facts (F0)–(F3') and (AF1)], they coincide with $\bar{h}^+, \bar{h}^-$ for all t, $0 \geqslant t \geqslant -\alpha_1$.

For simplicity, we shall also assume that the constant c_1, which one derives from (A3) for a uniform bound on $|\nu B^+(\nu,t)|/|C^+(\nu,t)|$ along the cuts, $\nu \geqslant N, 0 \geqslant t \geqslant -\alpha_1$, is such that $(2m^2/c_1)^{1/2} \geqslant \alpha_1$ (this is verified in reality).

According to the preceding section, we know that there exists an $\bar{\alpha}$, $0<\bar{\alpha}\leqslant\alpha_1$, such that, for $0<|t|<\bar{\alpha}$, $\bar{h}^{\pm}(\nu,t)\equiv h^{\pm}(\nu,t)$; we do not know, however, the value of $\bar{\alpha}$. Because of (AF1) and (F3), the only possibility for h^+,h^- to acquire common zeros is from infinity. In other words, if the construction ceases to be valid at $t=-\bar{\alpha}$, a sequence of points (ν_n,t_n) must exist such that, as $n\to\infty$, $|\nu_n|\to\infty, t_n\to-\bar{\alpha}, t_n<-\bar{\alpha}$, and $h^+(\nu_n,t_n)=0$.

At the same time, we know from (A4') that a constant c_R' and a neighborhood $U_{\bar{\alpha}}$ of t values exist, such that $|\mathrm{Re}\,C^+|<c_R'\,\mathrm{Im}\,C^+$ for all ν sufficiently large on the cut. Indeed, according to (b), $\phi_c(\bar{\alpha})\neq 0$. As a consequence, there exists a circle of radius N_1, which we can choose larger than N of (c), with the property that $|C^+(\nu,t)|$ does not vanish outside $|\nu|=N_1$ for $t\in U_{\bar{\alpha}}$, and is even bounded from below by const $|\nu|$ (cf. Sec. VI). So, $\nu B^+/C^+$ is holomorphic and polynomially bounded outside $|\nu|=N_1$. Consider then $m_{N_1}(t)=\max_{|\nu|=N_1}|\nu B^+|/|C^+|$. It is a continous function of t, $t\in U_{\bar{\alpha}}$.

On the other hand, from the vanishing of $h^+(\nu_n,t_n)$, where ν_n is such that $|\nu_n|\geqslant N_1$ and $t_n\in U_{\bar{\alpha}}$ we deduce that

$$(\nu B^+/C^+)(\nu_n,t_n)=\frac{(4m^2-t_n)\nu_n}{2m(\sqrt{|t_n|})(\sqrt{\nu_n^2-\bar{\nu}^2(t)})},\tag{8.1}$$

and so

$$\frac{|\nu B^+(\nu_n,t_n)|}{|C^+(\nu_n,t_n)|}\geqslant\frac{2m}{\sqrt{|t_n|}}\frac{|\nu_n|}{(\sqrt{|\nu_n|^2+\bar{\nu}^2(t)})}.\tag{8.2}$$

In particular, we deduce from the Phragmen–Lindelöf theorem, at $t=-\bar{\alpha}$,

$$2m/\sqrt{\bar{\alpha}}\leqslant\max(m_{N_1}(\bar{\alpha}),c_1),\tag{8.3}$$

where c_1 is the uniform bound on $|\nu B^+|/|C^+|$ along the cuts, $|\nu|\geqslant N$. However, using (a) for $t\geqslant-\alpha$, we also know from the maximum modulus theorem that

$$\max(m_{N_1}(\bar{\alpha}),c_1)\leqslant\max(m_N(\bar{\alpha}),c_1).\tag{8.4}$$

But, from (b) above, we learn that $m_N(\bar{\alpha})\leqslant 2m/\sqrt{|\alpha_1|}<2m/\sqrt{|\bar{\alpha}|}$. Using the inequality for c_1, we see a contradiction with (8.3). In other words, the hypothesis that the construction loses its validity for $-\bar{\alpha}\geqslant t\geqslant-\alpha_1$ is inconsistent with the other assumptions.

We can apply the same reasoning to an interval $a_1\geqslant t\geqslant 0$, using the D^+ amplitude (instead of C^+), and an assumption (AF2), similar to (AF1), but referring to $t>0$. Condition (c) can be abandoned and (A4') is not necessary.

In this way, we have answered question (i) at the beginning of this section. To sum up, we assumed that functions have been constructed according to the formulas of the preceding sections (i.e., under the assumption that h^+,h^- or S,A have no common zeros) and conditions have been given which are sufficient for them to be true amplitudes, on some interval of t values, $-\alpha_1\leqslant t\leqslant a_1$.

Question (ii) concerns the stability of the construction. Consider the class of pairs of functions $\{h_+,h_-\}$, defined by assumptions (A1)–(A4') and facts (F0)–(F3) and (AF1) with the correct analyticity properties and such that the corresponding combination $D^+(\nu,t)$ has a monotonically increasing imaginary part (at fixed s) in $0\leqslant t<4\mu^2$. Consider also a specific pair $\{h_+,h_-\}$ in this class and a sequence $\{h_{+},h_{-}\}$ so that the (six) observables constructed from each of its terms, at every point of $D=\{\nu,t\in R\,||\nu|\geqslant\nu_{\mathrm{th}},-\alpha_1\leqslant t\leqslant a_1\}$, tend to those constructed from $\{h_{+},h_{-}\}$ in some sense, in all of D. {Notice, we suppose that even in the unphysical region the quantities $K_{\pm},0$ [Eqs. (2.20) and (2.21)] and $\bar{K}_{\pm,0}$ [Eqs. (2.24) and (2.25)] are well approximated}. Can we then infer that the pairs $\{h_{+n},h_{-n}\}$ approximate $\{h_+,h_-\}$ increasingly well as $n\to\infty$?

There are three steps to be analyzed:

(α) If the "data" are close to each other, are the zeros and poles of $R_n(\nu,t)$, or $\bar{R}_n(\nu,t)$ [corresponding to (h_{+n},h_{-n})] close to those of $R_n,\bar{R}_n$?

(β) If the zeros are close to each other, is the result of the construction of Sec. IV–VII stable under small variations of the observables?

(γ) If the construction is stable, are the criteria of its validity, as shown in this section, insensitive to small errors?

Questions (β) and (γ) are mild if h^+,h^- have only a finite number of zeros in the complex s plane (this is the only case we shall consider—see below). For example, one can show[17] that the ratio $E^+(h_n^+)/E^+(h^+)$ [see Eq. (5.7)] approaches 1 at any interior point of the complex s plane, uniformly on compact subsets, if $\ln|h_n^+(\xi)|-\ln|h^+(\xi)|$ tends to zero in $L^2(0,2\pi)$, for $|\xi|=1$: here ξ is the variable used in Eq. (5.7). On the cuts themselves ($|\xi|=1$), the logarithm of the ratio goes to zero in $L^2(0,2\pi)$. Pointwise convergence can be achieved at those points where $\ln|h^+(\xi)|$ obeys a Hölder continuity condition. Similar considerations apply to the other integral expressions.

As for (γ), an apparent difficulty is that $\nu B^+(\nu,t)$ is determined with increasing errors as $t\to 0$. This is a consequence of the $\sqrt{t}$ factor multiplying B in Eq. (2.6). However, we could have formulated a weaker condition instead of (b) in this section, namely (b'): for each t in $0\geqslant t\geqslant-\alpha_1$, $(\sqrt{|t|})\times\sup_{|\nu|=N}|\nu\bar{B}^+(\nu,t)|/|\bar{C}^+(\nu,t)|\leqslant 2m-\delta,\delta>0$. The proof can be done then in the same way, by comparing two estimates of $\max_{|\nu|=N}|\nu B^+|/|C^+|$ at $t=-\bar{\alpha}$, obtained by continuity from $t>-\bar{\alpha}$ and $t<-\bar{\alpha}$, respectively. If $h,^+h^-$ are known with finite errors in the whole ν plane, so are C^+, $(\sqrt{|t|})\nu B^+$, and condition (b') can be checked reasonably.

Question (α) is at first sight more difficult. The reason is that we cannot fix the number of zeros and poles of $\bar{R}(\nu,t)$ from its boundary values at fixed t only. Indeed, we can always place one zero and one pole in $\bar{R}(\nu,t)$, with a sufficiently small distance between their positions, and in this way disturb the boundary values of $\bar{R}$ only within arbitrarily small errors. As we change the value of t, then, according to (A1), such pairs of zeros and poles can appear either from infinity or from the zeros of $h^+(\nu,t)$ and $h^-(\nu,t)$ lying on the cut. However, according to (A1)–(A4') and (F0)–(F3), we are not free to let the zeros of h^+ reach infinity at an arbitrary t value. This could, namely, cause a discontinuity in t of $\max_{|\nu|=N}(\sqrt{|t|})|\nu B^+(\nu,t)|/|C^+(\nu,t)|$ for some circle of radius N, as explained at the beginning of this section.

So the relevant question is whether we can control the zeros and poles of $\bar{R}(\nu,t)$ in an interval of t values,

$0 > t > -\alpha$, where conditions (a), (b), and (c) are satisfied. At these t values, a constant $\bar{c}(t)$ exists such that

$$|h^+(\nu,t)| \geqslant \bar{c}(t)|\nu| \tag{8.5}$$

for all $|\nu|$ sufficiently large in the complex ν plane.

For simplicity, we shall take the following "experimental fact" to be true (see Sec. IX):

(SF1): there exist constants $\bar{c}_- > 0, 0 < \gamma < 1$, such that $|h^-(\nu,t)| < \bar{c}_- \nu^\gamma, \nu \geqslant \nu_{\text{th}}, -\alpha \leqslant t \leqslant 0$.

Inequality (SF1) is valid for all $|\nu|$ sufficiently large in the complex plane, by the Phragmen–Lindelöf theorem. As a consequence of (SF1), (F'2), and Eq. (8.4), the amplitudes $h_+ = \frac{1}{2}(h^+ + h^-), h_- = \frac{1}{2}(h^+ - h^-)$ do not vanish for $|\nu|$ sufficiently large in the complex plane, $0 \geqslant t \geqslant -\alpha$. Even more, they increase like $|\nu|$ if $|\nu|$ is large enough.

If one further assumes [this assumption is only partly independent of (SF1)] that (SA1) the phases of $h_\pm(\nu,t)$ at fixed t can be defined along the cut and have the same properties stated in (A2) for the phase of $S_t(\nu)$, then one can write a decomposition like

$$h_\pm(\nu,t) = B_\pm(\nu,t)E_\pm(\nu,t), \tag{8.6}$$

with similar notations and meanings as in Eq. (1.10)

Further, as a consequence of (A1), we know that, as we move down in t, $h_+(\nu,t), h_-(\nu,t)$ acquire (or lose) zeros only at those points of the cut where $h_\pm(\nu,t) = 0$.

We can take the following to be true.

(SF2): The zeros of h_+, h_- along the cut are isolated, i.e., if $h_+(\tilde{s}_0,\tilde{t}_0) = 0$, $\tilde{s}_0,\tilde{t}_0$ real, there exists an r, such that for no other point (s,t), s,t real in $|s - \tilde{s}_0| < r, |t - \tilde{t}_0| < r$, is $h_+(s,t) = 0$.

As a consequence, using (F'2), there are only a finite number of zeros in $\nu \geqslant \nu_{\text{th}}, 0 \geqslant t \geqslant -\alpha_1$. They all lie below a certain finite $\nu_+(\nu_-)$. Even more is in fact true [The facts (SF1–3) are not necessary in order to produce a stable construction. They shorten, however, the discussion.]

(SF3): There is no $t, -t_0 \leqslant t \leqslant 0$, so that $h_+(s_{01},t) = 0, h_-(s_{02},t) = 0$, s_{01}, s_{02} are real numbers larger than $(m+\mu)^2$.

Then consider the ratio

$$r(\nu,t) = \frac{h_-(\nu,t)}{h_+(\nu,t)} = \frac{1 - \bar{R}(\nu,t)}{1 + \bar{R}(\nu,t)}. \tag{8.7}$$

We can divide out a function $E_r(\nu)$, having the modulus of $r(\nu,t)$ and no zeros in the cut ν plane. Clearly, $E_r(\nu) = E_-(\nu)/E_+(\nu)$. We are left with

$$r_1(\nu,t) = \frac{r(\nu,t)}{E_r(\nu,t)} = \frac{B_-(\nu,t)}{B_+(\nu,t)}. \tag{8.8}$$

The total variation $\Delta(t)$ of the phase of $r_1(\nu,t)$ allows one to determine $N_n - N_d$, where $N_n - N_d$ are the zeros of the numerator and denominator, respectively, in (8.8). From (SF1) and (SF2), it follows that jumps in $\Delta(t)$ appear only at those t values where $h_-(\nu,t), h_+(\nu,t)$ vanish on the cut at some point. From (SF3), one sees that one can determine successively N_n, N_d by finding the jump in $\Delta(t)$ at those t values. Thus, the number of parameters appearing in (8.8) is fixed.

The determination of the finite number of zeros and poles of $r(\nu,t)$ can be done in such a way that the errors of their positions tend to zero as the errors of the boundary values tend to zero, at least on open subintervals of $0 \geqslant t \geqslant -\alpha$. One must exclude the possible finite number of t values, where the zeros of the numerator happen to coincide with zeros of the denominator. At all other t values, $0 \geqslant t \geqslant -\alpha$, there exists a minimal distance $d_{\min}(t) \neq 0$ between the zeros of B_- and $B_+(\nu,t)$.

Assume then that stability was not achieved at such a t value: this means a number d existed such that, for any ε, we could find approximants

$$\left| \frac{B_{n-}(\nu)}{B_{n+}(\nu)} - \frac{B_-(\nu)}{B_+(\nu)} \right| < \varepsilon \tag{8.9}$$

for ν on the cuts, so that the number of zeros of $B_{n-}(\nu), B_{n+}(\nu)$ is the same as that of $B_-(\nu), B_+(\nu)$, respectively, but some of them staying a distance d away from the true ones. Let then ν_- be a zero of $B_-(\nu)$, which is not approximated by corresponding zeros of $B_{n-}(\nu)$. So, for any ε, $B_{n-}(\nu) > \text{const} > 0$. On the other hand,

$$|B_{n-}(\nu)B_+(\nu) - B_-(\nu)B_{n+}(\nu)| < \varepsilon \tag{8.10}$$

for ν on the cuts, and since the expression between bars is bounded by a constant for $\nu \to \infty$ in the complex plane, (8.10) holds in the whole complex plane. In particular, at ν_-, we get $|B_{n-}(\nu)| < \varepsilon/|B_+(\nu_-)|$. This contradicts the earlier inequality for ε sufficiently small.

For $t \geqslant 0$, the domain of interest is doubly connected, and this procedure must be slightly modified. One divides out a function $\Omega(\omega)$, similar to (4.20), containing the phase of $r(\omega) = H(-\omega)/H(\omega)$, along the two cuts [under a corresponding assumption (SA1) for the phase of $H(\omega)$] and one fits the resulting real analytic function, known along parts of the real axis, with a ratio of polynomials of known degrees.

This concludes the discussion of stability.

IX. THE SITUATION IN PRACTICE

In this section we state in more detail the evidence for the "experimental facts" (or assumptions) used in the text and give the actual numerical values of some of the quantities that occur in the construction of amplitudes described in this paper.

(a) We start with assumption (A3), Sec. VI. It states essentially that the amplitudes $A^+(\nu,t)$ and $\nu B^+(\nu,t)$ should not cancel each other when they form the combination $A^+ + \nu B^+$. This occurence is allowed by unitarity, since the Froissart bounds for A^+ and νB^+ are actually[21]

$$|A^+(s,t=0)| \leqslant Cs(\ln s)^3,$$

$$|\nu B^+(s,t=0)| \leqslant Cs(\ln s)^3, \tag{9.1}$$

but

$$|D^+(s,t=0)| \leqslant Cs(\ln s)^2. \tag{9.2}$$

We can get, however, independent information about the amplitudes from spin rotation experiments which measure the Wolfenstein parameter R, e.g., in $\pi^+ p$ scattering. The latter is related to h_+, g_+ in Eq. (2.6) by means of

$$R_+ \frac{d\sigma}{d\Omega} = \text{Re}\,(h_+ g_+^* e^{-i\theta_N})/(64\pi^2 s), \tag{9.3}$$

where θ_N is the recoil angle of the nucleon in the laboratory

frame. Together with $P_{\pm,0}$ and $(d\sigma/d\Omega)_{\pm,0}$ it clearly determines $h_{\pm}, g_{\pm}$ and so all amplitudes up to a common phase and a twofold ambiguity at any specific point (s,t) (if continuity of the area as a function of energy is used). The latter ambiguity is resolved by a measurement of the parameter A_+ or of R in $\pi^- p$ scattering. Practically, the charge exchange differential cross section is much smaller than the elastic one at high energies and small $|t|$, so that one can verify (A3) at these energies directly from the results of Refs. 29,30, and 31. One deduces $|A^+|/|D^+| \lesssim 0.5$ and decreasing for $t = -0.2 \div -0.5$ GeV^2 and $k \geqslant 6$ GeV/c.

This justifies (A3) in a sufficient way for $t \leqslant 0$. Information about $|A^+|, |D^+|$ at $t \geqslant 0$ can be obtained by simple extrapolation of these results in angle. The quality of the data does not allow this, but the fact that $|A^+|/|D^+|$ is so small at $t \leqslant 0$ makes (A3) plausible also for $t \geqslant 0$. We point out it is actually possible to believe[32,33] that not only $A^+(\nu,t)/(\nu B^+(\nu,t)) \to 0$ at high energies (s — channel helicity conservation) but even $A^+ \to 0$ (decoupling of the pomeron and f trajectories from A^+). There is no strong counterargument to this belief so far.[34]

(b) Statement (F'2) is implied by the fact that the differential cross section $d\sigma/dt$ of $\pi^{\pm} p$ scattering at high energies (see, e.g., Ref. 35) stay larger than a constant for $t \geqslant -0.5$ GeV^2 and the polarization in elastic scattering decreases in absolute value at fixed t (see Refs. 36 and 37); e.g., at $t = -0.2$, it decreases from $|P_{\pm}| \sim 0.2$ at $k = 6$ GeV/c to $|P_{\pm}| \simeq 0.05$ at $k = 45$ GeV/c. Since $|h^+|^2 = 2[|h_+|^2 + |h_-|^2] - |h^-|^2$ and the contribution of $|h^-|$ is negligible,[38] we can conclude $|h^+| \geqslant c_+ |\nu|$ for high ν. In fact, we just need $|h^+(\nu,t)| \geqslant c_+ |\nu|^{\gamma_1} |\nu| \geqslant \nu_0$ for some $\gamma_1 > 0$. For stability we prefer $\gamma_1 > \gamma$ in (SF1).

Statement (SF1) is related to the absence of a subtraction constant in the dispersion relation for C^-/ν. A discussion of this is given in Ref. 39. For (F2), see especially the same reference.

(c) The absence of symmetrical zeros of the forward amplitudes can be inferred from Ref. 40. There exists so far no published calculation of the zeros of $C_{\pm}$ $(t = 0, \nu)$. However, one finds in Ref. 40 the position of the single zero in the ν^2 plane of $C^+(t = 0, \nu^2)$ and one can check from the low-energy expansion of the same Ref. 40 (see also Ref. 41) that $C^-(\nu,t)$ is nonvanishing there. There is just one zero of $C^+(\nu^2,t)$ in the forward direction (there cannot be more than two; for the detailed pattern of zeros at low $|t|$, near $t = 0$, see Ref. 42).

The determination of the area [in (F0)] at high energies by continuity from low energies is made possible by the remarkably simple appearance of the lines where the isospin bound is saturated. At high s, they run approximately at $Z_s = \cos\theta = \text{const}$ (see Ref. 18).

(d) Reference 3 produces amplitudes only at $t \leqslant 0$. To make use of the discussion in Secs. IV and VI, we extrapolate them by means of phase shifts and check whether they are consistent with the correct analyticity properties. We do this by expanding $B^{\pm}, C^{\pm}$ (multiplied by suitable factors) at $t \geqslant 0$ in a Fourier series in the variable z, which maps the cut ν plane onto the unit disk, and computing the sum $S^{(-)}$ of squares of the coefficients with negative frequency. The latter do not vanish identically because of the noise in s of the functions $C^{\pm}(s_{j_1},t), B^{\pm}(s_{j_1},t)$, furnished by Ref. 3 (see Introduction), but one expects that they should be in some sense small compared to the corresponding sum $S^{(+)}$ of coefficients of the positive frequency terms. One obtains the ratios 0.003, 0.007, 0.02, and 0.05 as averages in the range $0 \leqslant t \leqslant \mu^2$ for sums of the first four coefficients.

(e) The amplitudes of Ref. 3 satisfy assumptions (A1)–(A4') and facts (F1)–(F3). So, we know there exists an interval $-\alpha \leqslant t \leqslant a$ where they are the only ones compatible with the observables they generate (including the unphysical region). One can check that $|\nu B^+|/|C^+| \sim 1$ on a circle of radius $N \simeq 10$ GeV^2, for $\mu^2 \geqslant t \geqslant -0.5$ GeV^2. Further, the limiting phase $\phi_c(t)$ of $C^+(\nu,t)$ increases slowly as t decreases, and is approximately $2\pi/3$ at $t = -0.5$. So, according to Sec. VIII, they are the only ones satisfying (A1)–(A4') and (F0)–(F3) and reproducing the observables on $\mu^2 \geqslant t \geqslant -0.5$ GeV^2. The amplitudes of Ref. 3 satisfy the construction of Secs. IV and V identically.

(f) The stability of the construction is good if zeros of $h_+(\nu,t), h_-(\nu,t)$ in the complex ν plane stay far apart and is worse if the distance between them is small.

Figure 1 shows the variation of the positions of the zeros of h_+, h_- in the complex ν plane, as a function of real t, as determined from Ref. 3. One sees that a zero of $h_-(\nu,t)$ moves close to one in $h_+(\nu,t)$ at $t = -0.04$ GeV^2 and stays so until $t = -0.16$ GeV^2. To get a measure of the dipole formed in h_+/h_- by the two almost coincident zeros, we divide h_+/h_- by a function having its modulus on the cut $|\nu| \geqslant \mu + t/4m$ and no zeros in the cut ν plane. The relevant information coming from h_+/h_- is represented by a ratio of two Blaschke products, each containing as many factors as there are zeros in h_+, h_-. Those zeros of $h_-(\nu,t)$ which lie far away from the one in $h_+(\nu,t)$ lead to clear structures in the phase

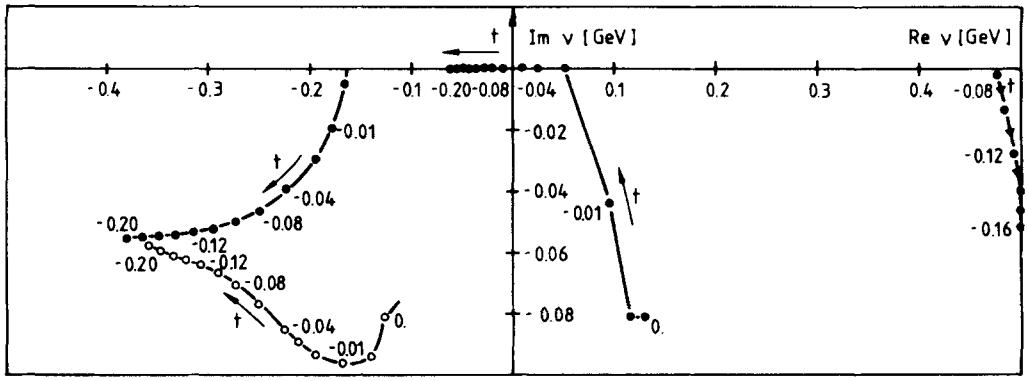


FIG. 1. The variation of the zeros of h_+ (black dots) and of h_- (open circles), as a function of $t, 0 > t > -0.2$ GeV^2, as determined from the amplitudes of Ref. 3.

and can be easily determined. We show in Fig. 2 the phase of the ratio of the two factors corresponding to the zeros that lie close to each other, for different t values. At $t = -0.12$, we see that the variation of the phase is small, and instabilities can occur. However, the present author believes the other curves show the existence of a domain where the structure of the phase of h_+/h_- cannot be obscured by small perturbations.

(g) We now turn to the interval $0 \leqslant t \leqslant 4\mu^2$, i.e., we study the zeros and poles of $R_t(\omega) = A_t(\omega)/S_t(\omega)$. Because of (F2), (A3), and positivity, $S_t(\omega^2)$ does not have zeros far away in the ν^2 plane, so that "dipoles" [almost coincident zeros of $S_t(\omega), A_t(\omega)$] can appear in $R_t(\omega)$ only in the low energy domain. A computation has been performed by means of the low energy expansion of Refs. 40 and 41 of the zeros of $S_t(w)$, $A_t(\omega)/\omega$, $w = \omega^2$ for the amplitudes of Ref. 3, for $t \geqslant 0$. The results are shown in Fig. 3 in a Re$w - t$ plot, which looks at first unnatural (the zeros lie mostly on the real axis of the w plane). A possibly dangerous dipole is seen near $w = -1$. One can estimate whether it is observable in the data for $R_t(\omega)/\omega$ by computing the residue at the zero of $S_t(\omega)$ and checking whether the pole gives an effect in the physical region larger than the resolution level. It turns out that this residue is rather large for $t \simeq 0$ (i.e., $\simeq -1.5$ in the w plane) but decreases quickly as one goes to $2\mu^2$. The effect at w = 1 is larger than 0.01 until $t = 1.7\mu^2$. So, at least in the range $0 \leqslant t \leqslant \mu^2$, one can construct the amplitudes of Ref. 3 by starting from the observables extrapolated in angle, at fixed energy, and allowing for small errors. [According to Sec. VIII, we should have studied the zeros and poles of $H(\omega)/H(-\omega)$, rather than of $A_t(\omega)/S_t(\omega)$. For the present purpose this is, however, irrelevant.]

(h) The last question we wish to discuss is the determination of the sign of $\overline{F}(\nu,t)$, Eqs. (3.4)–(3.6). Figure 4 shows the area as a function of angle along the unit circle of the w_s plane, Eq. (2.16), at $s = 3.54\ \text{GeV}^2$ as obtained in Ref. 3. The problem is whether one can exclude possible reflections along the dotted line $\overline{F} = 0$. To this end, we compute the Fourier coefficients a_n of the expansion

$$\overline{F}(w_s) = \sum_n a_n w_s^n + \sum a_n^* \frac{1}{w_s^n}, \tag{9.4}$$

and with them the quantity

$$\phi = \sum_n a_n^2 \left(r_1^{2n} + \frac{1}{r_1^{2n}}\right) \tag{9.5}$$

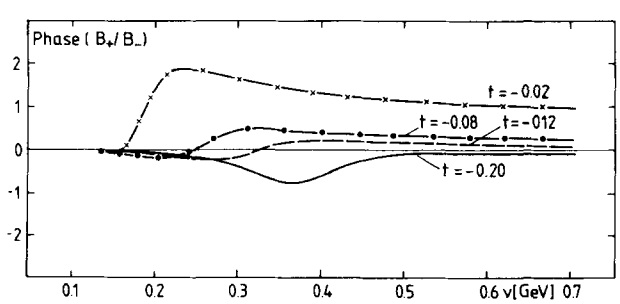


FIG. 2. The phase of the ratio of the Blaschke factors corresponding to the nearby zeros of h_+, h_- in Fig. 1, as a function $|\nu|$.

