Explicit evaluation of the representation functions of ISO(n)

M. K. F. Wong
Fairfield University, Fairfield, Connecticut 06430

Hsin-Yang Yeh
Moorhead State University, Moorhead, Minnesota 56560

(Received 30 July 1979; accepted for publication 3 October 1979)

All representation functions of ISO(n) have been found in explicitly closed form. They can all be expressed in terms of Bessel functions and Clebsch–Gordan coefficients of SO(n) involving the most degenerate representation $[k,\dot{0}]$.

I. INTRODUCTION

The inhomogeneous orthogonal group ISO(n) is of great importance in physics. ISO(3) is the Euclidean group in 3-space, and is the basic group for the time independent Schrödinger equation, which is the starting point for nonrelativistic scattering problems. ISO(4) is perhaps connected with ISO(3,1), which is the Poincaré group. It will therefore be useful to obtain the representation functions of ISO(n) in general, both for its intrinsic significance and its application to physics.

The irreducible representations of ISO(n) have been obtained by Chakrabarti,[1] who showed that an irreducible representation in ISO($2k$) and ISO($2k+1$) is characterized by k numbers: γ, which is continuous, and $[m_{n+1,2},m_{n+1,3},\ldots,m_{n+1,k}]$, which are discrete. The branching rules for $[m_{n+1,2},\ldots,m_{n+1,k}]$ show that they are equivalent to the labels characterizing an irreducible representation of SO($n-1$). The relationship between ISO(n) and SO(n,1) has been given by Wong and Yeh,[2] who have also obtained the eigenvalues for the invariant operators of ISO(n),[3] and the shift operators with their normalization constants.[4]

It has been known for a long time[5,6] that the representation functions of ISO(2) and ISO(3) are connected with ordinary Bessel functions $J_k(z)$ and spherical Bessel functions $j_k(z)$ respectively. It is our purpose to show that this relationship can be generalized to ISO(n). If one defines a Bessel function of "nth order," denoted by $J_k^{[n]}(z)$, as

$$J_k^{[n]}(z)=z^{1-n/2}J_{k-1+n/2}(z), \tag{1.1}$$

then one can make the following statement valid for all n: The representation functions of ISO(n), $n>2$, are expressible as a summation over k of $J_k^{[n]}(z)$, $k=0,1,\ldots,\infty$.

We are able to obtain this result through the theory of induced representations. For an excellent exposition, see, for example, the recent book by Barut and Rączka.[7] This approach was used by Wolf[8] to obtain the representation functions of ISO(n), as well as SO($n+1$) and SO(n,1), in terms of the d functions of SO(n). However, Wolf did not explicitly evaluate the integral he has obtained through the induced representation. We wish to show that the integral can be evaluated in a simple way, with the result that the representation functions of ISO(n) can be expressed as a sum over k of $J_k^{[n]}(\gamma\xi)$, multiplied by two CG coefficients of SO(n), involving the most degenerate representation $[k,\dot{0}]$ of SO(n).

In Sec. II, we rederive the results of ISO(2) and ISO(3) in terms of our method, and discuss briefly Wigner's contraction process. In Sec. III we obtain the representation function of ISO(4) explicitly. In Sec. IV we obtain the general representation functions for all ISO(n). All these expressions are in explicitly simple and closed form.

II. REPRESENTATION FUNCTIONS OF ISO(2) and ISO(3)

We derive all our results from the basic integral obtained by Wolf,[8]

$$\begin{aligned}{}^I d^{\gamma,L}_{JL'J'}(\xi)=&\frac{(\dim_n J\dim_n J')^{1/2}}{\dim_{n-1}L\dim_{n-1}L'}\frac{(\mathrm{Vol}H)^2}{\mathrm{Vol}G\,\mathrm{Vol}K}\\&\times\sum_M\dim_{n-2}M\int_0^\pi\sin^{n-2}\theta\,d\theta\,\overline{d^J_{LML'}(\theta)}\,\exp(i\gamma\xi\cos\theta)\,d^{J'}_{LML'}(\theta),\end{aligned} \tag{2.1}$$

where

$$\frac{(\mathrm{Vol}H)^2}{\mathrm{Vol}G\,\mathrm{Vol}K}=\frac{\Gamma(n/2)}{\pi^{1/2}\Gamma(\frac{1}{2}(n-1))}. \tag{2.2}$$

Equation (2.1) is valid for $n>2$. For $n=2$, the limit of integration goes from 0 to 2π, and the integral should be divided by 2. So for ISO(2), we obtain

$${}^I d^{\gamma}_{m_{12}m'_{12}}(\xi)=(2\pi)^{-1}\int_0^{2\pi}d\theta e^{-im_{12}\theta}e^{im'_{12}\theta}e^{i\gamma\xi\cos\theta}. \tag{2.3}$$

Now we use the well known expression[5] (Ref. 5, p. 73)

 0022-2488/80/010001-05$01.00

$$e^{i\gamma\xi\cos\theta} = \sum_{l=-\infty}^{\infty} i^l J_l(\gamma\xi)\, e^{il\theta}. \tag{2.4}$$

Substituting (2.4) into (2.3) and performing the integration, we obtain

$$^I d^{\gamma}_{m_{12}m'_{12}}(\xi) = i^{m_{12}-m'_{12}} J_{m_{12}-m'_{12}}(\gamma\xi). \tag{2.5}$$

This result was obtained by Wigner[6] through the contraction of the d function of SO(3), i.e.,

$$\lim_{l\to\infty} d^{\,l}_{mn}(\gamma\xi/l) = J_{m-n}(\gamma\xi). \tag{2.6}$$

This result can be derived from the relationship between the Bessel function and the limit of Jacobi polynomials $P_n^{(\alpha,\beta)}(\cos z/n)$ as $n\to\infty$,[9] i.e. (Ref. 9, Eq. 41, p. 173),

$$\lim_{n\to\infty} n^{-\alpha} P_n^{(\alpha,\beta)}(\cos z/n) = \lim_{n\to\infty} n^{-\alpha} P_n^{(\alpha,\beta)}(1 - z^2/2n^2) = (z/2)^{-\alpha} J_\alpha(z). \tag{2.7}$$

For ISO(3), we have

$$^I d^{\gamma,m_{24}}_{\substack{m_{13}m'_{13}\\ m_{12}m_{12}}}(\xi) = [(2m_{13}+1)(2m'_{13}+1)]^{1/2}\Gamma(3/2)\pi^{-1/2}\int_0^{\pi} \sin\theta\, d\theta\, d^{\,m_{13}}_{m_{24}m_{12}}(\theta)\, d^{\,m'_{13}}_{m_{24}m_{12}}(\theta) e^{i\gamma\xi\cos\theta} \tag{2.8}$$

This time we use the expansion (Ref. 10, p. 128)

$$e^{i\gamma\xi\cos\theta} = \sum_{L=0}^{\infty} i^L (2L+1)\, j_L(\gamma\xi) P_L(\cos\theta). \tag{2.9}$$

In terms of the d functions of SO(3), we have,

$$P_L(\cos\theta) = d^{\,L}_{00}(\theta). \tag{2.10}$$

Next we write

$$d^{\,m'_{13}}_{m_{24}m_{12}}(\theta)\, d^{\,L}_{00}(\theta) = \sum_l C\begin{pmatrix} m'_{13} & L & l \\ m_{24} & 0 & m_{24}\end{pmatrix} C\begin{pmatrix} m'_{13} & L & l \\ m_{12} & 0 & m_{12}\end{pmatrix} d^{\,l}_{m_{24}m_{12}}(\theta) \tag{2.11}$$

From the orthogonality property of the d functions of SO(3),[11] we obtain

$$\int_0^{\pi} \sin\theta\, d\theta\, \overline{d^{\,m_{13}}_{m_{24}m_{12}}(\theta)}\, d^{\,l}_{m_{24}m_{12}}(\theta) = 2(2l+1)^{-1}\delta_{l,m_{13}}. \tag{2.12}$$

Substituting (2.9), (2.10), (2.11), and (2.12) into (2.8), we obtain

$$^I d^{\gamma m_{24}}_{\substack{m_{13}m'_{13}\\ m_{12}m_{12}}}(\xi) = [(2m_{13}+1)(2m'_{13}+1)]^{1/2} \sum_{L=0}^{\infty} i^L j_L(\gamma\xi)\, C\begin{pmatrix} m'_{13} & L & m_{13} \\ m_{24} & 0 & m_{24}\end{pmatrix} C\begin{pmatrix} m'_{13} & L & m_{13} \\ m_{12} & 0 & m_{12}\end{pmatrix}. \tag{2.13}$$

This result is in agreement with Miller[5] and Wigner,[6] with the additional remark that in both Miller's equation (6.41) and Wigner's equation (12.24), $m = n$.

Again Wigner has shown that this result can be obtained from contracting the d function of SO(4). This is done as follows: Write the d function of SO(4) as

$$\begin{aligned}
d^{\,m_{14}m_{24}}_{\substack{m_{13}m'_{13}\\ m_{12}m_{12}}}(\theta) &= \sum_m [(2m_{13}+1)(2m'_{13}+1)]^{1/2} \begin{pmatrix} \tfrac12(m_{14}+m_{24}) & \tfrac12(m_{14}-m_{24}) & m_{13} \\ m & m_{12}-m & -m_{12}\end{pmatrix} \\
&\quad \times \begin{pmatrix} \tfrac12(m_{14}+m_{24}) & \tfrac12(m_{14}-m_{24}) & m'_{13} \\ m & m_{12}-m & -m_{12}\end{pmatrix} e^{-2im\theta} \\
&= \times \sum_m \sum_K (-1)^{(m_{14}-m_{24})/2+K+m_{12}-m}(2K+1) \begin{Bmatrix} m_{13} & \tfrac12(m_{14}+m_{24}) & \tfrac12(m_{14}-m_{24}) \\ \tfrac12(m_{14}+m_{24}) & m'_{13} & K\end{Bmatrix} \\
&\quad \times \begin{pmatrix} \tfrac12(m_{14}+m_{24}) & \tfrac12(m_{14}+m_{24}) & K \\ -m & m & 0\end{pmatrix} \begin{pmatrix} m_{13} & m'_{13} & K \\ m_{12} & -m_{12} & 0\end{pmatrix} [(2m_{13}+1)(2m'_{13}+1)]^{1/2} \\
&= \sum_K (-1)^{K+m_{12}+m_{14}}(2K+1)[(2m_{13}+1)(2m'_{13}+1)]^{1/2} \begin{Bmatrix} m_{13} & \tfrac12(m_{14}+m_{24}) & \tfrac12(m_{14}-m_{24}) \\ \tfrac12(m_{14}+m_{24}) & m'_{13} & K\end{Bmatrix} \\
&\quad \times \begin{pmatrix} m_{13} & m'_{13} & K \\ m_{12} & -m_{12} & 0\end{pmatrix} H_{(m_{14}+m_{24})/2,K}(\theta),
\end{aligned} \tag{2.14}$$

where

$$H_{(m_{14}+m_{24})/2,K}(\theta)=\sum_m(-1)^{(m_{14}+m_{24})/2-m}e^{-2im\theta}\begin{pmatrix}\frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m_{14}+m_{24}) & K\\ -m & m & 0\end{pmatrix}$$

$$=K!\left(\frac{(m_{14}+m_{24}-K)!}{(m_{14}+m_{24}+K+1)!}\right)^{1/2}(2i\sin\theta)^K C^{K+1}_{m_{14}+m_{24}-K}(\cos\theta). \tag{2.15}$$

The contraction process goes as follows: Write θ as $\gamma\xi/m_{14}$ and take the limit as $m_{14}\to\infty$. Then we have

$$\lim_{m_{14}\to\infty}\begin{Bmatrix} m_{13} & m'_{13} & K\\ \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m_{14}-m_{24})\end{Bmatrix}\to\frac{(-1)^{2m'_{13}+m_{14}+m_{24}}}{(m_{14}+m_{24})^{1/2}}\begin{pmatrix} m_{13} & m'_{13} & K\\ -m_{24} & m_{24} & 0\end{pmatrix} \tag{2.16}$$

and

$$\lim_{m_{14}\to\infty}(m_{14}+m_{24})^{-1/2}H_{(m_{14}+m_{24})/2,K}(\gamma\xi/m_{14})=(-i)^K j_K(\gamma\xi), \tag{2.17}$$

where (2.17) is a direct consequence of (2.7) and the relationship between Jacobi polynomials and Gegenbauer polynomials,[9] i.e., (Ref. 9, p. 174, Eq. 4),

$$(\lambda+\tfrac{1}{2})_n C^\lambda_n(x)=(2\lambda)_n P^{(\alpha\alpha)}_n(x),\quad \alpha=\lambda-\tfrac{1}{2}. \tag{2.18}$$

Thus Wigner's result can be summarized as follows:

$$\lim_{m_{14}\to\infty} d^{\,m_{14}m_{24}}_{\substack{m_{13}m'_{13}\\ m_{12}m_{12}}}(\gamma\xi/m_{14})={}^I d^{\,\gamma m_{24}}_{\substack{m_{13}m'_{13}\\ m_{12}m_{12}}}(\xi), \tag{2.19}$$

where the left-hand side is the d function of SO(4), and the right-hand side is the d function of ISO(3).

III. REPRESENTATION FUNCTION OF ISO(4)

For ISO(4), we have

$${}^I d^{\,\gamma m_{25}}_{\substack{m_{14}m_{24}m'_{14}m'_{24}\\ m_{13}m_{13}}}(\xi)=\frac{[(m_{14}+m_{24}+1)(m_{14}-m_{24}+1)(m'_{14}+m'_{24}+1)(m'_{14}-m'_{24}+1)]^{1/2}}{(2m_{25}+1)(2m_{13}+1)}$$

$$\times\frac{1}{\pi^{1/2}\Gamma(3/2)}\sum_M\int_0^\pi \sin^2\theta\, d\theta\, \overline{d^{\,m_{14}m_{24}}_{\substack{m_{25}m_{13}\\ MM}}(\theta)}\, d^{\,m'_{14}m'_{24}}_{\substack{m_{25}m_{13}\\ MM}}(\theta)e^{i\gamma\xi\cos\theta}. \tag{3.1}$$

Now write

$$e^{i\gamma\xi\cos\theta}=\sum_n k_n(\gamma\xi)(\gamma\xi)^{-1}C^1_n(\cos\theta)(n+1)^{-1}, \tag{3.2}$$

where $C^1_n(\cos\theta)$ is the Gengenbauer polynomial.

To obtain $k_n(\gamma\xi)$, we multiply both sides of Eq. (3.2) by $\sin^2\theta C^1_n(\cos\theta)$ and integrate over $d\theta$ from 0 to π. Then due to the orthogonality property of the Gegenbauer polynomials,[12] the right side becomes (Ref. 12, p. 462, Eq. 5)

$$\sum_n\frac{k_n(\gamma\xi)(\gamma\xi)^{-1}}{(n+1)}\int_0^\pi C^m_n(\cos\theta)C^{m'}_{n'}(\cos\theta)\sin^{2m}\theta\, d\theta$$

$$=\sum_n\delta_{mm'}\delta_{nn'}\frac{\pi\Gamma(n+2m)}{2^{2m-1}(m+n)n![\Gamma(m)]^2}k_n(\gamma\xi)(\gamma\xi)^{-1}(n+1)^{-1}. \tag{3.3}$$

In our case here, $m=1$ in Eq. (3.3).

The left-hand side becomes a Bessel function through the following formula (Ref. 9, p. 178, Eq. 38):

$$\Gamma(\lambda+\tfrac{1}{2})(2\lambda)_n i^n(\gamma\xi)^{-\lambda}J_{\lambda+n}(\gamma\xi)(n!)^{-1}2^\lambda\pi^{1/2}=\int_0^\pi e^{i\gamma\xi\cos\theta}C^\lambda_n(\cos\theta)\sin^{2\lambda}\theta\, d\theta. \tag{3.4}$$

In our case here, $\lambda=1$. Equating (3.3) and (3.4), we obtain

$$k_n(\gamma\xi)=2(n+1)^2\pi^{-1}J_{n+1}(\gamma\xi). \tag{3.5}$$

Now from the representation theory of SO(4), (See, e.g., Freedman and Wang[13]), we have

$$d^{\,n0}_{\substack{00\\ 00}}(\theta)=(n+1)^{-1}C^1_n(\cos\theta). \tag{3.6}$$

Next we write

$$d^{\,m'_{14}m'_{24}}_{\substack{m_{25}m_{13}\\ MM}}(\theta)\, d^{\,n0}_{\substack{00\\ 00}}(\theta)=\sum_{P,Q}d^{\,PQ}_{\substack{m_{25}m_{13}\\ MM}}(\theta)\, C\begin{pmatrix} m'_{14}\,m'_{24} & n\,0 & PQ\\ m_{25} & 0 & m_{25}\\ M & 0 & M\end{pmatrix}C\begin{pmatrix} m'_{14}\,m'_{24} & n\,0 & PQ\\ m_{13} & 0 & m_{13}\\ M & 0 & M\end{pmatrix}, \tag{3.7}$$

where the C's are CG coefficients of SO(4), whose value has been obtained by Biedenharn[14] in terms of 9-j symbols of SO(3). In our present case, because of the presence of a zero term, the 9-j symbols are actually reduced to 6-j symbols.

Next we use the orthogonality relation for the d functions of SO(4),[13]

$$\sum_M \int_0^\pi d\theta \sin^2\theta \, \overline{d^{\,m_{14}\,m_{24}}_{\substack{m_{25}\,m_{13}\\ M\,M}}(\theta)}\, d^{\,PQ}_{\substack{m_{25}\,m_{13}\\ M\,M}}(\theta) = \delta_{m_{14}P}\,\delta_{m_{24}Q}\,\frac{\pi(2m_{25}+1)(2m_{13}+1)}{2(m_{14}+m_{24}+1)(m_{14}-m_{24}+1)}. \tag{3.8}$$

Substituting Eqs. (3.2), (3.5), (3.6), (3.7), and (3.8) into Eq. (3.1) and collecting the results above, we obtain the final result for ISO(4):

$$\begin{aligned}
&{}^I d^{\,\gamma\, m_{25}}_{\substack{m_{14}m_{24}\, m'_{14}m'_{24}\\ m_{13}\, m_{13}}}(\xi) \\
&= (-1)^{m'_{14}+m'_{24}+n+m_{14}-m_{24}+m_{25}+m_{13}}\,[(m_{14}+m_{24}+1)(m_{14}-m_{24}+1)(m'_{14}+m'_{24}+1)(m'_{14}-m'_{24}+1)]^{1/2} \\
&\quad \times \sum_{n=0}^{\infty} 2\pi(n+1)(\gamma\xi)^{-1} J_{n+1}(\gamma\xi)
\begin{Bmatrix} \frac{1}{2}(m'_{14}-m'_{24}) & \frac{1}{2}(m_{14}-m_{24}) & n/2 \\ \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m'_{14}+m'_{24}) & m_{52} \end{Bmatrix}
\begin{Bmatrix} \frac{1}{2}(m'_{14}-m'_{24}) & \frac{1}{2}(m_{14}-m_{24}) & n/2 \\ \frac{1}{2}(m_{14}+m_{24}) & \frac{1}{2}(m'_{14}+m'_{24}) & m_{13} \end{Bmatrix},
\end{aligned} \tag{3.9}$$

where $\{\cdots\}$ is a 6-j symbol as defined in Edmonds.[11]

IV. REPRESENTATION FUNCTIONS OF ISO(n)

The procedures used in the previous two sections are now generalized to obtain the representation functions of all ISO(n). First we use the following expansion:

$$e^{i\gamma\xi\cos\theta} = \sum_k K_k(\gamma\xi) C_k^{1-n/2}(\cos\theta)(\gamma\xi)^{1-n/2}. \tag{4.1}$$

To obtain $K_k(\gamma\xi)$, we multiply both sides of (4.1) by $C_k^{1-n/2}(\cos\theta)\sin^{n-2}\theta$ and integrate over $d\theta$ from 0 to π. Using Eq. (3.3), we obtain the left-hand side as

$$(k!)^{-1}2^{1-n/2}\pi^{1/2}\Gamma((n-1)/2)(n-2)_k i^k(\gamma\xi)^{1-n/2}J_{k-1+n/2}(\gamma\xi). \tag{4.2}$$

Using the orthogonality relations for the Gegenbauer polynomials on the right, we obtain, for the right-hand side,

$$K_k(\gamma\xi)(\gamma\xi)^{1-n/2}\frac{\pi^{1/2}\Gamma(k+n-2)}{2^{n-3}(k-1+n/2)k![\Gamma(-1+n/2)]^2}. \tag{4.3}$$

Therefore we find:

$$K_k(\gamma\xi) = \{2^{-4+3n/2}\Gamma((n-1)/2)i^k(k-1+n/2)[\Gamma(-1+n/2)]^2\}\,[(n-1)!\Gamma(\tfrac{1}{2})]^{-1}J_{k-1+n/2}(\gamma\xi). \tag{4.4}$$

However, it has been found by Vilenkin[12] that the representation functions of SO(n) for the most degenerate representation $[k,\dot{0}]$ are expressible as Gegenbauer polynomials, i.e.,

$$d^{\,[k,\dot{0}]}_{[\dot{0}][\dot{0}]}(\theta) = C_k^{-1+n/2}(\cos\theta)/C_k^{-1+n/2}(1), \tag{4.5}$$

where

$$C_k^{-1+n/2}(1) = \frac{\Gamma(n+k-2)}{k!\Gamma(n-2)}. \tag{4.6}$$

We can then combine the two d functions through CG coefficients of SO(n):

$$\begin{aligned}
&d^{\,[m'_n]}_{\substack{[m_{n+1,2\cdots}]\,[m_{n-1}]\\ [m_{n-2}]\,[m_{n-2}]}}(\theta)\, d^{\,[k,\dot{0}]}_{\substack{[\dot{0}]\,[\dot{0}]\\ [\dot{0}]\,[\dot{0}]}}(\theta) = \sum_{[m''_n]} d^{\,[m''_n]}_{\substack{[m_{n+1,2\cdots}]\,[m_{n-1}]\\ [m_{n-2}]\,[m_{n-2}]}}(\theta) \\
&\qquad \times C\left(\begin{array}{cc|c} [m'_n] & [k,\dot{0}] & [m''_n] \\ [m_{n+1,2\cdots}] & [\dot{0}] & [m_{n+1,2\cdots}] \\ [m_{n-2}] & [\dot{0}] & [m_{n-2}] \end{array}\right)
C\left(\begin{array}{cc|c} [m'_n] & [k,\dot{0}] & [m''_n] \\ [m_{n-1}] & [\dot{0}] & [m_{n-1}] \\ [m_{n-2}] & [\dot{0}] & [m_{n-2}] \end{array}\right).
\end{aligned} \tag{4.7}$$

Integrating over θ, using the orthogonality property of the d functions of SO(n) [Ref. 8, Eq. (3.13)], we obtain a Kronecker delta $\delta_{[m''_n][m_n]}$. Thus the sum over $[m''_n]$ in Eq. (4.7) can be performed. The final result for the representation functions of ISO(n) is:

$$\begin{aligned}
&{}^I d^{\,\gamma,[m_{n+1,2\cdots}]}_{\substack{[m_n]\,[m'_n]\\ [m_{n-1}]\,[m_{n-1}]}}(\xi) \\
&= \left(\frac{\dim_n[m'_n]}{\dim_n[m_n]}\right)^{1/2} \sum_k K_k(\gamma\xi)(\gamma\xi)^{1-n/2} C_k^{-1+n/2}(1) \\
&\quad \times C\left(\begin{array}{cc||c} [m'_n] & [k,\dot{0}] & [m_n] \\ [m_{n+1,2\cdots}] & [\dot{0}] & [m_{n+1,2\cdots}] \end{array}\right)
\times C\left(\begin{array}{cc||c} [m'_n] & [k,\dot{0}] & [m_n] \\ [m_{n-1}] & [\dot{0}] & [m_{n-1}] \end{array}\right),
\end{aligned} \tag{4.8}$$

where $K_k(\gamma\xi)$ amd $C_k^{-1+n/2}(1)$ are given in Eqs. (4.4) and (4.6) respectively, and

$$C\begin{pmatrix} [m_n'] & [k,0] & \bigg|\bigg| & [m_n] \\ [\cdots] & [\dot{0}] & & [\cdots] \end{pmatrix}$$

is an isoscalar factor of SO(n).

This gives us a good reason to calculate the CG coefficients of SO(n) involving the most degenerate representation $[k,\dot{0}]$. In this paper we have made no attempt to calculate these coefficients. It is worth noting that these coefficients for SO(n), $n\geqslant 5$, are not multiplying free. Therefore the calculation of these coefficients is also intimately connected with the solution of the multiplicity problem of the CG coefficients of SO(n). Gavrilik and Klimyk[15] have claimed that all multiplicity-free CG coefficients of SO(n) can be obtained in principle. As far as we know, no explicit expressions for $k>1$ have been obtained for SO(n), $n\geqslant 5$. We urge the workers in this field to obtain an explicit expression for the CG coefficients in Eq. (4.8).

ACKNOWLEDGMENT

We wish to thank Moorhead State University for financial support of this work.

[1]A. Chakrabarti, J. Math. Phys. **9**, 2087 (1968).
[2]M.K.F. Wong and H.Y. Yeh, J. Math. Phys. **16**, 800 (1975).
[3]M.K.F. Wong and H.Y. Yeh, J. Math. Phys. **20**, 247 (1979).
[4]M.K.F. Wong and H.Y. Yeh, J. Math. Phys. **20**, 629 (1979).
[5]W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968).
[6]J.D. Talman, *Special Functions: A Group Theoretic Approach* (Benjamin, New York, 1968).
[7]A.O. Barut and R. Rączka, *Theory of Group Representations and Applications* (Polish Scientific Publishers, Warszawa, Poland, 1977).
[8]K.B. Wolf, J. Math. Phys. **12**, 197 (1971).
[9]A. Erdelyi, W. Magnus, F. Oberhetting, and F.G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2.
[10]G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University, Cambridge, England, 1922).
[11]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University, Princeton, New Jersey, 1960).
[12]N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations, Mathematical Monographs Vol.* 22, translated by V.N. Singh (American Mathematical Society, Providence, R.I., 1968).
[13]D.Z. Freedman and J.M. Wang, Phys. Rev. **160**, 1560 (1967).
[14]L.C. Biedenharn, J. Math. Phys. 7, 433 (1961).
[15]A.M. Gavrilik and A.U. Klimyk, "The Representation Matrix Elements and Clebsch–Gordan Coefficients of the Rotation Group SO(n)," Preprint (1977).

Asymptotic approximations for modified Bessel functions

P. Kasperkovitz

Institut für Theoretische Physik, TU Wien, A-1040 Karlsplatz 13, Austria

(Received 5 January 1979)

The behavior of a $q_\nu(x) = I_\nu(x)/I_\nu^a(x)$, where I_ν is a modified Bessel function with integral or half-integral index ν and I_ν^a the leading term of its asymptotic series, is investigated for $x \gg 1$. It is shown that $q_\nu(x)$ may be approximated by $e_\nu(x) = \exp(-\nu^2/2x)$, the difference $r_\nu(x) = q_\nu(x) - e_\nu(x)$ being of order $x^{-1/4}$. Bounds for $r_\nu(x)$ depending only on x are derived for each of the two classes of ν's and an application of these results in scattering theory is indicated.

1. INTRODUCTION

The modified Bessel functions I_ν are well known special functions[1] closely related to the ordinary Bessel functions J,

$$I_\nu(x) = e^{-i\nu(\pi/2)} J_\nu(ix), \quad x \text{ real}. \tag{1}$$

Their asymptotic behavior for large positive x and moderate $|\nu|$ ($\ll x$) is given by the asymptotic series

$$I_\nu(x) \sim I_\nu^a(x)\left\{1 - \frac{4\nu^2-1}{8x} + \frac{(4\nu^2-1)(4\nu^2-9)}{2!(8x)^2} - \cdots\right\}, \tag{2}$$

$$I_\nu^a(x) = (2\pi x)^{-1/2} e^x,$$

and for $|\nu|$ comparable to or larger than x ($\gg 1$) by the uniform asymptotic expansion

$$I_\nu(x) \sim (2\pi\nu)^{-1/2}(1 + x^2/\nu^2)^{-1/4} e^{\nu\eta}\left\{1 + \frac{3t - 5t^3}{24\nu} + \cdots\right\},$$

$$\eta = (1 + x^2/\nu^2)^{1/2} + \ln\frac{x}{\nu + (\nu^2 + x^2)^{1/2}},$$

$$t = \frac{\nu}{(\nu^2 + x^2)^{1/2}}. \tag{3}$$

The series on the rhs of Eqs. (2) and (3) are both semiconvergent and the number of terms needed for the best approximation of the lhs depends both on x and ν. However, taking only a glance at Eq. (2) one could be tempted to replace the curly bracket by $\exp(-\nu^2/2x)$, putting up with an error of order x^{-1}.

The main result of this paper is that such an approximation is actually possible if the index ν is an integral or half-integral number. Let $[\nu]$ and $r_\nu(x)$ be defined by

$$[\nu] = [0] \Longleftrightarrow \nu \text{ integer}, \qquad [\nu] = [\tfrac{1}{2}] \Longleftrightarrow \nu \text{ half-integer}, \tag{4}$$

$$r_\nu(x) = I_\nu(x)/I_\nu^a(x) - \exp(-\nu^2/2x). \tag{5}$$

Then as is shown in Sec. 2 the following relations hold true:

$$|r_\nu(x)| < r^{[\nu]}(x), \tag{6}$$

$$r^{[0]}(x) = 1{,}07x^{-1/4} \quad \text{for } x \geqslant 1, \tag{7}$$

$$r^{[1/2]}(x) = 23{,}06x^{-1/4} \quad \text{for } x \geqslant 2. \tag{8}$$

Furthermore, let the real function $F_x^{[\nu]}$, $F = A$ or Z, be defined by

$$A_x^{[0]}(\alpha) = (x/\pi)^{1/4} e^{x(\cos\alpha - 1)} = (x/\pi)^{1/4} e^{-x} \sum_m I_m(x) e^{im\alpha}, \tag{9}$$

$$Z_x^{[0]} = (4\pi^3 x)^{-1/4} \sum_m e^{-(m^2/2x) + im\alpha}, \tag{10}$$

$$A_x^{[1/2]}(t) = 2x^{1/2} e^{x(t-1)} = (2\pi)^{1/2} e^{-x} \sum_l I_{l+1/2}(x)(2l+1)P_l(t), \tag{11}$$

$$Z_x^{[1/2]}(t) = x^{-1/2} \sum_l e^{-(l+1/2)^2/2x}(2l+1)P_l(t), \tag{12}$$

and let $\|\cdot\|$ be the L^2 norm of the functions defined on $-\pi < \alpha < \pi$ ($[\nu] = [0]$) and $-1 < t < 1$, ($[\nu] = [\frac{1}{2}]$), respectively. Then it is also shown in Sec. 2 that

$$\|A_x^{[0]} - Z_x^{[0]}\| < 0{,}80x^{-1/2} \quad \text{for } x \geqslant 1, \tag{13}$$

$$\|A_x^{[1/2]} - Z_x^{[1/2]}\| < 20{,}17x^{-1/2} \quad \text{for } x \geqslant 2. \tag{14}$$

Equation (13) implies Eq. (7) but not vice versa. Equations (14) and (8) neither imply each other, but their proofs coincide to a large extent.

Equations (13) and (14) are not only stated because of their close connection with Eqs. (7) and (8), respectively, but even more since they are of use in a special problem of scattering theory. This problem, the determination of the angular distribution of a scattered Gaussian wave packet, is briefly outlined in Sec. 3.

2. PROOFS

A. The general method

Denote by $\langle .,. \rangle$ the usual scalar product in $L^2(-a,a)$ where $a = \pi$ for $[\nu] = [0]$ and $a = 1$ for $[\nu] = [\frac{1}{2}]$, and by F one of the following orthonormalized basis functions:

$$[\nu] = [0]: \quad \nu = m, \quad F_m(\alpha) = (2\pi)^{-1/2} e^{im\alpha}, \tag{15}$$

$$[\nu] = [\tfrac{1}{2}]: \quad \nu = l + \tfrac{1}{2}, \quad F_{l+1/2}(t) = (2l+1)^{1/2} P_l(t). \tag{16}$$

The quantities to be compared with each other, namely $q_\nu(x) = I_\nu(x)/I_\nu^a(x)$ and $e_\nu(x) = \exp(-\nu^2/2x)$, are then proportional to the scalar products $\langle A_x^{[\nu]}, F_\nu \rangle$ and $\langle Z_x^{[\nu]}, F_\nu \rangle$ [see Eqs. (9)–(12)], the common factor $\gamma^{[\nu]}(x)$ depending only on $[\nu]$ and x,

$$\gamma^{[0]}(x) = (\pi x)^{1/4}, \quad \gamma^{[1/2]}(x) = x^{1/2}. \tag{17}$$

Bounds of the difference $r_\nu(x) = q_\nu(x) - e_\nu(x)$ are obtained by means of Schwarz's inequality,

$$\begin{aligned} |I_\nu(x)/I_\nu^a(x) - \exp(-\nu^2/2x)| &= |\gamma^{[\nu]}(x)\langle A_x^{[\nu]}, F_\nu \rangle - \gamma^{[\nu]}(x)\langle Z_x^{[\nu]}, F_\nu \rangle| \\ &= \gamma^{[\nu]}(x)\,|\langle A_x^{[\nu]} - Z_x^{[\nu]}, F_\nu \rangle| \\ &\leqslant \gamma^{[\nu]}(x)\|A_x^{[\nu]} - Z_x^{[\nu]}\|. \end{aligned} \tag{18}$$

 0022-2488/80/010006-08$01.00

Since it is difficult to calculate $\|A_x^{[\nu]} - Z_x^{[\nu]}\|$ directly a number of auxiliary functions $B_x^{[\nu]}, C_x^{[\nu]}, \cdots$ are introduced and the rhs of (18) is estimated by means of the triangle inequality,

$$\|A_x^{[\nu]} - Z_x^{[\nu]}\| \leqslant \|A_x^{[\nu]} - B_x^{[\nu]}\| + \cdots + \|Y_x^{[\nu]} - Z_x^{[\nu]}\|. \quad (19)$$

This inequality and a suitable choice of auxiliary functions yield the bounds (13) and (14). Since the rhs of (13) is of order $x^{-1/2}$ and $\gamma^{[0]}(x)$ is proportional to $x^{1/4}$, Eq. (7) is obtained from Eq. (13) by multiplying both sides of (13) by $\gamma^{[0]}(x)$. For $[\nu] = [\frac{1}{2}]$ the situation is less satisfactory since some of the terms on the rhs of (19), say $\|C_x^{[\nu]} - D_x^{[\nu]}\|$, are of order $x^{-1/2}$ which is proportional to $\gamma^{[1/2]}(x)^{-1}$. Fortunately it is possible to derive in these cases bounds of the difference of the scalar products $\langle C_x^{[\nu]}, F_\nu \rangle$ and $\langle D_x^{[\nu]}, F_\nu \rangle$ independent of ν and of order $x^{-p}, p > \frac{1}{2}$. Hence, a combination of Schwarz's and triangle inequalities gives

$$\begin{aligned} &|I_\nu(x)/I_\nu^a(x) - \exp(-\nu^2/2x)| \\ &\quad \leqslant \gamma^{[\nu]}(x)\{\|A_x^{[\nu]} - B_x^{[\nu]}\| + \cdots + \max|\langle C_x^{[\nu]}, F_\nu\rangle \\ &\qquad - \langle D_x^{[\nu]}, F_\nu\rangle| + \cdots + \|Y_x^{[\nu]} - Z_x^{[\nu]}\|\}, \end{aligned} \quad (20)$$

finally yielding the bound (8) for $[\nu] = [\frac{1}{2}]$.

B. Estimation of sums

In the following it is often necessary to derive bounds for sums of the form $\Sigma f(k)$ where f is a real function and $a \leqslant k$ (integer) $\leqslant b$. Suppose that f is concave in $[a,b]$ and define $g(\lambda)$ and $d(k)$ by

$$g(\lambda) = \int_{\lambda - 1/2}^{\lambda + 1/2} d\kappa f(\kappa), \quad d(k) = f(k) - g(k). \quad (21)$$

Then, as a consequence of the concavity of f,

$$0 < d(k) < \tfrac{1}{2} f(k) - \tfrac{1}{4} f(k - \tfrac{1}{2}) - \tfrac{1}{4} f(k + \tfrac{1}{2}),$$
$$\text{for all } k \in [a + \tfrac{1}{2}, b - \tfrac{1}{2}]. \quad (22)$$

Repeated use of (22) gives

$$\begin{aligned} &a + \tfrac{1}{2} \leqslant k \leqslant b - 1:\ d(k) < g(k) - \tfrac{1}{2} f(k - \tfrac{1}{2}) - \tfrac{1}{2} g(k + \tfrac{1}{2}), \\ &a + 1 \leqslant k \leqslant b - 1:\ d(k) < g(k) - \tfrac{1}{2} g(k - \tfrac{1}{2}) - \tfrac{1}{2} g(k + \tfrac{1}{2}), \quad (23) \\ &a + 1 \leqslant k \leqslant b - \tfrac{1}{2}:\ d(k) < g(k) - \tfrac{1}{2} g(k - \tfrac{1}{2}) - \tfrac{1}{2} f(k + \tfrac{1}{2}), \end{aligned}$$

and hence, if $a + \frac{1}{2} \leqslant k' < k'' \leqslant b - \frac{1}{2}$,

$$\begin{aligned} 0 < \sum_{k=k'}^{k''} d(k) &< g(k') - \tfrac{1}{2} f(k' - \tfrac{1}{2}) - \tfrac{1}{2} g(k' + \tfrac{1}{2}) \\ &\quad + \sum_{k=k'+1}^{k''-1} [g(k) - \tfrac{1}{2}(k - \tfrac{1}{2}) - \tfrac{1}{2} g(k + \tfrac{1}{2})] \\ &\quad + g(k'') - \tfrac{1}{2} g(k'' - \tfrac{1}{2}) - \tfrac{1}{2} f(k'' + \tfrac{1}{2}) \\ &= \int_{k'-1/2}^{k'} d\kappa [f(\kappa) - f(k' - \tfrac{1}{2})] \\ &\quad + \int_{k''}^{k''+1/2} d\kappa [f(\kappa) - f(k'' + \tfrac{1}{2})]. \end{aligned} \quad (24)$$

Next suppose f to be differentiable in $[a,b]$ and let $\delta_\mp(c)$ be bounds satisfying

$$\delta_\mp(c) \geqslant |f'(\kappa)|, \quad \text{for } \kappa \pm \tfrac{1}{4} \in [c - \tfrac{1}{4}, c + \tfrac{1}{4}]. \quad (25)$$

Then inequality (24) can be simplified to

$$0 < \sum_{k=k'}^{k''} f(k) - \int_{k'-1/2}^{k''+1/2} dk f(k) \leqslant \tfrac{1}{8}[\delta_-(k') + \delta_+(k'')]. \quad (26)$$

The rhs of (26) gives an upper bound if the sum is replaced by the integral provided that f is concave. Equation (26) also shows the integral to be greater than the sum if f is convex ($-f$ concave). If f is neither concave nor convex, its domain may be divided into parts where f is either concave or convex, leaving out intervals of unit length containing the points of inflection ($f''(\kappa_i) = 0, \quad i = 1,2,\ldots$). Let k_i be the integer belonging to the ith of these intervals, $k_i \in [\kappa_i, \kappa_i + 1)$, f be monotone and differentiable in $[\kappa_i, \kappa_i + 1)$, and

$$\delta(k_i) \geqslant |f'(\kappa)|, \text{ for } \kappa \in [\kappa_i, \kappa_i + 1]. \quad (27)$$

Then $|d(k_i)| \leqslant \frac{1}{8}\delta(k_i)$ and

$$\begin{aligned} \sum_{k=p}^{q} f(k) \leqslant &\int_{p-1/2}^{p+1/2} d\kappa f(\kappa) \\ &+ \tfrac{1}{8} \sum_{i=1}^{n} [\delta(k_{2i-1}) + \delta_-(k_{2i-1} + 1) \\ &+ \delta_+(k_{2i} - 1) + \delta(k_{2i})]. \end{aligned} \quad (28)$$

The index is running over the number of intervals contained in $[p,q]$ where f is concave. If f is concave in $[p - \frac{1}{2}, p + \frac{1}{2}]$ (and/or in $[q - \frac{1}{2}, q + \frac{1}{2}]$) $\delta(k_1)$ [and/or $\delta(k_{2n})$] may be dropped in (28).