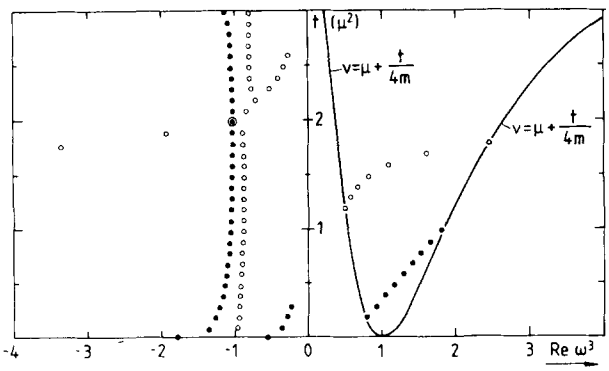


FIG. 3. The variation of the zeros of $S_t(w)$ (black dots), $\tilde{A}_t(w) \equiv A_t(\omega)/\omega$ (open circles), $w = \omega^2$, in the w plane, as a function of $t, t \geqslant 0$, as determined from the low energy expansion of Ref. 41.

with $r_1 < r_s, r_s$ given by Eq. (2.17). In the limit of infinitely many Fourier coefficients, this quantity should be finite only for the correct choice of reflections. [A change $\overline{F} \to -\overline{F}$ is excluded by the known forward amplitude.] There are clearly many possible combinations of such reflections; to give an idea, we quote just the values of ϕ for the plot of Fig. 4 and for the possible reflections of the negative pieces I, II, III of Fig. 4. We get $\phi: \phi_I: \phi_{II}: \phi_{III} = 40: 612: 3270: 97$ for $r_1 = 0.99\, r_s$. The difference between 40 and 97 might not seem convincing; it is important that this effect is the same over large intervals of energy.

X. CONCLUSIONS

It was the purpose of this paper to provide a theoretical description of the πN phase shift analysis based on isospin invariance and fixed t analyticity. There were two reasons to start this investigation. The first one is given by the rather surprising stability in time of the results of such an analysis; this is *a priori* not expected, in view of the large ambiguities at fixed energy which can appear. Moreover, its results show much similarity[42] to those obtained by other methods (Ref. 4), so that one can believe that they do represent the truth approximately and one would like to understand why. A

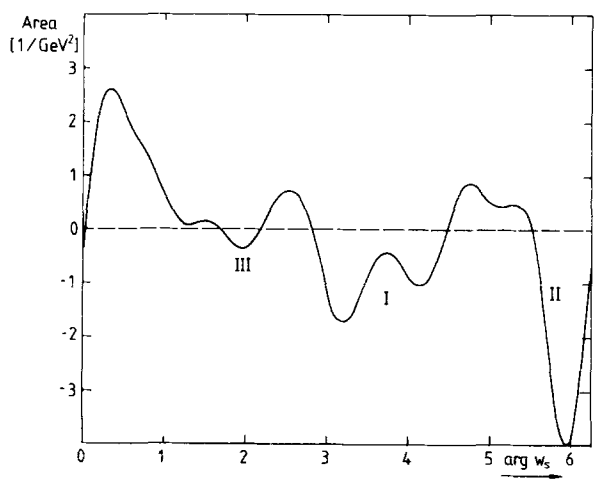


FIG. 4. The area of the isospin triangle constructed from the transversity amplitudes[18] $F(+), F(-)$, as a function of arg w_s in (radians), Eq. (2.14), at $s = 3.54\ \text{GeV}^2$.

second reason is that no clear statement was known to the author concerning the assumptions that get into the technically very involved procedure of phase shift analysis, in order to get the unique final result.

The problem of this paper might seem artificial to those concerned with a study of real data; we have assumed that the analysis of Ref. 3 provides a correct representation of the data on differential cross sections and polarizations of the elastic and charge exchange reactions (including their extrapolations to unphysical regions) in an interval $4\mu^2 \geqslant t \geqslant -t_0$ for all energies. We have, in particular, ignored all problems related to gaps in the data, incompatibilities between experiments, etc. However, we did not need the assumption that the input data are free of errors and the present author regards this point as important. We have, namely, argued in Sec. VIII that all calculations required are also feasible.

Apart from a reference to positivity in Sec. VI, we have completely ignored the constraint of unitarity, which is nevertheless the most important one for phase shift analysis at low energies. In particular, the poles of the ratios $H(\omega,t)/H(-\omega,t), h_-(\nu,t)/h_+(\nu,t)$, which all lie in the low energy domain, are much more precisely determined by this requirement than shown in Sec. VIII, from considerations of analyticity and isospin invariance. We expect the study of this paper to be of use for analyses at higher energies, where many partial waves occur and the constraint of unitarity is ineffective.

We next collect and list again the assumptions and "facts" that have been introduced in various parts of the text, to delimit a class of amplitudes in which, according to this paper, uniqueness can be obtained on an interval $-\alpha \leqslant t \leqslant a$.

(A1) The amplitudes $A_+(s,t), B_+(s,t)$ are continuous in two variables in the region $D = \{4\mu^2 > t > -t_0, t_0 > 0, s, u \geqslant (m+\mu)^2\}$ and $\mathrm{Im}A_+(s,t)$, $\mathrm{Im}\, B_+(s,t)$ are polynomially bounded in D.

(A2) The phase of $S_t(\nu)$ [and of $h^+(\nu,t)$] can be defined at fixed $t > 0$ ($t < 0$) by continuity along the cut, except for a finite number of finite jumps, is piecewise Hölder continuous, and has a finite limit as $\nu \to \infty$, which is attained in such a way that $\nu^\beta(\delta(\nu) - \delta(\infty)) \to$ const, for some $\beta > 0$.

(A3) There exist constants $\nu_0, c > 0$ such that, for $\nu \geqslant \nu_0$, $|A^+(\nu,t)|/|A^+(\nu,t) + \nu B^+(\nu,t)| < c$, for $-\alpha' \leqslant t \leqslant a'$.

(A4) There exist constants $c_R, \nu_0' > 0$ such that, for $\nu \geqslant \nu_0'$, $|\mathrm{Re}\, C^+(\nu,t)| \leqslant c_R\, \mathrm{Im}\, C^+(\nu,t)$, for $-\alpha'' \leqslant t \leqslant 0$.

Assumption (A4) is implied by (A4'), apparently more innocent looking, and knowledge of the forward amplitude.

(A4') The limit of the phase of $C^+(\nu,t)$ exists at each fixed t, as $\nu \to \infty$ along the cut and the limit is uniform in $t, 0 \geqslant t \geqslant -\alpha$.

The facts that we used are

(F0) The area of the isospin triangles is a known function, including its sign, at each ν and t in D.

(F1) There is no point in D where the isospin triangles reduce to points (we used in the text the weaker assumption that this occurs only in the physical region).

(F2) There exists a constant c_σ and an energy s_σ such that

$$\sigma_{\mathrm{tot}}(\pi^+ p) + \sigma_{\mathrm{tot}}(\pi^- p) > c_\sigma, \quad s \geqslant s_\sigma.$$

(F'2) There exist constants c_+, ν_0, so that $|h^+(\nu)| \geqslant c_+|\nu|$, for $\nu \geqslant \nu_0$.

(F3) There are no symmetrical zeros in the forward amplitude $C_\pm(\nu, t=0)$.

A stable construction of the amplitudes from the observables, and thus a determination of the interval of validity of the analysis, $a \geqslant t \geqslant -\alpha$, can be done in the subclass defined by

(SF1) There exist constants $0 < \gamma < 1, c_- > 0$ such that $|h^-(\nu)| < c_- \nu^\gamma$, for $\nu \geqslant \nu_{\mathrm{th}}$, $0 \geqslant t \geqslant -t_0$ [or $|A_t(\nu)| < c_- \nu^\gamma, 4\mu^2 > t \geqslant 0$].

(SF2) The zeros of h_+, h_- are isolated in D [as are those of $H(\pm\omega)$].

(SF3) h_+, h_- do not vanish at the same t value in D.

(SA1) The phases of $h_\pm(\nu,t)$ [and of $H(\nu,t)$] obey (A2).

This list of assumptions and facts is sufficient for uniqueness and stability and by no means necessary. They are not all independent. The interval of validity of phase shift analysis in the class of amplitudes obeying the usual analyticity requirements in t and s, this list of assumptions and facts, and having the observables interpolated by Ref. 3, is $\simeq \mu^2 \gtrsim t \gtrsim -0.5\ \mathrm{GeV}^2$ (see Sec. IX).

We recall that the domain D contains unphysical points, and we assume it is possible to perform an extrapolation from data in the unphysical region to all of D, with finite errors. This may well be the weakest point of this approach, but the author does not see any alternative to it.

One can wonder why the isospin constraint can help in the resolution of ambiguities of the analysis based on fixed t analyticity, whereas it is known[11] to be ineffective as a constraint at fixed energy. The reason is mainly experimental; the ambiguity which is allowed by isospin invariance consists, namely, of coincident zeros of h^+, h^- in the complex ν plane, for large intervals of t; the forward amplitudes $C^\pm(\nu, t=0)$ have no such coincident zeros and it is a very unlikely event that two of them appear through the cut on the physical sheet. In πN scattering, it does not happen. Positivity and (A4) forbid these zeros to migrate from infinity.

We point out that within the class of amplitudes defined by the assumptions and facts listed above, the uncertainty in the phase at a given point $(s,t), t \geqslant -\alpha$, can be estimated explicitly according to Sec. V, if errors for the extrapolated observables are prescribed in the low energy unphysical region. One expects the error of the phase at higher energies to be insensitive to reasonable variations of these prescriptions. However, it is clear that the errors thus obtained would be seriously overestimated because unitarity has been mostly ignored in the present treatment.

Finally, we point out that the formulas of Sec. IV and V allow a construction of the πN scattering amplitudes only in terms of a few parameters, which are the zeros and poles of h_+/h_- or of $H(\omega)/H(-\omega)$ (it is a piece of luck that there are so few of these). The computation involves otherwise only integrals (possibly with principal value) over known quantities. The author will report elsewhere about the possible simplifications for practical phase shift analysis which might result from this.

ACKNOWLEDGMENTS

I would like to thank Professor D. Atkinson for discussing with me the contents of this paper and for suggesting improvements of the orginal text. I am indebted to Dr. E. Borie and Dr. R. Koch for indicating to me a number of useful references and to Dr. E. Pietarinen for pointing out a mistake in an early version of this paper. Most of my (limited) knowledge of πN scattering stems from work done with Professor G. Höhler, to whom I wish to express my gratitude. The research described in this paper was performed under a grant of the Deutsche Forschungsgemeinschaft.

APPENDIX A: COMPLEMENT TO SECTION IV

We study the situation when $\kappa(\zeta)$, Eq. (4.19), has zeros on $(\zeta_{\rm th},\lambda)$. Let $\bar\zeta_1,\bar\zeta_2,\cdots,\bar\zeta_n$ be those zeros [a finite number, according to (A2)]. We part them into two classes: those of $S_t(\zeta)$ and those of $S_t(-\zeta)$. Knowledge of $R(\omega)$, Eq. (4.14), does allow such a partition since, if $R(\omega)$ gets infinite at some point ω_0 of the cut, then necessarily $S_t(\omega_0)=0$. Indeed, we show in Sec. VI that, under assumption (A1), it follows from (B1) that there exists an interval $[0,a]$ of t values, so that $S_t(\omega)$, $A_t(\omega)$ do not simultaneously vanish in the complex ω plane, including the cuts. So, let $\bar\zeta_i'$, be the zeros of $S(\zeta)$ and $\bar\zeta_i''$ those of $S(-\zeta)$. Assume first that, for any $\bar\zeta_i'$, there is no $\bar\zeta_j''$ so that $\bar\zeta_i'=\bar\zeta_j''$. We define then the discontinuity of the phase $\varphi(\zeta)$ of $\kappa(\zeta)$ as lying between 0 and 2π at points ζ_i' and between -2π and 0 at points ζ_i''. With this definition and the conventions $\varphi(\zeta_{\rm th})=0$ or π, we can define by continuity the phase $\varphi(\zeta)$ of $\kappa(\zeta)$ along all of $[\zeta_{\rm th},\lambda]$ (except for the points $\bar\zeta_i$, for which we have given a prescription). We call $\varphi_\kappa(\zeta)=\varphi(\zeta)$ if $\varphi(\zeta_{\rm th})=0$ and $\varphi_\kappa(\zeta)=\varphi(\zeta)-\pi$ if $\varphi(\zeta_{\rm th})=\pi$. Let the absolute value of the discontinuities at $\bar\zeta_i''$ be $\alpha_i, i=1,\cdots,n''$. We define then the functions $\varphi_2(\zeta)=\Sigma_{j=0}^{i}\alpha_j$ on $(\bar\zeta_i'',\bar\zeta_{i+1}'')$, $i=0,1,\cdots,n''$, $\bar\zeta_0''=0$, $\bar\zeta_{n+1}''=\lambda$ and $\varphi_1(\zeta)=\varphi_\kappa(\zeta)+\varphi_2(\zeta)$. We construct

$$\Omega_1(\zeta)=\exp\left[\frac{1}{\pi}\int_{\zeta_{\rm th}}^{\lambda}\frac{\varphi_1(\zeta')}{\zeta'-\zeta}\,d\zeta'\right], \tag{A1}$$

$$\Omega_2(\zeta)=\exp\left[\frac{1}{\pi}\int_{-\lambda}^{-\zeta_{\rm th}}\frac{\varphi_2(\zeta')}{\zeta'-\zeta}\,d\zeta'\right], \tag{A2}$$

and, as in Sec. IV

$$S_1(\zeta)=S(\zeta)/(\Omega_1(\zeta)\Omega_2(\zeta)), \tag{A3}$$

$$\kappa_1(\zeta)=\kappa(\zeta)/(\Omega_1(\zeta)\Omega_2(\zeta)\Omega_1(-\zeta)\Omega_2(-\zeta)). \tag{A4}$$

The function $S_1(\zeta)$ can still have zeros on $(\zeta_{\rm th},\lambda)$, $(-\lambda,-\zeta_{\rm th})$, however, such that the phase can jump there only by multiples of 2π, as we move along the cut. Further, the function $\kappa_1(\zeta)$ has a constant sign on $[\zeta_{\rm th},\lambda]$ but can vanish there. Equation (4.22) is satisfied by $S_1(\zeta)$. Let again $\varphi_1^0(\zeta)$ be the phase (defined by continuity, mod 2π) of a solution $S_1^0(\zeta)$ of (4.22) and let $\Omega_0(\zeta)$ be defined by (4.23). We conclude then, as in Sec. IV, that the function

$$S_2(\zeta)=S_1(\zeta)(\zeta-\lambda)^m/\left(\prod_{k=1}^{n}(\zeta-\zeta_k)\prod_{j=1}^{n_1}(\zeta-\bar\zeta_j)^2\right) \tag{A5}$$

is even in ζ. In (A5) we notice, in contrast to (4.26), the appearance of $\Pi_{j=1}^{n_1}(\zeta-\bar\zeta_j)^2$ corresponding to the known jumps of the phase by 2π at some of the points $\bar\zeta_i$ (n_1 of them). The further determination of $S_2(\zeta)$ proceeds as in Sec. IV.

If, for some $\bar\zeta_i'$, there exists a $\bar\zeta_j''$ so that $\bar\zeta_i'=\bar\zeta_j''=\zeta_0$, then we cannot find the discontinuity of the phase from knowledge of $\kappa(\zeta)$. We can nevertheless find its magnitude (mod 2π) from $R(\omega)$, Eq. (4.14), by looking[26] for the smallest value of α for which $\lim_{\zeta\to\zeta_0}\Omega_\alpha(\zeta)R(\omega(\zeta))$ is finite, where $\Omega_\alpha(\zeta)$ is the analog of (A1) with $\varphi_1(\zeta_0+0)=\varphi_1(\zeta_0-0)+\alpha$.

APPENDIX B: USE OF THE IMAGINARY PART IN THE UNPHYSICAL REGION $t<0$

We denote by γ_1,γ_2 the unphysical parts of the cuts of the complex ν plane at fixed t, and by Γ_1, Γ_2 the physical region. Then, if we assume the imaginary parts of B and C are obtained at $|\nu|<\bar\nu(t)$ from low energy phase shift analysis, and take into account that the roots in Eq. (2.6) are real, we have to solve the following problem: find all functions $h_+(\nu,t)$, $h_-(\nu,t)$, holomorphic and polynomially bounded in the cut ν plane, such that

$$|h_+(\nu)|_{\Gamma_1\cup\Gamma_2}=f_+(\nu),\quad |h_-(\nu)|_{\Gamma_1\cup\Gamma_2}=f_-(\nu), \tag{B1}$$

$$\operatorname{Im}h_+(\nu)|_{\gamma_1\cup\gamma_2}=d_+(\nu),\quad \operatorname{Im}h_-(\nu)|_{\gamma_1\cup\gamma_2}=d_-(\nu), \tag{B2}$$

$$|h_+(\nu)-h_-(\nu)|_{\Gamma_1\cup\Gamma_2}=f_0(\nu), \tag{B3}$$

where f_+,f_-,f_0,d_+,d_- are known continuous functions of ν. If we assume we know the area $\bar F(\nu,t)$, Eq. (3.5), then we can determine the ratio $\bar R(\nu)=h_+(\nu)/h_-(\nu)$ for all ν on $\Gamma_1\cup\Gamma_2$, by (5.2) and (5.3). The function $\bar R(\nu)$ is meromorphic in the cut ν plane and can be extended analytically to $\gamma_1\cup\gamma_2$. Its knowledge there can be used to solve for $\alpha_+=\operatorname{Re}h_+(\nu)$, $\alpha_-=\operatorname{Re}h_-(\nu)$, from

$$G'\equiv\operatorname{Re}\bar R(\nu)=\operatorname{Re}\frac{\alpha_++id_+}{\alpha_-+id_-}=\frac{\alpha_+\alpha_-+d_+d_-}{\alpha_-^2+d_-^2}, \tag{B4}$$

$$G''\equiv\operatorname{Im}\bar R(\nu)=\operatorname{Im}\frac{\alpha_++id_+}{\alpha_-+id_-}=\frac{d_+\alpha_--\alpha_+d_-}{\alpha_-^2+d_-^2}. \tag{B5}$$

Equations (B4) and (B5) yield an equation of the third order for α_-. It reads

$$G''\alpha_-^3+\alpha_-^2(G'd_--d_+)+G''d_-^2\alpha_-+G'd_-^3-d_-^2d_+=0. \tag{B6}$$

Standard algebraic tests reveal that if $d_-\neq0, G''\neq0$, this equation has just one real root. So, we can determine in principle h_+,h_- completely on $\gamma_1\cup\gamma_2$, independently of whether they have or not common zeros. Clearly, the analytic continuation of h_+,h_- from $\gamma_1\cup\gamma_2$ to the whole of $\Gamma_1\cup\Gamma_2$ is very unstable; it is not needed if h_+,h_- do not have common zeros. Indeed, in this case, their moduli are known on all $\gamma_1\cup\gamma_2\cup\Gamma_1\cup\Gamma_2$ and their zeros are the zeros and poles of $\bar R(\nu)$. The situations $d_-=0$, $G''=0$ are discussed in Ref. 43.

APPENDIX C: THE CONTINUITY OF THE AMPLITUDE AT POINTS OF THE PHYSICAL REGION

We wish to show the following. If a function $f(s,t)$

(a) is real holomorphic in s for $-\bar t_0\leqslant t\leqslant a$ in the s plane cut along $(s_{\rm th},\infty)$ and satisfies an unsubtracted dispersion relation in s for every t in $-\bar t_0\leqslant t\leqslant a$;

(b) $f(s,t)$ is continuous with respect to both Re s and Re t for s on the cuts and $-\bar{t}_0 \leqslant t \leqslant a$;

(c) $|\mathrm{Im} f(s,t)| \leqslant g(s) < cs^N$, $-\bar{t}_0 \leqslant t \leqslant a$, s real, with $g(s)$ continuous and such that $G(s) = \int_{s_{\mathrm{th}}}^{\infty} g(s')ds'/(s'-s)$ exists and is holomorphic in the s plane cut along $(s_{\mathrm{th}}, \infty)$, then $f(s,t)$ is continuous with respect to Re s, Im s, and Re t at points (s_0,t_0) with s_0 lying on the cuts and t_0 real.

Let $D_1 = \{s,t \in R \mid s \geqslant s_{\mathrm{th}}, -\bar{t}_0 \leqslant t \leqslant a\}$. We measure the distance between points (s,t) with Im $s \neq 0$ and points (s_0,t_0) of D_1 by means of

$$\|(s,t) - (s_0,t_0)\|^2 = \mathrm{Re}^2(s-s_0) + \mathrm{Im}^2(s-s_0) + (t-t_0)^2. \tag{C1}$$

We show that, for any $\epsilon > 0$ and any (s_0,t_0) in D_1, there exists $\delta(\epsilon,s_0,t_0)$ such that, if $\|(s,t) - (s_0,t_0)\| \leqslant \delta$, then $|f(s,t) - f(s_0,t_0)| \leqslant \epsilon$. To this end let s be such that $|s - s_0| \leqslant R$. Clearly

$$|f(s,t) - f(s_0,t_0)| \leqslant |f(s,t) - f(s,t_0)| + |f(s,t_0) - f(s_0,t_0)| \tag{C2}$$

and

$$\begin{aligned} |f(s,t) - f(s,t_0)| \leqslant & \left| \int_{s_{\mathrm{th}}}^{b(s_0,\epsilon)} ds' \frac{\mathrm{Im}\,(f(s',t) - f(s',t_0))}{s'-s} \right| \\ & + \left| \int_{b(s_0,\epsilon)}^{\infty} ds' \frac{\mathrm{Im} f(s',t_0)}{s'-s} \right| \\ & + \left| \int_{b(s_0,\epsilon)}^{\infty} ds' \frac{\mathrm{Im} f(s',t)}{s'-s} \right|, \end{aligned} \tag{C3}$$

where $b(s_0,\epsilon)$ is chosen such that $\int_{b(s_0,\epsilon)}^{\infty} g(s')ds'/(s'-s_0+R) \leqslant (\epsilon/8)$. In view of (c) above, the sum of the last two terms in (C3) is less than $\epsilon/4$. But the function

$$F(s,t) = (s-b)\int_{s_{\mathrm{th}}}^{b(s_0,\epsilon)} \frac{\mathrm{Im} f(s',t) - \mathrm{Im} f(s',t_0)}{s'-s} ds' \tag{C4}$$

is continuous with respect to s and t on the compact domain $-\bar{t}_0 \leqslant t \leqslant a s_{\mathrm{th}} \leqslant s \leqslant b$ because of (b) and the fact that $(s-b)\int_{b(s_0,\epsilon)}^{\infty} \mathrm{Im} f(s',t)\,ds'/(s'-s)$ is itself continuous at these points [as a consequence of (c)]. Further, $F(s,t_0) = 0$. On the other hand, from the maximum modulus principle, we get

$$|F(s,t)| \leqslant \sup_{s_{\mathrm{th}} \leqslant s' \leqslant b} |F(s',t)|. \tag{C5}$$

Since $F(s',t)$ is continuous on a compact set, it is uniformly continuous and so we can choose $\delta_1(b(s_0,\epsilon),t_0)$ so that, for $|t-t_0| \leqslant \delta_1$, $|F(s',t)| \leqslant \epsilon/4$ for all $s_{\mathrm{th}} \leqslant s' \leqslant b$ [since $F(s',t_0) = 0$ for all such s']. Thus, for all t in $|t-t_0| \leqslant \delta_1$, $|f(s,t) - f(s,t_0)| \leqslant \epsilon/2, |s-s_0| \leqslant R$.

We now turn to the second term in (C2). We map to this end the s plane cut along $s \geqslant s_{\mathrm{th}}$ onto the unit disk of a variable z. By writing a N times subtracted dispersion relation, we conclude that the imaginary part of $\tilde{f}(z) = (1+z)^{2N+1}(1-z)/(4z)^{N+1}(f(s(z)) - P_N(s(z))$, where $P_N(s)$ is a polynomial in s, can be represented as a Poisson integral of a certain function, continuous on $|z| = 1$. It follows that the real part can also be represented as a Poisson integral over its boundary values (Ref. 17, p. 54). But the values of such integrals converge to the boundary values Re $\tilde{f}(e^{i\theta})$, Im $\tilde{f}(e^{i\theta})$ at all points of continuity of the latter (Ref. 17, p.4). We still have to study $z = 1 (s = s_{\mathrm{th}})$. It is sufficient to notice that, by (a), $\bar{f}(z) \equiv \tilde{f}(z)/(z-1)$ is continuous at $z = 1$ along $|z| = 1$ and therefore bounded in the whole unit disk (by Phragmen–Lindelöf). Then $\bar{f}(z)$ can be represented as a Poisson integral over its continuous boundary values and this proves our assertion completely.

To generalize this statement to the case of a finite number of subtractions and of a left hand cut, one derives bounds for $|\mathrm{Im} B^{\pm}|, |\mathrm{Im} D^{\pm}|$ in terms of $(\partial/\partial t)(\mathrm{Im} D^{+}(1+z_s))$, $\mathrm{Im} D^{+}(1+z_s)$, respectively, calculated at $t = a$. This shows the continuity of all amplitudes with respect to Re s, Im s, Re t at points of D.

[1] R. Ayed. P. Bareyre, and Y. Lemoigne, CEN- Saclay preprint 1972 and Proceedings of the Aix-en-Provence Conference, 1973; R. Ayed, Thesis (Paris 1976) and Report CEA-N-1921, Centre d'Etudes Nucleaires de Saclay.

[2] E. Barrelet, Thesis (Paris 1970); Nuovo Cimento A **8**, 331 (1972); D. Chew, Proceedings of the 1976 Oxford Topical Conference on Baryon Resonances, edited by R. T. Ross, D. H. Saxon (Rutherford Laboratory, Chilton, United Kingdom, 1976), p. 156.

[3] E. Pietarinen, Nucl. Phys. B **107**,21 (1976).

[4] R. E. Cutkosky, C. P. Forsyth, R. E. Hendrick, and R. L. Kelly, Preprint COO-3066-117, LBL-8553 (1979); D. P. Hodgkinson *et al.*, Proceedings of the 1976 Conference on Baryonic Resonances (Rutherford Laboratory, Chilton, United Kingdom, 1976), p. 41.