C. Integral index

Here only one auxiliary function is needed, namely

$$B_x(\alpha) = \left(\frac{x}{\pi}\right)^{1/4} e^{-2x(\alpha/2)^2} \quad (29)$$

(the index $[\nu] = [0]$ is dropped throughout this subsection). To calculate $\|A_x - B_x\|$ we note that both A_x and B_x are real even functions and that $A_x(\alpha) \geqslant B_x(\alpha)$ since $\sin(\alpha/2) \leqslant \alpha/2$ for $\alpha \in [0,\pi]$. Hence $\langle A_x, B_x \rangle = \langle B_x, A_x \rangle > \|B_x\|^2$ and

$$\|A_x - B_x\|^2 < \|A_x\|^2 - \|B_x\|^2. \quad (30)$$

$\|A_x\|^2$ and $\|B_x\|^2$ reduce to well-known integrals[2] defining the special functions I_0 and erf,

$$\|A_x\|^2 = (4\pi x)^{1/2} e^{-2x} I_0(2x), \quad (31)$$
$$\|B_x\|^2 = \operatorname{erf}\pi x^{1/2} = 1 - \operatorname{erfc}\pi x^{1/2}; \quad (32)$$

whereas the error function erf and hence $\|B_x\|^2$ approach unity from below very rapidly $\|A_x\|^2$ exceeds 1 by a quantity of order x^{-1} (as can be seen from the numerical tables[3] of I_0) so that the bound

$$\|A_x - B_x\| < [(4\pi x)^{1/2} e^{-2x} I_0(2x) - \operatorname{erf}\pi x^{1/2}]^{1/2} \quad (33)$$

is of order $x^{-1/2}$.

To calculate $\|B_x - Z_x\|$ we introduce the functions f: $\mathbb{Z} \to \mathbb{C}$ defined by the Fourier coefficients of the functions F: $(-\pi,\pi) \to \mathbb{C}$,

$$f(m) = (2\pi)^{-1/2} \int_{-\pi}^{\pi} d\alpha\, F(\alpha) e^{-im\alpha}, \quad (34)$$

and define for them a scalar product and a norm by

$$\langle f, g \rangle = \sum_m f^*(m) g(m) = \langle F, G \rangle, \quad \|f\| = \|F\|. \quad (35)$$

This makes

$$\|B_x - Z_x\|^2 = \operatorname{erf}\pi x^{1/2} + \|z_x\|^2 - 2\langle b_x, z_x \rangle, \quad (36)$$

since z_x and b_x are real functions according to (10) and[4]

$$b_x(m) = \mathrm{Re}(\pi x)^{-1/4} e^{-(m^2/2x)}\left[\mathrm{erf}\left(i\frac{m}{\sqrt{2x}} + \pi\sqrt{\frac{x}{2}}\right) - \mathrm{erf}\left(i\frac{m}{\sqrt{2x}}\right)\right]. \tag{37}$$

Now Re $\mathrm{erf}(im/\sqrt{2x}) = 0$ and[5]

$$\begin{aligned}\mathrm{erf}\left(\frac{1}{\sqrt{2}}\left[\pi\sqrt{x} + i\frac{m}{\sqrt{x}}\right]\right) &= 1 - \frac{1}{\pi}e^{-(\pi^2/2)x + (m^2/2x) - i\pi m} \\ &\times\left\{\frac{\sqrt{2}\Gamma(\frac{1}{2})}{\pi\sqrt{x} + im/\sqrt{x}} - R_1\left(\frac{1}{\sqrt{2}}\left[\pi\sqrt{x} + im/\sqrt{x}\right]\right)\right\},\end{aligned} \tag{38}$$

$$\begin{aligned}b_x(m) = z_x(m) + (-1)^m\frac{1}{\pi}e^{-(\pi^2/2)x} &\times\mathrm{Re}\left\{\frac{\sqrt{2}\Gamma(\frac{1}{2})}{\pi\sqrt{x} + im/\sqrt{x}} - R_1\left(\frac{1}{\sqrt{2}}\left[\pi\sqrt{x} + im/\sqrt{x}\right]\right)\right\},\end{aligned} \tag{39}$$

where[5]

$$|R_1(z)| < \Gamma(\tfrac{3}{2})/|z| \; \mathrm{Re}z. \tag{40}$$

Therefore,

$$\begin{aligned}&\|B_x - Z_x\|^2 = \mathrm{erf}\pi x^{1/2} - \|z_x\|^2 - 2R_x, \\ &R_x = (\pi x)^{-1/2}\sum_m e^{-(m^2/2x)}(-1)^m\frac{1}{\pi}e^{-(\pi^2/2)x}\,\mathrm{Re}\{\cdots\},\end{aligned} \tag{41}$$

and because of the triangle inequality, $\mathrm{Re}z \leqslant |z|$, and (40),

$$\begin{aligned}|R_x| &< (\pi x)^{-1/2}\sum_m e^{-(m^2/2x)-(\pi^2/2)x}\frac{1}{\pi} \times\left\{\frac{2^{1/2}\Gamma(\frac{1}{2})}{\pi x^{1/2}} + \frac{2^{3/2}\Gamma(\frac{3}{2})}{(\pi^2 x + (m^2/x))\pi x^{1/2}}\right\} \\ &< 2^{1/2}\pi^{-2}x^{-1}e^{-(\pi^2/2)x}\left(1 + \frac{1}{\pi^2 x}\right) \times\left[(2\pi x)^{1/2} + \tfrac{1}{2}(ex)^{-1/2}\right],\end{aligned} \tag{42}$$

where the bracket [⋯] is the bound of $\Sigma\exp(-m^2/2x)$ obtained according to (28). This method is also used to derive the bounds

$$1 - (2\pi e)^{-1/2}x^{-1} < \|z_x\|^2 < 1 + (2\pi e)^{-1/2}x^{-1}. \tag{43}$$

Magnifying erf to 1 and combining (41), (42), and (43), one obtains

$$\begin{aligned}\|B_x - Z_x\|^2 &< (2\pi e)^{-1/2}x^{-1} + 2\pi^{-3/2}(1 + \pi^{-2}x^{-1})[1 + (8\pi e)^{-1/2}x^{-1}] \\ &\times x^{-1/2}e^{-(\pi^2/2)x}.\end{aligned} \tag{44}$$

For $x \geqslant x_0 > 0$ it is possible to bound from above all functions appearing in (44) by multiples of x^{-1}. This yields

$$\|B_x - Z_x\| < 0{,}492x^{-1/2} \quad \text{for } x \geqslant 1. \tag{45}$$

Inequality (45) supplemented by

$$\|A_x - B_x\| < 0{,}306x^{-1/2} \quad \text{for } x \geqslant 1 \tag{46}$$

(which may be derived from the numerical tables[3,6] of I_0 and $\mathrm{erfc} = 1 - \mathrm{erf}$) yields the estimate (13) from which inequality (7) is obtained by multiplying both sides of (13) by $(\pi x)^{1/4}$ (cf. Sec. 2A).

D. Half-integral index

In order to prove the estimates (8) and (14) five auxiliary functions are introduced: $B_x,\ldots,F_x$ (here too the index $[\nu]$ ($=[\frac{1}{2}]$) is omitted in the following). These functions are defined via their expansion coefficients with respect to the orthonormalized basis $\{(2l+1)^{1/2}P_l : P_l = \text{Legendre polynomial}, l = 0,1,\cdots\}$:

$$a_x(l) = [2\pi(2l+1)]^{1/2}e^{-x}I_{l+1/2}(x) \quad [cf.\ \text{Eq. (11)}], \tag{47}$$

$$b_x(l) = \sigma(2x - 3l)a_x(l), \tag{48}$$

$$c_x(l) = \sigma(2x - 3l)\left(\frac{2l+1}{x}\right)^{1/2}\left[1 - e^{-2x}\sum_{k=0}^{l}\frac{(2x)^k}{k!}\right]\left(1 - \frac{l+1}{2x}\right), \tag{49}$$

$$d_x(l) = \sigma(2x - 3l)\left(\frac{2l+1}{x}\right)^{1/2}\left(1 - \frac{l+1}{2x}\right), \tag{50}$$

$$e_x(l) = \sigma(2x - 3l)\left(\frac{2l+1}{x}\right)^{1/2}e^{-l(l+1)/2x}, \tag{51}$$

$$f_x(l) = \left(\frac{2l+1}{x}\right)^{1/2}e^{-l(l+1)/2x}, \tag{52}$$

$$z_x(l) = \left(\frac{2l+1}{x}\right)^{1/2}e^{-(l+\frac{1}{2})^2/2x}. \tag{53}$$

The function σ appearing in (48)–(51) is the step function:

$$\sigma(y) = 1 \quad \text{for } y \geqslant 0, \quad \sigma(y) = 0 \quad \text{for } y < 0. \tag{54}$$

If G and H are defined via their expansion coefficients $g(l)$ and $h(l)$, respectively, then

$$\|G - H\|^2 = \sum_l |g(l) - h(l)|^2. \tag{55}$$

Therefore, to estimate $\|A_x - B_x\|$, $\|B_x - C_x\|$, etc. [which is necessary for deriving (14)] one has to find l-dependent bounds for $|a_x(l) - b_x(l)|$, $|b_x(l) - c_x(l)|$, etc. that allow the summation in (55) to be performed. To derive (8) it is, on the other hand, necessary to find bounds for $|a_x(l) - b_x(l)|$, etc., holding for all values of l. The norms $\|A_x - B_x\|$, etc. provide such bounds but they turn out to be too weak in two cases so that both l-independent bounds that allow summation and l-independent bounds have to be found there.

To calculate $\|A_x - B_x\|$ we first use the integral representation[7] of $I_{l+\frac{1}{2}}(x)$ to write $a_x(l)$ as

$$a_x(l) = [x(2l+1)]^{1/2}e^{-x}\left(\frac{x}{2}\right)^l\frac{1}{l!}\int_{-1}^{1}dt\, e^{xt}(1-t^2)^l. \tag{56}$$

Since $(1-t^2)^{l+k} \leqslant (1-t^2)^l$ for $t \in [-1,1]$ and $k \geqslant 1$, one has

$$
\begin{aligned}
a_x(l+k)
&= [x(2l+2k+1)]^{1/2} e^{-x}\left(\frac{x}{2}\right)^{l+k}\frac{1}{(l+k)!}\\
&\quad\times\int_{-1}^{1} dt\, e^{xt}(1-t^2)^{l+k}\\
&\leqslant \frac{x}{2(l+1)}\cdots\frac{x}{2(l+k)}\left(\frac{2l+2k+1}{2l+1}\right)^{1/2} a_x(l)\\
&\leqslant \left(\frac{x}{2l+2}\right)^{k-1/2}\left(\frac{x}{2l+2k}\right)^{1/2}\\
&\quad\times\left(\frac{2l+2k+1}{2l+1}\right)^{1/2} a_x(l)\\
&\leqslant \left(\frac{x}{2l+2}\right)^{k-1/2}\left(\frac{x}{2l+2}\right)^{1/2}\left(\frac{2l+2}{2l+1}\right)^{1/2}\\
&\quad\times\left(\frac{2l+2k+1}{2l+2k}\right)^{1/2} a_x(l)\\
&\leqslant \left(\frac{x}{2l+2}\right)^{k}\left(1+\frac{1}{2+1}\right) a_x(l).
\end{aligned}
\tag{57}
$$

Obviously, inequality (57) holds true also for $k=0$. Because of

$$
\|A_x - B_x\|^2 = \sum_{l>(2x/3)-1} a_x^2(l), \tag{58}
$$

the coefficients of interest are those for which $x/2l < 3/4$. For these coefficients

$$
\begin{aligned}
a_x(l+k) &< \left(\frac{x}{2l}\right)^k\left(1+\frac{1}{2l}\right) a_x(l)\\
&< \left(\tfrac{3}{4}\right)^k\left(1+\frac{3}{4x}\right) a_x(l).
\end{aligned}
\tag{59}
$$

Now let l'' be the smallest integer larger than $2x/3$. Then

$$
\begin{aligned}
\|A_x - B_x\|^2 &= \sum_{k\geqslant 0} a_x^2(l''+k)\\
&< \tfrac{16}{7}(1+3/4x)\, a_x^2(l'')
\end{aligned}
\tag{60}
$$

and, since $a_x(l)>0$ [cf. Eq. (56)],

$$
\|A_x - B_x\| < \left(\tfrac{16}{7}\right)^{1/2}(1+3/4x)\, a_x(l''). \tag{61}
$$

To estimate $a_x(l'')$ we use again the integral representation (56) and the inequality

$$
1 - z \leqslant e^{-z} \tag{62}
$$

with $z = t^2$. This yields

$$
a_x(l) < [x(2l+1)]^{1/2} e^{-x}\left(\frac{x}{2}\right)^l \frac{1}{l!}\int_{-1}^{1} dt\, e^{xt - lt^2}. \tag{63}
$$

Since the integrand is positive for all $t\in\mathbb{R}$ one may extend the integration from $(-1,1)$ to $(-\infty,\infty)$ to get a closed expression for the integral.[8] This and Stirling's inequality[9]

$$
\frac{1}{n!} < (2\pi n)^{-1/2}\left(\frac{e}{n}\right)^n \tag{64}
$$

gives for $l > 2x/3$

$$
\begin{aligned}
a_x(l) &< \left(\frac{x}{l+1}\cdot\frac{l+1}{l}\cdot\frac{2l+1}{2l}\right)^{1/2} e^{-x+(x^2/2l)}\left(\frac{ex}{2l}\right)^l\\
&< \left(\frac{3}{2}\right)^{1/2}\left(1+\frac{3}{2x}\right) e^{-x+3x/8}\left(\frac{3e}{4}\right)^l.
\end{aligned}
\tag{65}
$$

Now $l'' < (2x/3)+1$ so that

$$
\begin{aligned}
a_x(l'') &< \left(\frac{3}{2}\right)^{1/2}\left(1+\frac{3}{2x}\right) e^{-x(1-3/8-2/3-(2/3)\ln(3/4))}\\
&\quad\times\left(\frac{3e}{4}\right)\\
&< \left(\frac{3}{2}\right)^{1/2}\left(\frac{3e}{4}\right)\left(1+\frac{3}{2x}\right) e^{-x/7},
\end{aligned}
\tag{66}
$$

$$
\|A_x - B_x\| < 4\left(1+\frac{3}{4x}\right)\left(1+\frac{3}{2x}\right) e^{-x/7}. \tag{67}
$$

In calculating $\|B_x - C_x\|$ one has to consider the coefficients

$$
\begin{aligned}
b_x(l) &= \left(\frac{2l+1}{2x}\right)^{1/2}\frac{1}{l!}\int_0^{2x} dy\, e^{-y} y^l\\
&\quad\times\left(1-\frac{y}{2x}\right)^l \quad [=a_x(l)],
\end{aligned}
\tag{68}
$$

$$
\begin{aligned}
c_x(l) &= \left(\frac{2l+1}{2x}\right)^{1/2}\frac{1}{l!}\int_0^{2x} dy\, e^{-y} y^l\\
&\quad\times\left(1-\frac{l+1}{2x}\right)^l,
\end{aligned}
\tag{69}
$$

with $l \leqslant 2x/3$ only $[\sigma(2x-3l)=1]$. Equation (68) is once more the integral representation (56) of $a_x(l)$ with $y = x(1-t)$. Equation (69) is easily checked using the relation[10]

$$
\int dy\, e^{-y} y^n/n! = -e^{-y}\sum_{k=0}^{n} y^k/k! + \text{const}. \tag{70}
$$

Splitting the range of integration and using the bounds

$$
(a-b)lb^{l-1} \leqslant a^l - b^l \leqslant (a-b)la^{l-1} \quad \text{for } a\geqslant b>0,\ l\geqslant 0 \tag{71}
$$

one finds for $l+1<2x$, i.e.

$x > \frac{3}{4}$:

$$
\begin{aligned}
&b_x(l) - c_x(l)\\
&= \left(\frac{2l+1}{x}\right)^{1/2}\frac{1}{l!}\left\{\int_0^{l+1} dy\, e^{-y} y^l\right.\\
&\quad\times\left[\left(1-\frac{y}{2x}\right)^l - \left(1-\frac{l+1}{2x}\right)^l\right] - \int_{l+1}^{2x} dy\, e^{-y} y^l\\
&\quad\left.\times\left[\left(1-\frac{l+1}{2x}\right)^l - \left(1-\frac{y}{2x}\right)^l\right]\right\}\\
&\geqslant \left(\frac{2l+1}{x}\right)^{1/2}\frac{1}{l!}\left\{\int_0^{l+1} dy\, e^{-y} y^l\right.\\
&\quad\times\frac{1}{2x}(l+1-y)l\left(1-\frac{l+1}{2x}\right)^{l-1} - \int_{l+1}^{2x} dy\, e^{-y} y^l\\
&\quad\left.\times\frac{1}{2x}(y-l-1)l\left(1-\frac{l+1}{2x}\right)^{l-1}\right\}\\
&= \left(\frac{2l+1}{x}\right)^{1/2}\frac{1}{2x(l-1)!}\int_0^{2x} dy\left[\frac{d}{dy}e^{-y}y^{l+1}\right]\\
&\quad\times\left(1-\frac{l+1}{2x}\right)^{l-1} \geqslant 0;
\end{aligned}
\tag{72}
$$

$x > \frac{3}{4}$:

$$
\begin{aligned}
&b_x(l) - c_x(l)\\
&\leqslant \left(\frac{2l+1}{x}\right)^{1/2}\frac{1}{2x(l-1)!}\int_0^{2x} dy\left[\frac{d}{dy}e^{-y}y^{l+1}\right]\\
&\quad\times\left(1-\frac{y}{2x}\right)^{l-1}.
\end{aligned}
\tag{73}
$$

Equations (72) and (73) show that

$$b_x(0) - c_x(0) = 0,$$

$$b_x(1) - c_x(1) = \left(\frac{3}{x}\right)^{1/2} 2xe^{-2x}, \quad (74)$$

which also follows from the definitions (47) and (49). Using (62) with z replaced by $y/2x$, extending the range of integration to infinity, and evaluating the integral by means of (70), one finds for $l \geqslant 2$,

$x > \frac{3}{4}$:

$$\begin{aligned} b_x(l) - c_x(l) &\leqslant \left(\frac{2l+1}{x}\right)^{1/2} \frac{1}{4x^2(l-2)!} \int_0^{2x} dy e^{-y} y^{l+1}\left(1 - \frac{y}{2x}\right)^{l-2} \\ &\leqslant \left(\frac{2l+1}{x}\right)^{1/2} \frac{1}{4x^2(l-2)!} \int_0^{\infty} dy e^{-y} y^{l} e^{-[(l-2)/2x]y} \\ &\leqslant \left(\frac{2l+1}{x}\right)^{1/2} \frac{1}{4x^2(l-2)!} \left(1 + \frac{l-2}{2x}\right)^{-l-2} (l+1)! \\ &\leqslant \left(\frac{2l+1}{x}\right)^{1/2} \frac{(l-1)l(l+1)}{4x^2} e^{-(l^2/4x)+(1/x)}. \end{aligned} \quad (75)$$

In the final step the inequality

$$(1+z)^{-1} \leqslant e^{-z/2} \quad \text{for } 0 \leqslant z < 1, \quad (76)$$

has been used for $z = (l-2)/2x$. This inequality can be proved by means of the series expansion of $\ln(1+z)$. Since $l+1 \leqslant 3l/2$ and $2l+1 \leqslant 5l/2$ for $l \geqslant 2$ and ($z = l^2$)

$$z \geqslant 0: \quad z^p e^{-qz} \leqslant \left(\frac{p}{qe}\right)^p, \quad (77)$$

inequalities (72) and (75) may be weakened to

$$\begin{aligned} |b_x(l) - c_x(l)| &\leqslant \left(\frac{5l}{2x}\right)^{1/2} \frac{3l^3}{8x^2} e^{-(l^2/4x)+(1/x)} \\ &< 3{,}11 x^{-3/4} \quad \text{for } x > \tfrac{3}{4} \end{aligned} \quad (78)$$

which holds true also for $l = 0, 1$ because of (74). Estimate (78) is used in the derivation of (8). To obtain a bound for $\|B_x - C_x\|$ (75) is simplified to

$x > \frac{3}{4}$:

$$b_x(l) - c_x(l) < \left(\frac{2l+1}{x}\right)^{1/2} \frac{l^2(l+1)}{4x^2} e^{-(l^2/4x)+(1/x)}. \quad (79)$$

Since (79) holds true also for $l = 0, 1$ one has for

$x > \frac{3}{4}$:

$$\begin{aligned} \|B_x - C_x\|^2 &< \frac{e^{2/x}}{16x^5} \sum_{l \geqslant 0} (2l+1) l^4 (l+1)^2 e^{-l^2/2x} \\ &< 6(1 + \tfrac{2}{5} x^{-1/2})^5 e^{2/x} x^{-1} \end{aligned} \quad (80)$$

To obtain the last result the sums $\Sigma l^p \exp(-l^2/2x)$, $4 \leqslant p \leqslant 7$, have been estimated according to (28) and the resulting polynomial has been replaced by a simpler one with larger coefficients. (80) is equivalent to

$$\|B_x - C_x\| < 6^{1/2}(1 + \tfrac{2}{5}x^{-1/2})^{5/2} e^{1/x} x^{-1/2} \quad \text{for } x > \tfrac{3}{4}. \quad (81)$$

To estimate $\|C_x - D_x\|$ we first note that here too all the coefficients vanish for $l > 2x/3$ and that

$$d_x(l) = \left(\frac{2l+1}{x}\right)^{1/2} \frac{1}{l!} \int_0^{\infty} dy\, e^{-y} y^l \left(1 - \frac{l+1}{2x}\right)^l \quad (82)$$

because of (70). Hence,

$$\begin{aligned} 0 &< d_x(l) - c_x(l) \\ &= \left(\frac{2l+1}{x}\right)^{1/2} \frac{1}{l!} \int_{2x}^{\infty} dy e^{-y} y^l \left(1 - \frac{l+1}{2x}\right)^l \\ &= \left(\frac{2l+1}{x}\right)^{1/2} \frac{(2x)^l}{l!} e^{-2x}\left[1 + \frac{l}{2x} + \frac{l(l-1)}{(2x)^2} + \cdots + \frac{l!}{(2x)^l}\right]\left(1 - \frac{l+1}{2x}\right)^l \\ &< \left(\frac{2l+1}{x}\right)^{1/2} \frac{(2x)^l}{l!} e^{-2x} \sum_{k=0}^{\infty} \left(\frac{l}{2x}\right)^k \left(1 - \frac{l+1}{2x}\right)^l. \end{aligned} \quad (83)$$

Using the inequalities $[1 - (l/2x)]^{-1} < 2$, (64), and

$$0 < \lambda < 2x: \left(\frac{2x}{\lambda} - 1\right)^{\lambda} \leqslant \left(\frac{2x}{\lambda_0} - 1\right)^{\lambda_0} < (5-1)^{x/2} = 2^x, \quad (84)$$

where λ_0 is the solution of the transcendental equation $\ln[(2x/\lambda_0) - 1] = [1 - (\lambda_0/2x)]^{-1}$, allow (83) to be replaced by

$$\begin{aligned} d_x(l) - c_x(l) &< 2\left(\frac{2l+1}{x}\right)^{1/2} \frac{(2x)^l}{l!} e^{-2x}\left(1 - \frac{l}{2x}\right)^l \\ &< 2\left(\frac{2l+1}{x}\right)^{1/2} e^{-2x} (2\pi l)^{-1/2} e^l \left[\frac{2x}{l}\left(1 - \frac{l}{2x}\right)\right]^l \\ &< \left(\frac{6}{\pi}\right)^{1/2} e^l x^{-1/2} 2^x e^{-2x}, \end{aligned} \quad (85)$$

provided that $l \geqslant 1$. For $l = 0$ Eq. (82) yields

$$d_x(0) - c_x(0) = x^{-1/2} e^{-2x}. \quad (86)$$

Therefore, combining (86) and (85) and denoting by l' the largest integer $\leqslant 2x/3$, one finally obtains

$$\begin{aligned} \|C_x - D_x\|^2 &< x^{-1} e^{-4x} + \frac{6}{\pi} x^{-1} e^{-4x} 4^x \sum_{1 \leqslant l \leqslant l'} e^{2l} \\ &< x^{-1} e^{-4x} + \frac{6}{\pi(1 - e^{-2})} x^{-1} e^{-2(2 - \ln 2 - 2/3)x} \\ &< 4(1 + \tfrac{1}{8} e^{-3x})^2 x^{-1} e^{-x}, \end{aligned} \quad (87)$$

$$\|C_x - D_x\| < 2(1 + \tfrac{1}{8} e^{-3x}) x^{-1/2} e^{-x/2}. \quad (88)$$

To find a bound for $\|D_x - E_x\|$ the coefficients $d_x(l)$ and $e_x(l)$ have to be considered for $l \leqslant 2x/3$. Definitions (50) and (51) imply

$$d_x(0) - e_x(0) = 0. \quad (89)$$

For $l \geqslant 1$ inequality (62) and, if $x > 3/4$, $(l+1)/2x < 1$, also the inequalities

$$z < 1, \; l \geqslant 0: \quad -(1-z)^l \leqslant lz - 1 \quad \text{(Bernoulli)}, \quad (90)$$

$$0 \leqslant z < 1, \; l \geqslant 0: \quad -[e^z(1-z)]^l \leqslant -\left(1 - \frac{z^2}{2} - \frac{z^3}{2}\right)^l, \quad (91)$$

can be used to derive the relations

$x>\frac{3}{4}$:

$$
\begin{aligned}
0&<e_x(l)-d_x(l)\\
&=\left(\frac{2l+1}{x}\right)^{1/2}e^{-l(l+1)/2x}\\
&\quad\times\left\{1-\left[e^{(l+1)/2x}\left(1-\frac{l+1}{2x}\right)\right]^l\right\}\\
&<\left(\frac{2l+1}{x}\right)^{1/2}e^{-l(l+1)/2x}\left\{1-\left[1-\frac{1}{2}\left(\frac{l+1}{2x}\right)^2\right.\right.\\
&\quad\left.\left.-\frac{1}{2}\left(\frac{l+1}{2x}\right)^3\right]^l\right\}\\
&<\left(\frac{2l+1}{x}\right)^{1/2}\frac{l}{2}\left(\frac{l+1}{2x}\right)^2\left[1+\left(\frac{l+1}{2x}\right)\right]e^{-l(l+1)/2x}.
\end{aligned}
\tag{92}
$$

Since $l+1\leqslant 2l$, $2l+1\leqslant 3l$ for $l\geqslant 1$ (92) may be weakened to

$x>\frac{3}{4}$:

$$e_x(l)-d_x(l)<\frac{3^{1/2}}{2x^{5/2}}\,l^{7/2}\left(1+\frac{l}{x}\right)e^{-l^2/2x}. \tag{93}$$

Use of (89), (92), (93), and (77) with $z=l^2$, yields

$$|d_x(l)-e_x(l)|<1{,}35(1+2x^{-1/2})x^{-3/4}, \tag{94}$$

holding true also for $l=0$. (94) is used in the derivation of (8). To calculate the contribution of $\|D_x-E_x\|$ to the bound (14) Eqs. (89) and (92) may be used to derive the estimate

$x>\frac{3}{4}$:

$$
\begin{aligned}
&\|D_x-E_x\|^2\\
&\quad<\frac{1}{64x^5}\sum_{l\geqslant 0}(2l+1)l^2(l+1)^4\left(1+\frac{l+1}{2x}\right)e^{-l^2/x}\\
&\quad<\tfrac{3}{32}(1+\tfrac{2}{3}x^{-1/2})^{11}x^{-1}.
\end{aligned}
\tag{95}
$$

Here too the sums have been evaluated according to (28) and the resulting polynomial has been replaced by a simpler one. (95) implies

$$\|D_x-E_x\|<(\tfrac{3}{32})^{1/2}(1+\tfrac{2}{3}x^{-1/2})^{11/2}x^{-1/2}\quad\text{for } x>\tfrac{3}{4}. \tag{96}$$

The method described in Sec. 2B is also used to find a bound for

$$\|E_x-F_x\|^2=\frac{1}{x}\sum_{l>l''}(2l+1)e^{-l(l+1)/x}. \tag{97}$$

The lower bound of the integral, $l''-\frac{1}{2}$, is replaced by $(2x/3)-\frac{1}{2}$ and the derivatives are bounded by $\max g'(\lambda)$, $g(\lambda)=(2\lambda+1)\exp(-\lambda[\lambda+1]/x)$. Furthermore, x is chosen to be so large that $g'(\lambda)=0$ for some $\lambda>0$.

$x>\frac{9}{8}$:

$$
\begin{aligned}
\|E_x-F_x\|^2&<e^{-(4x/9)+(1/4x)}+\frac{3}{2x}e^{-3x/2+1/4}\\
&\leqslant e^{1/4}\left(1+\frac{3}{2x}e^{-x}\right)e^{-4x/9},
\end{aligned}
\tag{98}
$$

$$\|E_x-F_x\|<e^{1/8}\left(1+\frac{3}{4x}e^{-x}\right)e^{-2x/9}\quad\text{for } x>\tfrac{9}{8}. \tag{99}$$

The final step is the derivation of a bound for

$$\|F_x-Z_x\|=(1-e^{-1/8x})\|F_x\|. \tag{100}$$

This is reducible to the following estimate obtained in a similar way as (98).

$x>\frac{9}{8}$:

$$
\begin{aligned}
\|F_x\|^2&=\frac{1}{x}+\frac{1}{x}\sum_{l\geqslant 1}(2l+1)e^{-l(l+1)/x}\\
&<\frac{1}{x}+1+\frac{1}{8x}\left(2-\frac{1}{x}\right)\\
&\quad+\frac{1}{x}e^{-3x/2+1/4}<1+\frac{2}{x},
\end{aligned}
\tag{101}
$$

$$\|F_x-Z_x\|<(1+1/x)(1-e^{-1/8x})\quad\text{for } x>\tfrac{9}{8}. \tag{102}$$

The bound (8) follows from (67), (78), (88), (94), (99), and (101), by multiplying all bounds by $x^{1/2}$ and magnifying all functions of x to multiples of $x^{-3/4}$ (which is of course only possible for $x\geqslant x_0=2$). If a function does not decrease monotonically already for $x\geqslant 2$, relation (77) is used to minimize the error if this function is replaced by a multiple of $x^{-1/4}$. Estimate (14) is a straight consequence of (67), (81), (88), (96), (99), and (102), if all functions are magnified to multiples of $x^{-1/2}$.

3. AN APPLICATION IN SCATTERING THEORY

Nonrelativistic time-dependent scattering theory deals with square-integrable solutions of the Schrödinger equation. The behavior of such a solution (in the following called wave packet) is well controlled for $t\to\pm\infty$ if the potential V belongs to a certain class of functions[11] [e.g., $V\in L^1(\mathbb{R}^3)\cap L^2(\mathbb{R}^3)$]. In the following it is also assumed that the potential is repulsive ($V\geqslant 0$) so that the spectrum of the Hamiltonian is purely continuous. (This assumption is only made to reduce the mathematical amount; it is not a necessary precondition for applying the results of Sec. 1). Including also two-dimensional problems ($n=2$) the wave packet is then given by ($\hbar=m=1$)

$$\psi_t(\mathbf{x})=\int_{\mathbf{R}^n}d^nk\,\tilde{\psi}_0(\mathbf{k})e^{-i(k^2/2)t}f(\mathbf{x},\mathbf{k}), \tag{103}$$

where $\tilde{\psi}_0\in L^2(R^n)$ and the generalized eigenfunction

$$f(\mathbf{x},\mathbf{k})=(2\pi)^{-n/2}e^{i\mathbf{k}\cdot\mathbf{x}}+g(\mathbf{x},\mathbf{k}) \tag{104}$$

is the solution of the time-independent Schrödinger equation studied in time-independent scattering theory. The scattering wave g is assumed to have the asymptotic form

$$g(\mathbf{x},\mathbf{k})\to x^{(1-n)/2}e^{ikx}g_k(\boldsymbol{\xi},\boldsymbol{\kappa})\quad\text{for } x\to\infty, \tag{105}$$

where the convention $\mathbf{a}=a\boldsymbol{\alpha}$, $a\geqslant 0$, $\boldsymbol{\alpha}^2=1$ had been adopted for $\mathbf{a}=\mathbf{k},\mathbf{x}$. Equation (105) entails that for $t\to-\infty$ the wave packet ψ_t coincides with a free wave packet φ_t obtained from (103) and (104) by putting $g=0$. The function $\tilde{\psi}_0$ appearing in (103) is then equal to $\tilde{\varphi}_0$, the Fourier transform of φ_0. The behavior of the scattered wave packet ψ_t for $t\to\infty$ depends on $\tilde{\psi}_0$ and the asymptotic form of the generalized eigenfunction, especially on the scattering amplitude $g_k(\boldsymbol{\xi},\boldsymbol{\kappa})$. If the potential is rotationally invariant, g_k has the following form:

$n=2$:

$$g_k(\boldsymbol{\xi},\boldsymbol{\kappa})=(2\pi^3)^{-1/2}k^{-1/2}e^{i(\pi/4)}\sum_m\sin\delta_m(k)e^{i\delta_m(k)}e^{im(\xi-\kappa)}, \tag{106}$$

$\boldsymbol{\xi} = (\cos\xi, \sin\xi), \quad \boldsymbol{\kappa} = (\cos\kappa, \sin\kappa)$

$n = 3$:

$$g_k(\boldsymbol{\xi},\boldsymbol{\kappa}) = (2\pi)^{-3/2}k^{-1}\sum_{l\geqslant 0}\sin\delta_l(k)e^{i\delta_l(k)} \times (2l+1)P_l(\boldsymbol{\xi}\cdot\boldsymbol{\kappa}). \tag{107}$$

The phase shifts $\delta_p(k)$, $p = m$ or l, depend on the details of V and can be calculated explicitly for a number of examples. They can be used to define in $L^2(\mathbb{R}^n)$ a unitary operator $\tilde{S}$ (unitarily equivalent to the usual scattering operator) via

$$n = 2: \quad \tilde{S}\tilde{\psi}_0(\mathbf{k}) = \sum_m \tilde{\psi}_{0m}(k)e^{im\kappa + 2i\delta_m(k)}$$

$$\tilde{\psi}_{0m}(k) = (2\pi)^{-1}\int d\kappa\tilde{\psi}_0(\mathbf{k})e^{-im\kappa}, \tag{108}$$

$$n = 3: \quad \tilde{S}\tilde{\psi}_0(\mathbf{k}) = \sum_{lm} \tilde{\psi}_{0lm}(k)Y_{lm}(\boldsymbol{\kappa})e^{2i\delta_l(k)}$$

$$\tilde{\psi}_{0lm}(k) = (4\pi)^{-1}\int d\boldsymbol{\kappa}\tilde{\psi}_0(\mathbf{k})Y^*_{lm}(\boldsymbol{\kappa}), \tag{109}$$

the integrations running over the whole range of $\boldsymbol{\kappa}$. $\tilde{S}$ proves to be useful in calculating the probability of finding, in the limit $t\to\infty$, the particle within the cone

$$C = C(\mathbf{n},c) = \{\mathbf{x}:\mathbf{x}\cdot\mathbf{n}\geqslant cx\}, \tag{110}$$

characterized by $\mathbf{n}$ and c ($\mathbf{n}^2 = 1 \geqslant c$). For the scattering-into-cones theorem of Dollard[12] states

$$\lim_{t\to\infty}\int_C d^nx\,|\psi_t(\mathbf{x})|^2 = \int_C d^nk\,|\tilde{S}\tilde{\psi}_0(\mathbf{k})|^2. \tag{111}$$

Equation (111) may be simplified if the initial state (i.e., the $t\to\infty$ limit) is of such a nature that $\tilde{\psi}_0$ may be factorized,

$$\tilde{\psi}_0(\mathbf{k}) = R_{k_0,\gamma}(k)A(\boldsymbol{\kappa}), \tag{112}$$

and if the radial function $R_{k_0,\gamma}$ has a pronounced peak of width $(2\gamma)^{-1/2}$ at $k = k_0$. The physical meaning of

$$2\gamma k_0^2 \gg 1 \tag{113}$$

is that the mean wavelength ($\propto k_0^{-1}$) is assumed to be much smaller than the minimal extension of the wave packet in ordinary space ($\propto \sqrt{2\gamma}$). For a free particle ($V = 0$, $\tilde{S} = 1$-operator) the separation (112) carries over to $|\tilde{\psi}_0|^2 = |\tilde{S}\tilde{\psi}_0|^2$. If $\|\psi_t\| = \|\tilde{\psi}_0\| = 1$ the angular function may be chosen to be normalized on the n-sphere,

$$\int d\boldsymbol{\kappa}\,|A(\boldsymbol{\kappa})|^2 = 1, \tag{114}$$

and Eq. (111) becomes

$$\lim_{t\to\infty}\int_C d^nx\,|\psi_t(\mathbf{x})|^2 = \int_{\boldsymbol{\kappa}\cdot\mathbf{n}\geqslant c} d\boldsymbol{\kappa}\,\rho(\boldsymbol{\kappa}), \quad \rho(\boldsymbol{\kappa}) = |A(\boldsymbol{\kappa})|^2 \tag{115}$$

since the integration over k may be performed. In case of scattering an exact factorization of $\tilde{S}\tilde{\psi}_0$ is impossible since the phase shifts $\delta_p(k)$ depend both on p and k. However, condition (113) ensures the approximate separation

$$\tilde{S}\tilde{\psi}_0(\mathbf{k}) \simeq R_{k_0,\gamma}(k)\bar{A}(\boldsymbol{\kappa}), \tag{116}$$

$$n = 2: \quad \bar{A}(\boldsymbol{\kappa}) = \sum_m A_m e^{im\kappa + 2i\delta_m(k_0)},$$

$$A_m = (2\pi)^{-1}\int d\kappa\,A(\boldsymbol{\kappa})e^{-im\kappa}, \tag{117}$$

$$n = 3: \quad \bar{A}(\boldsymbol{\kappa}) = \sum_{lm} A_{lm}Y_{lm}(\boldsymbol{\kappa})e^{2i\delta_l(k_0)},$$

$$A_{lm} = (4\pi)^{-1}\int d\boldsymbol{\kappa}\,A(\boldsymbol{\kappa})\,Y^*_{lm}(\boldsymbol{\kappa}), \tag{118}$$

to make sense since the main contribution to the rhs of Eq. (111) comes from the intersection of C with a spherical shell around $k = k_0$ whose thickness $(2\gamma)^{-1/2}$ is so small that the variation of the phase shifts may be ignored. Physically this corresponds to a situation where the delay time is negligible. In this case

$$\lim_{t\to\infty}\int_C d^nx\,|\psi_t(\mathbf{x})|^2 = \int_{\boldsymbol{\kappa}\cdot\mathbf{n}\geqslant c} d\boldsymbol{\kappa}\,\bar{\rho}(\boldsymbol{\kappa}), \quad \bar{\rho}(\boldsymbol{\kappa}) = |\bar{A}(\boldsymbol{\kappa})|^2. \tag{119}$$

For a normalized Gaussian wave packet

$$\tilde{\psi}_0(\mathbf{k}) = \left(\frac{2\gamma}{\pi}\right)^{n/4}e^{-\gamma(\mathbf{k}-\mathbf{k}_0)^2}. \tag{120}$$

Although this function cannot be factorized into a radial and an angular part it can be approximated (in the mean) by the separable function

$$\tilde{\psi}_0^N(\mathbf{k}) = \left(\frac{2\gamma}{\pi}\right)^{1/4}k^{(1-n)/2}e^{-\gamma(k-k_0)^2}A(\mathbf{k})$$

$$A(\mathbf{k}) = \left(\frac{2\gamma}{\pi}\right)^{(n-1)/4}k_0^{(n-1)/2}e^{2\gamma k_0^2(\boldsymbol{\kappa}\cdot\boldsymbol{\kappa}_0-1)}. \tag{121}$$

By techniques similar to those used in Sec. 2 it can be shown that

$$n = 2: \quad \|\tilde{\psi}_0 - \tilde{\psi}_0^N\| < 0{,}91(2\gamma k_0^2)^{-1/2}, \tag{122}$$

$$n = 3: \quad \|\tilde{\psi}_0 - \tilde{\psi}_0^N\| < 0{,}67(2\gamma k_0^2)^{-1/2}, \tag{123}$$

so that $\tilde{\psi}_0^N$ is indeed a good approximation of $\tilde{\psi}_0$ if (113) holds. The coefficients needed to calculate the final angular density $\bar{\rho}$ are [cf. Eqs. (117), (118), (121), and Eqs. (9), (10)]

$n = 2$:

$$A_m = (8\pi^3\gamma k_0^2)^{-1/4}\frac{I_m(2\gamma k_0^2)}{I_m^a(2\gamma k_0^2)}e^{-im\kappa_0}, \tag{124}$$

$n = 3$:

$$A_{lm} = \left(\frac{2\pi}{\gamma k_0^2}\right)^{1/2}\frac{I_{l+1/2}(2\gamma k_0^2)}{I_{l+1/2}^a(2\gamma k_0^2)}Y^*_{lm}(\boldsymbol{\kappa}_0). \tag{125}$$

In principle $\bar{\rho}$ could be calculated from Eqs. (124), (117), or (125), (118), and (119). Looking at Eq. (124) or (125) it seems to be difficult to estimate the number of terms that have to be taken into account in the series expansion of $\bar{A}$. But if (113) holds it is legitimate to replace the quotients in (124) and (125) by simple exponentials, thereby making the calculation of $\bar{A}$ simpler and more transparent. The final result

$$n = 2: \quad \bar{A}(\boldsymbol{\kappa}) \simeq (8\pi^3\gamma k_0^2)^{-1/4}\sum_m \exp(-m^2/4\gamma k_0^2 + 2i\delta_m(k_0) + im(\kappa - \kappa_0)), \tag{126}$$

$$n = 3: \quad \bar{A}(\boldsymbol{\kappa}) \simeq (8\pi\gamma k_0^2)^{-1/2}\sum_{l\geqslant 0}\exp[-(l+\tfrac{1}{2})^2/4\gamma k_0^2$$

$$+ 2i\delta_l(k_0)](2l+1)P_l(\boldsymbol{\kappa}\cdot\boldsymbol{\kappa}_0), \tag{127}$$

is also obtained if the original angular function A (Eq. (121)) determining the angular distribution in the far past,

$$\lim_{t\to-\infty}\int_C d^n x\,|\psi_t(\mathbf{x})|^2 = \int_{\boldsymbol{\kappa}\cdot\mathbf{n}\geqslant c} d^n k\,\rho(-\boldsymbol{\kappa}),$$

$$\rho(\boldsymbol{\kappa}) = |A(\boldsymbol{\kappa})|^2 \tag{128}$$

is replaced by functions proportional to (10) or (12). The bounds (13) and (14) allow to estimate the error made by this substitution which is seen to be a good approximation if (113) holds.