[5] J. H. Crichton, Nuovo Cimento A **45**, 256 (1966).

[6] F. A. Berends, Nucl. Phys. B **64**, 236 (1973).

[7] J. E. Bowcock and H. Burkhardt, Rep. Prog. Phys. **38**, 1099 (1975).

[8] D. Atkinson, P. Johnson, and R. Warnock, Commun. Math. Phys. **28B** 133 (1972).

[9] G. R. Bart, P. W. Johnson, and R. L. Warnock, J. Math. Phys. **14**, 1558 (1973).

[10] G. R. Bart, P. W. Johnson, and R. L. Warnock, Nucl. Phys. B **72**, 329 (1974).

[11] D. Atkinson, G. Mahoux, and F. J. Yndurain, Nucl. Phys. B **66**, 429 (1973).

[12] D. Atkinson, A. C. Heemskerk, and S. D. Swiestra, Nucl. Phys. B **109**, 322 (1976).

[13] D. Atkinson, L. P. Kok, and M. de Roo, Phys. Rev. D **17**, 2492 (1978).

[14] H. Burkhardt and A. Martin, Nuovo Cimento A **29**, 141 (1975).

[15] E. Pietarinen, Phys. Scripta **14**, 11 (1976).

[16] A. Martin, *Unitarity, Analyticity and Crossing, Lecture Notes in Physics* (Springer, Berlin, 1969).

[17] H. P. Duren, *Theory of H^p Spaces* (Academic, New York, 1970).

[18] G. Höhler, F. Kaiser, R. Koch, and E. Pietarinen, Handbook of Pion Nucleon Scattering, ZAED Report 12-1 (1979).

[19] G. Mahoux and A. Martin, Phys. Rev. **174**, 2140 (1968).

[20] G. Sommer, Nuovo Cimento A **52**, 373 (1967).

[21] G. Sommer, Fortschr. Phys. **18**, 577 (1970).

[22] L. Michel, Nuovo Cimento **22**, 203 (1961).

[23] M. G. Doncel, L. Michel, and P. Minnaert, Phys. Lett. B **38**, 42 (1972).

[24] S. M. Roy, Phys. Rep. **5**, 125 (1972).

[25] G. Höhler, F. Kaiser, H. Kohl, and W. Schmidt, Lett. Nuovo Cimento **5**, 413 (1972).

[26] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).

[27] H. Cornille and A. Martin, Nucl. Phys. B **48**, 104 (1972).

[28] W. F. Osgood, *Lehrbuch der Funktionentheorie, Bd. II* (Chelsea, New York, 1965).

[29] A. de Lesquen *et al.*, Phys. Lett. B **40**, 277 (1972).

[30] J. Pierrard *et al.*, Phys. Lett. B **57**, 393 (1975); **61**, 107 (1976).

[31] J. Pierrard, Thesis, Note CEA-N-208, May 1978.

[32]G. Höhler and R. Strauss, Z. Phys. **232**, 205 (1970).
[33]F. J. Gilman, J. Pumplin, A. Schwimmer, and L. Stodolsky, Phys. Lett. B **31**, 387 (1970).
[34]G. Höhler, H. P. Jakob, and F. Kaiser, Preprint TKP 11/72, unpublished.
[35]C. Bruneton *et al.*, Nucl. Phys. B **124**, 391 (1977).
[36]M. Borghini *et al.*, Phys. Lett. B **31**, 405 (1970); B **36**, 493 (1971); J.A. Scheid *et al.*, Phys. Rev. D **8**, 1263 (1973).
[37]A. Gaidot *et al.*, Phys. Lett. B **57**, 389 (1975); B **61**, 103 (1976); M. Fujisaki *et al.*, Nucl. Phys. B **151**, 206 (1979); I. P. Auer *et al.*, Phys. Rev. Lett. **39**, 313 (1977); see also E. Borie and H. P. Jakob, Phys. Lett. B **78**, 323 (1978).
[38]W. D. Apel *et al.*, Phys. Lett. B **72**, 132 (1977).
[39]G. Höhler and F. Kaiser, Review and Tables of Pion Nucleon Forward Amplitudes, KfK Report No. 3027, August 1980.
[40]G. Höhler, H. P. Jakob and R. Strauss, Nucl. Phys. B **39**, 237 (1972).
[41]R. Koch, Thesis, Karlsruhe 1978 (unpublished).
[42]E. Borie, W. Gampp, G. Höhler, R. Koch, and I. S. Stefanescu, Z. Phys. C **4**, 333 (1980); Preprint TKP 78/18, to be published.
[43]I. S. Stefanescu, Preprint TKP 78/20.

Complete set of SU(5) monopole solutions

I. G. Koh[a)]
Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, Illinois 60510
Physics Department, Sogang University, C.P.O. Box 1142, Seoul, Korea

Y. D. Kim, Y. J. Park, C. H. Kim, and Y. S. Kim
Physics Department, Sogang University, C.P.O. Box 1142, Seoul, Korea

(Received 4 November 1981; accepted for publication 29 January 1982)

The manifestly nonsingular asymptotic forms of the most general possible pointlike SU(5) monopoles are found by using the magnetic symmetry introduced by Cho. New magnetic tensor representations greatly expedite the computations. These magnetic tensors explicitly exhibit the full homotopy class of the mappings classifying the arbitrary allowed magnetic charges.

PACS numbers: 14.80.Hv, 11.30.Ly, 02.40. + m

The nontrivial topological structures[1] of the monopoles have been much studied in the gauge field theory. No regular monopole solutions exist for the abelian gauge group. However for nonabelian SU(2) gauge group, Wu and Yang have explicitly found regular monopole solutions except the singularity at the origin.[2] The remaining singularity at the origin was removed by 't Hooft to give regular solutions everywhere with finite mass for the monopole.[3] Monopoles in the minimal grand unified gauge group SU(5) get wide attention due to the cosmological implications.[4] To suppress the production rate of monopoles, Guth and Tye[5] and Langacker and Pi[6] have introduced several different scenarios. In these circumstances, it would be very interesting to find a complete set of monopole solutions for the grand unified group SU(5). Wilkinson,[7] Bais and Weldon,[8] Wilkinson and Bais,[9] Brandt and Neri,[10] Weinberg,[11] Daniel, Lazarides, and Shafi,[12] and Dokos and Tomaras[13] have studied monopoles in the nonabelian gauge group.

Cho[14] has obtained all the monopole solutions for the SU(3) gauge group by examining the topological structure, instead of using the usual methods of solving the equations of motion. The magnetic symmetry, as a set of self-consistent Killing vector fields, plays a crucial role in finding a complete set of monopole solutions. Recently, we have obtained a complete set of monopole solutions in the SU(4) gauge group.[15] Cho and we have used the generators in adjoint representations, and the calculations become quite complicated for large SU(N) group. In this paper, we obtained the manifestly nonsingular asymptotic forms of the most general possible SU(5) monopoles by using 5×5 tensor representation of SU(5) instead of the 24-dimensional representation for the Killing vectors. These tensor representations considerably expedite the computations.

The monopole solutions are classified by the homotopy group,[16] and the nontrivial homotopy groups for SU(5) are

$$\pi_2[\mathrm{SU}(5)/\mathrm{U}(1)\times\mathrm{U}(1)\times\mathrm{U}(1)\times\mathrm{U}(1)] = \pi_1[\mathrm{U}(1)\times\mathrm{U}(1)\times\mathrm{U}(1)\times\mathrm{U}(1)] = Z\times Z\times Z\times Z, \quad (1)$$

$$\pi_2[\mathrm{SU}(5)/\mathrm{U}(2)\times\mathrm{U}(1)\times\mathrm{U}(1)] = \pi_1[\mathrm{U}(2)\times\mathrm{U}(1)\times\mathrm{U}(1)] = Z\times Z\times Z, \quad (2)$$

$$\pi_2 = [\mathrm{SU}(5)/\mathrm{U}(3)\times\mathrm{U}(1)] = \pi_1[\mathrm{U}(3)\times\mathrm{U}(1)] = Z\times Z, \quad (3)$$

$$\pi_2 = [\mathrm{SU}(5)/\mathrm{U}(4)] = Z, \quad (4)$$

where $\mathrm{U}(1)\times\mathrm{U}(1)\times\mathrm{U}(1)\times\mathrm{U}(1)$, $\mathrm{U}(2)\times\mathrm{U}(1)\times\mathrm{U}(1)$, $\mathrm{U}(3)\times\mathrm{U}(1)$, and U(4) are subgroups associated with λ_3-, λ_8-, λ_{15}-, and λ_{24}-like symmetry, respectively. Therefore the monopoles must now be classified by four, three, two, and one integers, respectively.

The magnetic charges must satisfy the generalized quantization condition[17] given by

$$\exp[4\pi i g(\tfrac{1}{2}\lambda_3 g_1 + \tfrac{1}{2}\lambda_8 g_2 + \tfrac{1}{2}\lambda_{15} g_3 + \tfrac{1}{2}\lambda_{24} g_4)] = 1, \quad (5)$$

where λ_3, λ_8, λ_{15}, and λ_{24} are self-commuting generators of the SU(5) group. To satisfy the above condition, one can take

$$g_1 = g^{-1}(n_1 - \tfrac{1}{2}n_2 - \tfrac{1}{2}n_3 - \tfrac{1}{2}n_4), \quad (6)$$

$$g_2 = g^{-1}\tfrac{1}{2}\sqrt{3}\left[n_2 + \tfrac{1}{3}(n_3 + n_4)\right], \quad (7)$$

$$g_3 = g^{-1}(2/3)^{1/2}[n_3 + \tfrac{1}{4}n_4], \quad (8)$$

$$g_4 = g^{-1}(10^{1/2}/4)n_4, \quad (9)$$

where n_1, n_2, n_3, and n_4 are arbitrary integers. The value of the magnetic charges depends on the topological configurations which are expressed in Eqs. (1)–(4). When n_4 is set to zero, the monopole changes are reduced to the SU(4) case.[15] Here additional zeroes for n_3 would give SU(3) results.[14] Subsequent zeroes for n_2 will finally give SU(2) results.

To find a complete set of the monopole solutions for every set of integers n_1, n_2, n_3, and n_4, the Killing symmetry assumptions[14] introduced by Cho are required to be

$$D_\mu m_1 = 0, \quad D_\mu m_2 = 0, \quad D_\mu m_3 = 0, \text{ and } D_\mu m_4 = 0, \quad (10)$$

where

$$D_\mu = \partial_\mu - ig[B_\mu, \quad \text{and} \quad B_\mu = B^i_\mu \lambda^i/2. \quad (11)$$

The gauge potential that satisfies the above Killing symmetry assumption should be given by the following form:

$$B_\mu = A^1_\mu m_1 + A^2_\mu m_2 + A^3_\mu m_3 + A^4_\mu m_4 - ig^{-1}[m_1,\partial_\mu m_1] - ig^{-1}[m_2,\partial_\mu m_2] - ig^{-1}[m_3,\partial_\mu m_3] - ig^{-1}[m_4,\partial_\mu m_4], \quad (12)$$

where A^1_μ, A^2_μ, A^3_μ, and A^4_μ are the components not fixed by the condition (10).

[a)]Senior Fulbright Fellow.

 0022-2488/82/061210-03$02.50

The magnetic tensors m_1, m_2, m_3, and m_4 are chosen to exhibit the full homotopy class of the mappings (1)–(3). These are found by the gauge transformation such that

$$m_1 = U\tfrac{1}{2}\lambda_3 U^\dagger, \quad m_2 = U\tfrac{1}{2}\lambda_8 U^\dagger,$$
$$m_3 = U\tfrac{1}{2}\lambda_{15} U^\dagger, \quad \text{and} \quad m_4 = U\tfrac{1}{2}\lambda_{24} U^\dagger, \tag{13}$$

where

$$\begin{aligned} U = \exp\{ & in_4\tfrac{1}{2}\phi\,[\tfrac{1}{2}\lambda_3 - \sqrt{3}/2\lambda_8 + (3/2)^{1/2}\lambda_{15} \\ & - 3/10^{1/2}\lambda_{24}]\}\exp(-i\tfrac{1}{2}\theta\lambda_{19})\exp\{in_3\tfrac{1}{2}\phi\,[\tfrac{1}{2}\lambda_3 \\ & - \sqrt{3}/2\lambda_8 + (3/2)^{1/2}\lambda_{15}]\}\exp(-i\tfrac{1}{2}\theta\lambda_{10})\exp\{in_2\tfrac{1}{2}\phi\,[\tfrac{1}{2}\lambda_3 \\ & - \sqrt{3}/2\lambda_8]\}\exp(-i\tfrac{1}{2}\theta\lambda_7)\exp[-i\tfrac{1}{2}\phi\lambda_3(n_1 - \tfrac{1}{2}n_2 \\ & - \tfrac{1}{2}n_3 - \tfrac{1}{2}n_4)]\exp(-i\tfrac{1}{2}\theta\lambda_2). \end{aligned} \tag{14}$$

Here $\lambda_i (i = 1,\ldots,24)$ are the fundamental representations of the SU(5) generators. Explicit forms of magnetic tensors are shown in Table I. These m_1, m_2, m_3, and m_4 represent the homotopic mappings (1), (2), (3), and (4), respectively. These magnetic tensors in Table I are regular except at the origin. Since the ϕ part in the magnetic tensors accompanies $\sin\theta$, their derivatives are also regular except at the origin.

The magnetic tensors are invariant under the additional gauge transformation by the diagonal generators, i.e.,

$$m_1 = U\tfrac{1}{2}\lambda_3 U^\dagger = U'\tfrac{1}{2}\lambda_3 U'^\dagger, \tag{15}$$
$$m_2 = U\tfrac{1}{2}\lambda_8 U^\dagger = U'\tfrac{1}{2}\lambda_8 U'^\dagger, \tag{16}$$
$$m_3 = U\tfrac{1}{2}\lambda_{15} U^\dagger = U'\tfrac{1}{2}\lambda_{15} U'^\dagger, \tag{17}$$

and

$$m_4 = U\tfrac{1}{2}\lambda_{24} U^\dagger = U'\tfrac{1}{2}\lambda_{24} U'^\dagger \tag{18}$$

with

$$U' = U\exp[-i\tfrac{1}{2}(\alpha\lambda_3 + \beta\lambda_8 + \gamma\lambda_{15} + \delta\lambda_{24})] \tag{19}$$

for arbitrary α, β, γ, and δ.

Finally the unrestricted A^1_μ, A^2_μ, A^3_μ, and A^4_μ in Eq. (12) are chosen to obtain the desired monopole solutions as

$$\begin{aligned} A^1_\mu &= -(1/g)\{[(\tfrac{1}{2}n_2 + n_3 + n_4) + \tfrac{1}{4}n_4\cos\theta\,] \\ &\quad \times \sin^2\theta\,\partial_\mu\phi + \tfrac{1}{2}(n_3 + n_4)\phi\sin\theta\,\partial_\mu\theta\}, \\ A^2_\mu &= (1/g)\sqrt{3}/4\sin\theta\,[n_4\sin\theta\,\partial_\mu\phi + (2n_3/3)\phi\,\partial_\mu\theta\,], \\ A^3_\mu &= (1/g)(2/3)^{1/2}(2n_3 + n_4)\phi\sin\theta\,\partial_\mu\theta, \\ A^4_\mu &= 0 \end{aligned} \tag{20}$$

with

$$\begin{aligned} \alpha &= [-\tfrac{1}{2}(n_3 + n_4)\cos\theta + (\tfrac{1}{2}n_2 + n_3 + n_4)]\phi, \\ \beta &= -\sqrt{3}[\tfrac{1}{6}n_3\cos\theta + (\tfrac{1}{6}n_3 + \tfrac{1}{2}n_4)]\phi, \\ \gamma &= [(2/3)^{1/2}(2n_3 + n_4)\cos\theta + 6^{1/2}(\tfrac{1}{6}n_3 + \tfrac{1}{4}n_4)]\phi, \\ \delta &= (-10^{1/2}/20)n_4\phi. \end{aligned} \tag{21}$$

Earlier results for Eqs. (20) and (21) in SU(4), SU(3), and SU(2) can be easily recovered by setting n_4, n_3, and n_2 to zero subsequently.

Since the magnetic Killing tensors in Table I, their derivatives, and the A^1, A^2_μ, A^3_μ, and A^4_μ in Eq. (20) are smooth, the potential (12) is regular everywhere except at the origin. If the gauge potential (12) is expressed in terms of m_1,

TABLE I. Explicit forms of the magnetic tensors with $a = \sin\theta$, $b = \cos\theta$, $c \equiv \sin(\theta/2)$, and $d = \cos(\theta/2)$.

$$m_1 = \tfrac{1}{2}\begin{bmatrix} b(1+b) & \tfrac{1}{2}(1+b)a\exp[-i(n_1-n_2-n_3-n_4)\phi] & \tfrac{1}{2}a^2\exp[-in_4\phi] & \tfrac{1}{2}abexp[i(n_3+n_4)\phi] & \tfrac{1}{2}a^2dexp[-i(n_1-n_2-n_3)\phi] \\ \tfrac{1}{2}(1+b)a\exp[i(n_1-n_2-n_3-n_4)\phi] & -\tfrac{1}{2}b(1+b)^2 & -\tfrac{1}{2}abexp[-i(n_2+n_3+n_4)\phi] & -\tfrac{1}{2}b(1-b) & -\tfrac{1}{2}(1-b)a\exp[i(n_1+n_3+n_4)\phi] \\ \tfrac{1}{2}a^2\exp[in_4\phi] & -\tfrac{1}{2}abdexp[i(n_2+n_3+n_4)\phi] & -\tfrac{1}{2}a^2dexp[-i(n_1-n_2)\phi] & \tfrac{1}{2}(1-b)a\exp[-i(n_1+n_3+n_4)\phi] & \tfrac{1}{2}b(1-b) \\ \tfrac{1}{2}abexp[-i(n_1+n_4)\phi] & -\tfrac{1}{2}a^2dexp[-i(n_1-n_2-n_3)\phi] & -\tfrac{1}{2}abcexp[-i(n_1+n_2+n_3)\phi] & \tfrac{1}{2}a^2cexp[i(n_1-n_2-n_3)\phi] & -\tfrac{1}{2}(1-b^2)b \\ \tfrac{1}{2}a^2dexp[i(n_1-n_2-n_3)\phi] & -\tfrac{1}{2}ab(1+b)\exp(in_4\phi) & -\tfrac{1}{2}abcexp[-i(n_2+n_3)\phi] & \tfrac{1}{2}a^2cexp[i(n_1-n_2-n_3)\phi] & -\tfrac{1}{2}(1-b^2)b \end{bmatrix}$$

$$m_2 = \frac{1}{2\sqrt{3}}\begin{bmatrix} \tfrac{1}{2}(1+b) & 0 & 0 & \tfrac{1}{2}a\exp[i(n_3+n_4)\phi] & 0 \\ 0 & \tfrac{1}{4}(3b-1)(b+1) & \tfrac{1}{4}da\exp[-i(n_2+n_3+n_4)\phi] & 0 & \tfrac{1}{4}a(3b-1)\exp[-in_4\phi] \\ 0 & \tfrac{3}{4}da\exp[i(n_2+n_3+n_4)\phi] & -\tfrac{1}{4}(1+3b) & 0 & \tfrac{3}{4}ac\exp[i(n_2+n_3)\phi] \\ \tfrac{1}{2}a\exp[-i(n_3+n_4)\phi] & 0 & \tfrac{1}{4}a(3b-1)\exp(in_4\phi) & \tfrac{1}{2}(1-b) & 0 \\ 0 & \tfrac{1}{4}a(3b-1)\exp(in_4\phi) & \tfrac{3}{4}ac\exp[-i(n_2+n_3)\phi] & 0 & -\tfrac{1}{4}(3b-1)(b-1) \end{bmatrix}$$

$$m_3 = \frac{1}{2\cdot 6^{1/2}}\begin{bmatrix} 2b-1 & 0 & 0 & 2a\exp[i(n_3+n_4)\phi] & 0 \\ 0 & \tfrac{1}{2}(1+b) & 0 & 0 & \tfrac{1}{2}a\exp[-in_4\phi] \\ 0 & 0 & 1 & 0 & 0 \\ 2a\exp[-i(n_3+n_4)\phi] & 0 & 0 & -2b-1 & 0 \\ 0 & \tfrac{1}{2}a\exp(in_4\phi) & 0 & 0 & \tfrac{1}{2}(1-b) \end{bmatrix}$$

$$m_4 = \frac{1}{2\cdot 10^{1/2}}\begin{bmatrix} - & 0 & 0 & 0 & 0 \\ 0 & -\tfrac{3}{2}+\tfrac{5}{2}b & 0 & 0 & \tfrac{5}{2}a\exp(-in_4\phi) \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \tfrac{5}{2}a\exp(in_4\phi) & 0 & 0 & -\tfrac{3}{2}-\tfrac{5}{2}b \end{bmatrix}$$

m_2, m_3, and m_4 and of the pure gauge terms as

$$B_\mu = (g_1 m_1 + g_2 m_2 + g_3 m_3 + g_4 m_4) \times \cos\theta\, \partial_\mu \phi - i g^{-1} U' \partial_\mu U'^{\dagger}, \quad (22)$$

the string singularity in the first term of Eq. (22) is cancelled by the second term. In the λ gauge, the potential (22) reduces to the standard Dirac potential form with the string singularity as

$$(g_1\lambda_3/2 + g_2\lambda_8/2 + g_3\lambda_{15}/2 + g_4\lambda_{24}/2)\cos\theta\partial_\mu\phi. \quad (23)$$

But it is to be emphasized that the potential (12) with the regular magnetic tensors in Table I and with Eq. (20) is regular everywhere except at the origin.

This potential B_μ describes the desired solutions

$$G_{\mu\nu} = -(g_1 m_1 + g_2 m_2 + g_3 m_3 + g_4 m_4) \times \sin\theta\, (\partial_\mu\theta\partial_\nu\phi - \partial_\nu\theta\partial_\mu\phi), \quad (24)$$

using $G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu - ig\,[B_\mu, B_\nu]$. The magnetic charges g_m^1, g_m^2, g_m^3, and g_m^4 of the solutions can then be defined as

$$g_m^1 = 2\int dS^{\mu\nu}\mathrm{Tr}(m_1 G_{\mu\nu}) = 4\pi g^{-1}(n_1 - \tfrac{1}{2}n_2 - \tfrac{1}{2}n_3 - \tfrac{1}{2}n_4),$$

$$g_m^2 = 2\int dS^{\mu\nu}\mathrm{Tr}(m_2 G_{\mu\nu}) = 4\pi g^{-1}(\sqrt{3}/2)(n_2 + \tfrac{1}{3}n_3 + \tfrac{1}{3}n_4), \quad (25)$$

$$g_m^3 = 2\int dS^{\mu\nu}\mathrm{Tr}(m_3 G_{\mu\nu}) = 4\pi g^{-1}(2/3)^{1/2}(n_3 + \tfrac{1}{4}n_4),$$

$$g_m^4 = 2\int dS^{\mu\nu}\mathrm{Tr}(m_4 G_{\mu\nu}) = 4\pi g^{-1}(10^{1/2}/4)n_4.$$

The potential (12) with magnetic tensors in Table I and with Eq. (20) indeed describes all the homotopically inequivalent pointlike SU(5) monopoles.

At this point, it is appropriate to comment why we take such a long computation to find the gauge potential (22). To obtain the field tensor (24) for monopole, one would simply write down Eq. (23) with help of the generalized quantization condition (5). But there is no easy practical way to find the regular gauge potential (12) or (22) from the singular one (23). But the new method of magnetic tensor described in this paper is quite straightforward to generalize the regular gauge potential of Wu and Yang[2] in SU(2) for SU(N). The remaining singularity of gauge potential (22) at the origin can be eliminated by introducing Higgs fields as in the SU(2) by 't Hooft[3] and Polyakov.[3] This is much more ambitious than we are aiming for in this paper.

ACKNOWLEDGMENTS

We are grateful for Professor Y. M. Cho's critical comments and encouragements. One of us (I.G.K.) would like to thank members of the Fermilab Theory Group, in particular, Professor W. A. Bardeen, Dr. C. K. Zachos, and Dr. J. F. Schonfeld, for helpful discussion and also for warm hospitality. We appreciate the help of W. S. l'Yi and S. H. Lee in computation. These calculations are also checked independently by symbolic computation program REDUCE[18]; for this program we thank Professor A. C. Hearn. This research was supported in part by the Korea Science and Engineering Foundation and the Korea Traders Foundation.

[1] J. Arafune, P. G. G. Freund, and C. G. Goebel, J. Math. Phys. **16**, 433 (1975); S. Coleman, in *New Phenomena in Subnuclear Physics*, Proceedings of the 14th Course of the International School of Subnuclear Physics, Erice, 1975, edited by A. Zichichi (Plenum, New York, 1977).
[2] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[3] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974); A. M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)].
[4] J. P. Preskill, Phys. Rev. Lett. **43**, 1365 (1979).
[5] A. Guth and S.-H. Tye, Phys. Rev. Lett. **44**, 631 (1980); M. B. Einhorn, D. L. Stein, and D. Toussaint, Univ. of Michigan Preprint UM HE 80-1 (University of Michigan, 1980).
[6] P. Langacker and S. Pi, Phys. Rev. Lett. **45**, 1 (1980).
[7] D. Wilkinson, Nucl. Phys. B **125**, 423 (1977).
[8] F. A. Bais and H. A. Weldon, Phys. Rev. Lett. **41**, 601 (1978).
[9] D. Wilkinson and F. A. Bais, Phys. Rev. D **19**, 2410 (1979).
[10] R. A. Brandt and F. Neri, Nucl. Phys. B **186**, 84 (1981).
[11] E. J. Weinberg, Nucl. Phys. B **167**, 500 (1980).
[12] M. Daniel, G. Lazarides, and Q. Shafi, Nucl. Phys. B **170**, 156 (1980).
[13] C. P. Dokos and T. N. Tomaras, Phys. Rev. D **21**, 2940 (1980).
[14] Y. M. Cho, Phys. Rev. Lett. **44**, 1115 (1980); Y. M. Cho, Phys. Rev. D **21**, 1080 (1980).
[15] Y. D. Kim, I. G. Koh, and Y. J. Park, Phys. Rev. D. **25**, 587 (1982).
[16] A. Sinha, Phys. Rev. D **14**, 2016 (1976); A. Actor, Rev. Mod. Phys. **51**, 461 (1979); E. Corrigan, D. Olive, D. B. Fairlie, and J. Nuyts, Nucl. Phys. B **106**, 475 (1976).
[17] Y. S. Tyupkin, V. A. Fateev, and A. S. Shvarts, Pis'ma Zh. Eksp. Teor. Fiz. **21**, 91 (1975) [JETP Lett. **21**, 42 (1975)]; M. I. Mastyrsky and A. M. Perelomov, Pis'ma Zh. Eksp. Teor. Fiz. **21**, 94 (1975) [JETP Lett. **21**, 43 (1975)]; F. Englert and P. Windey, Phys. Rev. D **14**, 2728 (1976); P. Goddard, J. Nuyts, and D. Olive, Nucl. Phys. B. **125**, 1 (1977).
[18] A. C. Hearn, *REDUCE 2, User's Manual* (University of Utah, Salt Lake City, 1973).