Equations (126), (127) turn out to be useful if the angular distribution of the scattered wave packet is discussed as a function of the ratio $\sqrt{2\gamma}/d$ where $\sqrt{2\gamma}$ is the diameter of the wave packet (at collision time $t=0$) and the impact parameter d characterizes the range of the potential ($\delta_p(k_0)\simeq 0$ for $p>d$). Examples of this kind will be discussed elsewhere. Moreover, if a large number of terms has to be considered (e.g., in the classical limit $k_0\to\infty$) and analytical approximations of the phase shifts are available it is easier to evaluate the sums (126), (127) by approximating them by integrals than to deal with sums containing the original coefficients (124), (125).

It should be noted that the considerations of this section are not restricted to quantum mechanics but apply also to classical wave theories (acoustics, optics) for which time-dependent scattering theories exist.[13] In these theories the frequency appearing in Eq. (103) is k ($c=1$) instead of $k^2/2$, ψ_t is a (complex) potential from which the (real) basic fields can be derived, and Eqs. (128) and (119) represent the asymptotic energy distributions for $|t|\to\infty$.

4. CONCLUSION

The essential results of this paper are the following:

(i) A simple approximation has been given for the ratio $I_\nu(x)/I_\nu^a(x)$ where $I_\nu(x)$ is a modified Bessel function with integral or half-integral index ν, $I_\nu^a(x)$ the leading term of its asymptotic series [Eqs. (2)], and x a large positive number. Error bounds independent of ν have been derived for this approximation and shown to be of order $x^{-1/4}$ [Eqs. (5)–(8)]. The approximation of the ratio may be transformed into an approximation for $I_\nu(x)$ the relative error being small for $2\nu^2 \ll x\ \ln x$.

(ii) Similar approximations and error bounds have been obtained for functions defined as series with coefficients proportional to $I_\nu(x)/I_\nu^a(x)$ [Eqs. (9)–(14)].

(iii) These approximations have been shown to simplify the calculation of angular distributions of scattered Gaussian wave packets.

The results mentioned in (i) pose some interesting mathematical questions. One is whether the order of the bounds found here is optimal (the larger numerical factors obtained for half-integral ν are probably due to the more indirect proof). The other, more fundamental one, is whether (and by what methods) the results obtained here can be extended to arbitrary real values of the index ν.

ACKNOWLEDGMENT

I thank M. Posch for critically reading a preliminary version of the manuscript.

[1]M. Abamowitz and I.A. Stegun (Eds.), *Handbook of Mathematical Functions*) (Dover, New York, 1965), pp. 374–8, 443–5.
[2]Ref. 1, p. 367, Eq. 9.6.34; p. 297, Eq. 7.1.1.
[3]Ref. 1, pp. 416–22, Table 9.8.
[4]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 307, Eq. 3.222.
[5]Ref. 4, p. 931, Eq. 8.254.
[6]Ref. 1, p. 316, Table 7.3.
[7]Ref. 1, p. 376, Eq. 9.6.18.
[8]Ref. 4, p. 307, Eq. 3.332.2.
[9]Ref. 1, p. 257, Eq. 6.1.38.
[10]Ref. 4, p. 92, Eq. 2.321.2.
[11]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, New Jersey, 1971).
[12]J.D. Dollard, Commun. Math. Phys. **12**, 193 (1969).
[13]C.H. Wilcox, *Scattering Theory for the d'Alembert Equation in Exterior Domains* (Springer, Berlin, 1975); P.D. Lax and R.S. Phillips, *Scattering Theory* (Academic, New York, 1967).

The Planck integral cannot be evaluated in terms of a finite series of elementary functions[a)]

Richard Pavelle
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts 02173

(Received 7 August 1979; accepted for publication 20 September 1979)

It is shown that the Planck integral cannot be evaluated in terms of a finite series of elementary functions. A relation is first established between the Planck and dilogarithm integrals. To prove nonintegrability the Risch decision procedure for elementary functions is then applied to the dilogarithm integral.

It is well known that the Planck integral in the wavelength domain takes the form

$$\int \frac{1}{\lambda^5(e^{1/\lambda}-1)}\,d\lambda, \tag{1}$$

where all physical parameters have been set to unity. A question which has been stated in the literature, yet apparently never settled, is whether this integral can be evaluated between two arbitrary points in finite terms[1,2] (in terms of a finite series of elementary functions). I shall now answer this question.

Transforming to the frequency domain by $\lambda \rightarrow x^{-1}$, Eq. (1) becomes

$$\int \frac{x^3}{e^x-1}\,dx. \tag{2}$$

For convenience I shall use an indirect approach whereby this integrand can be written as a power series and then integrated to give[3]

$$\int \frac{x^3}{e^x-1}\,dx = \int x^3 \sum_{n=1}^{\infty} e^{-nx}\,dx = -\sum_{n=1}^{\infty}\left(\frac{x^3}{n}+\frac{3x^2}{n^2}+\frac{6x}{n^3}+\frac{6}{n^4}\right)e^{-nx}. \tag{3}$$

The first series on the right-hand side can be summed and results in

$$\sum_{n=1}^{\infty}\frac{e^{-nx}}{n} = -\log(1-e^{-x}). \tag{4}$$

Notice that summing the second series on the right-hand side of Eq. (3) is equivalent to integrating $\log(1-e^{-x})$ over x. Similarly, summing the other series in Eq. (3) is equivalent to finding respective second and third integrals. Hence, since the four series on the right-hand side of Eq. (3) are independent, the nonintegrability in finite terms of Eq. (1) or (2) is equivalent to proving that $\log(1-e^{-x})$ is not integrable in finite terms. This may be seen as follows: We transform the integral with $1-e^{-x}\rightarrow x$ to

$$\int \log(1-e^{-x})\,dx = \int \frac{\log(x)}{1-x}\,dx. \tag{5}$$

This integral is known as Spence's function[4] (or Euler's dilogarithm), and it is known to experts in the field of integration that the function cannot be evaluated in finite terms. Since the proof does not seem to appear in the literature, I shall now demonstrate this fact. From Risch's algorithm one knows that if Eq. (5) is integrable in finite terms, then it must be representable in the form[5]

$$\int \frac{\log(x)}{1-x}\,dx = A2\,\log^2(x) + A1\,\log(x), \tag{6}$$

where $A1$ and $A2$ are rational functions of x. Differentiating Eq. (6) gives

$$\frac{\log(x)}{1-x} = A2_x\,\log^2(x) + \frac{2\,A2\,\log(x)}{x} + A1_x\,\log(x) + \frac{A1}{x}, \tag{7}$$

where the subscripts are partial derivatives. This leads immediately to $A2=$ constant. Then, collecting terms in $\log(x)$ and integrating with respect to x, it is seen that $A1(x)$ cannot be rational. Yet it must be rational if the integral exists. Hence, the contradiction tells us the Planck integral cannot be evaluated in terms of a finite series of elementary functions.

The Planck integral is simply one function in a wide class of functions in mathematical physics which contains logarithms and exponentials. It is clear that Risch's decision procedure is a very powerful tool for looking at the integrability in finite terms of functions in this class. One could perform a systematic study of other common functions to determine their integrability in finite terms. To this end it may be noted that the algorithm has been implemented on MACSYMA,[6] and this provides not only a rapid check on integrability in finite terms but, also, gives the closed form for the integral when it exists.

[a)]This work was sponsored by the Department of the Army. The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

[1]R.B. Johnson *et al.*, J. Opt. Soc. Am. **64**, 1445 (1974).
[2]R.D. Tippets *et al.*, Opt. Eng. **18**, 313 (1979).
[3]The second term in Eq. (3) arises as follows: Since $e^{x} > 1$,

$$(e^x-1)^{-1} = e^{-x}(1-e^{-x})^{-1} = e^{-x}\sum_{n=0}^{\infty} e^{-nx} = \sum_{n=1}^{\infty} e^{-nx}.$$

[4]M. Abramowitz *et al.*, *Handbook of Mathematical Functions* (U.S. Govt. Printing Office, Washington, D.C., 1964).
[5]R.H. Risch, Trans. Am. Soc. **139**, 167 (1969).
[6]MACSYMA is the symbolic manipulation computer system of the Laboratory for Computer Science, Massachusetts Institute of Technology.

 0022-2488/80/010014-01$01.00

Linearization stability and a globally singular gange of variables

Judith M. Arms
Department of Mathematics and Physics University of Utah Salt Lake City Utah 84112

(Received 26 February 1979; accepted 24 May 1979)

An example is given showing that global properties of a change of field variables can affect the validity of linear perturbation theory.

I. INTRODUCTION

Choice of variables can affect properties of the solution set of an equation. For example, the solution set for $x^2 - y^2 = 0$ is not a manifold at the origin. As one consequence of this, the equation is linearization unstable there: the linearized equation $2x\cdot\delta x - 2y\cdot\delta y = 0\cdot\delta x - 0\cdot\delta y = 0$ has arbitrary $(\delta x,\delta y)$ as a solution, but only perturbations such that $\delta y = \pm\delta x$ are actually tangent to the solution set of the nonlinear equation. The singular change of variables $y^2 \to y$ removes the singularity so that the equation, $x^2 - y = 0$, is linearization stable at the origin: $\{(\delta x,\delta y):2x\cdot\delta x - \delta y = 0\cdot\delta x - \delta y = 0\}$ is the tangent space to $\{(x,y):x^2 - y = 0\}$ so there are no longer any spurious linear perturbations.

A nonsingular change of variables should not affect such properties as linearization stability. However, it is easy to overlook the fact that a change of variables may appear nonsingular "locally" and have singularities "globally." This paper gives an example to show that linearization stability, which concerns the relationship between global solutions of the linear and nonlinear equations, can depend on global characteristics of a change of variables.

The example is a simple one: a real scalar field ϕ obeying the linear field equation

$$\Box\phi - \mu^2\phi = 0 \tag{1}$$

on a particular flat, two-dimensional, spatially compact background spacetime. Equation (1) is linear and therefore trivially linearization stable. Following Kuchar[1] this problem can be reformulated so as to appear nonlinear by use of embedding variables to parameterize the background. With the restriction that the embedded surfaces be spacelike, the problem is still linearization stable. However, a canonical transformation of the embedding variables suggested by Kuchar leads to a formulation of the problem that is linearization unstable. (The singularity in the transformation which causes this anomaly also occurs in the noncompact case. This paper considers only the case of spatial compactness because the linearization stability analysis is more delicate on noncompact manifolds.)

Thus linear perturbation results can depend more subtly on the choice of field variables than might first appear. In particular, global considerations play an important role. The example raises other questions. What other properties besides linearization stability are affected by global considerations? How does one recognize the correct set of variables for linearization and other purposes? Perhaps linearization stabilty can be used as one criterion for this choice in some situations. (However, as linearization instability is a type of bifurcation, it is probably not desirable to rule out all cases of such instability.)

II. THE EMBEDDING VARIABLES AND LINEARIZATION STABILITY

On the spacetime $S = \mathbb{R}\times U(1)$, let (T,X) be coordinates, T and X real numbers and $\exp(2\pi iX)\in U(1)$. A function on $U(1)$ may be represented by a periodic function of x with period 1. Take the metric to be

$$g_{\alpha\beta} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

in these coordinates. [Choosing any other constant diagonal matrix with signature $(-,+)$ would not materially affect what follows.]

This spacetime can be described by giving a time-dependent embedding

$$e : \mathbb{R}\times U(1) \to S : (t,x) \mapsto (T_t(x), X_t(x)),$$

where x is a real number so $\exp(2\pi ix)\in U(1)$. $T, X' = \partial X/\partial x$, and $\exp(2\pi iX)$ must be periodic with period 1, so integral changes in x must correspond to integral changes in $X(x)$. With π_ϕ, π_T, and π_x as the momenta for ϕ [in (1)], T, and X, respectively, the constraint and evolution equations for the field ϕ and the embedding are generated by the action

$$\int dt \int_{U(1)} dx\,(\pi_\phi\dot\phi + \pi_T\dot T + \pi_X\dot X - NH - N^1H_1),$$

where

$$H = X'\pi_T + T'\pi_X + \tfrac{1}{2}\pi_\phi^2 + \tfrac{1}{2}\phi'^2 + \mu^2(X'^2 - T'^2)\phi^2,$$

$$H_1 = T'\pi_T + X'\pi_X + \phi'\pi_\phi,$$

$N = (X'^2 - T'^2)^{-1/2}$ times the usual lapse function,

N^1 = usual shift vector

[thus $\dot e^\alpha = (X'^2 - T'^2)^{1/2}Nn^\alpha + N^1e^{\alpha\prime}$, where n^α = future pointing unit normal on the hypersurface t = time = constant], and dot and prime indicate differentiation with respect to t and x, respectively. (For discussion of notation and derivation of equations, see Kuchar[1] Sec. 2 and 3. His p_T, p_X, $\tilde N$, and $\tilde H$ are π_T, π_X, N, and H, respectively, in this work.)

The constraint equations on the initial data for the

 0022-2488/80/010015-04$01.00

fields are

$$\Phi = (H, H_1) = 0. \tag{2}$$

These are linearization stable, i.e., every solution to the linearized equations associated to (2) is tangent to a one parameter family of exact solutions to the nonlinear equations (2). (Linearization stability of the full set of field equations is equivalent to that of the constraint equations because the Cauchy problem is well posed.) The proof follows immediately from the implicit function theorem for Banach spaces if the linearized equations associated to (2) are in some appropriate sense elliptic and surjective (cf. discussion of linearization stability and its proof in Arms[2] or Fischer and Marsden[3]). A simple calculation (see Sec. IV) shows that these conditions are met whenever $X'^2 - T'^2 \neq 0$. The Hamiltonian formalism described above requires that the hypersurfaces t = constant be spacelike, so

$$X'^2 - T'^2 > 0. \tag{3}$$

Thus in the cases of interest, (2) is linearization stable.

III. THE CANONICAL TRANSFORMATION AND LINEARIZATION INSTABILITY

In order to obtain a Hamiltonian which is purely quadratic in the momenta, Kuchar suggests the following canonical transformation of the field variables for the case $S = \mathbb{R}\times\mathbb{R}$. This transformation,

$$L : (T, X, \pi_T, \pi_X) \rightarrow (\xi, \eta, \pi_\xi, \pi_\eta)$$

defined by

$$\xi = 2^{-1/2}\Big(T + \int^x \pi_X\Big), \quad \eta = 2^{-1/2}\Big(-T + \int^x \pi_X\Big), \tag{4a}$$

$$\pi_\xi = 2^{-1/2}(X' + \pi_T), \quad \pi_\eta = 2^{-1/2}(X' - \pi_T),$$

and with inverse given by

$$X = 2^{-1/2}\int^x (\pi_\xi + \pi_\eta), \quad T = 2^{-1/2}(\xi - \eta), \tag{4b}$$

$$\pi_X = 2^{-1/2}(\xi' + \eta'), \quad \pi_T = 2^{-1/2}(\pi_\xi - \pi_\eta),$$

can also be used in the spatially compact case. H and H_1 become

$$H = \tfrac{1}{2}\{\pi_\phi^2 + \pi_\xi^2 - \pi_\eta^2 + \phi'^2 + \xi'^2 - \eta'^2 + \mu^2\phi^2[(\pi_\xi + \pi_\eta)^2 - (\xi' - \eta')^2]\},$$

$$H_1 = \phi'\pi_\phi + \xi'\pi_\xi + \eta'\pi_\eta.$$

This transformation seems nonsingular for all reasonable purposes. It is well defined locally [on $U(1)$]. The kernels of L and L^{-1} contain $\{X = \text{constant}\}$ and $\{\xi + \eta = \text{constant}\}$, respectively, but this seems unimportant because only X', ξ', and η' appear in the field equations. However, a more serious problem arises because perturbation theory concerns global functions. If ξ and η are single valued, they must be periodic and so $\int_0^1 dx\, \pi_X$ must be zero. Even if ξ and η are coordinates on $U(1)$ (like x and X),

$$2^{-1/2}\int_0^1 dx\, \pi_X = \xi(1) + \eta(1) - \xi(0) - \eta(0)$$

must be an integer in order that $\exp(2\pi i\xi)$ and $\exp(2\pi i\eta)$ be single valued. Thus L will be undefined for many functions π_X. Similarly L^{-1} is underfined unless $2^{-1/2}\int_0^1 dx(\pi_\xi + \pi_\eta)$ is an integer. [By adjusting the coordinates X, ξ, and η on $U(1)$, $\int_0^1 dx\, \pi_X$ and $\int_0^1 dx\, (\pi_\xi + \pi_\eta)$ may take on any value, but the range of possible values will be discrete once the coordinates are chosen.] If the integral of a function must assume discrete values, then small perturbations of that function must integrate to zero. In other words, a small perturbation must have been a zero constant term in its Fourier series. Thus considering the kernels and domains of definition, L is nonsingular only if the canonically conjugate pairs of variables (X, π_X) and $(\xi + \eta, \pi_\xi + \pi_\eta)$ are considered modulo constants.