Scattering of electromagnetic waves from random media with multiple scattering included

Ronald L. Fante
Avco Systems Division, Wilmington, Massachusetts 01887

(Received 18 January 1982; accepted for publication 19 February 1982)

Closed-form results which include all orders of multiple scattering have been obtained for the field and intensity when an arbitrary electromagnetic wave is incident on a randomly inhomogeneous medium. In obtaining the results it is assumed that the magnitude of the permittivity fluctuations is small in comparison with unity and the spatial extent of the fluctuations is large in comparison with the wavelength. Simplified approximations, which are valid outside the backscatter cone, are then obtained. These approximations are found to depend only on the second moments of the field in the medium, and are valid if the mean free path between photon scatterings is large in comparison with the size of the inhomogeneities.

PACS numbers: 42.20. – y

I. INTRODUCTION

It is conventional[1–10] when calculating the scattering of electromagnetic waves from randomly inhomogeneous media to use the single scattering approximation (Born approximation). It happens that in most cases encountered in practice this approximation is acceptable, but there are some situations where it is not. For example, recent data[11–14] on lidar scattering from clouds has indicated that single scattering theory is inadequate. Some effort has been made to use radiative transport theory to include the effect of multiple scattering but this effort has been limited primarily to simple planar geometries,[15–23] and results are not generally available for arbitrary geometries. In this paper we will attempt to partially remedy this deficiency by obtaining an expression for the scattered intensity which includes all orders of multiple scattering within the medium, provided the following conditions are satisfied: (1) The magnitude of the mean-square fluctuation in relative permittivity in the medium is small in comparison with unity; (2) the size of the random inhomogeneities is large in comparison with a wavelength so that most of the scattering is in the forward direction and depolarization is not important; and (3) the mean free path between (photon) scatterings is large in comparison with the spatial size of the random inhomogeneities. When the aforementioned conditions are satisfied, we will demonstrate that it is possible to obtain relatively simple expressions for the ensemble-averaged field and intensity scattered by an arbitrary random medium. The expression for the scattered intensity reduces to the well-known Booker–Gordon formula (i.e., single scatter) when the spatial coherence length of the field inside the scattering medium is sufficiently large.

II. ANALYTICAL PRELIMINARIES

Consider an arbitrary field $\mathbf{e}_0(\mathbf{r})$ incident along the z axis on a scattering volume V with a relative permittivity distribution given by $\epsilon(\mathbf{r}) = 1 + \eta(\mathbf{r})$, where $\eta(\mathbf{r})$ is a random real function of the position $\mathbf{r}$, with the property $|\eta| < 1$. Then by a straightforward extension of the analysis presented by Tatarski,[4] it is readily shown that the electric field scattered by this medium is given, in the Fraunhofer zone, by

$$\mathbf{E}_s(\mathbf{m}) = \frac{k^2}{4\pi R}\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3r\,\eta(\mathbf{r})\{\mathbf{e}(\mathbf{r}) - \mathbf{m}[\mathbf{m}\cdot\mathbf{e}(\mathbf{r})]\} \times \exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR), \tag{1}$$

where k is the vacuum wavenumber of the radiation, $\mathbf{m}$ is a unit vector directed towards the observer (see Fig. 1) from an arbitrary origin within the scattering medium, R is the distance from the origin to the observer, and $\mathbf{e}(\mathbf{r})$ is the actual electric field inside the scattering medium. It can be demonstrated (see Appendix A) that the error made in using (1) is at most of order η^2, in comparison with (1). Note that we will *not* assume that $\mathbf{e}(\mathbf{r})$ is equal to the incident field $\mathbf{e}_0(\mathbf{r})$; this assumption would lead to the usual Born approximation.

We will now assume that the scale size (correlation length) of the random inhomogeneities is large in comparison with the wavelength of the incident field. In this case any cross polarization of the field can be ignored, so that the field inside the scattering medium is polarized in the same direction as the incident field. Consequently, if we assume that the incident field is linearly polarized along the x axis, we can rewrite (1) as

$$\mathbf{E}_s(\mathbf{m}) = \frac{k^2(\hat{x} - \mathbf{m}\cos\psi)}{4\pi R}\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3r\,\eta(\mathbf{r})e(\mathbf{r}) \times \exp(ikR - ik\mathbf{m}\cdot\mathbf{r}), \tag{2}$$

where $e(\mathbf{r})$ is now a scalar, $\cos\psi = \mathbf{m}\cdot\hat{x}$ and $\hat{x}$ is a unit vector along the x axis. From (2) we can readily obtain formal expressions for the average field and average intensity scattered from a randomly inhomogeneous medium. These are

$$\langle\mathbf{E}_s(\mathbf{m})\rangle = \frac{k^2(\hat{x} - \mathbf{m}\cos\psi)}{4\pi R}\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3r\,\langle\eta(\mathbf{r})e(\mathbf{r})\rangle \times \exp(ikR - ik\mathbf{m}\cdot\mathbf{r}), \tag{3}$$

 0022-2488/82/061213-10$02.50

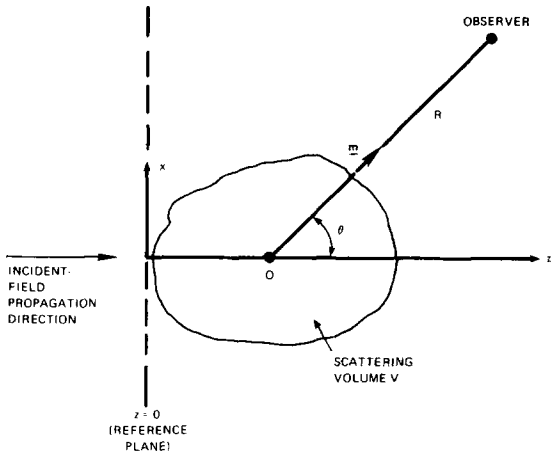


FIG. 1. Geometry for scattering by a randomly inhomogeneous volume.

$$P_s(\mathbf{m}) = \frac{k^4 \sin^2\psi}{(4\pi R)^2} \int_{-\infty}^{\infty} \cdots \int d^3r d^3 r' \times \langle \eta(\mathbf{r})\eta(\mathbf{r}')e(\mathbf{r})e^*(\mathbf{r}')\rangle \times \exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')], \tag{4}$$

where $\langle\ \rangle$ denotes an ensemble average and $P_s \equiv \langle \mathbf{E}_s \cdot \mathbf{E}_s^* \rangle$. In order to evaluate Eqs. (3) and (4), we need to calculate $\langle \eta(\mathbf{r})e(\mathbf{r})\rangle$ and $\langle \eta(\mathbf{r})\eta(\mathbf{r}')e(\mathbf{r})e^*(\mathbf{r}')\rangle$. These averages can be computed using the Novikov–Furutsu theorem. This states that if $\eta(\mathbf{r})$ is a zero-mean, Gaussian random variable and $f[\eta]$ is an arbitrary functional of η, then

$$\langle \eta(\mathbf{r}) f[\eta]\rangle = \int_{-\infty}^{\infty}\int\int d^3r' \langle \eta(\mathbf{r})\eta(\mathbf{r}')\rangle \left\langle \frac{\delta f}{\delta \eta(\mathbf{r}')} \right\rangle, \tag{5}$$

where $\delta / \delta\eta(\mathbf{r}')$ is a variational derivative. The proof of this theorem is presented in Refs. 9 and 24.

If we set $f[\eta] = e(\mathbf{r})$ and then apply Eq. (5) to (3), we obtain

$$\langle \mathbf{E}_s \rangle = \frac{-k^4(\hat{x} - \mathbf{m}\cos\psi)}{4\pi R} \int_{-\infty}^{\infty} \cdots \int d^3r d^3r' \langle e(\mathbf{r}')G(\mathbf{r},\mathbf{r}')\rangle \times \langle \eta(\mathbf{r})\eta(\mathbf{r}')\rangle \exp(ikR - ik\mathbf{m}\cdot\mathbf{r}). \tag{6}$$

In obtaining (6) we have used the result, derived in Appendix B, that

$$\frac{\delta e(\mathbf{r})}{\delta \eta(\mathbf{r}')} = -k^2 e(\mathbf{r}')G(\mathbf{r},\mathbf{r}'), \tag{7}$$

where $G(\mathbf{r}, \mathbf{r}')$ is the stochastic Green's function for the operator, $\nabla^2 + k^2(1+\eta)$.

The average required in Eq. (4) is somewhat more difficult to evaluate, and requires the application of Eq. (5) twice. We first set $f[\eta] = \eta(\mathbf{r}')e(\mathbf{r})e^*(\mathbf{r}')$. We then find using (5) that $T_1 \equiv \langle \eta(\mathbf{r})\eta(\mathbf{r}')e(\mathbf{r})e^*(\mathbf{r}')\rangle$ is given by

$$T_1 = \int_{-\infty}^{\infty}\int\int d^3r_2 \langle \eta(\mathbf{r})\eta(\mathbf{r}_2)\rangle \left\langle e(\mathbf{r})e^*(\mathbf{r}')\frac{\delta\eta(\mathbf{r}')}{\delta\eta(\mathbf{r}_2)} + \eta(\mathbf{r}')e^*(\mathbf{r}')\frac{\delta e(\mathbf{r})}{\delta\eta(\mathbf{r}_2)} + \eta(\mathbf{r}')e(\mathbf{r})\frac{\delta e^*(\mathbf{r}')}{\delta\eta(\mathbf{r}_2)} \right\rangle. \tag{8}$$

It is demonstrated elsewhere[24] that $\delta\eta(\mathbf{r})/\delta\eta(\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')$, where $\delta(\mathbf{r}-\mathbf{r}')$ is the three-dimensional Dirac delta function. Also, $\delta e/\delta\eta$ can be evaluated using (6). Consequently, (8) becomes

$$T_1 = \langle \eta(\mathbf{r})\eta(\mathbf{r}')\rangle\langle e(\mathbf{r})e^*(\mathbf{r}')\rangle - k^2 \int_{-\infty}^{\infty}\int\int d^3r_2 \langle \eta(\mathbf{r})\eta(\mathbf{r}_2)\rangle \times [\langle \eta(\mathbf{r}')G(\mathbf{r},\mathbf{r}_2)e(\mathbf{r}_2)e^*(\mathbf{r}')\rangle + \langle \eta(\mathbf{r}')e(\mathbf{r})G^*(\mathbf{r}',\mathbf{r}_2)e^*(\mathbf{r}_2)\rangle]. \tag{9}$$

We again use the result in Eq. (5) to evaluate each term within the integral on the right-hand side of (9). For example, we can evaluate $T_2 \equiv \langle \eta(\mathbf{r}')G(\mathbf{r},\mathbf{r}_2)e(\mathbf{r}_2)e^*(\mathbf{r}')\rangle$ by setting $f[\eta] = G(\mathbf{r},\mathbf{r}_2)e(\mathbf{r}_2)e^*(\mathbf{r}')$, and then using (5). The result is

$$T_2 = \int_{-\infty}^{\infty}\int\int d^3r_3 \langle \eta(\mathbf{r}')\eta(\mathbf{r}_3)\rangle \left\langle G(\mathbf{r},\mathbf{r}_2)e(\mathbf{r}_2)\frac{\delta e^*(\mathbf{r}')}{\delta\eta(\mathbf{r}_3)} + G(\mathbf{r},\mathbf{r}_2)e^*(\mathbf{r}')\frac{\delta e(\mathbf{r}_2)}{\delta\eta(\mathbf{r}_3)} + e(\mathbf{r}_2)e^*(\mathbf{r}')\frac{\delta G(\mathbf{r},\mathbf{r}_2)}{\delta\eta(\mathbf{r}_3)} \right\rangle. \tag{10}$$

If we now apply the result in (7) plus the relation, derived in Appendix B, that

$$\frac{\delta G(\mathbf{r},\mathbf{r}')}{\delta\eta(\mathbf{r}'')} = -k^2 G(\mathbf{r},\mathbf{r}'')G(\mathbf{r}'',\mathbf{r}'), \tag{11}$$

the evaluation of (10) is readily performed. The result is

$$T_2 = -k^2 \int_{-\infty}^{\infty}\int\int d^3r_3 \langle \eta(\mathbf{r}')\eta(\mathbf{r}_3)\rangle \langle G(\mathbf{r},\mathbf{r}_2)e(\mathbf{r}_2)e^*(\mathbf{r}_3)G^*(\mathbf{r}',\mathbf{r}_3) + G(\mathbf{r},\mathbf{r}_2)G(\mathbf{r}_2,\mathbf{r}_3)e^*(\mathbf{r}')e(\mathbf{r}_3) + e(\mathbf{r}_2)e^*(\mathbf{r}')G(\mathbf{r},\mathbf{r}_3)G(\mathbf{r}_3,\mathbf{r}_2)\rangle. \tag{12}$$

A similar procedure allows one to formally evaluate the second term within the integral on the right-hand side of (9). If we do this and then use (9) and (12) in (4), we obtain after considerable manipulation

$$P_s(m) = \frac{k^4 \sin^2\psi}{(4\pi R)^2} \int_{-\infty}^{\infty} \cdots \int d^3r d^3r'\, b(\mathbf{r},\mathbf{r}')\Gamma(\mathbf{r},\mathbf{r}') \times \exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')] + \frac{k^8 \sin^2\psi}{(4\pi R)^2} \int_{-\infty}^{\infty} \cdots \int d^3r d^3r_1 \times d^3r_2 d^3r_3 b(\mathbf{r}_1,\mathbf{r}_3)b(\mathbf{r},\mathbf{r}_2)\exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_1)] \times \{2\mathrm{Re}\langle G(\mathbf{r},\mathbf{r}_2)G(\mathbf{r}_2,\mathbf{r}_3)e^*(\mathbf{r}_1)e(\mathbf{r}_3) + G(\mathbf{r},\mathbf{r}_3)G(\mathbf{r}_3,\mathbf{r}_2)e(\mathbf{r}_2)e^*(\mathbf{r}_1)\rangle + \langle G(\mathbf{r},\mathbf{r}_2)G^*(\mathbf{r}_1,\mathbf{r}_3)e(\mathbf{r}_2)e^*(\mathbf{r}_3) + G(\mathbf{r},\mathbf{r}_3)G^*(\mathbf{r}_1,\mathbf{r}_2)e(\mathbf{r}_3)e^*(\mathbf{r}_2)\rangle\}, \tag{13}$$

where $\Gamma(\mathbf{r},\mathbf{r}') = \langle e(\mathbf{r})e^*(\mathbf{r}')\rangle$ is the mutual coherence function of the field, $b(\mathbf{r},\mathbf{r}_1) = \langle \eta(\mathbf{r})\eta(\mathbf{r}_1)\rangle$ is the correlation function of the relative permittivity fluctuations, and Re denotes "real part of." Note that it can be shown that the right-hand side of (13) is purely real (as expected, because P_s is a real quantity).

Equation (13) is the general expression for the scattered intensity and includes all orders of multiple scattering within the medium. However, it is still not in a really useful form, because it contains terms of the form $\langle GGee^*\rangle$, which, because G is a linear functional of e, implicitly involves the evaluation of the fourth moment of the stochastic Green's

function. All we have done so far is to decouple the fieldlike quantities (e and G) from the scattering medium characteristics (η). In the next section we will demonstrate that the quantities $\langle ee^*GG\rangle$, etc., can also be decoupled if $\langle\eta^2\rangle$ is sufficiently small in comparison with unity.

Before leaving this section, it should be noted that the form of the general result in (13) is considerably different from the approximate result obtained by deWolf[15] and Ito and Adachi[20] for the special case of backscatter from a planar slab in the multiple-forward-scatter, single-backscatter limit. We have been unable to determine the conditions under which (13) reduces to that limit.

III. DECOUPLING OF THE AVERAGES

It will now be demonstrated that if $kl \gg 1$ but $k^2 l^2 \langle\eta^2\rangle \ll 1$, where l is the correlation length of the permittivity fluctuations, the averages $\langle G(\mathbf{r}, \mathbf{r}_2) G(\mathbf{r}_2, \mathbf{r}_3) e^*(\mathbf{r}_1) e(\mathbf{r}_3)\rangle$ can be decoupled into $\langle G(\mathbf{r}, \mathbf{r}_2) G(\mathbf{r}_2, \mathbf{r}_3)\rangle \Gamma(\mathbf{r}_3, \mathbf{r}_1)$, where $\Gamma(\mathbf{r}_3, \mathbf{r}_1) = \langle e(\mathbf{r}_3) e^*(\mathbf{r}_1)\rangle$ is the mutual coherence function of the electric field. Because the results are lengthy, we will not present this proof for every term in (13), but only for a typical term. For example, let us consider the product in the second term of (13) which is of the form

$$V_1 = \langle G(\mathbf{r}, \mathbf{r}_2) G(\mathbf{r}_2, \mathbf{r}_3) e^*(\mathbf{r}_1) e(\mathbf{r}_3)\rangle. \tag{14}$$

Now, because $kl \gg 1$, it can be shown that we may express $e(\mathbf{r})$ using the extended Huygens–Fresnel principle as[25–27]

$$e(\mathbf{r}) = K \int_{-\infty}^{\infty} d^2\rho\, e_0(\boldsymbol{\rho}, z = 0) G_0(\mathbf{r}, \boldsymbol{\rho}) \times \exp[i\phi(\mathbf{r}, \boldsymbol{\rho})], \tag{15}$$

where $\boldsymbol{\rho} \equiv (x, y)$, $e_0(\boldsymbol{\rho}, 0)$ is the incident field distribution in some reference plane (denoted by $z = 0$ in Fig. 1) outside the scattering medium, K is a constant, $G_0(\mathbf{r}, \boldsymbol{\rho})$ is the vacuum Green's function for propagation from the point $(\boldsymbol{\rho}, 0) \equiv (x_0, y_0, 0)$ to the point $\mathbf{r} = (x, y, z)$ and $\phi(\mathbf{r}, \boldsymbol{\rho})$ is the additional phase, due to the presence of the randomly inhomogeneous medium, of a spherical wave propagating from $(\boldsymbol{\rho}, 0)$ to $\mathbf{r}$. We assume that ϕ is real because, as shown elsewhere,[28] its imaginary part does not contribute significantly to the field in a lossless medium. Note that (15) reduces to the standard[29] Huygens–Fresnel principle in the limit when $\phi = 0$.

The stochastic Green's function which appears in (14) can be written as[30]

$$G(\mathbf{r}, \mathbf{r}_2) \simeq G_0(\mathbf{r}, \mathbf{r}_2) \exp[i\phi(\mathbf{r}, \mathbf{r}_2)], \tag{16}$$

where $G_0(\mathbf{r}, \mathbf{r}_2)$ is the vacuum Green's function, $[4\pi|\mathbf{r} - \mathbf{r}_2|]^{-1} \exp[ik|\mathbf{r} - \mathbf{r}_2|]$, and $\phi(\mathbf{r}, \mathbf{r}_2)$ is the additional random phase, due to inhomogeneities, of a spherical wave propagating from $\mathbf{r}_2 = (x_2, y_2, z_2)$ to $\mathbf{r} = (x, y, z)$. It can be demonstrated that (16) correctly includes multiple forward scatterings (of photons by inhomogeneities) but ignores individual backscatterings. Consequently, it is valid only for $kl > 1$, and would be inappropriate for $kl \ll 1$.

We next assume ϕ is a zero-mean[31] Gaussian random variable. If (15) and (16) are substituted into (14) and the ensemble average then performed, we obtain

$$V_1 = \langle G(\mathbf{r}, \mathbf{r}_2) G(\mathbf{r}_2, \mathbf{r}_3)\rangle |K|^2 \times \int_{-\infty}^{\infty} \cdots \int d^2\rho d^2\rho'\, e_0(\boldsymbol{\rho}, 0) e_0^*(\boldsymbol{\rho}', 0) G_0(\mathbf{r}_3, \boldsymbol{\rho}) \times G_0^*(\mathbf{r}_1, \boldsymbol{\rho}') \chi(\mathbf{r}_3, \boldsymbol{\rho}; \mathbf{r}, \boldsymbol{\rho}') Q(\mathbf{r}_3, \boldsymbol{\rho}, \mathbf{r}_1, \boldsymbol{\rho}', \mathbf{r}_2), \tag{17}$$

where

$$\chi(\mathbf{r}_3, \boldsymbol{\rho}; \mathbf{r}, \boldsymbol{\rho}') = \exp\{-\tfrac{1}{2}\langle[\phi(\mathbf{r}_3, \boldsymbol{\rho}) - \phi(\mathbf{r}, \boldsymbol{\rho}')]^2\rangle\}, \tag{18}$$

$$Q = \exp\{-\langle[\phi(\mathbf{r}_3, \boldsymbol{\rho}) - \phi(\mathbf{r}, \boldsymbol{\rho}')] \times [\phi(\mathbf{r}, \mathbf{r}_2) + \phi(\mathbf{r}_2, \mathbf{r}_3)]\rangle\}. \tag{19}$$

In obtaining (17) we have used the result

$$\langle G(\mathbf{r}, \mathbf{r}_2) G(\mathbf{r}_2, \mathbf{r}_3)\rangle = G_0(\mathbf{r}, \mathbf{r}_2) G_0(\mathbf{r}_2, \mathbf{r}_3) \times \exp\{-\tfrac{1}{2}\langle[\phi(\mathbf{r}, \mathbf{r}_2) + \phi(\mathbf{r}_2, \mathbf{r}_3)]^2\rangle\}. \tag{20}$$

If it is recalled that

$$\langle e(\mathbf{r}_3) e^*(\mathbf{r}_1)\rangle = |K|^2 \int_{-\infty}^{\infty} \cdots \int d^2\rho d^2\rho' \times e_0(\boldsymbol{\rho}, 0) e_0^*(\boldsymbol{\rho}', 0) G_0(\mathbf{r}_3, \boldsymbol{\rho}) G_0^*(\mathbf{r}_1, \boldsymbol{\rho}') \chi(\mathbf{r}_3, \boldsymbol{\rho}; \mathbf{r}, \boldsymbol{\rho}') \tag{21}$$

it is evident from (17) that if $Q \simeq 1$, we can rewrite V_1 as

$$V_1 \simeq \langle G(\mathbf{r}, \mathbf{r}_2) G(\mathbf{r}_2, \mathbf{r}_3)\rangle \Gamma(\mathbf{r}_3, \mathbf{r}_1). \tag{22}$$

It can be seen from (19) and (17) that $Q \simeq 1$ if

$$|\langle[\phi(\mathbf{r}_3, \boldsymbol{\rho}) - \phi(\mathbf{r}, \boldsymbol{\rho}')][\phi(\mathbf{r}, \mathbf{r}_2) + \phi(\mathbf{r}_2, \mathbf{r}_3)]\rangle| \ll 1. \tag{23}$$

The condition expressed by (23) will be satisfied if each of the products, such as $\langle\phi(\mathbf{r}_3, \boldsymbol{\rho})\phi(\mathbf{r}, \mathbf{r}_2)\rangle$ in (23), is individually much less than 1/4 in magnitude. We will now determine the conditions when this is so.

The phases ϕ in (23) can be estimated by employing the geometric optics approximation; this method yields valid results for the real part of ϕ, even though it gives poor results for its imaginary part except in the limit when the propagation path is small in comparison with the Fresnel length of the smallest inhomogeneities. In the geometric optics approximation we can write[32]

$$\phi(\mathbf{r}_3, \boldsymbol{\rho}) \simeq \tfrac{1}{2} k \int_{\boldsymbol{\rho}}^{\mathbf{r}_3} ds\, \eta(\xi), \tag{24}$$

where $\eta(\xi)$ is the relative permittivity fluctuation at the point ξ along the ray path s between $\boldsymbol{\rho}$ and $\mathbf{r}_3$. By using (24) we obtain

$$\langle\phi(\mathbf{r}_3, \boldsymbol{\rho})\phi(\mathbf{r}, \mathbf{r}_2)\rangle \simeq \frac{k^2}{4} \int_{\boldsymbol{\rho}}^{\mathbf{r}_3} ds \int_{\mathbf{r}_2}^{\mathbf{r}} dt\, b(\xi, \tau), \tag{25}$$

where $b(\xi, \tau) \equiv \langle\eta(\xi)\eta(\tau)\rangle$. When $|\langle\eta^2\rangle \ll 1$, it can be shown[33] that the actual ray paths in (25) can be approximated by straight lines.

In order to estimate the magnitude of the right-hand side of (25), we now assume that the permittivity fluctuations are statistically stationary, so that $b(\xi, \tau) = b(\xi - \tau)$. We further assume, for purposes of estimating (25) only, that $b(\xi - \tau)$ is of order of $\langle\eta^2\rangle$ for $|\xi - \tau| < l$ and is negligibly small for $|\xi - \tau| > l$, where l is the effective correlation length of the permittivity fluctuations. Then, we find that

$$\langle\phi(\mathbf{r}_3, \boldsymbol{\rho})\phi(\mathbf{r}, \mathbf{r}_2)\rangle \simeq \tfrac{1}{4} k^2 \langle\eta^2\rangle l^2. \tag{26}$$

One can readily convince oneself of the validity of (26) by assuming that the vectors $\mathbf{r}_3 - \boldsymbol{\rho}$ and $\mathbf{r} - \mathbf{r}_2$ are arbitrarily directed straight lines, assuming $b(\xi - \tau) = \langle \eta^2 \rangle \exp[-(\xi - \tau)^2/l^2]$ and then actually performing the integrations in (25). Equation (26) is a good estimate of the magnitude of (25) except for the pathological case when the path $\mathbf{r}_3 - \boldsymbol{\rho}$ is parallel to $\mathbf{r} - \mathbf{r}_2$ and both lines are overlapping, or nearly overlapping, over distances much greater than l. In this case one finds $\langle \phi(\mathbf{r}_3, \boldsymbol{\rho}) \phi(\mathbf{r}, \mathbf{r}_2) \rangle \sim k^2 \langle \eta^2 \rangle L l/4$, where L is the distance over which the paths $\mathbf{r} - \mathbf{r}_2$ and $\mathbf{r}_3 - \boldsymbol{\rho}$ are overlapping. Because $kl \gg 1$, so that individual forward scatterings (of photons) by the inhomogeneities are much more likely than individual backscatterings, this special situation arises only when the observation point is in the backscatter cone. That is, the field at an observation point outside the backscatter cone is highly likely to be the result of multiple forward scatterings [Fig. 2(A)], a situation where $\mathbf{r}_3 - \boldsymbol{\rho}$ and $\mathbf{r} - \mathbf{r}_2$ are likely to overlap only for distances of order l [the segments $\mathbf{r}_3 - \boldsymbol{\rho}$ and $\mathbf{r} - \mathbf{r}_2$ may each consist of a number of the individual scatterings in Fig. 2(A)]. It is highly unlikely that this field is produced by multiple forward scatterings plus two (or any even number) backscatterings. This situation is depicted in Fig. 2(B). However, in the backscatter cone $\Delta\theta_B \sim (kl)^{-1}$, it is possible that the field is due to either multiple forward scatterings or multiple forward scatterings plus a single backscatter, as depicted in Fig. 2(C). In this latter case the paths (the forward scattering portion would correspond to $\mathbf{r}_3 - \boldsymbol{\rho}$ and the backscattering portion to $\mathbf{r} - \mathbf{r}_2$) may overlap over a distance of order of the size of the medium in the z direction. Consequently, for observation points in the backscatter cone, the phase correlation in (23) is of order $k^2 \langle \eta^2 \rangle L l/4$, which may be large.[34] Consequently, if the observation point lies in a cone of angle $\Delta\theta_B \sim (kl)^{-1}$ centered about the negative z axis, the phase correlation in (23) may be large and (26) is not a sufficient condition for decoupling the averages. In this case the general expression in (13) must be employed.