The proof of linearization stability fails to carry over because the transformation is singular. The linearized constraint equations are still in a suitable sense elliptic, but are not surjective at some solutions. (See calculations in Sec. V.) In fact, it is fairly easy to determine all such solutions. It is shown below that surjectivity fails exactly when there is a vector field on S which is a symmetry for all the fields. No such vector field exists for the embedding variables, because the flow along any nonzero vector field is a change in position in S, i.e., in (T, X). But for ξ and η, it is easily observed from (4) that there is exactly one symmetry, the vector field $\partial/\partial X$. If ϕ has the same symmetry, then

$$\phi = C\cos\mu(T - T_0). \tag{5}$$

All symmetric solutions may be obtained by choosing a Cauchy surface in S and determining the initial data there for ϕ, ξ, η, and their momenta.

Thus the method of choice for proving linearization stability—the implicit function theory—fails. The proof of linearization instability is the construction of a spurious linear perturbation. At symmetric solutions, the second order perturbation of the constraint equations gives rise to a condition on the first order perturbations alone; Sec. V exhibits a linear perturbation violating this second order condition.

IV. PROOF OF LINEARIZATION STABILITY

With respect to the embedding variables, the linearized constraint equations

$$D\Phi(q_\phi, q_T, q_X, p_\phi, p_T, p_X) = 0$$

are

$$DH = q_X'\pi_T + X'p_T + q_T'\pi_X + T'p_X + \pi_\phi p_\phi + \phi' q_\phi' + 2\mu^2(X'^2 - T'^2)\phi q_\phi + 2\mu^2\phi^2(X'q_X' - T'q_T') = 0$$

$$DH_1 = q_T'\pi_T + T'p_T + q_X'\pi_X + X'p_X + q_\phi'\pi_\phi + \phi' p_\phi = 0,$$

where D stands for linearization (functional derivative) and q_ϕ, q_T, q_X, p_ϕ, p_T and p_X are perturbations of ϕ, T, X, π_ϕ, π_T, and π_X, respectively. The adjoint operator, $D\Phi^*$, defined by the equation

$$\int_{U(1)} dx\, D\Phi(q_\phi, q_T, q_X, p_\phi, p_T, p_X)\cdot(N, N^1) = \int_{U(1)} dx\, (q_\phi, q_T, q_X, p_\phi, p_T, p_X)\cdot D\Phi^*(N, N^1), \tag{6}$$

is given by

$$D\Phi^*(N,N^1)$$
$$= \{ -(N\phi')' + 2\mu^2\phi(X'^2 - T'^2)N - (N^1\pi_\phi)',$$
$$- [N(\pi_X + 2\mu^2\phi^2T')]' - (N^1\pi_T)',$$
$$- [N(\pi_T - 2\mu^2\phi^2X')]' - (N^1\pi_X)', N\pi_\phi + N^1\phi',$$
$$NX' + N^1T', NT' + N^1X'\}.$$

The weighting system devised by Douglis, Nirenberg, and Hörmander (cf. Hörmander[4]) allows the choice of the principal part of $D\Phi^*$ as the first order part of the first three components and the zeroeth order or algebraic part of the last three. With this weighting, $D\Phi^*$ is elliptic, i.e., has injective principal symbol, because the last two components of

$$\text{symbol } (D\Phi^*)(N,N^1) = 0$$

are

$$NX' + N^1T' = 0,$$
$$NT' + N^1X' = 0,$$

which, using (3), imply $(N,N^1) = (0,0)$.

This ellipticity implies the splitting

$$\text{Codomain } D\Phi = \ker D\Phi^* \oplus \text{Im } D\Phi, \tag{7}$$

so $D\Phi$ is surjective if and only if $D\Phi^*$ is injective. As the argument for injectivity of the symbol of $D\Phi^*$ depends on the algebraic components, the operator is injective by the same reasoning. Thus $D\Phi$ is surjective and by the implicit function theorem (2) is linearization stable.

V. PROOF OF LINEARIZATION INSTABILITY

With respect to the transformed variables, the linearized equations become

$$DH = \phi' q'_\phi + \mu^2\phi\,[(\pi_\xi + \pi_\eta)^2 - (\xi' - \eta')^2]q_\phi$$
$$+ [\xi' - \mu^2\phi^2(\xi' - \eta')]q'_\xi$$
$$+ [\mu^2\phi^2(\xi' - \eta') - \eta']q'_\eta + \pi_\phi p_\phi$$
$$+ [\pi_\xi + \mu^2\phi^2(\pi_\xi + \pi_\eta)]p_\xi$$
$$+ [\mu^2\phi^2(\pi_\xi + \pi_\eta) - \pi_\eta]p_\eta = 0 \tag{8a}$$

and

$$DH_1 = q'_\phi\pi_\phi + q'_\xi\pi_\xi$$
$$+ q'_\eta\pi_\eta + \phi' p_\phi + \xi' p_\xi + \eta' p_\eta = 0, \tag{8b}$$

where the notation is analogous to that of Sec. IV. The adjoint operator, defined as in (6), is

$$D\Phi^*(N,N^1)$$
$$= \{ -(N\phi')' + \mu^2\phi\,[(\pi_\xi + \pi_\eta)^2$$
$$- (\xi' - \eta')^2]N - (N^1\pi_\phi)', ([\mu^2\phi^2(\xi' - \eta')$$
$$- \xi']N)' - (N^1\pi_\xi)', ([\eta' - \mu^2\phi^2(\xi' - \eta')]N)'$$
$$- (N^1\pi_\eta)', \pi_\phi N + \phi' N^1,$$
$$[(1 + \mu^2\phi^2)\pi_\xi + \mu^2\phi^2\pi_\eta]N + \xi' N^1,$$
$$[\mu^2\phi^2\pi_\xi + (\mu^2\phi^2 - 1)\pi_\eta]N + \eta' N^1\}. \tag{9}$$

It is still elliptic. Again weighting so that the principal part is first order in the first three components and algebraic in the last three, the principal symbol is

$$\sigma(D\Phi^*)(N,N^1)$$
$$= \{ -\phi' N - \pi_\phi N^1, [\mu^2\phi^2(\xi' - \eta') - \xi']N$$
$$- \pi_\xi N^1, [\eta' - \mu^2\phi^2(\xi' - \eta')]N - \pi_\eta N^1,$$
$$\pi_\phi N + \phi' N^1, [(1 + \mu^2\phi^2)\pi_\xi + \mu^2\phi^2\pi_\eta]N$$
$$+ \xi' N^1, [\mu^2\phi^2\pi_\xi + (\mu^2\phi^2 - 1)\pi_\eta]N + \eta' N^1\}.$$

Suppose $(N,N^1)\not\equiv(0,0)$ is in $\ker\sigma(D\Phi^*)$. At any point on $U(1)$ where $N^1 = 0$ but $N\neq 0$, the last two components of $\sigma(D\Phi^*) = 0$ force

$$\pi_\xi = \pi_\xi = 0. \tag{10}$$

At any point where $N = 0$ but $N^1\neq 0$, the second and third components of $\sigma(D\Phi^*) = 0$ again lead to (10). Finally, if N and N^1 are both nonzero, solving the second, third, fifth, and sixth components of the equation for the ratio N/N^1, setting these four quantities equal to each other, and combining the equations appropriately yields

$$(\pi_\eta + \pi_\xi)^2 = (\eta' - \xi')^2. \tag{11}$$

But in terms of the transformed variables, (3) becomes

$$(\pi_\xi + \pi_\eta)^2 > (\eta' - \xi')^2,$$

which contradicts (10) and (11). The contradiction proves $\sigma(D\Phi^*)$ is injective so $D\Phi^*$ is elliptic. Thus the splitting (7) holds in this case also.

However, $D\Phi^*$ itself is not always injective (and so $D\Phi$ is not always surjective). For example, if ϕ, ξ, η, and π_ϕ are constant, then $H_1 = 0$ and the equation $H = 0$ has a solution with π_ξ and π_η constant. Then $(N,N^1) = (0,1)\in\ker D\Phi^*$, as is easily seen from (9).

In fact, up to choice of initial embedding this is the only case in which $D\Phi^*$ is not injective. This follows because elements of $\ker D\Phi^*$ may be identified with simultaneous symmetries of all the fields (and, as discussed in Sec. III, there is essentially only one instance of such a symmetry). From the action

$$\int dt \int_{U(1)} dx(\pi_\phi\dot\phi + \pi_\xi\dot\xi + \pi_\eta\dot\eta - NH - N^1H_1)$$

the equations of motions are

$$\frac{\partial}{\partial t}(\phi, \xi, \eta, \pi_\phi, \pi_\xi, \pi_\eta) = J\circ D\Phi^*(N,N^1), \tag{12}$$

where

$$J = \begin{pmatrix} 0 & Id \\ -Id & 0 \end{pmatrix}.$$

[This form of the evolution equations follows directly from varying the action with respect to the dynamic variables and generalizes Hamilton's equations:

$$\frac{\partial}{\partial t}(q,p) = J\circ DH^*(1).$$

For a more detailed discussion of this form for evolution equations, cf. Fischer and Marsden.[5]] Equation (12) gives the change in the variables along the flow of a timelike vector field

$$\frac{\partial}{\partial t} = \tfrac{1}{2}[(\pi_\xi + \pi_\eta)^2 - (\xi' - \eta')^2]^{1/2}N\eta^\alpha + N^1e^{\alpha\prime} \tag{13}$$

on S. By linearity, (12) defines the change along any vector field given by (13), timelike or not. From (12) it is clear that a simultaneous symmetry for all the fields must give rise, by the decomposition (13), to $(N,N^1)\in\ker D\Phi^*$. Conversely, suppose $(N,N^1)\in\ker D\Phi^*$. Define $Z^\alpha = \partial/\partial t$ as in (13) on the Cauchy surface. No matter how Z^α is extended off the surface, by (12)

$$\frac{\partial \xi}{\partial \lambda} = \frac{\partial \eta}{\partial \lambda} = \frac{\partial \pi_\xi}{\partial \lambda} = \frac{\partial \pi_\eta}{\partial \lambda} = 0 \tag{14}$$

holds on the surface, where λ = parameter for the flow along Z^α. From (14) and (4b),

$$Z^\alpha = (\text{const})\cdot \frac{\partial}{\partial X} \tag{15}$$

on the surface. Extend Z^α to S by (15); then (14) will hold on S, so Z^α will be a symmetry for ξ and η. Now consider $\partial\phi/\partial\lambda = Z^\alpha\phi_{,\alpha}$, which satisfies the scalar wave equation (1) (using the fact that Z^α is a Killing vector). From (12), $\partial\phi/\partial\lambda$ and its normal derivative on the Cauchy surface are zero, so by standard arguments, $\partial\phi/\partial\lambda = 0$ on S. Thus $(N,N^1)\in\ker D\Phi^*$ gives rise to a simultaneous symmetry of all fields.

At a solution with symmetry and at the most convenient embedding, T = const and $X = x$, the symmetry (N,N^1) is (0,1). Consider the second order perturbation of the constraint equations (2):

$$D^2\Phi\left((q_\phi,q_\xi,q_\eta,p_\phi,p_\xi,p_\eta),(q_\phi,q_\xi,q_\eta,p_\phi,p_\xi,p_\eta)\right) + D\Phi(Q_\phi,Q_\xi,Q_\eta,P_\phi,P_\xi,P_\eta) = 0, \tag{16}$$

where Q_ϕ, Q_ξ, Q_η, P_ϕ, P_ξ, and P_η represent the quadratic perturbations of the fields. Integrating the inner product of (16) and (0,1) over $U(1)$ yields

$$2\int_{U(1)} dx\,(q'_\phi\, p_\phi + q'_\xi\, p_\xi + q'_\eta\, p_\eta) = 0, \tag{17}$$

where the second term in (16) has dropped out by integration by parts because $(0,1)\in\ker D\Phi^*$.

To prove linearization instability it suffices to exhibit a linear perturbation $(q_\phi,q_\xi,q_\eta,p_\phi,p_\xi,p_\eta)$ satisfying the linearized constraints (8) and violating the second order condition (17). At the chosen embedding the symmetric solution is given by constant ϕ, ξ, η, and π_ϕ, with $\pi_\xi + \pi_\eta = 2^{1/2}$ [from (4b)]. Solving the constraint $H = 0$ gives the equations

$$\pi_\xi = 2^{-1/2}(1 - \mu^2\phi^2 - \tfrac{1}{2}\pi_\phi^2)$$

and

$$\pi_\eta = 2^{-1/2}(1 + \mu^2\phi^2 + \tfrac{1}{2}\pi_\phi^2).$$

Equations (8) reduce to

$$2\mu^2\phi q_\phi + \pi_\phi\, p_\phi + 2^{-1/2}(\mu^2\phi^2 - \tfrac{1}{2}\pi_\phi^2 + 1)\,p_\xi + 2^{-1/2}(\mu^2\phi^2 - \tfrac{1}{2}\pi_\phi - 1)\,p_\eta = 0 \tag{18}$$

and

$$\pi_\phi q'_\phi + \pi_\xi\, q'_\xi + \pi_\eta q'_\eta = 0. \tag{19}$$

The desired perturbation is the following. Take any nonzero function $f(x)$ on $U(1)$ such that $\int_{U(1)} dx\, f(x) = 0$. If $\pi_\phi = 0$, let $q'_\phi = p_\phi = f(x)$, $q_\xi = q_\eta = 0$, and choose p_ξ and p_η so that (18) is satisfied. If $\pi_\phi \neq 0$ but π_ξ (or π_η) = 0, let $q'_\xi = p_\xi = f(x)$ [or $q'_\eta = p_\eta = f(x)$] and solve (18) for p_ϕ, letting all other perturbations be zero. In both these cases the integral in (17) becomes

$$2\int_{U(1)} dx\, f^2(x) \neq 0.$$

If none of the momenta are zero, suppose $\mu^2\phi^2 - \tfrac{1}{2}\pi_\phi^2 + 1 \neq 0$. (Otherwise $\mu^2\phi^2 - \tfrac{1}{2}\pi_\phi^2 - 1 \neq 0$, and the roles of ξ and η and their associated momenta and perturbations should be interchanged in the following.) Let $q'_\phi = p_\xi = f(x)$ and $q_\xi = p_\eta = 0$, and solve (18) for p_ϕ and (19) for q_η. The integral in (17) becomes

$$\begin{aligned} 2\int_{U(1)} dx\, q'_\phi\, p_\phi &= -\frac{2}{\pi_\phi}\int_{U(1)} dx\, q'_\phi\left[2\mu^2\phi q_\phi + 2^{-1/2}(\mu^2\phi^2 - \tfrac{1}{2}\pi_\phi^2 + 1)\,p_\xi\right] \\ &= -\frac{2^{1/2}}{\pi_\phi}\int_{U(1)} dx\,(\mu^2\phi^2 - \tfrac{1}{2}\pi_\phi^2 + 1) f^2(x) \neq 0. \end{aligned}$$

Thus (17) is violated, which proves linearization instability.

ACKNOWLEDGMENTS

I thank Professor Kuchar for suggesting this problem and for many helpful conversations. This research was partially supported by NSF grants PHY78-05368 and MCS78-01460.

[1]K. Kuchar, J. Math. Phys. **19**, 390–400 (1978).
[2]J.M. Arms, J. Math. Phys. **18**, 830–33 (1977).
[3]A.E. Fischer and J.E. Marsden, Duke Math J. **42**, 519–47 (1975).
[4]L.Hörmander, Ann. Math. **83**, 129–209 (1966).
[5]A.E. Fischer and J.E. Marsden, "Topics in the Dynamics of General Relativity," Sec. 2, in *Proceedings of the Summer School of the Italian Physical Society* (1976), edited by J. Ehlers (to appear).

Conservation laws of the Benjamin–Ono equation

H. H. Chen and D. J. Kaup[a)]

Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 6 April 1979; accepted for publication 14 September 1979)

We report here an empirical algorithm to construct conservation laws of the Benjamin–Ono equation.

I. INTRODUCTION

Recently,[1] one of us (HHC) in collaborating with Lee and Pereira have found multisoliton solutions of the so-called Benjamin–Ono equation[2,3]

$$q_t + 2qq_x + Hq_{xx} = 0, \tag{1}$$

where H is the Hilbert transform operator defined by

$$Hq(x) = \frac{P}{\pi}\int \frac{q(z)}{z-x}\,dz. \tag{2}$$

These multisolitons are represented by N-pairs of poles moving in the complex x plane. The dynamics of these poles are found to be exactly the well known Integrable Calogero–Moser–Sutherland N-body problem.[4] Therefore, a complete description of the N-soliton motion is known. However, N-soliton solutions are only special solutions of the Benjamin–Ono equation. There is in general, a nonsoliton part which behaves like a linear wave packet. It disperses and spreads. To our knowledge, there is yet no way to describe the time evolution of this nonsoliton wave packet. On the other hand, it is obvious that conservation laws provide information about the time-evolution of the nonsoliton solution. The more number of conservation laws we know, the more knowledge we have in predicting the evolution of the solution. As a matter of fact, for an integrable nonlinear system like the Korteweg–de Vries equation,[5] it is believed that conservation laws are closely connected to the inverse scattering scheme that solves the equation.[6] Therefore, we shall study in this paper the conservation laws of the Benjamin–Ono equation.

II. EXISTING CONSERVATION LAWS

Ono[3] seems to be the first one to study the conservation laws of the Benjamin–Ono equation. He presented four conservation laws

$$C_1 = \int_{-\infty}^{\infty} q\,dx, \tag{3}$$

$$C_2 = \int_{-\infty}^{\infty} q^2\,dx, \tag{4}$$

$$C_3 = \tfrac{4}{3}\int_{-\infty}^{\infty} (q^3 + \tfrac{3}{2}qHq_x)\,dx, \tag{5}$$

and

$$B_1 = \frac{d}{dt}\int_{-\infty}^{\infty} xq\,dx. \tag{6}$$

In fact, the last one, B_1, is not an independent conserved quantity. It is equivalent to C_2. Therefore, Ono has indeed found only three conservation laws, C_1, C_2, and C_3. Meiss and Pereiria[7] subsequently found two more conservation laws,

$$C_4 = 2\int_{-\infty}^{\infty} (q^4 + 3q^2Hq_x + 2q_x^2)\,dx \tag{7}$$

and

$$C_5 = \tfrac{16}{5}\int_{-\infty}^{\infty} [q^5 + \tfrac{10}{3}q^3Hq_x + \tfrac{5}{2}q^2H(qq_x) + \tfrac{5}{2}q(Hq_x)^2 + \tfrac{15}{2}qq_x^2 + \tfrac{5}{2}q_xHq_{xx}]\,dx. \tag{8}$$

In fact, it is these two conservation laws that inspired the discovery of multisoliton solutions.