Arguments similar to those in Eqs. (24)–(26) can be used to show that each of the other three terms in (23) is also of order $k^2 \langle \eta^2 \rangle l^2/4$, except in the backscatter cone. Consequently, if $k^2 \langle \eta^2 \rangle l^2 \ll 1$, we may approximate Eq. (14) by (22) except in the backscatter cone. A similar proof may be applied to the other terms in (13). Hence we find that if[35]

$$k^2 \langle \eta^2 \rangle l^2 \ll 1, \tag{27}$$

then, except in the cone $\Delta\theta_B$ centered about the negative z axis, we may approximate (13) by

$$\begin{aligned} P_s(\mathbf{m}) = {} & \frac{k^4 \sin^2\psi}{(4\pi R)^2} \int_{-\infty}^{\infty} \cdots \int d^3r d^3r'\, b(\mathbf{r}, \mathbf{r}') \Gamma(\mathbf{r}, \mathbf{r}') \\ & \times \exp[-ik\mathbf{m}\cdot(\mathbf{r} - \mathbf{r}')] \\ & + \frac{k^8 \sin^2\psi}{(4\pi R)^2} \int_{-\infty}^{\infty} \cdots \int d^3r d^3r_1 d^3r_2 d^3r_3\, b(\mathbf{r}_1, \mathbf{r}_3) \\ & \times b(\mathbf{r}, \mathbf{r}_2) \exp[-ik\mathbf{m}\cdot(\mathbf{r} - \mathbf{r}_1)] \{2\mathrm{Re}[\langle G(\mathbf{r}, \mathbf{r}_2) \\ & \times G(\mathbf{r}_2, \mathbf{r}_3) \rangle \Gamma(\mathbf{r}_3, \mathbf{r}_1) + \langle G(\mathbf{r}, \mathbf{r}_3) G(\mathbf{r}_3, \mathbf{r}_2) \rangle \Gamma(\mathbf{r}_2, \mathbf{r}_1)] \\ & + \langle G(\mathbf{r}, \mathbf{r}_2) G^*(\mathbf{r}_1, \mathbf{r}_3) \rangle \Gamma(\mathbf{r}_2, \mathbf{r}_3) \\ & + \langle G(\mathbf{r}, \mathbf{r}_3) G^*(\mathbf{r}_1, \mathbf{r}_2) \rangle \Gamma(\mathbf{r}_3, \mathbf{r}_2) \}. \end{aligned} \tag{28}$$

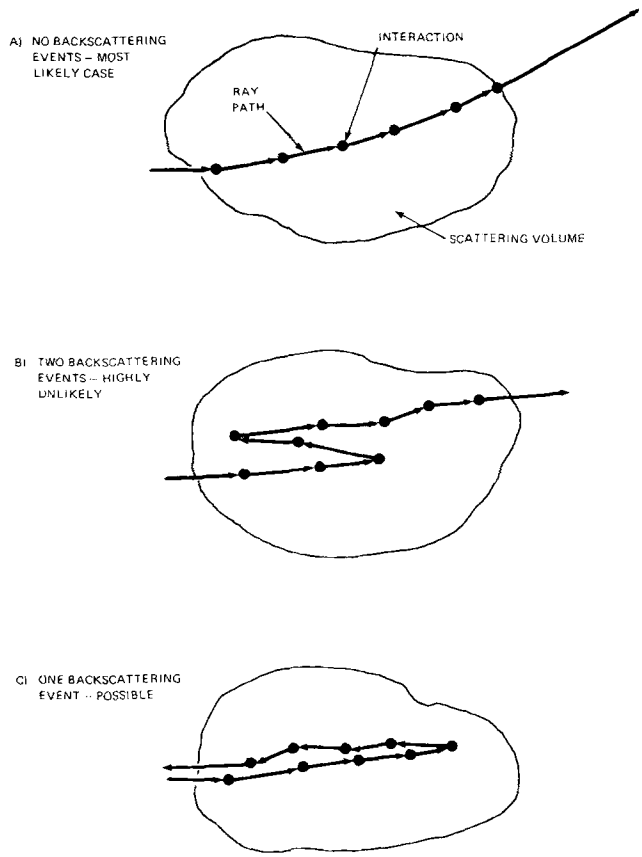


FIG. 2. Ray paths for some potential scattering events.

Thus we have now expressed the scattered intensity in terms of the correlation and Green's functions of the scattering medium and the mutual coherence function $\Gamma(\mathbf{r}_1, \mathbf{r}_2) \equiv \langle e(\mathbf{r}_1) e^*(\mathbf{r}_2) \rangle$ of the field within the medium. Equation (28) includes multiple scatter within the medium and is valid for the intensity scattered in any direction except the backscatter cone $\Delta\theta_B$. Within the cone $\Delta\theta_B$, the condition $k^2 \langle \eta^2 \rangle l^2 \ll 1$ is not a sufficient one to decouple the ensemble averages, and one must then use the exact results in (13), which involves the evaluation of the fourth moment $\langle GGee^* \rangle$ rather than the second moment $\langle GG \rangle \langle ee^* \rangle$. In the next section we will show that the first term in (28) dominates the scattering everywhere except in the backscatter cone, $\Delta\theta_B$, and in a cone around the forward-scatter direction. We will also show that only the first and fourth terms in (28) are important for forward scatter.

A similar analysis applied to the result in (6) shows that if $k^2 \langle \eta^2 \rangle l^2/4 \ll 1$ we can approximate (6) by

$$\begin{aligned} \langle E_s \rangle = {} & \frac{-k^4 (\hat{x} - \mathbf{m}\cos\psi)}{4\pi R} \int_{-\infty}^{\infty} \cdots \int d^3r d^3r'\, b(\mathbf{r}, \mathbf{r}') \\ & \times \langle G(\mathbf{r}, \mathbf{r}') \rangle \langle e(\mathbf{r}') \rangle \exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR). \end{aligned} \tag{29}$$

IV. SIMPLIFICATION OF EQUATION (28)

In this section we will show that if $k^2 \langle \eta^2 \rangle l^2 \ll 1$, the contribution from the first term in (28) dominates the intensity scattered in any direction except within the forward scatter-

ing (and, as noted earlier, the backscatter cone $\Delta\theta_B$). In order to prove this, we shall assume that $b(\mathbf{r},\mathbf{r}')\equiv\langle\eta(\mathbf{r})\eta(\mathbf{r}')\rangle$ is locally stationary (statistically) and can be written as

$$b(\mathbf{r},\mathbf{r}')\simeq\langle\eta^2\rangle C((\mathbf{r}+\mathbf{r}')/2)\,B(\mathbf{r}-\mathbf{r}'), \tag{30}$$

where $B(\mathbf{r}-\mathbf{r}')$ changes significantly over distances $|\mathbf{r}-\mathbf{r}'|$ of order l, but $C((\mathbf{r}+\mathbf{r}')/2)$ changes significantly over distances $|\mathbf{r}+\mathbf{r}'|$ which are very much larger than l. Also, $C(0)\equiv B(0)\equiv 1$. The mutual coherence function $\Gamma(\mathbf{r},\mathbf{r}')$ may also be written as a locally stationary function. In particular, we write

$$\Gamma(\mathbf{r},\mathbf{r}')\simeq\tilde{\Gamma}((\mathbf{r}+\mathbf{r}')/2,\mathbf{r}-\mathbf{r}')\exp[ik\hat{z}\cdot(\mathbf{r}-\mathbf{r}')], \tag{31}$$

where we have explicitly extracted the propagating portion of Γ.

Finally, for the purpose of estimating the relative sizes of the various terms in (28), we will use the approximation

$$\begin{aligned}\langle G(\mathbf{r},\mathbf{r}_2)G(\mathbf{r}_2,\mathbf{r}_3)\rangle &\simeq\langle G(\mathbf{r},\mathbf{r}_2)\rangle\langle G(\mathbf{r}_2,\mathbf{r}_3)\rangle\\ &\simeq G_0(\mathbf{r}-\mathbf{r}_2)G_0(\mathbf{r}_2-\mathbf{r}_3)\exp(-|\mathbf{r}-\mathbf{r}_2|/\\ &\quad\times\rho_\phi-|\mathbf{r}_2-\mathbf{r}_3|/\rho_\phi).\end{aligned} \tag{32}$$

It is demonstrated in Appendix C that this approximation may be used provided $k^2\langle\eta^2\rangle l^2/4\ll 1$.

If (30)–(32) are substituted into (28) and then a number of coordinate transformations are made, we obtain

$$P_s(\mathbf{m})=\frac{\langle\eta^2\rangle k^4\sin^2\psi}{(4\pi R)^2}\times[I_1(\mathbf{m})+I_2(\mathbf{m})+I_3(\mathbf{m})+I_4(\mathbf{m})+I_5(\mathbf{m})], \tag{33}$$

where, provided $B(-\xi)=B(\xi)$, we have

$$I_1(\mathbf{m})=\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3u\,C(u)\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3v\,B(\mathbf{v})\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}), \tag{34}$$

$$\begin{aligned}I_2(\mathbf{m})&=2\langle\eta^2\rangle k^4\,\mathrm{Re}\int_{-\infty}^{\infty}\!\cdots\!\int d^3\xi d^3\tau\,B(\tau)\bar{G}_0(\tau)\bar{G}_0(\xi)\\ &\quad\times\exp[-ik\mathbf{m}\cdot\tau+ik\boldsymbol{\beta}\cdot\boldsymbol{\xi}]\\ &\quad\times\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3u\,C(\mathbf{u})C(\mathbf{u}-\mathbf{v}/2+\tau/2-\xi)\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3v\\ &\quad\times B(\mathbf{v})\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}),\end{aligned} \tag{35}$$

$$\begin{aligned}I_3(\mathbf{m})&=2\langle\eta^2\rangle k^4\,\mathrm{Re}\int_{-\infty}^{\infty}\!\cdots\!\int d^3\xi d^3\tau B(\tau+\xi/2)\\ &\quad\times\bar{G}_0(\xi)\bar{G}_0(\tau-\xi/2)\exp[-ik\mathbf{m}\cdot(\tau+\xi/2)]\\ &\quad\times\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3u\\ &\quad\times C(\mathbf{u}+\tau/2-\xi/4)C(\mathbf{u}+\tau/2+\mathbf{v}/2+\xi/4)\\ &\quad\times\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3v\,B(\mathbf{v}+\tau-\xi/2)\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}),\end{aligned} \tag{36}$$

$$\begin{aligned}I_4(\mathbf{m})&=\langle\eta^2\rangle k^4\int_{-\infty}^{\infty}\!\cdots\!\int d^3\xi d^3\tau\,B(\tau)B(\xi)\bar{G}_0(\tau)\bar{G}_0^*(\xi)\\ &\quad\times\exp[-ik\mathbf{m}\cdot(\tau-\xi)]\int_{-\infty}^{\infty}\!\cdots\!\int d^3u d^3v\\ &\quad\times C(\mathbf{u}-\mathbf{v}/2+\xi/2)\\ &\quad\times C(\mathbf{u}+\mathbf{v}/2+\tau/2)\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}),\end{aligned} \tag{37}$$

$$\begin{aligned}I_5(\mathbf{m})&=\langle\eta^2\rangle k^4\int_{-\infty}^{\infty}\!\cdots\!\int d^3\xi d^3\tau\bar{G}_0(\tau)\bar{G}_0^*(\xi)\\ &\quad\times\exp[ik\,\boldsymbol{\beta}\cdot\tau+ik\,\boldsymbol{\xi}\cdot\mathbf{m}]\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3u\,C(\mathbf{u}+\xi/2)\\ &\quad\times C(\mathbf{u}-\tau/2)\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3v\,B(\mathbf{v}-\xi)B(\mathbf{v}-\tau)\\ &\quad\times\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}),\end{aligned} \tag{38}$$

and $\boldsymbol{\alpha}\equiv\mathbf{m}-\hat{z}$, $\boldsymbol{\beta}=\mathbf{m}-2\hat{z}$, and $\bar{G}_0(u)=G_0(u)\exp(-u/\rho_\phi)$.

Now because $C(\tau)$ varies much more slowly than $B(\tau)$ and $\bar{G}_0(\tau)$, we may approximate terms of the form $C(\mathbf{u}+\tau+\xi+\mathbf{v})B(\xi)B(\mathbf{v})\bar{G}_0(\tau)$ by $C(\mathbf{u})B(\xi)B(\mathbf{v})\bar{G}_0(\tau)$. Furthermore, because $C(\tau)$ is very slowly varying in comparison with the smaller of the medium correlation length l and the field correlation length, we see that [for purposes of estimating the relative sizes of the terms in (33), but not for precise calculations]

$$\left|\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3u\,C^2(\mathbf{u})F(\mathbf{u})\right|\sim\left|\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3u\,C(\mathbf{u})F(\mathbf{u})\right|, \tag{39}$$

where

$$F(\mathbf{u})=\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3v\,B(\mathbf{v})\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}). \tag{40}$$

By using the aforementioned approximations we can estimate I_2 to I_5 as

$$\begin{aligned}I_2(\mathbf{m})&\sim 2\langle\eta^2\rangle k^4 I_1(\mathbf{m})\,\mathrm{Re}\int_{-\infty}^{\infty}\!\cdots\!\int d^3\xi d^3\tau\\ &\quad\times B(\tau)\bar{G}_0(\tau)\bar{G}_0(\xi)\exp[-ik\mathbf{m}\cdot\tau+ik\,\boldsymbol{\beta}\cdot\boldsymbol{\xi}],\end{aligned} \tag{41}$$

$$\begin{aligned}I_3(\mathbf{m})&\sim 2\langle\eta^2\rangle k^4\,\mathrm{Re}\int_{-\infty}^{\infty}\!\cdots\!\int d^3\xi d^3\tau\,B(\tau+\xi/2)\\ &\quad\times\bar{G}_0(\xi)\bar{G}_0(\tau-\xi/2)\exp[-ik\mathbf{m}\cdot(\tau+\xi/2)]\\ &\quad\times\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3u\,C(\mathbf{u})\int_{-\infty}^{\infty}\!\!\int\!\!\int d^3v\,B(\mathbf{v}+\tau-\xi/2)\\ &\quad\times\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}),\end{aligned} \tag{42}$$

$$\begin{aligned}I_4(\mathbf{m})&\sim k^4\langle\eta^2\rangle\int_{-\infty}^{\infty}\!\cdots\!\int d^3\xi d^3\tau\,B(\tau)B(\xi)\\ &\quad\times\bar{G}_0(\tau)\bar{G}_0^*(\xi)\exp[-ik\mathbf{m}\cdot(\tau-\xi)]\end{aligned}$$

$$\times\int_{-\infty}^{\infty}\iiint d^3u\, C(u)\int_{-\infty}^{\infty}\iiint d^3v\, \tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}), \tag{43}$$

$$I_5(\mathbf{m})\sim k^4\langle\eta^2\rangle\int_{-\infty}^{\infty}\cdots\int d^3\xi d^3\tau\,\bar{G}_0(\tau)\bar{G}_0^*(\xi)$$
$$\times\exp[ik\boldsymbol{\beta}\cdot\boldsymbol{\tau}+ik\mathbf{m}\cdot\boldsymbol{\xi}]$$
$$\times\int_{-\infty}^{\infty}\iiint d^3u\, C(\mathbf{u})\int_{-\infty}^{\infty}\iiint d^3v\, B(\mathbf{v}-\boldsymbol{\xi})B(\mathbf{v}-\boldsymbol{\tau})$$
$$\times\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}). \tag{44}$$

We will now employ (41)–(44) to prove that $I_n/I_1\ll 1$ (for $n=2,3,4,5$) for all scattering directions except the forward scattering cone, provided $k^2\langle\eta^2\rangle l^2\ll 1$.

A. Ratio I_2/I_1

We begin with an estimate of I_2/I_1. In order to do this, we recall that $\bar{G}_0(\tau)=(4\pi\tau)^{-1}\exp(ik\tau-\tau/\rho_\phi)$, where $\tau=|\boldsymbol{\tau}|$, assume that $B(\tau)=\exp(-\tau/l)$ and then use the results in Appendix D to calculate the integrals on $d^3\xi$ and $d^3\tau$ in (41). We obtain, after taking the real part

$$\frac{I_2}{I_1}\sim\frac{\delta[1-\cos\theta+\delta^2/4+\langle\eta^2\rangle/32]}{4[(1-\cos\theta+\delta^2/4)^2+\delta^2/4]}, \tag{45}$$

where

$$\delta\equiv\langle\eta^2\rangle kl. \tag{46}$$

From (45) it is evident that, for all values of the scattering angle θ, the ratio I_2/I_1 is of order $\delta=\langle\eta^2\rangle kl$. Consequently, as long as $\langle\eta^2\rangle kl\ll 1$, we may ignore I_2 in comparison with I_1. Because $kl\gg 1$ the condition $\delta\ll 1$ is automatically satisfied if our previous assumption that $\langle\eta^2\rangle k^2l^2\ll 1$ holds.

B. Ratio I_3/I_1

The integrals in (42)–(45) are not as simple to evaluate as the one in (41) and to estimate the ratio I_n/I_1, for $n=3$, 4, and 5 we will need to consider separately the limits when the coherence length ρ_0 of the mutual coherence function $\tilde{\Gamma}$ is small and large in comparison with the correlation length l of the permittivity fluctuations. In the limit when $\rho_0\ll l$, we can approximate $B(\mathbf{v}+\boldsymbol{\tau}-\boldsymbol{\xi}/2)\tilde{\Gamma}(\mathbf{u},\mathbf{v})$ by $B(\boldsymbol{\tau}-\boldsymbol{\xi}/2)\tilde{\Gamma}(\mathbf{u},\mathbf{v})$, etc. Consequently, (42) can be approximated by

$$I_3(\mathbf{m})\simeq 2\langle\eta^2\rangle k^4 I_1(\mathbf{m})\mathrm{Re}\int_{-\infty}^{\infty}\cdots\int d^3\xi d^3\tau$$
$$\times B(\boldsymbol{\tau}+\boldsymbol{\xi}/2)B(\boldsymbol{\tau}-\boldsymbol{\xi}/2)\bar{G}_0(\boldsymbol{\xi})\bar{G}_0(\boldsymbol{\tau}-\boldsymbol{\xi}/2)$$
$$\times\exp[-ik\mathbf{m}\cdot(\boldsymbol{\tau}+\boldsymbol{\xi}/2)], \tag{47}$$

where

$$I_1(\mathbf{m})\simeq\int_{-\infty}^{\infty}\iiint d^3u\, C(\mathbf{u})\int_{-\infty}^{\infty}\iiint d^3v\,\tilde{\Gamma}(\mathbf{u},\mathbf{v})$$
$$\times\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}). \tag{48}$$

In the opposite limit when $\rho_0\gg l$ we may approximate $\tilde{\Gamma}(\mathbf{u},\mathbf{v}+\boldsymbol{\tau})B(\boldsymbol{\tau})$ by $\tilde{\Gamma}(\mathbf{u},\mathbf{v})B(\boldsymbol{\tau})$ and $\tilde{\Gamma}(\mathbf{u},\mathbf{v})B(\mathbf{v})$ by $\tilde{\Gamma}(\mathbf{u},0)B(\mathbf{v})$, etc. Consequently, in this limit, (42) becomes

$$I_3(\mathbf{m})\sim 2k^4\langle\eta^2\rangle\mathrm{Re}\int_{-\infty}^{\infty}\cdots\int d^3\xi d^3\tau$$
$$\times B(\boldsymbol{\tau}+\boldsymbol{\xi}/2)\bar{G}_0(\boldsymbol{\xi})\bar{G}_0(\boldsymbol{\tau}-\boldsymbol{\xi}/2)\exp(-ik\,\boldsymbol{\xi}\cdot\mathbf{s}-ik\boldsymbol{\tau}\cdot\hat{z})$$
$$\times\int_{-\infty}^{\infty}\iiint d^3u\, C(\mathbf{u})\tilde{\Gamma}(\mathbf{u},\boldsymbol{\tau}-\boldsymbol{\xi}/2)$$
$$\times\int_{-\infty}^{\infty}\iiint d^3v\, B(\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}), \tag{49}$$

where $\mathbf{s}=\mathbf{m}-\hat{z}/2$. Also, when $\rho_0\gg l$, (34) can be approximated by

$$I_1(\mathbf{m})\simeq\int_{-\infty}^{\infty}\iiint d^3v B(\mathbf{v})\exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v})\int_{-\infty}^{\infty}\iiint d^3u C(\mathbf{u})\tilde{\Gamma}(\mathbf{u},0). \tag{50}$$

We would now like to express $I_3(\mathbf{m})$ in terms of $I_1(\mathbf{m})$ in (49) so that we can estimate the ratio I_3/I_1. Therefore, for purposes of estimation only we assume that $\tilde{\Gamma}(\mathbf{u},\mathbf{v})$ can be written as $\tilde{\Gamma}(\mathbf{u},\mathbf{v})\simeq\Gamma_1(\mathbf{u})\Gamma_2(\mathbf{v})$. If this is done, we obtain from (49) and (50) the result

$$I_3/I_1\sim 2\langle\eta^2\rangle k^4\,\mathrm{Re}\int_{-\infty}^{\infty}\cdots\int d^3\xi d^3\tau$$
$$\times B(\boldsymbol{\tau}+\boldsymbol{\xi}/2)\bar{G}_0(\boldsymbol{\xi})\bar{G}_0(\boldsymbol{\tau}-\boldsymbol{\xi}/2)$$
$$\times\Gamma_2(\boldsymbol{\tau}-\boldsymbol{\xi}/2)\exp(-ik\,\boldsymbol{\xi}\cdot\mathbf{s}-ik\boldsymbol{\tau}\cdot\hat{z}). \tag{51}$$

The integrals in (47) and (51) can be evaluated by substituting $B(\boldsymbol{\xi})=\exp(-\xi/l)$, $\bar{G}_0(\tau)=(4\pi\tau)^{-1}\exp(ik\tau-\tau/\rho_\phi)$, and then using the approximations in Appendix E. The details are quite laborious so we present only the conclusion: For all values of θ and for either $\rho_0\ll l$ or $\rho_0\gg l$, we find

$$|I_3/I_1|\lesssim k^2\langle\eta^2\rangle l^2. \tag{52}$$

Consequently, when $k^2\langle\eta^2\rangle l^2\ll 1$ we may ignore $I_3(\mathbf{m})$ in comparison with $I_1(\mathbf{m})$ for all scattering directions.

C. Ratio I_4/I_1

In estimating the magnitude of $I_4(\mathbf{m})$, we again consider separately the limiting cases when $\rho_0\ll l$ and $\rho_0\gg l$. If $\rho_0\ll l$, we readily obtain from (43) the result that

$$I_4(\mathbf{m})/I_1(\mathbf{m})\simeq k^4\langle\eta^2\rangle\left|\int_{-\infty}^{\infty}\iiint d^3\tau B(\boldsymbol{\tau})\bar{G}_0(\boldsymbol{\tau})\exp(-ik\mathbf{m}\cdot\boldsymbol{\tau})\right|^2. \tag{53}$$

By using the results in Appendix D the integrals in (53) can be evaluated: we find for all values of θ that

$$|I_4/I_1|\lesssim k^2\langle\eta^2\rangle l^2. \tag{54}$$

The opposite limit when $\rho_0\gg l$ is more difficult to evaluate, and we must consider separately the subcases when $\rho_0\ll V^{1/3}$ and $\rho_0\gg V^{1/3}$. The details of the analysis are fairly involved so we will only summarize the results. These are:

(i) If $V^{1/3} \gg \rho_0 \gg l$ and $\theta > (kl)^{-1}$, then $|I_4/I_1| \sim \langle \eta^2 \rangle (kl)^2 (l/\rho_0) \ll k^2 l^2 \langle \eta^2 \rangle$;

(ii) if $V^{1/3} \gg \rho_0 \gg l$ and $\theta < (k\rho)^{-1}$, then $|I_4/I_1| \sim (kl)^2 \langle \eta^2 \rangle (\rho_0/l)^2$;

(iii) if $\rho_0 \gg V^{1/3}$ and $\theta > (kV^{1/3})^{-1}$, then $|I_4/I_1| \sim (l/V^{1/3})(kV^{1/3})^{-4} \ll 1$;

(iv) if $\rho_0 \gg V^{1/3}$ and $\theta \ll (kV^{1/3})^{-1}$, then $|I_4/I_1| \sim \langle \eta^2 \rangle (k^3 V)(kl)^{-1}$.

From the aforementioned results, along with (54), we see that if $\langle \eta^2 \rangle k^2 l^2 \ll 1$, we can ignore $\Gamma_4(\mathbf{m})$ in comparison with $I_1(\mathbf{m})$ in all directions *except* the forward-scatter cone. In the forward-scatter cone I_4 may be much larger than I_1.