III. AN EMPIRICAL ALGORITHM

The conservation laws of the Benjamin–Ono equation are in fact global laws, in contrast to the Korteweg–de Vries conservation laws that are local. In other words, if we now define

$$C_n \equiv \int_{-\infty}^{\infty} \sigma_n\,dx, \tag{9}$$

then

$$\sigma_{n,t} = (\beta_{n+1})_x + \delta_{n+1} \quad \text{for} \quad n = 1,2,3,4, \text{ and } 5. \tag{10}$$

Here δ_{n+1} represents the collection of those terms that cannot be written as total derivatives. We demand also that δ_{n+1} be of the form

$$\delta_{n+1} = \sum_i f_iHg_i + g_iHf_i, \tag{11}$$

in order to uniquely define δ_{n+1}. Note that Eq. (11) guarantees that $\int_{-\infty}^{\infty} \delta_{n+1}\,dx = 0$ due to the integral theorem

$$\int_{-\infty}^{\infty} fHg\,dx = -\int_{-\infty}^{\infty} gHf\,dx. \tag{12}$$

Otherwise, σ_n would not be conservation laws. For example, we have

$$\sigma_{2,t} = -\sigma_{3,x} + \delta_3 \tag{13}$$

with

$$\sigma_2 = q^2 + Hq_x,$$
$$\sigma_3 = \tfrac{4}{3}(q^3 + \tfrac{3}{2}qHq_x + \tfrac{3}{4}H\sigma_{2,x}), \tag{14}$$

[a)]Permanent address: Department of Physics, Clarkson Institute of Technology, Potsdam, New York 13676.

and

$$\delta_3 = 2q_x Hq_x.$$

Obviously, δ_3 is not a total derivative. But, according to Eq. (12) we see immediately that

$$\int_{-\infty}^{\infty} q_x Hq_x \, dx = - \int_{-\infty}^{\infty} q_x Hq_x \, dx = 0. \tag{15}$$

As a matter of fact, for the five known conservation laws, we have

$$\beta_n \equiv -\sigma_n, \quad n = 2,3,4,5, \tag{16}$$

or the empirical algorithm

$$\sigma_{n,t} = -\sigma_{n+1,x} + \delta_{n+1} \quad \text{for } n = 1,2,3,4. \tag{17}$$

The quantities σ_n, δ_n can be written out explicitly:

$$\sigma_1 = q, \tag{18}$$

$$\sigma_2 = q^2 + \{H\sigma_{1,x}\}, \tag{19}$$

$$\sigma_3 = \tfrac{4}{3}(q^3 + \tfrac{3}{2}qHq_x) + \{H\sigma_{2,x}\}, \tag{20}$$

$$\sigma_4 = \tfrac{8}{4}[q^4 + 3q^2Hq_x + 2q_x^2] + 2[qH(q^2)_x - q^2Hq_x + \tfrac{1}{2}(Hq_x)^2 - qq_x - \tfrac{3}{2}q_x^2 + \tfrac{1}{2}H\sigma_{3,x}], \tag{21}$$

$$\begin{aligned}\sigma_5 = \tfrac{16}{5}[q^5 &+ \tfrac{10}{3}q^3Hq_x + \tfrac{5}{2}q^2H(qq_x) + \tfrac{5}{2}q(Hq_x)^2 \\ &+ \tfrac{15}{2}qq_x^2 + \tfrac{5}{2}q_xHq_{xx}] + \{-\tfrac{8}{3}(q^3Hq)_x \\ &- 4[q(Hq)(Hq_x)]_x - 8(q^2q_x)_x - 2(qHq_x)_{xx} \\ &+ 8[qH(q^2q_x) + q^2q_xHq] \\ &+ 4[qH(qHq_{xx}) + q(Hq_{xx})(Hq)] \\ &+ 4[(Hq_x)(Hqq_x) - qq_x^2] \\ &+ 4[(Hq)q_xHq_x + qH(q_xHq_x)]\},\end{aligned} \tag{22}$$

and

$$\delta_1 = \delta_2 = 0, \tag{23}$$

$$\delta_3 = 2q_xHq_x, \tag{24}$$

$$\delta_4 = 4qq_xHq - 4qH(qq_x), \tag{25}$$

$$\begin{aligned}\delta_5 &= 8q^2q_xHq_x + 8q_xH(q^2q_x) + 4q_x(Hq_x)^2 + 4q_xH(q_xHq_x) \\ &+ 4q_xH(qHq_{xx}) + 4(Hq_x)(qHq_{xx}) + 4q_{xx}Hq_{xx}.\end{aligned} \tag{26}$$

We do not yet have a proof of Eq. (17) for general n. But we shall be very surprised to find it wrong for higher n.

IV. A NEW CONSERVATION LAW

Indeed, we tried the case for $n = 5$ in (17) and we found a new conservation law,

$$\begin{aligned}C_6 = \tfrac{32}{6}\int_{-\infty}^{\infty} [q^6 &+ \tfrac{5}{12}q^4Hq_x + 5q^3H(qq_x) \\ &+ \tfrac{15}{2}q(Hq_x)(Hqq_x) + \tfrac{15}{4}q^2(Hq_x)^2 + \tfrac{75}{4}q^2q_x^2 \\ &+ 15qq_xHq_{xx} + \tfrac{15}{2}q_x^2Hq_x + 3q_{xx}{}^2]dx.\end{aligned} \tag{27}$$

We have not applied our empirical algorithm (17) to $n = 6$ or higher. The algebra becomes more and more messy. But we strongly believe that infinite number of conservation laws exist. These conservation laws should reduce to the conservation laws of the Calogero–Moser–Sutherland many-body problems when we specialized to the pure N-soliton solutions. Recent work[6] shows close connection between conservation laws and the inverse scattering scheme. We hope that the eventual verification of our empirical algorithm (17) would reveal this connection for the Benjamin–Ono equation also.

ACKNOWLEDGMENT

This research is sponsored by grants from NSF and ONR.

[1] H.H. Chen, N.R. Pereira, and Y.C. Lee, Phys. Fluids **22**, 187 (1979).
[2] T.B. Benjamin, J. Fluid Mech. **29**, 559 (1967); **26**, 241 (1966).
[3] H. Ono, J. Phys. Soc. Japan **39**, 1082 (1975).
[4] J. Moser, Adv. Math. **16**, 197 (1975).
[5] C. Gardner, J. Greene, M. Kruskal, and R. Miura, Commun. Pure Appl. Math. **27**, 97 (1974).
[6] H.H. Chen, Y.C. Lee, and C.S. Liu, Phys. Scr. **20**, 490 (1979).
[7] J.D. Meiss and N.R. Pereira, Phys. Fluids. **20**, 700 (1978).

Asymptotic regularizations as an alternative to distributions for the study of singular hypersurfaces

P. Gravel
Mathematiques, Universite de Montreal, Montreal, P.O., Canada H3C3J7

(Received 18 June 1979; accepted for publication 23 August 1979)

A technique designed to give shock wave surface and propagation equations of quasilinear differential systems is given which can be used when the distribution method does not lead to any practicable results. The technique, called "asymptotic regularizations", uses a smoothing process of the discontinuous functions which becomes negligible as a parameter ω is made arbitrarily large, thus revealing the behavior of the discontinuities on the singular hypersurface. This paper is also, in a way, the answer to a conjecture formulated by Lichnérowicz to the effect that there should be a theorem relating distribution theory to asymptotic expansions on manifolds (as defined by Choquet-Bruhat) which could explain the formal similarity of their respective results when they are applied to the differential systems of some relativistic fluids.

1. INTRODUCTION

Distributions in R^n have been generalized to distributions on Riemannian manifolds by Lichnerowicz (a summary of this method can be found in Ref. 1). This author has also shown how these distributions could be used to obtain equations for infinitesimal shock propagation and for the hypersurfaces on which they occur in the case of source-free Einstein–Maxwell (E–M) equations and those of relativistic magnetohydrodynamics (MHD) (cf. Refs. 2 and 3).

However this method fails to give infinitesimal shock propagation equations when we consider for example, a noninductive, heat-current free relativistic fluid.

This failure to obtain propagation equations is due to the fact that in distribution theory a continuous tensor T, defined on an open set Ω, which has discontinuous derivatives as it crosses a hypersurface Σ satisfies on Ω a distribution equation of the type

$$\delta[\nabla_\alpha \nabla_\beta T] = \nabla_\alpha l_\beta \bar{\delta} T + l_\alpha \nabla_\beta \bar{\delta} T + l_\beta \nabla_\alpha \bar{\delta} T + l_\alpha l_\beta \bar{T} \quad (1)$$

(cf. Ref. 1), in which $\bar{T}$ is a tensor distribution, of the same tensorial type as T, which is not given by distribution theory: It is only known to exist. "Propagation equations" given by distribution theory are thus of little interest if it cannot be shown that contributions from those $\bar{T}$'s amount to zero. Although this can be done for the fluids mentioned above (source-free E–M and relativistic MHD) it is not so, for example, for the noninductive, heat-current free relativistic fluids (to be considered in a later publication).

What we intend to show here is that there exists an alternative way to find propagation equations in which *no* function of indeterminate form appears.

On the other hand, the conjecture of A. Lichnérowicz to the effect that his distribution theory should be mathematically related in some way to asymptotic expansions (as defined and used by Y. Choquet–Bruhat in Ref. 4) is disproved and replaced by a mathematical correspondence with asymptotic regularizations, the former similarity being due in fact to the similitudes encountered in computations occurring both in asymptotic expansions and asymptotic regularizations. Asymptotic expansions do *not* always give results similar to those of distributions for *all* types of shocks and fluids, whereas asymptotic regularizations do.

2. REGULARIZATIONS OF THE METRIC TENSOR

Suppose we are given an open neighborhood Ω of a manifold V_l on which are defined the local chart (x^α) and the regular function ϕ of the coordinates of this chart. Assume also that the equation $\phi \equiv 0$ defines a regular hypersurface Σ which divides Ω into two nonempty open sets Ω_+ and Ω_-, on which $\phi > 0$ and $\phi < 0$, respectively. Then we have either one of the following possibilities:

A. Regular metric

By this we mean that the metric together with its first and second derivative are continuous on Ω.

(A) Consider first a tensor field ${}^0T(x)$ which is continuous on both Ω_+ and Ω_-, but regularly discontinuous at Σ (i.e., $\mathrm{Lim}_{\phi\to 0^+}\ {}^0T$ and $\mathrm{Lim}_{\phi\to 0^-}\ {}^0T$ are both tensor valued functions defined on Σ.

We shall denote the discontinuity field on Σ of a tensorial quantity A by

$$[A] \equiv \mathrm{Lim}_{\phi\to 0^-} A - \mathrm{Lim}_{\phi\to 0^+} A \equiv A_- - A_+,$$

where the $-$ and $+$ subscripts should have obvious meanings.

Let us construct a "regularized" field T in the following way. We shall have

$$T(x,\omega\phi) \equiv {}^0T(x) + {}^1T(x,\omega\phi), \quad (2)$$

where ${}^1T$ is chosen in such a way that T and its first derivatives are continuous on Ω. Of course, this implies that ${}^1T$ is itself discontinuous at Σ. Actually we must have

$$[{}^1T] \equiv - [{}^0T]$$

on Σ.

If we restrict ${}^1T$ to have its support in Ω_+, we then have

$${}^1T \equiv 0 \quad \text{if } \phi(x) \leqslant 0$$

 0022-2488/80/010021-03$01.00

and

$$^1T_+ \equiv \operatorname*{Lim}_{\phi\to 0^-} {}^1T = [{}^0T].$$

We shall also demand that given ω and x_0 on Σ, $^1T(x,\omega\phi)$ be absolutely integrable with respect to the variable $\xi \equiv \omega\phi$, these integrals being uniformly bounded on Σ. The scalar function

$$\|A\| \equiv \sup_{\substack{x_0\in\Sigma \\ \alpha,\cdots,\delta=0,\cdots,l}} \int_{I_{x_0}} |A^{\alpha\cdots\beta}_{\gamma\cdots\delta}|\, d\phi,$$

in which I_{x_0} is that integral path of the field $l_\alpha \equiv \partial\phi/\partial x^\alpha$ which crosses Σ at x_0, is then a norm when restricted to the set of tensors which are continuous on Ω_+ and Ω_- (expressions such as "almost everywhere" are of no use here). It follows that as ω goes to infinity, T converges to ^0T with respect to this norm. Moreover, if we take the covariant derivative of both sides of Eq. (2) we get, e.g.,

$$\begin{aligned}\nabla_\alpha T_{\beta\gamma} &= \nabla_\alpha {}^0T_{\beta\gamma} + \nabla_\alpha {}^1T_{\beta\gamma}\\ &= \nabla_\alpha {}^0T_{\beta\gamma} + \omega {}^1T'_{\beta\gamma} l_\alpha\\ &\quad + {}^1T_{\beta\gamma,\alpha} - {}^1T_{\lambda\gamma}\Gamma^\lambda_{\beta\alpha} - {}^1T_{\beta\lambda}\Gamma^\lambda_{\gamma\alpha},\end{aligned}$$

where $^1T' \equiv \partial\, {}^1T/\partial\xi$ and the $\Gamma^\alpha_{\beta\gamma}$'s are the connection coefficients.

Since $\nabla_\alpha T_{\beta\gamma}$ must be continuous, we get by taking the limit as ϕ goes to 0 from both sides of Σ and subtracting the two limits on Σ,

$$\begin{aligned}0 &= [\nabla_\alpha T_{\beta\gamma}]\\ &= [\nabla_\alpha {}^0T_{\beta\gamma}] + [\omega {}^1T'_{\beta\gamma} l_\alpha]\\ &\quad + [{}^1T_{\beta\gamma,\alpha} - {}^1T_{\lambda\gamma}\Gamma^\lambda_{\beta\alpha} - {}^1T_{\beta\lambda}\Gamma^\lambda_{\gamma\alpha}]\\ &= [\nabla_\alpha {}^0T_{\beta\gamma}] + \{\omega {}^1T'_{\beta\gamma+}\, l_\alpha\\ &\quad + {}^1T_{\beta\gamma,\alpha+} - {}^1T_{\lambda\gamma+}\Gamma^\lambda_{\beta\alpha} - {}^1T_{\beta\lambda+}\Gamma^\lambda_{\gamma\alpha}\},\end{aligned}$$

whence for a non-null hypersurface ($l_\alpha l^\alpha \neq 0$)

$$[\nabla_\alpha {}^0T_{\beta\gamma}] = l_\alpha\left(\frac{{}^1T_{\beta\gamma+,\lambda}\, l^\lambda}{l_\mu l^\mu} + \omega {}^1T'_{\beta\gamma+}\right) + {}^+\nabla_\alpha [{}^0T_{\beta\gamma}], \quad (3)$$

in which $^+\nabla$ denotes the covariant derivative defined on Σ by the restriction to this hypersurface of the metric and its connection.

We note now that

$$\delta[\nabla_\alpha {}^0T] = \nabla_\alpha(\delta[{}^0T]) + l_\alpha \bar\delta T, \quad (4)$$

the result given by distribution theory for that type of discontinuity (cf. Ref. 5) is quite similar to (3). We shall suppose from now on that $^1T_{\beta\gamma+,\lambda}\, l^\lambda = 0$ on Σ. (3) and (4) then have the same meaning when (4) is restricted to Σ. Moreover, an elementary development shows that $\omega {}^1T'_{\alpha\beta}$ constitutes a sequence, with respect to ω, of regular distributions defined on Ω which converges to a singular one with support on Σ.

(B) If we are given instead a ^0T continuous on Ω with discontinuous derivatives as it crosses Σ, then it is known from Hadamard that there exist quantities A, of the same tensorial type as T on Σ, such that

$$[{}^0T_{,\alpha}] = A l_\alpha.$$

Thus, if we construct the regularization

$$T(x,\omega\phi) = {}^0T(x) + \frac{1}{\omega}\, {}^2T(x,\omega\phi), \quad (5)$$

where ^2T has support in Ω_+ (the left superscripts 0,1,2, etc. referring to the order in ω^{-1} of the respective expressions once the norm has been computed) and require that T together with its first derivatives be continuous and absolutely integrable on Ω, it then follows that

$$\operatorname*{Lim}_{\phi\to 0^-} {}^2T' = A$$

and

$$\operatorname*{Lim}_{\phi\to 0^-} {}^2T_{,\alpha} = 0 \quad \text{on } \Sigma. \quad (6)$$

Hence, we get, since $\nabla_\alpha T_\beta$ is to be continuous,

$$\begin{aligned}0 &= [\nabla_\alpha T_\beta]\\ &= [\nabla_\alpha {}^0T_\beta] + \left[{}^2T'_\beta + \frac{{}^2T_{\beta,\rho} l^\rho}{l_\mu l^\mu}\right] l_\alpha\\ &\quad + \frac{1}{\omega}\left[{}^2T_{\beta,\alpha} - \frac{{}^2T_{\beta,\rho} l^\rho}{l_\mu l^\mu} - {}^2T_\lambda \Gamma^\lambda_{\beta\alpha}\right].\end{aligned}$$

If we assume, as in the previous case, that $^2T_{\beta,\rho} l^\rho$ vanishes, we have

$$\begin{aligned}[\nabla_\alpha {}^0T_\beta] &= -[{}^2T'_\beta] l_\alpha - \frac{1}{\omega}[{}^2T_{\beta,\alpha} - {}^2T_\lambda \Gamma^\lambda_{\beta\alpha}]\\ &= {}^2T'_{\beta+}\, l_\alpha\end{aligned}$$

from (6) and the continuity of $^2T_\lambda$ and $\Gamma^\alpha_{\beta\lambda}$ on Ω. This is again to be paralleled by

$$\delta[\nabla_\alpha {}^0T_\beta] = l_\alpha \bar\delta\, {}^0T_\beta$$

which is the corresponding result from distribution theory (we can go no further than just pointing out this similarity, since the form of $\bar\delta\, {}^0T$ is never exactly known, although it should be obvious that on Σ they are identical, since the two equations are Hadamard's condition for that type of discontinuity on Σ). It should be noticed that $\operatorname{Lim}_{\omega\to\infty} {}^2T'$ has support in Σ only.

It is more interesting to look at $[\nabla_\alpha\nabla_\beta {}^0T]$, supposing $\nabla_\alpha\nabla_\beta T$ is continuous on Ω and that $\operatorname{Lim}_{\phi\to 0^-} {}^2T'_{\mu,\nu} l^\nu$ vanishes. We get

$$\begin{aligned}[\nabla_\alpha\nabla_\beta {}^0T_\gamma] &= \omega\, {}^2T''_{\gamma+}\, l_\alpha l_\beta + l_\alpha {}^+\nabla_\beta {}^2T'_{\gamma+}\\ &\quad + l_\beta {}^+\nabla_\alpha {}^2T'_{\gamma+} + {}^2T'_{\gamma+}\, {}^+\nabla_\alpha l_\beta\\ &\quad + O(\omega^{-2}),\end{aligned}$$

where $O(\omega^{-2})$ means "terms of order 2 or greater in ω^{-1}," with respect to the norm defined above. This result should be compared with (1). Actually, they have the same meaning on Σ, provided we identify $\bar T_\beta$ and $(-\omega\, {}^2T''_\beta)$, and still associate $\bar\delta\, {}^0T$ with $^2T'$ as we did above. Again, it can be easily shown that the limit of the sequence of "regular distributions" $\{\omega\, {}^2T''_\beta\}$ as ω goes to infinity is an irregular, i.e., "true," distribution which can be considered a multiple of the usual Dirac δ, on each integral path I_{x_0} of the field l_α, with support $x_0 \in \Sigma$.

B. Metric with discontinuous derivatives

In this case, we have from Hadamard

$$[{}^0g_{\alpha\beta,\gamma}] = l_\gamma a_{\alpha\beta}. \tag{7}$$

Since g itself is continuous, we do as in (5) and construct

$$g_{\alpha\beta} = {}^0g_{\alpha\beta}(x) + \frac{1}{\omega}\,{}^2g_{\alpha\beta}(x,\omega\phi) \tag{8}$$

as asymptotic regularization to ${}^0g_{\alpha\beta}$, ${}^2g_{\alpha\beta}$ having support in Ω_+. As we require that g be continuous and have continuous first derivatives, we obtain by the usual process

$$\lim_{\phi\to 0^-} {}^2g_{\alpha\beta} = 0,$$

$${}^2g_{\alpha\beta} = 0 \quad \text{if } \phi < 0,$$

$${}^2g'_{\alpha\beta} = 0 \quad \text{if } \phi < 0,$$

$$\lim_{\phi\to 0^+} {}^2g'_{\alpha\beta} = a_{\alpha\beta}, \quad {}^2g'_{\alpha\beta} \text{ integrable along } \xi.$$

The two results we have from distribution theory are

$$\begin{aligned} g^{\mu\nu}\delta[g_{\alpha\beta,\mu\nu}] &= 2l^\lambda\partial_\lambda\bar{\delta}g_{\alpha\beta} + g^{\mu\nu}l_{\mu,\nu}\bar{\delta}g_{\alpha\beta}, \\ \delta[F_{\beta,\alpha}] &= l_\alpha\bar{\delta}F_\beta, \end{aligned} \tag{9}$$

where $F^\alpha = {}^0g^{\mu\nu}\,{}^+\Gamma^\alpha_{\mu\nu}$ are the so-called "harmonic quantities." On the other hand, using the asymptotic regularization given by (8), we get first

$$g^{\alpha\beta} = {}^0\bar{g}^{\alpha\beta} - \frac{{}^2\bar{g}^{\alpha\beta}}{\omega} + O(\omega^{-2}),$$

where the bar indicates indices raised with ${}^0\bar{g}^{\alpha\beta}$, $({}^0\bar{g}^{\alpha\beta}\,{}^0g_{\alpha\lambda} = \delta^\beta_\lambda)$. We then get by subtraction of the two limits as $\phi\to 0^-$ and $\phi\to 0^+$

$$\begin{aligned} 0 &= [g^{\mu\nu}g_{\alpha\beta,\mu\nu}] \\ &= {}^0\bar{g}^{\mu\nu}[{}^0g_{\alpha\beta,\mu\nu}] + \omega l_\mu\bar{l}^\mu[{}^2g''_{\alpha\beta}] + 2\bar{l}^\mu[{}^2g'_{\alpha\beta,\mu}] \\ &\quad + {}^0\bar{g}^{\mu\nu}l_{\mu,\nu}[{}^2g'_{\alpha\beta}] + O(\omega^{-2}), \end{aligned}$$

that is,

$$\begin{aligned} {}^0\bar{g}^{\mu\nu}[{}^0g_{\alpha\beta,\mu\nu}] &= 2\bar{l}^\mu\,{}^2g'_{\alpha\beta+,\mu} + {}^0\bar{g}^{\mu\nu}l_{\mu,\nu}\,{}^2g'_{\alpha\beta+} \\ &\quad + \omega l_\mu\bar{l}^\mu({}^2g''_{\alpha\beta})_+ + O(\omega^{-2}). \end{aligned}$$

Hence we have to order one in ω^{-1}, a result which has the same meaning as (9) on Σ, provided

$$l_\mu\bar{l}^\mu({}^2g''_{\alpha\beta})_+ = 0.$$

3. REGULARIZATION OF PRODUCTS OF DISCONTINUOUS FUNCTIONS

It may happen that the discontinuity of a product of discontinuous functions is required. When this happens, it is an easy matter to obtain it using asymptotic regularizations.