D. Ratio I_5/I_1

In the limit when the field correlation length ρ_0 is much less than the index-of-refraction correlation length l, we may approximate $\tilde{\Gamma}(\mathbf{u},\mathbf{v})B(\mathbf{v}-\xi)B(\mathbf{v}-\tau)$ by $\tilde{\Gamma}(\mathbf{u},v)B(-\xi)B(-\tau) = \tilde{\Gamma}(\mathbf{u},\mathbf{v})B(\xi)B(\tau)$ in Eq. (44). Consequently, upon using (48) for $I_1(\mathbf{m})$, we may approximate $I_5(\mathbf{m})$ as

$$I_5(\mathbf{m})/I_1(\mathbf{m}) \simeq k^4 \langle \eta^2 \rangle \int_{-\infty}^{\infty} \cdots \int d^3\xi d^3\tau B(\xi) B(\tau) \times \bar{G}_0(\tau)\bar{G}_0(\xi) \exp[ik\boldsymbol{\beta}\cdot\boldsymbol{\tau} + ik\mathbf{m}\cdot\boldsymbol{\xi}]. \tag{55}$$

The integrals in (55) can be evaluated by using the results in Appendix D. It is found that for all θ

$$|I_5/I_1| \lesssim \langle \eta^2 \rangle k^2 l^2. \tag{56}$$

We next consider the limit when $\rho_0 \gg l$. If we make the coordinate transformation $\mathbf{v} = \boldsymbol{\phi} + \boldsymbol{\tau}$, we may rewrite (44) as

$$I_5(\mathbf{m}) = k^4 \langle \eta^2 \rangle \int_{-\infty}^{\infty} \cdots \int d^3\xi d^3\tau\, \bar{G}_0(\tau)\bar{G}_0^*(\xi) \times \exp[-ik\hat{z}\cdot\boldsymbol{\tau} + ik\,\boldsymbol{\xi}\cdot\mathbf{m}] \times \int\int\int d^3u C(\mathbf{u}) \int\int\int d^3\phi B(\boldsymbol{\phi}) B(\boldsymbol{\phi}+\boldsymbol{\tau}-\boldsymbol{\xi}) \times \tilde{\Gamma}(\mathbf{u},\boldsymbol{\phi}+\boldsymbol{\tau}) \exp(-ik\boldsymbol{\alpha}\cdot\boldsymbol{\phi}). \tag{57}$$

Because $\rho_0 \gg l$, we may approximate $B(\boldsymbol{\phi})B(\boldsymbol{\phi}+\boldsymbol{\tau}-\boldsymbol{\xi})\tilde{\Gamma}(\mathbf{u},\boldsymbol{\phi}+\boldsymbol{\tau})$ by $B(\boldsymbol{\phi})B(\boldsymbol{\tau}-\boldsymbol{\xi})\tilde{\Gamma}(\mathbf{u},\boldsymbol{\tau})$. If, for purposes of estimation only, we again write $\tilde{\Gamma}(\mathbf{u},\boldsymbol{\tau}) = \Gamma_1(\mathbf{u})\Gamma_2(\boldsymbol{\tau})$ and use (50), it is possible to rewrite (57) as

$$I_5(\mathbf{m})/I_1(\mathbf{m}) \sim k^4 \langle \eta^2 \rangle \int_{-\infty}^{\infty} \cdots \int d^3\tau d^3\xi \bar{G}_0(\tau)\bar{G}_0^*(\xi) B(\boldsymbol{\tau}-\boldsymbol{\xi}) \Gamma_2(\boldsymbol{\tau}) \times \exp[-ik\hat{z}\cdot\boldsymbol{\tau} + k\mathbf{m}\cdot\boldsymbol{\xi}]. \tag{58}$$

By using the results in Appendices D and E the integrals in (58) can be approximately evaluated. For all θ it is found that

$$|I_5/I_1| \sim kl\,\langle \eta^2 \rangle. \tag{59}$$

Because $kl \gg 1$, so that $kl\,\langle \eta^2 \rangle \ll k^2 l^2 \langle \eta^2 \rangle$, we may therefore state that for both $\rho_0 \gg l$ and $\rho_0 \ll l$ the term $I_5(\mathbf{m})$ will be negligible in comparison with $I_1(\mathbf{m})$ if $\langle \eta^2 \rangle k^2 l^2 \ll 1$.

V. SUMMARY AND DISCUSSION

The results presented in Eqs. (45)–(59) have demonstrated that if $kl \gg 1$ but

$$k^2 \langle \eta^2 \rangle l^2 \ll 1, \tag{60}$$

we may ignore I_2, I_3, I_4, and I_5 in comparison with I_1 in Eq. (33), provided that the observation point does not lie either in the backscatter cone $\Delta\theta_B$ or in the forward-scatter cone. The forward scatter cone is defined as $\theta < \Delta\theta_f$ where $\Delta\theta_f = (kV^{1/3})^{-1}$ for the case when $\rho_0 \gg V^{1/3}$ and $\Delta\theta_f = (k\rho_0)^{-1}$ when $\rho_0 \ll V^{1/3}$. It is shown elsewhere that Eq. (60) is physically equivalent to the requirement that the mean free path between scatterings is large in comparison with the size of the random inhomogeneities. Therefore, when the aforementioned conditions are satisfied, we may use (33) to write the scattered intensity as

$$P_s(\boldsymbol{\alpha}) = [\langle \eta^2 \rangle k^4 \sin^2\psi/(4\pi R)^2] \int_{-\infty}^{\infty} \cdots \int d^3u d^3v \times C(\mathbf{u})B(\mathbf{v})\tilde{\Gamma}(\mathbf{u},\mathbf{v}) \exp(-ik\boldsymbol{\alpha}\cdot\mathbf{v}), \tag{61}$$

provided the correlation function $b(\mathbf{r},\mathbf{r}')$ can be decomposed into the form of Eq. (30) and Γ can be written in the form of (31). When this is not possible, one must write

$$P_s(\mathbf{m}) = [k^4 \sin^2\psi/(4\pi R)^2] \int_{-\infty}^{\infty} \cdots \int d^3r d^3r' \times b(\mathbf{r},\mathbf{r}')\Gamma(\mathbf{r},\mathbf{r}') \exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}')]. \tag{62}$$

Within the forward-scatter cone, Eqs. (61) and (62) are not valid because $I_4(\mathbf{m})$ is then not negligible in comparison with $I_1(\mathbf{m})$ in (33). However, I_2, I_3, and I_5 are still negligible, provided that $k^2 l^2 \langle \eta^2 \rangle \ll 1$. Consequently, inside the forward-scatter cone we must retain both the first and fourth terms in Eqs. (28) or (33). If we apply Eq. (32) to the fourth term in (28), we have (for $\theta < \Delta\theta_f$)

$$\begin{aligned} P_s(\mathbf{m}) = {} & [k^4 \sin^2\psi/(4\pi R)^2] \int_{-\infty}^{\infty} \cdots \int d^3r d^3r_1 \times b(\mathbf{r},\mathbf{r}_1)\Gamma(\mathbf{r},\mathbf{r}_1) \exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_1)] \\ & + [k^8 \sin^2\psi/(4\pi R)^2] \int_{-\infty}^{\infty} \cdots \int d^3r d^3r_1 d^3r_2 d^3r_3 \times b(\mathbf{r}_1,\mathbf{r}_3) b(\mathbf{r},\mathbf{r}_2) \langle G(\mathbf{r},\mathbf{r}_2) \rangle \times \langle G^*(\mathbf{r}_1,\mathbf{r}_3) \rangle \Gamma(\mathbf{r}_2,\mathbf{r}_3) \exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_1)]. \end{aligned} \tag{63}$$

Equations (63) and (28) can also be used to compute the quantity $\Delta P_s = \langle [E_s(\mathbf{m}) - \langle E_s \rangle]^2 \rangle$. The result is (for $\theta < \Delta\theta_f$)

$$\begin{aligned} \Delta P_s = {} & [k^4 \sin^2\psi/(4\pi R)^2] \int_{-\infty}^{\infty} \cdots \int d^3r d^3r_1 \times b(\mathbf{r},\mathbf{r}_1)\Gamma(\mathbf{r},\mathbf{r}_1) \exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_1)] \\ & + [k^8 \sin^2\psi/(4\pi R)^2] \int_{-\infty}^{\infty} \cdots \int d^3r d^3r_1, d^3r_2 d^3r_3 \end{aligned}$$

$$\times b(\mathbf{r},\mathbf{r}_2)b(\mathbf{r}_1,\mathbf{r}_3)\langle G(\mathbf{r},\mathbf{r}_2)\rangle\langle G^*(\mathbf{r}_1,\mathbf{r}_3)\rangle$$
$$\times[\Gamma(\mathbf{r}_2,\mathbf{r}_3)-\langle e(\mathbf{r}_2)\rangle\langle e^*(\mathbf{r}_3)\rangle]$$
$$\times\exp[-ik\mathbf{m}\cdot(\mathbf{r}-\mathbf{r}_1)]. \tag{64}$$

Note that, in the limit when the field coherence length is large in comparison with l and $V^{1/3}$, we have $\Gamma(\mathbf{r}_2,\mathbf{r}_3)\simeq\langle e(\mathbf{r}_2)\rangle\langle e^*(\mathbf{r}_3)\rangle$ so that the second term in (64) is negligible.

Equations (61)–(64) include all orders of multiple scattering within the medium. That is, the results in (28) and (33) are equivalent to the sum of all terms in a Born series approximation to the integral equations for scattering, and (61)–(64) are valid approximations to (28) and (33), provided that the observation point does not lie in the backscatter cone $\Delta\theta_B$. When the observation point lies in the backscatter cone, one must use the most general expression for the scattered intensity given by Eq. (13). In order to evaluate (13), one must calculate fourth moments of the field quantities, as opposed to second moments for other scattering directions.

It is sometimes convenient to express $B(\mathbf{v})$ and $\tilde{\Gamma}(\mathbf{u},\mathbf{v})$ as Fourier transforms. We therefore write

$$\langle\eta^2\rangle B(\mathbf{v})=\int_{-\infty}^{\infty}\int\int d^3\kappa\,\Phi(\boldsymbol{\kappa})\exp(i\boldsymbol{\kappa}\cdot\mathbf{v}), \tag{65}$$

$$\tilde{\Gamma}(\mathbf{u},\mathbf{v})=\int_{-\infty}^{\infty}\int\int d^3\kappa\,\hat{\Gamma}(\mathbf{u},\boldsymbol{\kappa})\exp(i\boldsymbol{\kappa}\cdot\mathbf{v}), \tag{66}$$

where $\Phi(\boldsymbol{\kappa})$ is the wavenumber spectrum of the relative permittivity fluctuations. Note that $\Phi(\boldsymbol{\kappa})=4\Phi_n(\boldsymbol{\kappa})$, where $\Phi_n(\boldsymbol{\kappa})$ is the wavenumber spectrum of the index-of-refraction fluctuations which is used by many authors. If (65) and (66) are used in (61), the result is

$$P_s(\alpha)=(\pi k^4\sin^2\psi/2R^2)\int_{-\infty}^{\infty}\cdots\int d^3u\,d^3\kappa$$
$$\times C(\mathbf{u})\Phi(k\boldsymbol{\alpha}-\boldsymbol{\kappa})\hat{\Gamma}(\mathbf{u},\boldsymbol{\kappa}). \tag{67}$$

The single-scatter limit can be recovered from (51) by realizing that for sufficiently large correlation lengths $\hat{\Gamma}(\mathbf{u},\boldsymbol{\kappa})$ will peak around $\boldsymbol{\kappa}=0$. Consequently, we may expand $\Phi(k\boldsymbol{\alpha}-\boldsymbol{\kappa})$ in a Taylor series around $\boldsymbol{\kappa}=0$ as

$$\Phi(k\boldsymbol{\alpha}-\boldsymbol{\kappa})=\Phi(k\boldsymbol{\alpha})-[(\boldsymbol{\kappa}\cdot\nabla_p)\Phi(\mathbf{p})]_{\mathbf{p}=k\boldsymbol{\alpha}}+\cdots. \tag{68}$$

If (68) is used in (67), the result is

$$P_s(\alpha)=(\pi k^4\sin^2\psi/2R^2)\Phi(k\boldsymbol{\alpha})\int_{-\infty}^{\infty}\int\int d^3u$$
$$\times C(\mathbf{u})\tilde{\Gamma}(\mathbf{u},0)+\cdots, \tag{69}$$

where we have used (66) to replace $\int\int\int d^3\kappa\,\hat{\Gamma}(\mathbf{u},\boldsymbol{\kappa})$ by $\tilde{\Gamma}(\mathbf{u},0)$. Consequently, we see that the usual[1–5] single-scatter (Booker–Gordon formula) result is simply the first term in a Taylor-series expansion of our more general result.

Although we have discussed primarily continuum inhomogeneities, the generalization to the case of discrete particles is straightforward. In this case the correlation length l is replaced by the particle size D, etc.

VI. AN EXAMPLE

It is interesting to examine the form of the results for the case when a plane wave is incident along the z axis on a volume V containing statistically homogeneous turbulence satisfying a Kolmogorov spectrum,

$$\Phi(\boldsymbol{\kappa})=0.132Cn^2/\kappa^2+1/L_0^2)^{11/6}. \tag{70}$$

Then it can be shown that the mutual coherence function is given by[28]

$$\Gamma(\mathbf{r},\mathbf{r}_1)\simeq\tilde{\Gamma}(\mathbf{u},\mathbf{v})\exp[ik\hat{z}\cdot(\mathbf{r}-\mathbf{r}_1)], \tag{71}$$

where

$$\tilde{\Gamma}(\mathbf{u},\mathbf{v})=\exp[-\beta_0|\mathbf{v}_\perp|^{5/3}u_z-\gamma|v_z|], \tag{72}$$

and $\mathbf{u}=(\mathbf{r}+\mathbf{r}_1)/2$, $\mathbf{v}=\mathbf{r}-\mathbf{r}_1$, $\mathbf{v}_\perp$ is the component of $\mathbf{v}$ transverse to the z axis, $\beta_0=1.46\,k^2Cn^2$, $\gamma=0.39k^2C_n^2L_0^{5/3}$, C_n^2 is the index-of-refraction structure constant for the turbulence, and L_0 is the outer scale size of the turbulent eddies. For Kolomogorov turbulence, the condition $\langle\eta^2\rangle k^2l^2\ll1$ becomes $k^2C_n^2L_0^{8/3}\ll1$. If (72) is used in (61) and we make use of the fact that, because $\tilde{\Gamma}(\mathbf{u},\mathbf{v})$ is multiplied by $B(\mathbf{v})$, the term $\gamma|v_z|$ in (69) is at most of order $0.39k^2Cn^2L_0^{8/3}\ll1$ (because $k^2\langle\eta^2\rangle l^2=0.39k^2Cn^2L_0^{8/3}\ll1$ by assumption), we find, provided that the observation point does not lie in either the forward-or backward-scatter cones,

$$P_s(\alpha)=(k^4\sin^2\psi/4R^2)\int_{-\infty}^{\infty}\int\int d^3uC(\mathbf{u})\int_{-\infty}^{\infty}\int d\kappa_x\,d\kappa_y$$
$$\times\Phi(\kappa_x,\kappa_y,k\alpha_z)H(u_z,|\boldsymbol{\kappa}_\perp-k\boldsymbol{\alpha}_\perp|), \tag{73}$$

where

$$H(u_z,t)=\int_0^\infty v\,dv\,J_0(tv)\exp(-\beta_0u_zv^{5/3}) \tag{74}$$

and $J_0(\cdots)$ is the zero-order Bessel function. In a future paper, (73) will be evaluated and compared with the single scatter result in Eq. (69), for a variety of conditions.

APPENDIX A

Here we will derive the scattering formula in Eq. (1). The vector Maxwell-wave equation can be written as

$$(\nabla^2+k^2)\mathbf{e}(\mathbf{r})=-\eta k^2\mathbf{e}(\mathbf{r})-\nabla[(\mathbf{e}\cdot\nabla\eta)/(1+\eta)], \tag{A1}$$

where the relative permittivity $\epsilon/\epsilon_0=1+\eta(\mathbf{r})$. Equation (A1) can be rewritten in integral form as

$$\mathbf{e}(\mathbf{r}')=(1/4\pi)\int_V\int\int d^3r\,\{k^2\eta\mathbf{e}+\nabla[(\mathbf{e}\cdot\nabla\eta)/(1+\eta)]\}$$
$$\times\exp(ik|\mathbf{r}-\mathbf{r}'|)/|\mathbf{r}-\mathbf{r}'|. \tag{A2}$$

In the Fraunhofer zone of the scattering volume shown in Fig. 1, Eq. (A2) becomes

$$\mathbf{E}_s(\mathbf{m})=(1/4\pi R)\int\int\int d^3r\{k^2\eta\mathbf{e}+\nabla[(\mathbf{e}\cdot\nabla\eta)/(1+\eta)]\}$$
$$\times\exp(ikR-ik\mathbf{m}\cdot\mathbf{r}), \tag{A3}$$

where $\mathbf{m}$ and R are shown in Fig. 1. We now set $u=(1+\eta)^{-1}\mathbf{e}\cdot\nabla\eta$ and $v=\exp(-ik\mathbf{m}\cdot\mathbf{r})$ and then use $\nabla(uv)=u\nabla v+v\nabla u$, plus the theorem from vector integral calculus

$$\int\int_V \int \nabla(uv)d^3r = \int\int_S uv\mathbf{n}\,dS, \tag{A4}$$

where S is any surface enclosing the volume V and $\mathbf{n}$ is the unit normal to S. Upon using these results, along with the fact that $\eta(\mathbf{r}) = 0$ on S, we find that (A3) becomes

$$E_s(\mathbf{m}) = (k/4\pi R)\iiint d^3r\{\eta k\mathbf{e} + i\mathbf{m}[(\mathbf{e}\cdot\nabla\eta)/(1+\eta)]\} \times\exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR). \tag{A5}$$

We next set $\mathbf{w} = \mathbf{e}\exp(-ik\mathbf{m}\cdot\mathbf{r})$ and $\psi = \ln(1+\eta)$, and then use the vector identity $\nabla\cdot(\mathbf{w}\,\psi) = \psi\nabla\cdot\mathbf{w} + \mathbf{w}\cdot\nabla\psi$ along with Gauss' theorem

$$\iiint_V \nabla\cdot\mathbf{a}\,d^3r = \iint_S \mathbf{a}\cdot\mathbf{n}\,dS. \tag{A6}$$

Upon applying these results to the second term in the square bracket is (A5), along with the fact that $\eta = \ln(1+\eta) = 0$ on S, we obtain, after some manipulation,

$$E_s(\mathbf{m}) = (k/4\pi R)\iiint d^3r \times[\eta k\mathbf{e} - k\psi\mathbf{m}(\mathbf{m}\cdot\mathbf{e}) + i\mathbf{m}\psi(\mathbf{e}\cdot\nabla\psi)] \times\exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR). \tag{A7}$$

We next expand $\psi = \ln(1+\eta)$ in a Taylor series in $\eta \ll 1$ as $\psi \simeq \eta - \eta^2/2 + \cdots$. If this is done, we obtain, after some differentiations, the result

$$E_s(\mathbf{m}) = (k^2/4\pi R)\iiint d^3r\,\eta[\mathbf{e} - \mathbf{m}(\mathbf{m}\cdot\mathbf{e})] \times\exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR) + (ik\mathbf{m}/8\pi R)\exp(ikR)\iiint d^3r\mathbf{e}\cdot\nabla \times[\eta^2\exp(-ik\mathbf{m}\cdot\mathbf{r})] + \text{terms of higher order in } \eta. \tag{A8}$$

If we apply the result in (A4), along with the fact that $\eta = 0$ on S, to the last term in Eq. (A8) the result is

$$\mathbf{E}_s(\mathbf{m}) = (k/4\pi R)\iiint d^3r \times\{k\eta[\mathbf{e} - \mathbf{m}(\mathbf{m}\cdot\mathbf{e})] - \tfrac{1}{2}i\mathbf{m}\eta^2\nabla\cdot\mathbf{e}\} \times\exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR) + \cdots. \tag{A9}$$

Finally we use the Maxwell Equation $\nabla\cdot\epsilon\mathbf{e} = 0$ to rewrite

$$\nabla\cdot\mathbf{e} = (1-\eta)^{-1}\mathbf{e}\cdot\nabla\eta \simeq \mathbf{e}\cdot\nabla\eta.$$

Therefore, Eq. (A9) becomes

$$\mathbf{E}_s(\mathbf{m}) = (k^2/4\pi R)\iiint d^3r\,\eta(\mathbf{r})[\mathbf{e}(\mathbf{r}) - \mathbf{m}(\mathbf{m}\cdot\mathbf{e})] \times\exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR) + (ik\mathbf{m}/8\pi R)\iiint d^3r\,\eta^2(\mathbf{r})\mathbf{e}\cdot\nabla\eta \times\exp(-ik\mathbf{m}\cdot\mathbf{r} + ikR) + \cdots. \tag{A10}$$

The second term in (A10) is of order $\eta^2(kl)^{-1}$ in comparison with the first, where l is the spatial size of the permittivity fluctuations. Because $|\eta| \ll 1$ we see that unless $kl \ll 1$, the second term in (A10) is negligible in comparison with the first term. Consequently, this is all that we have retained in Eq. (1).

APPENDIX B

The scalar wave equation for propagation in a randomly inhomogeneous medium is

$$(\nabla^2 + k^2[1 + \eta(\mathbf{r})])e(\mathbf{r}) = 0, \tag{B1}$$

where $e(\mathbf{r})$ is the field strength, k is the vacuum wavenumber, and $\eta(\mathbf{r})$ is the fluctuation in relative permittivity. Also, the Green's function for propagation in this medium satisfies

$$\{\nabla^2 + k^2[1 + \eta(\mathbf{r})]\}G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \tag{B2}$$

where $\delta(\mathbf{r} - \mathbf{r}')$ is the three-dimensional Dirac delta function.

Now we operate on (B1) with the variational derivative $\delta/\delta\eta(\mathbf{r}')$, and use the result that

$$\frac{\delta\eta(\mathbf{r})}{\delta\eta(\mathbf{r}')} = \delta(\mathbf{r} - \mathbf{r}'). \tag{B3}$$

Then (B1) becomes

$$\{\nabla^2 + k^2[1 + \eta(\mathbf{r})]\}\frac{\delta e(\mathbf{r})}{\delta\eta(\mathbf{r}')} = -k^2 e(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}'). \tag{B4}$$

Upon comparing (B4) with (B2) it is evident that

$$\frac{\delta e(\mathbf{r})}{\delta\eta(\mathbf{r}')} = -k^2 e(\mathbf{r}')G(\mathbf{r},\mathbf{r}'). \tag{B5}$$

Similarly, we can take the variational derivative of (B2) to obtain

$$\{\nabla^2 + k^2[1 + \eta(\mathbf{r})]\}\frac{\delta G(\mathbf{r},\mathbf{r}')}{\delta\eta(\mathbf{r}'')} = -k^2 G(\mathbf{r},\mathbf{r}')\delta(\mathbf{r} - \mathbf{r}''). \tag{B6}$$

Upon comparing (B6) with (B2), and using the fact that $G(\mathbf{r},\mathbf{r}')\delta(\mathbf{r} - \mathbf{r}'') = G(\mathbf{r}'',\mathbf{r}')\delta(\mathbf{r} - \mathbf{r}'')$, we see that

$$\frac{\delta G(\mathbf{r},\mathbf{r}')}{\delta\eta(\mathbf{r}'')} = -k^2 G(\mathbf{r},\mathbf{r}'')G(\mathbf{r}'',\mathbf{r}'). \tag{B7}$$

APPENDIX C

By using (16) the quantity $g_1 = \langle G(\mathbf{r},\mathbf{r}_2)G(\mathbf{r}_2,\mathbf{r}_3)\rangle$ in (22) can be rewritten as

$$g_1 = G_0(\mathbf{r},\mathbf{r}_2)G_0(\mathbf{r}_2,\mathbf{r}_3) \times\langle\exp[i\phi(\mathbf{r},\mathbf{r}_2) + i\phi(\mathbf{r}_2,\mathbf{r}_3)]\rangle. \tag{C1}$$

If ϕ is assumed to be a zero-mean Gaussian random variable, the ensemble average in (C1) can be performed to give

$$g_1 = \langle G(\mathbf{r},\mathbf{r}_2)\rangle\langle G(\mathbf{r}_2,\mathbf{r}_3)\rangle \times\exp[-\langle\phi(\mathbf{r},\mathbf{r}_2)\phi(\mathbf{r}_2,\mathbf{r}_3)\rangle], \tag{C2}$$

where

$$\langle G(\mathbf{r},\mathbf{r}_2)\rangle = G_0(\mathbf{r},\mathbf{r}_2)\exp(-\tfrac{1}{2}\langle\phi^2(\mathbf{r},\mathbf{r}_2)\rangle). \tag{C3}$$

The exponent in (C2) can be calculated by using the result in (26). We see that unless the path from $\mathbf{r}_2$ to $\mathbf{r}$ completely overlaps the path from $\mathbf{r}_3$ to $\mathbf{r}_2$ the quantity $\langle\phi(\mathbf{r},\mathbf{r}_2)\phi(\mathbf{r}_2,\mathbf{r}_3)\rangle \simeq k^2\langle\eta^2\rangle l^2/4 \ll 1$. Consequently,

$$g_1 = \langle G(\mathbf{r},\mathbf{r}_2)\rangle\langle G(\mathbf{r}_2,\mathbf{r}_3)\rangle. \tag{C4}$$

It is also desirable to calculate the average $\langle\phi^2(\mathbf{r},\mathbf{r}_2)\rangle$ which appears in (C3). By using (24) we find

$$\langle \phi^2(\mathbf{r},\mathbf{r}_2)\rangle = \tfrac{1}{4}k^2 \int_{\mathbf{r}_2}^{\mathbf{r}} ds \int_{\mathbf{r}_2}^{\mathbf{r}} dt\, b(\xi,\tau). \tag{C5}$$

If we use (30) in (C5) and assume $C[(\xi+\tau)/2]$ is nearly constant and equal to unity over distances of order of $|\mathbf{r}-\mathbf{r}_2|$, we find

$$\langle \phi^2(\mathbf{r},\mathbf{r}_2)\rangle \simeq \tfrac{1}{4}k^2\langle\eta^2\rangle |\mathbf{r}-\mathbf{r}_2| \int_{-\infty}^{\infty} d\xi B(\xi)$$

$$\simeq \tfrac{1}{4}k^2\langle\eta^2\rangle l\, |\mathbf{r}-\mathbf{r}_2|. \tag{C6}$$

Consequently, if we define $\rho_\phi = 8(k^2\langle\eta^2\rangle l)^{-1}$, we can write

$$\langle G(\mathbf{r},\mathbf{r}_2)\rangle = G_0(\mathbf{r},\mathbf{r}_2)\exp(-|\mathbf{r}-\mathbf{r}_2|/\rho_\phi). \tag{C7}$$

Note that because $k^2\langle\eta^2\rangle l^2$ is assumed to be small in comparison with unity, the exponential term in (C7) can be approximated by unity when $|\mathbf{r}-\mathbf{r}_2|$ is less than or of order of the correlation length l.

APPENDIX D

Let us evaluate integrals of the form

$$S = \int\int\int_{-\infty}^{\infty} d^3t\, G_0(t)\exp(-\gamma t + ik\mathbf{p}\cdot\mathbf{t}). \tag{D1}$$

By substituting $G_0(t) = (4\pi t)^{-1}\exp(ikt)$ and letting $\mathbf{p}$ lie along the polar axis in a spherical coordinate system in $d^3t = t^2\sin\theta\, dt d\theta d\phi$ space, we can rewrite (D1) as

$$S = \tfrac{1}{2}\int_0^\infty t^2 dt \exp(ikt-\gamma t)\int_0^\pi d\theta \sin\theta \exp(ikpt\cos\theta), \tag{D2}$$

where $p = |\mathbf{p}|$. The integrals on θ and t are readily performed to give

$$S = [\gamma - ik(1+p)]^{-1}[\gamma - ik(1-p)]^{-1}. \tag{D3}$$

For $p \neq 1$ and $k \gg \gamma$, we have

$$S \simeq [k^2(p^2-1)]^{-1}. \tag{D4}$$

For $p = 1$ and $k \gg \gamma$, we get

$$S \simeq i(2\gamma k)^{-1}. \tag{D5}$$

APPENDIX E

In order to evaluate (49) and (51), we are required to evaluate integrals of the form

$$J = \int\int\int_{-\infty}^{\infty} d^3\tau \exp(-\tau/l - i2k\mathbf{m}\cdot\tau)$$

$$\times \int\int\int_{-\infty}^{\infty} d^3t \exp(ikt - t/l + ik|2\tau-\mathbf{t}|)/t|2\tau-\mathbf{t}|, \tag{E1}$$

where $\tau = |\tau|$ and $t = |\mathbf{t}|$. The integral in (E1) cannot be evaluated exactly, but its magnitude can be estimated by splitting the range of integration on dt into two domains: one from $|\mathbf{t}| = 0$ to $|2\tau|$ and the other from $t = |2\tau|$ to infinity. If this is done and we approximate $|2\tau - \mathbf{t}|$ by $|2\tau|$ for $0 < |\mathbf{t}| < |2\tau|$ and by t for $|\mathbf{t}| > |2\tau|$, we obtain

$$J \sim 4\pi \int\int\int_{-\infty}^{\infty} d^3\tau \exp(-\tau/l - i2k\mathbf{m}\cdot\tau)$$

$$\times\left[\int_0^{2\tau}(tdt/2\tau)\exp(ikt - t/l + i2k\tau)\right.$$

$$\left. + \int_{2\tau}^{\infty} dt \exp(i2kt - t/l)\right]. \tag{E2}$$

The integrations in (E2) are readily performed to give an order of magnitude estimate of the value of J.

[1] H. Booker and W. Gordon, Proc. IRE **38**, 401 (1950).
[2] F. Villars and V. Weisskopf, Phys. Rev. **94**, 232 (1954).
[3] E. Salpeter and S. Trieman, J. Geophys. Res. **69**, 869 (1964).
[4] V. Tatarski, *Wave Propagation in a Turbulent Medium* (Dover, New York, 1961).
[5] A. Wheelon, J. Res. NBS D **63**, 205 (1959).
[6] D. Wort, J. Phys. A **8**, 75 (1969).
[7] J. Sheffield, *Plasma Scattering of Electromagnetic Radiation* (Academic, New York, 1975).
[8] K. Hardy and I. Katz, Proc. IEEE **57**, 468 (1969).
[9] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
[10] H. Bremmer, J. Res. NBS D **68**, 967 (1964).
[11] J. Milton *et al.*, Appl. Opt. **11**, 697 (1972).
[12] R. Anderson and E. Browell, Appl. Opt. **11**, 1345 (1972).
[13] B. Brinkworth, Appl. Opt. **12**, 427 (1973).
[14] A. Cohen, Appl. Opt. **14**, 2873 (1975).
[15] D. deWolf, IEEE Trans. Antennas Propagation **AP-19**, 254 (1971).
[16] D. deWolf, IEEE Trans. Antennas Propagation **AP-20**, 805 (1972).
[17] D. Feinstein *et al.*, Phys. Fluids **15**, 1641 (1972).
[18] V. Granatstein, Appl. Opt. **11**, 1217 (1972).
[19] R. Fante, IEEE Trans. Antennas Propagation **AP-21**, 750 (1973).
[20] S. Ito and S. Adachi, IEEE Trans. Antennas Propagation **AP-25**, 205 (1977).
[21] A. Ishimaru and K. Painter, Radio Sci. **15**, 87 (1980).
[22] L. Tsang and J. Kong, Radio Sci. **13**, 753 (1978); **14**, 359 (1979).
[23] L. Tsang and J. Kong, J. Appl. Phys. **50**, 6600 (1979).
[24] V. Tatarskii, *The Effect of the Turbulent Atmosphere on Wave Propagation* (U.S. Dept. Commerce, Springfield, Va., 1971).
[25] Z. Feizulen and Y. Kravtsov, Radiophys. Quant. Electron. **10**, 33 (1967).
[26] R. Lutomirski and H. Yura, Appl. Opt. **10**, 1652 (1971).
[27] R. Fante, Proc. IEEE **63**, 1669 (1975).
[28] R. Fante, Radio Sci. **15**, 757 (1980).
[29] M. Born and E. Wolf, *Principles of Optics* (Macmillan, New York, 1964).
[30] V. Tatarskii, Ref. 24, Sec. 37.

[31] The zero-mean assumption is not necessary and is done only for mathematical simplicity. The analysis can also be done for $\langle\phi\rangle \neq 0$.

[32] V. Tatarski, Ref. 4, Chap. 6.
[33] V. Tatarskii, Ref. 24, Sec. 39.

[34] It can be shown that the mean free path l_m between individual scattering events is $l_m \simeq (k^2\langle\eta^2\rangle l)^{-1}$. Thus, if the extent of the scattering volume in the z direction is much greater than one free path for scattering, the possibility will exist for $\langle\phi\phi\rangle \sim k^2\langle\eta^2\rangle Ll$ to be much greater than unity. In the limit when the scattering medium is smaller than l_m, $\langle\phi\phi\rangle$ will always be less than unity. In this case the usual Born approximation is valid.
[35] The requirement expressed by (27) can be shown to be equivalent to the requirement that the mean free path between individual scatterings in the medium is large in comparison with the spatial size l of the random inhomogeneities.

Shift operator techniques for the classification of multipole–photon states. X. P_l^0 eigenstate and eigenvalue determination in G_2

H. E. De Meyer,[a] G. Vanden Berghe, and J. Van der Jeugt[b]
Seminarie voor Wiskundige Natuurkunde, R.U.G., Krijgslaan 271-S9, B-9000 Gent, Belgium

(Received 4 June 1981; accepted for publication 19 February 1982)

On account of previously derived relations between quadratic shift operator products in the group G_2, part of the eigenvalue spectrum of the scalar shift operator P_l^0 is derived in closed form. The corresponding eigenstates which are closely related to the octupole–phonon states are defined in terms of angular momentum lowering shift operator actions upon the maximum angular momentum state.

PACS numbers: 63.20.Dj, 02.20.Nq

I. INTRODUCTION

Previously we have established in two separate papers[1,2] (to be referred to as IV and IX, respectively) relations which connect quadratic products of the shift operators P_l^k ($|k|<5$) made up with G_2 group generators. In IV we considered scalar product operators of the form $P_{1+k}^{-k}P_1^{+k}$ ($|k|<5$) together with the G_2 Casimir V^*, whereas the nonscalar extension leading to 25 independent relations among products of the form $P_l^{+j}P_l^{+k}$ ($|j|,|k|<5$; $-10<j+k<0$) has been treated in IX.

In the present paper we want to derive formulae expressing the P_1^0 eigenvalues and eigenstates by appropriate use of both the scalar and the nonscalar relations. Exactly as for the R(5) case[3] and the SU(3) case,[4,5] it is the latter type of relations which permits a step-by-step determination of consecutive eigenvalues and eigenvectors, hence avoiding cumbersome tree generation. Moreover, in the R(5) treatment[3] it has been shown that even the l-multiplicity of states can be obtained on account of shift operator calculations. Here, for the sake of brevity, we shall not complete the argumentation to that point and therefore we shall merely assume at each step that the precise number of degenerated states is known in advance. Also we shall be brief at those places where the reasoning is analogous to the one expounded in the context of R(5).[3]

Looking at the symmetric irreducible representations of G_2, which are closely related to the nuclear octupole–phonon states, one is struck by the abundancy of l-degeneracies even for the higher angular momentum states. This fact necessitates a compact notation by which redundant repetition in defining eigenstates from shift operator actions can be avoided. Therefore, let us introduce in a general way coefficients $a_{v,l-k}^{(j)}[l,(i)]$ by means of the following formula:

$$P_l^{-k}|v,l(i)\rangle = \sum_j a_{v,l-k}^{(j)}[l,(i)]\,|v,l-k,(j)\rangle \quad (0<k<5). \tag{1.1}$$

Herein the states $|v,l,(i)\rangle(|v,l-k,(j)\rangle)$ where i (j) takes on integer values between one and the number indicating the l-multiplicity of states with seniority v and angular momentum l ($l-k$), form a set of orthonormalized eigenstates of P_l^0 (P_{l-k}^0). The additional label i between parentheses shall be omitted whenever $|v,l>$ is not degenerated. We also define

$$P_l^0|v,l,(i)\rangle = \alpha_{v,l}^{(i)}|v,l,(i)\rangle, \tag{1.2}$$

showing that $\alpha_{v,l}^{(i)}$ denotes the eigenvalue of P_l^0 with respect to the state $|v,l,(i)\rangle$. As a consequence of Hermiticity properties of the shift operators it has been demonstrated by Hughes and Yadegar[6] that

$$\sum_j |a_{v,l-k}^{(j)}[l,(i)]|^2 = (\beta_{k,l-k})^{-1}\langle v,l,(i)|P_{l-k}^{+k}P_l^{-k}|v,l,(i)\rangle \tag{1.3}$$

with

$$\beta_{k,l} = \frac{2l+1}{2l+2k+1}. \tag{1.4}$$

In the following sections we shall derive closed expressions for all eigenvalues $\alpha_{v,l}^{(i)}$ ($3v-5\leq l\leq 3v$) starting at the highest possible angular momentum $l=3v$ and proceeding thereafter by shifting downward the l-value.

II. THE HIGH ANGULAR MOMENTUM STATES AND THEIR EIGENVALUES

The determination of the eigenvalues $\alpha_{v,3v}$ of the nondegenerate maximum angular momentum states $|v,3v\rangle$ ($v=0,1,2,\cdots$) is straightforward if we assume that $|v,3v-1\rangle$ states are nonexistent. Indeed, when we let Eqs. (IV.3.2) and (IV.3.3) act upon $|v,3v\rangle$, use the property that for all acceptable l-values (and possible i-values)

$$V^*|v,l,(i)\rangle = -v(v+5)/3, \tag{2.1}$$

and multiply on the left by $\langle v,3v|$, we immediately obtain, on account of the normalization of states, a system of two quadratic equations in the unknown $\alpha_{v,3v}$. This yields as a unique solution the expression

$$\alpha_{v,3v} = -\tfrac{2}{7}\sqrt{3}\,v(v+1)(2v+1)(3v+1)(3v+2)(6v+5) \quad (v>0). \tag{2.2}$$

The unambiguous derivation of $\alpha_{v,3v-2}$ eigenvalues is the first example where relations between nonscalar product operators play a significant role. The nondegenerate and normalized state $|v,3v-2\rangle$ is proportional to the action of P_{3v}^{-2} upon $|v,3v\rangle$. Following (1.1), (1.3), and (1.4) the proportionality factor $a_{v,3v-2}[3v]$ satisfies

[a] Research Associate N.F.W.O. (Belgium).
[b] Research Assistant N.F.W.O. (Belgium).

 0022-2488/82/061223-05$02.50

$$|a_{v,3v-2}[3v]| = \left[\frac{6v+1}{6v-3}\langle v,3v|P^{+2}_{3v-2}\,P^{-2}_{3v}|v,3v\rangle\right]^{1/2}. \quad (2.3)$$

Next, applying Eq. (IX.2.6) on $|v,3v\rangle$, we obtain after some cancellations

$$\begin{aligned}&[3(v+1)(2v+1)(3v+2)(6v+5)P^0_{3v-2}\,P^{-2}_{3v}\\&\quad-(2v-7)(3v-1)(3v+10)(6v+1)P^{-2}_{3v}\,P^0_{3v}\\&\quad+(2^3/7)\sqrt{3}\,v(v+1)(2v+1)(3v-1)(3v+1)\\&\quad\times(3v+2)(6v-1)(6v+1)(6v+5)\,P^{-2}_{3v}]|v,3v\rangle=0.\end{aligned} \quad (2.4)$$

Since $P^0_{3v}|v,3v\rangle = \alpha_{v,3v}|v,3v\rangle$ and $\alpha_{v,3v}$ is already known, it directly follows from Eq. (2.4) that

$$\begin{aligned}\alpha_{v,3v-2} = &-(2\sqrt{3}/21)v(3v-1)(3v+1)\\&\times(6v+1)(6v^2+23v-74)\quad(v\geqslant 1).\end{aligned} \quad (2.5)$$

Finally, the action of Eq. (IV.3.4) upon $|v,3v\rangle$ delivers, after substitution of the expression for $P^0_{3v}|v,3v\rangle$ and multiplication on the left by $\langle v,3v|$, the matrix element

$\langle v,3v|P^{+2}_{3v-2}\,P^{-2}_{3v}|v,3v\rangle$, i.e.,

$$\begin{aligned}\langle v,3v|P^{+2}_{3v-2}\,P^{-2}_{3v}|v,3v\rangle &= (2^3 3^4/5)v^4(v-1)\\&\times(2v+1)^2(3v-1)^3(3v+1)^2(3v+2)^2(6v+1)\quad(v\geqslant 1).\end{aligned} \quad (2.6)$$

Hence, the normalization factor $a_{v,3v-2}[3v]$ is in absolute value fully determined by (2.3) and (2.6). If we impose the reality of coefficients, the choice of sign is arbitrary. So let us put

$$a_{v,3v-2}[3v] = +|a_{v,3v-2}[3v]|. \quad (2.7)$$

For a fixed seniority number $v\geqslant 2$ the state $|v,3v-3\rangle$ is always nondegenerate. There are, however, two ways to arrive at this state by shift operator actions which lower the angular momentum. Indeed, we can either let P^{-3}_{3v} act upon $|v,3v\rangle$ or P^{-1}_{3v-2} upon $|v,3v-2\rangle$. The corresponding normalization factors are easily seen from (1.3) and (1.4) to satisfy

$$\begin{aligned}&|a_{v,3v-3}[3v]|\\&\quad=\left[\frac{6v+1}{6v-5}\,\langle v,3v|P^{+3}_{3v-3}\,P^{-3}_{3v}|v,3v\rangle\right]^{1/2},\end{aligned} \quad (2.8)$$

$$\begin{aligned}&|a_{v,3v-3}[3v-2]|\\&\quad=\left[\frac{6v-3}{6v-5}\,\langle v,3v-2|P^{+1}_{3v-3}\,P^{-1}_{3v-2}|v,3v-2\rangle\right]^{1/2}.\end{aligned} \quad (2.9)$$

The calculation of the eigenvalue expression for $\alpha_{v,3v-3}$ is straightforward and proceeds in exactly the same manner as for $\alpha_{v,3v-3}$. The relevant equation is now (IX.2.11), from which, on substitution of the expression for $P^0_{3v}|3v\rangle$, it follows that

$$\begin{aligned}\alpha_{v,3v-3} = &-(2\sqrt{3/7})v(2v-1)(6v-1)\\&\times(9v^3-18v^2+11v-282)\quad(v>2).\end{aligned} \quad (2.10)$$

It suffices to substitute expression (2.10) and also consecutively (2.7), (2.3), and (2.6) into Eq. (IX.2.10), acting upon $|v,3v\rangle$, in order to obtain a linear homogeneous relation between $a_{v,3v-3}[3v]$ and $a_{v,3v-3}[3v-2]$, namely

$$\begin{aligned}&10^{1/2}\,a_{v,3v-3}[3v] + 3v(2v+1)(3v+1)(3v+2)\\&\quad\times[(v-1)(3v-1)/(2v-1)]^{1/2}\,a_{v,3v-3}[3v-2]=0.\end{aligned} \quad (2.11)$$

It is clear from this relation that by making the choice

$$a_{v,3v-3}[3v] = +|a_{v,3v-3}[3v]|, \quad (2.12)$$

one necessarily has that

$$a_{v,3v-3}[3v-2] = -|a_{v,3v-3}[3v-2]|. \quad (2.13)$$

To complete the analysis of the $|v,3v-3\rangle$ states, there remains to determine the matrix elements occurring on the right-hand sides of (2.8) and (2.9). This can again be accomplished by using relations among scalar product operators. More precisely, from the action of Eq. (IV.3.7) upon $|v,3v\rangle$ it is straightforward to prove that

$$\begin{aligned}&\langle v,3v|P^{+3}_{3v-3}\,P^{-3}_{3v}|v,3v\rangle\\&\quad=(2^5 3^5/5)v^4(v-1)(2v-1)(2v+1)^2(3v-1)^2(3v+1)^2\\&\quad\times(3v-2)^2(6v+1)(6v-1),\end{aligned} \quad (2.14)$$

whereas from the action of Eq. (IV.3.3) upon $|v,3v-2\rangle$ it follows that

$$\begin{aligned}&\langle v,3v-2|P^{+1}_{3v-3}\,P^{-1}_{3v-2}|v,3v-2\rangle\\&\quad=2^6 3^2 v^2(v-2)(2v-1)(3v-1)\\&\quad\times(3v-2)^2(6v-1)(6v+1)^2.\end{aligned} \quad (2.15)$$

The reader can easily verify that the substitution of (2.14), (2.15), respectively, in (2.8), (2.9) leads, on account of the sign convention (2.12), (2.13), to expressions for $a_{v,3v-3}[3v]$ and $a_{v,3v-3}[3v-2]$ which satisfy Eq. (2.11). Also, it has to be noticed that the matrix elements in (2.14) and (2.15) vanish when $v=2$, which shows that the particular state $|2,3\rangle$ does not exist. For this reason, $v=2$ has been excluded already in formula (2.10).

III. ANALYSIS OF THE $l = 3v-4$ STATES

It is known that for $v>3$ there exist in general two independent orthonormal states $|v,3v-4,(1)\rangle$ and $|v,3v-4,(2)\rangle$. On the other hand, we can indicate three possible ways to generate linear combinations of these states by l-lowering shift operator actions, namely $P^{-4}_{3v}|v,3v\rangle$, $P^{-2}_{3v-2}|v,3v-2\rangle$, and $P^{-1}_{3v-3}|v,3v-3\rangle$. Equation (IX.2.14) upon $|v,3v\rangle$ produces the following relationship among $\alpha^{(i)}_{v,3v-4}$, $a^{(i)}_{v,3v-4}[3v]$, and $a^{(i)}_{v,3v-4}[3v-2]$ for $i=1$ and $i=2$:

$$\begin{aligned}&[(6v-1)\,\alpha^{(i)}_{v,3v-4} + (2\sqrt{3/7})(3v-1)(3v-2)(6v-5)(12v^4-92v^3+129v^2-55v+426)]\,a^{(i)}_{v,3v-4}[3v]\\&\quad+2^4 3^3 5\,v(2v-1)(3v-1)(3v-2)(6v+1)[2(v-1)(3v-1)/3.5(2v-1)]^{1/2}\,a^{(i)}_{v,3v-4}[3v-2]=0\quad(i=1,2).\end{aligned} \quad (3.1)$$

Similarly, a relation among $\alpha^{(i)}_{v,3v-4}$, $a^{(i)}_{v,3v-4}[3v]$, and $a^{(i)}_{v,3v-4}[3v-3]$ is obtained from the action of Eq. (IX.2.15) upon $|v,3v\rangle$, i.e.,

$$[(6v-1)\,\alpha^{(i)}_{v,3v-4}+(2\sqrt{3}/7)(v-1)(3v-2)(v-2)(6v-1)(36v^3-204v^2-17v-225)]\,a^{(i)}_{v,3v-4}[3v]$$
$$+2^2 3^3 5v^2(3v-1)(3v-2)^2(3v+1)(6v+1)$$
$$\times[2.3(v-1)(v-2)(2v-1)(6v-1)/5(6v-5)]^{1/2}\,a^{(i)}_{v,3v-4}[3v-3]=0 \quad (i=1,2). \tag{3.2}$$

Finally, a third relation connecting the four unknowns is found either from the action of Eq. (IX.2.1) upon $|v,3v-3\rangle$ or from the action of Eq. (IX.2.6) upon $|v,3v-2\rangle$. Let us mention here the first of these:

$$v^2(3v-1)(3v-2)(3v+1)(6v-1)(6v+1)$$
$$\times[\alpha^{(i)}_{v,3v-4}+(2\sqrt{3}/7))2v-1)(6v-5)(9v^4+54v^3-133v^2-194v-1136)]$$
$$\times a^{(i)}_{v,3v-4}[3v-3]-2^5 5v(v-1)(2v-1)(3v-1)(3v+1)(3v+5)(6v+1)$$
$$\times[3(v-2)(3v-1)(6v-1)(6v-5)]^{1/2}$$
$$\times a^{(i)}_{v,3v-4}[3v-2]+2^3 5(v-1)(3v-2)(4v+3)(6v-5)$$
$$\times[2.3(v-1)(v-2)(2v-1)(6v-1)(6v-5)/5]^{1/2}\,a^{(i)}_{v,3v-4}[3v]=0$$
$$(i=1,2). \tag{3.3}$$

Equations (3.1), (3.2), and (3.3), which are homogeneous with respect to the $a^{(i)}_{v,3v-4}$'s for $i=1$ or 2, yield as an eliminant a quadratic equation to be satisfied by both $\alpha^{(i)}_{v,3v-4}$'s. Replacing Eq. (3.3) by the one obtained from Eq. (IX.2.6) reproduces the same eliminant, which reads

$$(\alpha^{(i)}_{v,3v-4})^2+(4\sqrt{3}/7)(3v-1)(3v-2)(12v^4-16v^3-147v^2-54v-145)\,\alpha^{(i)}_{v,3v-4}$$
$$+\tfrac{12}{49}(v-1)(v-2)(2v-1)(3v-1)^2(3v-2)^2(6v-5)(12v^4+20v^3-851v^2$$
$$+1219v+1980)=0 \quad (i=1,2). \tag{3.4}$$

Consequently, we find

$$\alpha^{(i)}_{v,3v-4}=-(2\sqrt{3}/7)(3v-1)(3v-2)[(12v^4-16v^3-147v^2$$
$$-54v-145)+7(-1)^{i-1}\Gamma^{1/2}],$$
$$\Gamma=144v^6-1152v^5+2568v^4-216v^3-839v^2$$
$$+1970+25 \quad (i=1,2)\ (v>3). \tag{3.5}$$

In order to obtain for each eigenvalue expression (3.5) the corresponding orthonormalization factors $a^{(i)}_{v,3v-4}[3v-k\,](k=0,2,3)$ at least two relations of the type (1.3) are needed. Hence, we must calculate matrix elements first. To this aim, we can again rely upon appropriate relations between scalar shift operator products. Without going into the details of the straightforward but lengthy calculations, we just mention the following results:

$$\langle v,3v|P^{+4}_{3v-4}\,P^{-4}_{3v}|v,3v\rangle=(2^4 3^7/5)v^4(v-1)^3(3v-1)^3(3v-2)^3(3v+1)^2(6v+1)(6v-1)(12v^2-40v+37), \tag{3.6}$$

$$\langle v,3v-2|P^{+2}_{3v-4}\,P^{-2}_{3v-2}|v,3v-2\rangle=(2^4 3^4/5)v^2(v-1)^2(3v-1)^2(3v-2)^3(6v-1)(36v^5-156v^4$$
$$+157v^3-40v^2-237v+730), \tag{3.7}$$

$$\langle v,3v-3|P^{+1}_{3v-4}\,P^{-1}_{3v-3}|v,3v-3\rangle$$
$$=(2^7 3^3/5)(v-1)^2(2v-1)(3v-1)(3v-2)(6v-1)^2(18v^4-48v^3+8v^2-43v-60). \tag{3.8}$$

With these expressions we have in addition to the homogeneous equations (3.1)–(3.3) and the eigenvalue solutions (3.5), three relations of the form

$$\sum_i |a^{(i)}_{v,3v-4}[3v-k\,]|^2$$
$$=[(6v-2k+1)/(6v-7)]\langle v,3v-k|P^{4-k}_{3v-4}\,P^{k-4}_{3v-k}|v,3v-k\rangle \quad (k=0,2,3). \tag{3.9}$$

Hence we can calculate also expressions for the coefficients $a^{(i)}_{v,3v-4}[3v-k\,](k=0,2,3)$:

$$|a^{(i)}_{v,3v-4}[3v]|^2=(2^3 3^7/5)v^4(v-1)^3(3v-1)^3(3v-2)^3(3v+1)^2(6v+1)(6v-1)/(6v-7)$$
$$\times[(12v^2-40v+37)+(-1)^{i-1}(-144v^5+1056v^4-2496v^3$$
$$+2756v^2-1167v-455)(\Gamma)^{-1/2}], \tag{3.10}$$

$$|a^{(i)}_{v,3v-4}[3v-2]|^2=(2^3 3^5/5)v^2(v-1)^2(2v-1)(3v-1)^2(3v-2)^3(6v-1)/(6v-7)$$
$$[(36v^5-156v^4+157v^3-40v^2-237v+730)$$
$$+(-1)^{i-1}(432v^8-3600v^7$$
$$+9768v^6-14\,952v^5+35\,663v^4$$
$$-64\,029v^3+56\,573v^2-35\,905v-8450)(\Gamma)^{-1/2}], \tag{3.11}$$

$$|a^{(i)}_{v,3v-4}\ [3v-3]|^2 = (2^6 3^3/5)(v-1)^2(2v-1)(3v-1)(3v-2)(6v-1)^2(6v-5)/(6v-7)$$
$$\times[(18v^4 - 48v^3 + 8v^2 - 43v - 60)] + (-1)^{i-1}(216v^7 - 1440v^6$$
$$+ 2598v^5 - 2238v^4 + 5632v^3 - 3463v^2 + 4345v$$
$$+ 600)(\Gamma)^{-1/2}], \tag{3.12}$$

whereby Γ has the same meaning as in (3.5). Again, there is a choice of sign to be made for one of these coefficients. It is convenient to require that $a^{(i)}_{v,3v-4}\ [3v]$ be positive, because then the signs of $a^{(i)}_{v,3v-4}\ [3v-2]$ and $a^{(i)}_{v,3v-4}\ [3v-3]$, respectively, follow from the homogeneous equations (3.1) and (3.2).

In order to complete the discussion of the $|v,3v-4\rangle$ eigenstates of P^0_{3v-4}, we have to draw attention upon the fact that the coefficients $a^{(2)}_{v,3v-4}\ [3v-k\ ]\ (k=0,2,3)$ vanish for $v=3$. This shows that there is a unique nondegenerate $|3,5\rangle$ state, the eigenvalue of which follows from formula (3.5) with $i=1$, namely,

$$\alpha_{3,5} = -2^5\cdot 3\cdot 5\sqrt{3}. \tag{3.13}$$

IV. ANALYSIS OF THE $l = 3v-5$ STATES

In case that $v>4$ there exist two orthonormal states with $l=3v-5$, which we denote $|v,3v-5,(1)\rangle$ and $|v,3v-5,(2)\rangle$. Five states with higher angular momentum value come into play for making combinations of the basis states by means of shift operator actions: $|v,3v\rangle$, $|v,3v-2\rangle$, $|v,3v-3\rangle$, $|v,3v-4,(1)\rangle$, and $|v,3v-4,(2)\rangle$. Hence, at first sight one could imagine the normal way to proceed being to set up five independent homogeneous equations in the unknowns $a^{(i)}_{v,3v-5}\ [3v-k\ ](k=0,2,3)$ and $a^{(i)}_{v,3v-5}\ [3v-4,(j)](j=1,2)$, where i takes either the value 1 or 2. Since the coefficients in these equations can contain the unknown $\alpha^{(i)}_{v,3v-5}$ at most linearly, the eliminant is of degree 5 or less with respect to $\alpha^{(i)}_{v,3v-5}$. Unless this equation reduces to a quadratic equation, parasitic solutions are expected, and these should be ruled out by verifying which solutions correspond to a zero eigenvector. The method outlined is rigorous and straightforward, but calculations are likely to become extremely tedious. However, things turn out to be much less involved, since it appears that we may forget completely about any knowledge concerning the eigenvalues and the construction of the two $|v,3v-4\rangle$ states. As a consequence, the analysis of $|v,3v-5\rangle$ states is quite analogous to that of the $|v,3v-4\rangle$ states in Sec. III. The three relations connecting nonscalar product operators which play a role here are (IX.2.17), (IX.2.19), and (IX.2.11). For the first two the action upon $|v,3v\rangle$ is considered, and for the last one the action upon $|v,3v-2\rangle$. In this way we end up with three linear homogeneous equations with respect to the three unknowns $a^{(i)}_{v,2v-5}\ [3v-k\ ](k=0,2,3)$ and $i=1$ or $i=2$) and the consistency relation is only quadratic in the eigenvalue $\alpha^{(i)}_{v,3v-5}$ ($i=1$ or $i=2$)! Performing the necessary calculations, we thus arrive at the result

$$\alpha^{(i)}_{v,3v-5} = -(2\sqrt{3}/3\cdot 7)(6v-5)[(54v^5 - 63v^4 - 1275v^3 + 175v^2 - 109v - 42) + 5.7(-1)^{i-1}\sqrt{\Psi}\,], \tag{4.1}$$

$$\Psi = 81v^6 - 2160v^5 + 24\,138v^4 - 51\,396v^3 + 47\,689v^2 - 20\,228v + 3172 \quad (v>4).$$

On the other hand, the relations between scalar product operators permit us again to calculate the relevant matrix elements which when substituted into equations of the type (1.3) provide us with nonhomogeneous but quadratic equations with respect to the $a^{(i)}_{v,3v-5}$'s. For the sake of completeness, we list the main results:

$$\langle v,3v|P^{+5}_{3v-5}\,P^{-5}_{3v}|v,3v\rangle$$
$$= 2^4 3^8 v^4(v-1)^3(v-2)(2v-1)(3v-1)^3(3v-2)^2(3v-4)^2(6v-1)(6v+1)(6v-5)(18v^2-51v+41), \tag{4.2}$$

$$\langle v,3v-2|P^{+3}_{3v-5}\,P^{-3}_{3v-2}|v,3v-2\rangle$$
$$= (2^6 3^3/5)(v-1)^2(v-2)(2v-1)(3v-1)^2(3v-2)^2(3v-4)^2(6v-1)(6v-5)(18v^5 - 123v^4 + 251v^3 - 136v^2 + 131v + 84), \tag{4.3}$$

$$\langle v,3v-3|P^{+2}_{3v-5}\,P^{-2}_{3v-3}|v,3v-3\rangle$$
$$= (2^3 3^4/5)(v-1)^2(2v-1)^2(3v-1)(3v-4)^2(6v-5)(81v^7 + 81v^6 - 3015v^5 + 4707v^4 + 5158v^3 - 5204v^2 + 7576v + 416), \tag{4.4}$$

$$|a^{(i)}_{v,3v-5}[3v]|^2 = [2^3 3^7 v^4(v-1)^3(v-2)(2v-1)(3v-1)^3(3v-2)^2(3v-4)^2(6v+1)^2(6v-1)(6v-5)/(2v-3)]$$
$$\times[(18v^2 - 51v + 41) + (-1)^i(162v^5 - 3267v^4 + 11\,259v^3 - 15\,699v^2 + 10\,163v - 2330)(\Psi)^{-1/2}] \quad (i=1,2), \tag{4.5}$$

$$|a^{(i)}_{v,3v-5}[3v-2]|^2 = [2^5 3^3(v-1)^2(v-2)(2v-1)^2(3v-1)^2(3v-2)^2(3v-4)^2(6v-1)(6v-5)/5(2v-3)]$$
$$\times[(18v^5 - 123v^4 + 251v^3 - 136v^2 + 131v + 84) + (-1)^{i-1}$$
$$\times(162v^8 - 4725v^7 + 31\,860v^6 - 74\,766v^5 + 62\,999v^4 - 22\,725v^3 + 6681v^2 - 12\,794v + 5208)(\Psi)^{-1/2}] \quad (i=1,2), \tag{4.6}$$

$$|a^{(i)}_{v,3v-5}[3v-3]|^2 = [2^2 3^3(v-1)^2(2v-1)^2(3v-1)(3v-4)^2(6v-5)^2/5(2v-3)]$$
$$\times[(81v^7 + 81v^6 - 3015v^5 + 4707v^4 + 5158v^3 - 5204v^2 + 7576v + 416) + (-1)^{i-1}(729v^{10} - 8991v^9 + 6966v^8 + 54486v^7$$
$$+ 687897v^6 - 3\,407\,511v^5 + 4\,480\,648v^4 - 2\,663\,904v^3 + 589\,440v^2 - 171\,600v + 79\,040)(\Psi)^{-1/2}] \quad (i=1,2), \tag{4.7}$$

where Ψ has been defined in (4.1). To finish the analysis of all $|v,3v-5\rangle$ states we have still to consider the two particular cases $v=3$ and $v=4$. Normally we should distinguish the treatment of both cases on account of the fact that there is only one $|3,5\rangle$ state and two $|4,8\rangle$ states. But all the results (4.2)–(4.7) have been derived without reference to $|v,3v-4\rangle$ states, and hence these formulae hold for $v=4$ as well as for $v=3$. It is then only a matter of direct numerical verification to check that

$$|a^{(2)}_{3,4}\ [9-k\ ]| = |a^{(2)}_{4,7}\ [9-k\ ]| = 0 \quad (k=0,2,3), \tag{4.8}$$

showing that the respective eigenvalues are $\alpha^{(1)}_{v,3v-4}$ $(v=3,4)$, or

$$\alpha_{3,4} = 2^6\cdot 5(13)/\sqrt{3}, \tag{4.9}$$

$$\alpha_{4,7} = -2^3 5(13)(19)\sqrt{3}. \tag{4.10}$$

V. DISCUSSION

By suitable combination of relations between scalar and nonscalar shift operator products, we have succeeded in setting up closed expressions for part of the eigenvalue spectrum of the scalar shift operator P^0_1, and in defining the corresponding eigenstates in terms of shift operator actions. It is not without reasons, however, that in the present paper we stopped our treatment at the case $l=3v-5$. First of all, for $l=3v-5$ we have come already dangerously close to the limits of the calculation possibilities on a pocket calculator. Secondly, for $l=3v-5$ there is general threefold degeneracy. In doing most of the calculations to find the $\alpha^{(i)}_{v,3v-6}$ eigenvalues, we noticed that, as in the case of SU(3),[4,5] we would end up with an equation of third degree for which it would not be possible to write the solutions in simple form. Of course, for the particular cases whereby the l-multiplicity is 1 or 2, serious simplifications arise. As examples we quote

$$\alpha_{3,3} = 2^5\cdot 3\cdot 5^2\sqrt{3}/7,$$

$$\alpha^{(i)}_{4,6} = 2^3\cdot 3\cdot 5\sqrt{3}[-9+(-1)^{i-1}(13\,953)^{1/2}](i=1,2).$$

The present study concerning the symmetric irreducible representations of G_2 should be viewed as an intermediate step for the classification of nuclear octupole–phonon states which are connected to the symmetric representations of R(7). In fact, our next concern is to find also part of the eigenvalue spectrum of the R(3) scalar shift operator O^0_1 built with the R(7) group generators. Since O^0_1 and P^0_1, although they do not commute, are related in many different ways, the present results will undoubtedly show their full significance in future contributions.

ACKNOWLEDGMENTS

We want to express our sincere gratitude to Professor Dr. C. C. Grosjean for his continuous interest.

[1] G. Vanden Berghe and H. E. De Meyer, J. Math. Phys. **21**, 1967 (1980).
[2] G. Vanden Berghe, H. E. De Meyer, and J. Van Der Jeugt, J. Math. Phys. **22**, 2332 (1981).
[3] H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 2635 (1980).
[4] H. E. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360 (1981).
[5] H. E. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2366 (1981).
[6] J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).

Bounds for effective conductivity of heterogeneous materials by the finite element method

Mitsunobu Nakamura and Makoto Mizuno
Department of Electronics, Tamagawa University, Machida, Tokyo 194, Japan

(Received 14 October 1981; accepted for publication 15 January 1982)

Bounds for the effective conductivity σ^* are derived using two standard variational principles established by Beran and the finite element method. We divide a two-dimensional material into triangular finite elements, use a linear local function as a trial function for each element, and apply the function to the variational principles. As a result a very simple method of computing the bounds is derived. As a numerical example, we apply the method to a model for heterogeneous materials and obtain very narrow gaps between bounds for σ^*.

PACS numbers: 72.10.Bg, 02.60. + y, 02.70. + d

I. INTRODUCTION

The variational method has been used to obtain the expressions for an upper and a lower bound of the effective physical constants or bulk modulus of heterogeneous materials. After Wiener's research for the bounds[1] Hashin and Shtrikman[2] derived the most restrictive bounds that can be given in terms of the phase properties and volume fractions with the aid of some variational principles. To improve Hashin and Shtrikman's (HS) bounds, one must take into consideration some phase geometries. Beran[3,4] developed the bounds on the effective bulk modulus of two-phase solids by two three-point correlation functions for material properties. To evaluate the Beran bounds, Miller[5,6] used the cell model for the phase geometries, and Corson determined the three-correlation functions by experimental methods.[7] Hori[8] expressed the bounds in terms of many-point correlation functions of the spatial variation of the material property, but it is difficult to concretely calculate all the correlation functions.

The purpose of this paper is to derive an upper and a lower bound on the effective physical constants of some random heterogeneous materials by the finite element method (FEM), which is a sort of variational principle. In the ordinary variational method a global trial function with a few variational parameters is used, and it is hard to select a trial function with a concrete form for problems with a complicate structure such as heterogeneous materials, while, in the FEM, local trial functions with a simple form and many variational parameters are used and it is easy to apply the FEM to the problems with a complicated structure. In a word, the FEM is a variational method for computers and the variational method is applied in a very simple and clear form.

In Sec. II we apply the FEM to the standard variational principle established by Beran[3] and derive expressions for the bounds on the effective physical constants of heterogeneous materials. In Sec. III we apply the method in Sec. II to a model for random media and compute the bounds for the effective conductivity. We compare the result with the HS bounds and the effective conductivity derived by the effective medium theory.[9–14]

In this paper we use only the term of the electric conduction but the obtained result holds for the thermal conductivity, permittivity, permeability, and diffusion constant.

II. UPPER AND LOWER BOUNDS FOR EFFECTIVE CONDUCTIVITY OF A HETEROGENEOUS MATERIAL

The conductivity σ of a heterogeneous material is a function of position and the effective conductivity or the observed conductivity σ^* is given by

$$[\mathbf{J}] = \sigma^*[\mathbf{E}], \tag{1}$$

where the brackets [ ] denote the spatial average

$$[\ ] = \frac{1}{V}\int\int\int dv \quad \text{for 3D},$$

$$= \frac{1}{S}\int\int ds \quad \text{for 2D}. \tag{2}$$

And also σ^* is given by

$$\sigma^* = \frac{[\sigma \mathbf{E}^2]}{[\mathbf{E}]^2} = \frac{[\mathbf{J}]^2}{[\mathbf{J}^2/\sigma]}, \tag{3}$$

with the equations

$$\mathbf{E} = -\operatorname{grad} U \tag{4}$$

and

$$\mathbf{J} = \operatorname{rot}\mathbf{A}, \tag{5}$$

where U and $\mathbf{A}$ are a scalar and a vector potential, respectively. If we treat only a 2D problem for simplicity, $\mathbf{A}$ is

$$\mathbf{A} = (0 \ 0 \ A). \tag{6}$$

If U^t and $\mathbf{A}^t$ are trial functions for U and $\mathbf{A}$, bounds for σ^* are derived by the variational principle from (3)–(5)[3]:

$$\frac{[\sigma(\operatorname{grad} U^t)^2]}{[\mathbf{E}]^2} \geqslant \sigma^* \geqslant \frac{[\mathbf{J}]^2}{[(\operatorname{rot}\mathbf{A}^t)^2/\sigma]}. \tag{7}$$

We divide a two-dimensional material into triangular finite elements and assign nodal numbers i,j,k in the counterclockwise order to the three vertexes of the nth element. If it is supposed that U^t and A^t are linear in the element, they are determined by the three nodal values U_i, U_j, U_k and A_i, A_j, A_k. Then we obtain

$$U^t = N_i U_i + N_j U_j + N_k U_k, \tag{8}$$

$$A^t = N_i A_i + N_j A_j + N_k A_k, \tag{9}$$

where

$$N_m = (1/2s_n)(a_m + b_m x + c_m y) \quad (m = i,j,k); \tag{10}$$

 0022-2488/82/061228-03$02.50

s_n is the area of the nth element and

$$a_i = x_j y_k - x_k y_j, \quad b_i = y_j - y_k, \quad c_i = x_k - x_j, \tag{11}$$

$a_j, b_j, c_j, a_k, b_k, c_k$ being obtained by the rotation of i, j, k in (11).

Let

$$P_1 = [\sigma(\text{grad}\, U^t)^2] \tag{12}$$

and

$$P_2 = [(\text{rot}\, \mathbf{A}^t)^2/\sigma] \tag{13}$$

and substitute (8) and (9) into (12) and (13), respectively, assuming that σ is uniform within an element. And then we have

$$P_1 = \frac{1}{4} \sum_n (\sigma_n/s_n)[(b_i U_i + b_j U_j + b_k U_k)^2 + (c_i U_i + c_j U_j + c_k U_k)^2] \tag{14}$$

and

$$P_2 = \frac{1}{4} \sum_n (1/s_n\sigma_n)[(b_i A_i + b_j A_j + b_k A_k)^2 + (c_i A_i + c_j A_j + c_k A_k)^2], \tag{15}$$

where σ_n is the conductivity of the nth element.

For a node p on a section of the boundary where a Dirichlet condition is specified,

$$U_p = g_p, \quad A_p = h_p, \tag{16}$$

where g_p and h_p are constants. But the Neumann condition cannot be satisfied. Substituting (16) into (14) and (15) and minimizing P_1 and P_2 with respect to all unknown U_i and A_i, we obtain two simultaneous equations:

$$[G]\mathbf{u} = \mathbf{i} \tag{17a}$$

and

$$[G']\mathbf{a} = \mathbf{i}' \tag{17b}$$

where $[G]$ and $[G']$ are positive definite symmetrical band matrices, $\mathbf{u}$ and $\mathbf{a}$ are columns of unknown U_i and A_i, and $\mathbf{i}$ and $\mathbf{i}'$ are known vectors. By the solutions of (17a) and (17b), (14), (15), and (7), bounds for σ^* can be calculated.

III. NUMERICAL EXAMPLE

For simplicity we use a square material with unit length and give boundary conditions

$$U = 0 \quad \text{and} \quad A = 0 \quad \text{on } x = 0, \qquad U = 1 \quad \text{and} \quad A = 1 \quad \text{on } x = 1, \tag{18}$$

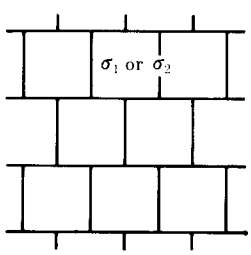


FIG. 1. Structure of a model for heterogeneous materials.

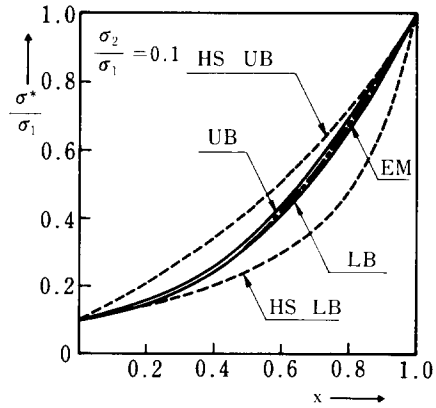


FIG. 2. Bounds of σ^*/σ_1 against the volume fraction x for $\sigma_2/\sigma_1 = 0.1$.

and

$$\frac{\partial U}{\partial y} = \frac{\partial A}{\partial y} = 0 \quad \text{on } y = 0 \text{ and } y = 1. \tag{19}$$

The Dirichlet conditions (18) are taken by substituting nodal values on the boundary into (14) and (15), while the Neumann conditions (19) cannot be satisfied. From (18) we obtain

$$[\mathbf{E}]^2 = 1 \quad \text{and} \quad [\mathbf{J}]^2 = 1. \tag{20}$$

As a model for random heterogeneous materials we use a material with a structure as shown in Fig. 1 and assign σ_1 and σ_2 $(\sigma_1 > \sigma_2)$ to the conductivity of each cell at random. Each cell is divided into eight equal triangular finite elements. We compute bounds for σ^*/σ_1 of the model with 900 cells by the method in Sec. II and plot them against the volume fraction x, taking σ_2/σ_1 as a parameter, in Figs. 2, 3, and 4. Also in these figures we show the two-dimensional HS bounds[2,15] and the results of the effective medium theory (EM),[9-14] and designate the upper and lower bounds by UB and LB, respectively. We give the two-dimensional HS bounds and the EM in (21) and (22):

$$\sigma_1 + \frac{1-x}{1/(\sigma_2 - \sigma_1) + x/2\sigma_1} \geqslant \sigma^* \geqslant \sigma_2 + \frac{x}{1/(\sigma_1 - \sigma_2) + (1-x)/2\sigma_2}, \tag{21}$$

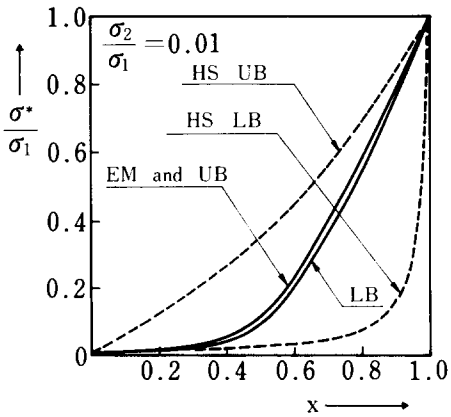


FIG. 3. Bounds of σ^*/σ_1 against the volume fraction x for $\sigma_2/\sigma_1 = 0.01$.

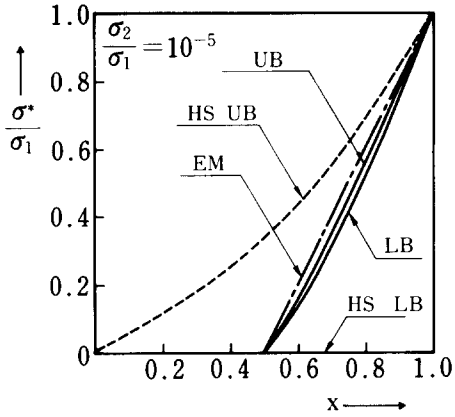


FIG. 4. Bounds of σ^*/σ_1 against the volume fraction x for $\sigma_2/\sigma_1 = 10^{-5}$.

$$x\frac{\sigma_1-\sigma^*}{\sigma_1+\sigma^*}+(1-x)\frac{\sigma_2-\sigma^*}{\sigma_2+\sigma^*}=0\,. \tag{22}$$

As shown in Figs. 2, 3, and 4, we can obtain very narrow gaps between the bounds for σ^* of a heterogeneous material with a definite structure. The upper bounds are very close to the results of the EM. If a concrete geometrical configuration of materials is determined, the bounds for σ^* are computed by a very simple method in this paper. In the conventional variational method it is hard to take into account definite geometrical configurations. Even if one of them is given, it is difficult to select a trial function or to concretely evaluate many-point correlation functions. We can properly take spatial fluctuations of physical properties into the FEM because local trial functions are used in the FEM.

With the increase of the capacity of computers it is no more laborious to construct (17a) and (17b) and to solve them than to execute algebraic calculations in the conventional variational method.

ACKNOWLEDGMENTS

The authors wish to express their great appreciation to Professor F. Yonezawa for useful suggestions and to Professor N. Asano for useful discussions and encouragement. Numerical calculations were performed on the Hitac 8800/8700 computer of the Computer Centre, University of Tokyo.

[1]O. Wiener, Abh. Math. -Phys. Kl. Kgl. Sächs. Ges. **32**, 509 (1912).
[2]Z. Hashin and S. Shtrikman, J. Appl. Phys. **33**, 3152 (1962).
[3]M. Beran, Nuovo Cimento **38**, 771 (1965).
[4]M. Beran and J. Molyneux, Quart. Appl. Math. **24**, 107 (1966).
[5]M. N. Miller, J. Math. Phys. **10**, 1988 (1969).
[6]M. N. Miller, J. Math. Phys. **10**, 2005 (1969).
[7]P. B. Corson, J. Appl. Phys. **45**, 3180 (1974).
[8]M. Hori, J. Math. Phys. **14**, 1942 (1973).
[9]D. A. G. Bruggeman, Ann. Phys. **25**, 645 (1936).
[10]C. F. Böttcher, Rec. Trav. Chim. Pays-Bas **64**, 47 (1945).
[11]R. Landauer, J. Appl. Phys. **27**, 824 (1973).
[12]M. Hori and F. Yonezawa, J. Math. Phys. **16**, 352 (1975).
[13]M. Hori and F. Yonezawa, J. Math. Phys. **16**, 365 (1975).
[14]M. Hori and F. Yonezawa, J. Phys. C **10**, 229 (1977).
[15]M. Beran and N. R. Silnutzer, J. Compos. Mater. **5**, 246 (1971).

ERRATA

Erratum: On the classification of Clifford algebras and their relation to spinors in n dimensions [J. Math. Phys. 23, 1 (1982)]

Nikos Salingaros
Department of Mathematics, The University of Iowa, Iowa City, Iowa 52242

(Received 2 March 1982; accepted for publication 12 March 1982)

PACS numbers: 02.20.Sv, 02.40.Ky, 99.10. + g

VI. MATRIX REPRESENTATIONS AND LIE ALGEBRAS:

The construction in this section was done explicitly for the Clifford algebras in Riemannian spaces of dimensions one to five. However, the properties of Clifford algebras in higher dimensions, as given by Eqs. (19)–(25), should follow instead from the expressions given here.

Theorem 8: The matrix representation space of each Clifford algebra is

$$N_{2k-1} \approx \mathbb{R}(2^k),$$
$$N_{2k} \approx H(2^{k-1}),$$
$$S_k \approx \mathbb{C}(2^k),$$
$$\Omega_{2k-1} \approx \Omega(2^k) \approx \mathbb{R}(2^k) \oplus \mathbb{R}(2^k),$$
$$\Omega_{2k} \approx (\Omega \otimes H)(2^{k-1}) \approx H(2^{k-1}) \oplus H(2^{k-1}).$$

Theorem 9: The Lie algebra corresponding to every Clifford algebra of Table I is given as

$$N_{2k-1} \sim \mathrm{SL}(2^k;\mathbb{R}),$$
$$N_{2k} \sim \mathrm{SL}(2^{k-1};H),$$
$$S_k \sim \mathrm{SL}(2^k;\mathbb{C}),$$
$$\Omega_{2k-1} \sim \mathrm{SL}(2^k;\Omega) \approx \mathrm{SL}(2^k;\mathbb{R}) \oplus \mathrm{SL}(2^k;\mathbb{R}),$$
$$\Omega_{2k} \sim \mathrm{SL}(2^{k-1};\Omega \otimes H) \approx \mathrm{SL}(2^{k-1};H) \oplus \mathrm{SL}(2^{k-1};H).$$

Therefore, the recursion relations for Lie algebras [Eq (25)] should read in general as

$$\mathrm{SL}(2^k;\mathbb{R}) \otimes \mathrm{SL}(1;H) \approx \mathrm{SL}(2^{k-1};H) \otimes \mathrm{SL}(2;\mathbb{R}) \approx \mathrm{SL}(2^k;H),$$
$$\mathrm{SL}(2^k;\mathbb{R}) \otimes \mathrm{SL}(2;\mathbb{R}) \approx \mathrm{SL}(2^{k-1};H) \otimes \mathrm{SL}(1;H) \approx \mathrm{SL}(2^{k+1};\mathbb{R}).$$

Erratum: Adjoints of nondensely defined Hilbert space operators [J. Math. Phys. 22, 1619 (1981)]

Peter Pfeifer
Laboratory of Physical Chemistry, ETH Zürich, CH-8092 Zürich, Switzerland

(Received 27 January 1981; accepted for publication 20 March 1981)

PACS numbers: 02.30.Tb, 99.10. + g

(1) The last expression in line 3 of the proof of Theorem 2 should read $\overline{P_{T^*} T P_T}$.

(2) The first three lines of Theorem 4 should read

Theorem 4: Let $T \in \mathscr{L}(\mathscr{H})$. Then

$$\mathscr{N}(T^*) = \mathscr{R}(T)^\perp, \quad \overline{\mathscr{R}(T^*)} = \mathscr{N}(\overline{P_{T^*} T})^\perp \cap \overline{\mathscr{D}(T)}. \quad (6)$$

$$\overline{\mathscr{D}(T^*)} = \left\{ \lim_{n\to\infty} T\psi_n \,|\, (\psi_n)_{n=1}^\infty \text{ is a sequence in } \mathscr{D}(T) \cdots \right.$$

(3) Line 6 of Theorem 6 should read

(d) T^* is bounded. (d') $\mathscr{R}(T^*) = \mathscr{N}(T)^\perp \cap \overline{\mathscr{D}(T)}$.

(4) The expression in the last line of Sec. 3 should be at the end of line 6 of Remark 7.

(5) Line 2 of Theorem 7 should start: $\mathscr{D}(T^*) \cap \overline{\mathscr{R}(T)}$, denoted by···.

(6) Eq. (12) in Theorem 7 should start: $\mathscr{R}((T^{-1})^*) = \cdots$.

(7) Line 2 of the proof of Theorem 7 should start: Since $\langle (T^{-1})^* \varphi \,|\, T\psi \rangle = \cdots$.