Suppose for instance that we are given two quantities ${}^0A$, ${}^0B$ which are continuous on Ω with discontinuities of their first derivatives at Σ. Then we have, for appropriate quantities a and b,

$$[{}^0A_{,\alpha}] = al_\alpha, \quad [{}^0B_{,\beta}] = bl_\beta.$$

Regularizations of ${}^0A$ and ${}^0B$ are of the form

$$A(x,\omega\phi) = {}^0A(x) + \frac{1}{\omega}\,{}^2A(x,\omega\phi),$$

$$B(x,\omega\phi) = {}^0B(x) + \frac{1}{\omega}\,{}^2B(x,\omega\phi),$$

for which we must have

$$\lim_{\phi\to 0^+} {}^2A' = a,$$

$$\lim_{\phi\to 0^+} {}^2B' = b$$

together with the usual integrability conditions. From this we get

$$\begin{aligned} &A_{,\alpha}B_{,\beta} \\ &= \left({}^0A_{,\alpha} + {}^2A'l_\alpha + \frac{1}{\omega}\,{}^2A_{,\alpha}\right)\left({}^0B_{,\beta} + {}^2B'l_\beta + \frac{1}{\omega}\,{}^2B_{,\beta}\right) \\ &= {}^0A_{,\alpha}\,{}^0B_{,\beta} + {}^0A_{,\alpha}\,{}^2B'l_\beta + {}^2A'\,{}^0B_{,\beta}l_\alpha \\ &\quad + {}^2A'\,{}^2B'l_\alpha l_\beta + O(\omega^{-2}) \end{aligned}$$

and from the continuity of $A_{,\alpha}B_{,\beta}$, we obtain

$$\begin{aligned} [{}^0A_{,\alpha}\,{}^0B_{,\beta}] &= {}^0A_{,\alpha+}\,{}^2B'_+l_\beta + {}^2A'_+\,{}^0B_{,\beta+}l_\alpha \\ &\quad + {}^2A'_+\,{}^2B'_+l_\alpha l_\beta. \end{aligned}$$

This has the advantage of being defined in terms of limits taken from the Ω_+ side of Σ only.

[1] A. Lichnérowicz, "Sur les ondes de choc gravitationnelles" C. R. Acad. Sci. Paris, Ser. A. **273**, 528–32 (1971).

[2] A. Lichnérowicz, "Sur les ondes de choc gravitationnelles et électromagnétiques," Colloques Internationaux C.N.R.S., 1973 **220**, 47–56 (1974).

[3] A. Lichnérowicz, "Ondes de choc en magnétohydrodynamique relativiste," Ann. Inst. H. Poincaré, Ser. A Vol. V, No. 1, 37–75 (1966).

[4] Y. Choquet-Bruhat, "Couplage d'ondes gravitationnelles et électromagnétiques à haute fréquence," Colloques Internationaux C.N.R.S., 1973 **220**, 85–100 (1974).

[5] A. Lichnérowicz, "Théorie des rayons en hydrodynamique et magnétohydrodynamique relativistes," Ann. Inst. H. Poincaré, Ser. A Vol. VII, No. 4, 271–302 (1967).

A note on completeness

T. A. Weber and C. L. Hammer

Ames Laboratory-DOE and Department of Physics, Iowa State University, Ames, Iowa 50011

(Received 23 April 1979; accepted for publication 26 September 1979)

Completeness relationships for eigenfunctions of second order differential equations are presented in a form which employs a contour integration rather than the usual integration and summation over eigenvalues. This technique which is particularly applicable for scattering problems simplifies the usual procedures and the proper weight functions are easily obtained. Some examples are given.

I. INTRODUCTION

In many applications of physics it is convenient to expand the general solution of a second order differential equation in terms of a complete set of eigenfunctions which are chosen to satisfy appropriate boundary conditions. In general we must obtain all the bound and continuum states and then determine the expansion coefficients, a complicated procedure indeed. This is also true for scattering problems even though the incident wave packet has no overlap with the scattering center. The bound-state contributions vanish only if the incident packet is at an infinite distance from the scattering center.

A technique is presented here which simplifies the usual approach. The method involves a single contour integration over complex values of a parameter. When the solutions satisfy the appropriate boundary conditions this parameter is just the eigenvalue. In this way, the bound and continuum states are treated in the same way.

Our proof depends only on the existence of L^2 integrable functions of a Hermitian operator. It is therefore more general than the proof given previously[1] which depended specifically upon that fact that the asymptotic scattering solutions were plane waves.

We find that our expansion represents both the *regular* and *irregular* solutions of the differential equation, even at the singular points. The completeness relationships for Hermite polynomials and for spherical Bessel functions are given as specific examples. The formulation is especially suitable for scattering solutions since the completeness relationship is manifestly written in terms of outgoing spherical waves. Such solutions are treated extensively in our previous work.[1]

II. COMPLETENESS RELATIONS

Equations of the form

$$(L-\lambda)\psi=0 \tag{1}$$

will be considered where L is an Hermitian operator. For the radial Schrödinger equation, which can be put in the form

$$L=-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+\frac{2mV(r)}{\hbar^2},\quad 0\leqslant r\leqslant\infty, \tag{2}$$

λ corresponds to the energy eigenvalues

$$\lambda=(2mE/\hbar^2)\equiv k^2. \tag{3}$$

One method for solving Eq. (1), with given boundary conditions, is by use of a Green's function which satisfies the inhomogeneous equation

$$(L-\lambda)G_\lambda(r,r_0)=-\delta(r-r_0). \tag{4}$$

Note that the l-dependence has been suppressed to eliminate notational clutter. The Green's function G_λ can also be written as an operator

$$G_\lambda=(\lambda-L)^{-1}, \tag{5}$$

called the resolvent. It follows from Hilbert's spectral theorem for Hermitian operators[2] that the resolvent is an analytic function of λ except for simple poles for λ in the discrete spectrum and a branch line for λ in the continuous spectrum. These occur for real values of λ.

Now consider integration of G_λ counterclockwise on a large circle in the physical sheet of the complex λ plane. If the operator is bounded then a circle can be chosen large enough to include all singularities. Otherwise, the radius of the circle must be allowed to approach infinity. Thus a Cauchy integral representation can always be found such that

$$(2\pi i)^{-1}\oint G_\lambda\, d\lambda=1. \tag{6}$$

This is an integral form for the completeness relationship for the eigensolutions of Eq. (1). In configuration space, this equation is

$$\int_C d\lambda\; G_\lambda(r,r_0)=2\pi i\delta(r-r_0), \tag{7}$$

where the circle is collapsed into the contour C shown in Fig. 1.

The Green's function can be constructed from solutions satisfying the usual boundary conditions at the origin and at infinity. At the origin, regular solutions for the operator defined in Eq. (2) are chosen.[3] These satisfy[4]

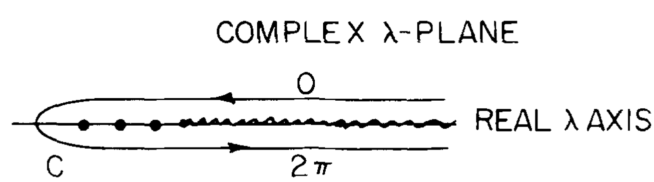


FIG. 1. Contour C in the complex λ plane.

$$\lim_{r\downarrow 0}[r^{-l-1}\phi_\lambda(r)]\rightarrow 1, \tag{8}$$

provided the potential behaves better than r^{-2} at the origin. Since the boundary condition is independent of λ, $\phi_\lambda(r)$ is an analytic function[5] of λ. The boundary condition at infinity depends on whether or not a continuum exists. It is convenient to treat the case where there is only a discrete spectrum before considering the more general case.

For potentials that are unbounded at infinity, such as the harmonic oscillator, which lead to discrete spectra only, solutions that approach zero as $r\uparrow\infty$ can be found. That such solutions, defined as $f_\lambda(r)$, exist can be seen from the asymptotic form obtained from the WKB method,

$$f_\lambda(r) \underset{r\uparrow\infty}{=} [U(r)-\lambda]^{-1/4}\exp\left\{-\int^r [U(\acute{r})-\lambda]^{1/2}\,d\acute{r}\right\}, \tag{9}$$

where $U(r)=2mV(r)/\hbar^2$ and λ is real. The Green's function may then be constructed as

$$\begin{aligned} G_\lambda(r,r_0) &= \frac{-\phi_\lambda(r)f_\lambda(r_0)}{L_\lambda}, \quad r<r_0,\\ &= -\frac{\phi_\lambda(r_0)f_\lambda(r)}{L_\lambda}, \quad r>r_0, \end{aligned} \tag{10}$$

where

$$\begin{aligned} L_\lambda &= f_\lambda(r)\phi_\lambda'(r)-f_\lambda'(r)\phi_\lambda(r)\\ &\equiv W[f_\lambda(r),\ \phi_\lambda(r)], \end{aligned}$$

is the Wronskian which is independent of r and the prime denotes differentiation with respect to the argument. For values of λ where the Wronskian is zero, the solutions $f_\lambda(r)$ and $\phi_\lambda(r)$ are clearly linearly related. This means that each function must satisfy the boundary condition for $r\downarrow 0$ as well as that for $r\uparrow\infty$. These must be the bound-state eigenfunctions which occur for the real values of λ which correspond to the simple poles of the integrand of Eq. (7). Thus Eq. (7) reduces to

$$\int_C d\lambda\,\frac{\phi_\lambda(r)f_\lambda(r_0)}{L_\lambda} = -2\pi i\,\delta(r-r_0), \tag{11}$$

where the contour C is counterclockwise around all the zeros of L_λ.

For the general case there is a continuous spectrum as well as a discrete spectrum. The WKB method shows that there exist asymptotic solutions such that

$$f_\lambda \equiv f(\sqrt{\lambda},r) \underset{r\uparrow\infty}{=} \exp[i\sqrt{\lambda}\,r], \tag{12}$$

provided the potential decreases faster than r^{-1} at ∞. If λ is chosen to have phases 0 to 2π as shown in Fig. 1, then $f(\sqrt{\lambda},r)$ is convergent for complex λ in the cut plane. If the cut is approached from above, outgoing wave conditions are obtained, while if the cut is approached from below incoming wave conditions are obtained.

It is convenient to make the transformation $k=\sqrt{\lambda}$. Then

$$\int_{C'} G_k(r,r_0)k\,dk = -\pi\delta(r-r_0), \tag{13}$$

where the contour is shown in Fig. 2. Here $G_k(r,r_0)$ is an

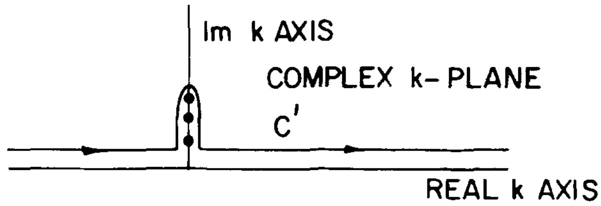


FIG. 2. Contour C' in the complex k plane.

analytic function of k in the upper half k-plane except for simple poles on the imaginary axis. The Green's function is

$$\begin{aligned} G_k(r,r_0) &= -\frac{\phi(k,r)f(k,r_0)}{L_+(k)}, \quad \text{for } r<r_0,\\ &= -\frac{\phi(k,r_0)f(k,r)}{L_+(k)}, \quad \text{for } r>r_0, \end{aligned} \tag{14}$$

where ϕ satisfies the boundary condition given in Eq. (8) and f satisfies Eq. (12). The Wronskian, which becomes

$$\begin{aligned} L_+(k) &= f(k,r)\phi'(k,r)-f'(k,r)\phi(k,r),\\ &= (2l+1)\lim_{r\downarrow 0}[r^l f(k,r)], \end{aligned}$$

is the Jost function. The function $f(k,r)$ is an analytic function of k in the upper half k plane[6] and the function $\phi(k,r)$ may be written as

$$\phi(k,r) = -(2ik)^{-1}[L_+(k)f(-k,r)-L_-(k)f(k,r)], \tag{15}$$

where

$$\begin{aligned} L_-(k) &= \phi'(k,r)f(-k,r)-\phi(k,r)f'(-k,r)\\ &= (2l+1)\lim_{r\downarrow 0}[r^l f(-k,r)]. \end{aligned}$$

It follows in the general case that Eq. (3) reduces to

$$\int_{C'}\frac{\phi(k,r)f(k,r_0)}{L_+(k)}\,k\,dk = i\pi\delta(r-r_0), \tag{16}$$

in parallel to Eq. (11). The proof of Eq. (16) for the discrete part of the spectrum is identical to that given previously. That is, $f(k,r)$ and $\phi(k,r)$ are linearly related for values of k where $L_+(k)=0$ thereby making $G_k(r,r_0)$ symmetric in r and r_0. The remainder of the proof follows from the fact that $G_k(r,r_0)$ has this same symmetry for values of k on the real axis. To see this consider that part of the C' integration that corresponds to the integral

$$\begin{aligned} \int_{-\infty}^{\infty} k\,dk\,\frac{\phi(k,r)f(k,r_0)}{L_+(k)} &= (i/2)\int_{-\infty}^{\infty} dk\,f(-k,r)f(k,r_0)\\ &\quad -(i/2)\int_{-\infty}^{\infty} dk\,[L_-(k)/L_+(k)]f(k,r)f(k,r_0). \end{aligned} \tag{17}$$

The change of integration variable from k to $-k$ in the first term of the rhs of Eq. (17) then gives, when $f(k,r)$ is factored from the two integrands,

$$\int_{-\infty}^{\infty} k\,dk\,\frac{\phi(k,r)f(k,r_0)}{L_+(k)} = \int_{-\infty}^{\infty} k\,dk\,\frac{\phi(k,r_0)f(k,r)}{L_+(k)}.$$

It follows from Eq. (16) that an arbitrary function $g(r)$ can be expanded either in terms of $\phi(k,r)$ or $f(k,r)$. For example, either

$$g(r) = \int_{C'} \frac{k\, dk\, a(k) f(k,r)}{L_+(k)}, \tag{18}$$

where

$$a(k) = (i\pi)^{-1} \int_0^\infty g(r)\phi(k,r)\, dr, \tag{19}$$

or f and ϕ can be interchanged.

Equation (18) is particularly useful if it is desired to expand $g(r)$ in terms of functions which asymptotically are spherical outgoing waves. Note also that, if $\phi(k,r)$ is sufficiently regular at the origin, $a(k)$ is finite even for function $g(r)$ that are singular at the origin. In the next section, as an example, we illustrate how spherical Hankel functions can be used to represent irregular functions in the closed interval 0 to ∞.

III. EXAMPLES

A. Completeness for simple harmonic oscillator

The equation for the simple harmonic oscillator is

$$\frac{d^2\psi}{dr^2} - \left(\frac{r^2}{4} - \lambda\right)\psi = 0, \tag{20}$$

where r is expressed in units $(2m\omega/\hbar)^{-1/2}$ and $\lambda = (E/\hbar\omega)$. The solutions are parabolic cylinder functions. The boundary condition on ϕ given by Eq. (8) gives an expansion for odd functions only. The appropriate solutions are[8]

$$\phi_\lambda(r) = y_2(-\lambda,r), \tag{21}$$

which is regular at the origin and

$$f_\lambda(r_0) = U(-\lambda,r_0), \tag{22}$$

which is bounded at infinity. The Wronskian

$$L_\lambda = W[f_\lambda(r),\phi_\lambda(r)] = \pi^{1/2}\left[2^{\lambda/2-1/4}/\Gamma(\tfrac{3}{4} - \tfrac{1}{2}\lambda)\right] \tag{23}$$

has zeros which occur for

$$\lambda_n = n + \tfrac{1}{2} \quad (n \text{ odd}).$$

For these values of λ,

$$\begin{aligned} y_2(-\lambda_n,r) \\ &= -\pi^{-1/2} 2^{-\lambda/2} n^{-1/4} \Gamma(\tfrac{1}{4} - \tfrac{1}{2}\lambda_n)\, U(-\lambda_n,r), \end{aligned} \tag{24}$$

where

$$U(-\lambda_n,r) = 2^{-n/2} e^{-r^2/4} \quad H_n(r/\sqrt{2}), \tag{25}$$

and H_n is the Hermite polynomial. The completeness relation, Eq. (11), can be used to obtain the sum of residues

$$\begin{aligned} \sqrt{2/\pi} \sum_{\substack{n=1\\ \text{odd}}}^{\infty} (n!)^{-1}(2)^{-n} e^{(r^2+r_0^2)/4} H_n(r/\sqrt{2}) \\ \times H_n(r_0/\sqrt{2}) = \delta(r - r_0), \end{aligned} \tag{26}$$

which is the standard result. It is apparent that the functions which are orthogonal for the range of integration $r = 0$ to $r = \infty$, are

$$\left(\frac{2}{\pi}\right)^{1/4} (n!)^{-1/2} (2)^{-n/2} e^{-r^2/4} H_n\left(\frac{r}{\sqrt{2}}\right).$$

The even functions can be obtained by requiring the boundary condition on ϕ to be

$$\phi(0) = 1,$$
$$\phi'(0) = 0.$$

The remainder of the analysis is identical to the "odd" case described above.

B. Completeness for spherical Bessel functions

For $V = 0$ the radial Schroedinger equation is

$$\{-(d^2/dr^2) + [l(l+1)/r^2] = k^2\psi_l. \tag{27}$$

The solutions of interest are spherical Bessel functions

$$\phi_l(k,r) = (2l+1)!!\, k^{-l} r j_l(kr),$$

and

$$f_l(k,r) = i^{(l+1)} kr\, h_l^{(1)}(kr), \tag{28}$$

where

$$L_l(k) = (2l+1)!!\, i^l k^{-l}. \tag{29}$$

The completeness relation, Eq. (16), becomes

$$\pi^{-1} \int_{-\infty}^{\infty} r j_l(kr) r_0 h_l^{(1)}(kr_0) k^2\, dk = \delta(r - r_0). \tag{30}$$

According to this relationship an arbitrary function $g(r)$ can be represented by spherical Bessel functions by choosing

$$a_l(k) = \pi^{-1} \int_0^\infty r\, dr\, g(r) j_l(kr),$$

$$g(r) = \int_{-\infty}^{\infty} k^2\, dk\, a_l(k)\, r h_l^{(1)}(kr). \tag{31}$$

If $\lim_{r\downarrow 0} r^\gamma g(r) = \text{const}$, $\gamma > 0$, the integer l must be chosen to satisfy $l \geqslant \gamma - 2$, if $a(k)$ is to exist. The expansion as written in Eq. (31) will represent $g(r)$ everywhere except possibly at the origin. To assure that $g(r)$ has the appropriate behavior at the origin, we choose a limiting process such that

$$g(r) = r^{-2} \int_{-\infty}^{\infty} p^2\, dp\, a(p/r) h_l^{(1)}(p), \tag{32}$$

which requires $r \downarrow 0$ following $|k| \uparrow \infty$. That the expression, Eq. (32), represents $g(r)$ even at the origin can be seen by substituting $g(r) \sim r^{-\gamma}$ into Eq. (31). The result is

$$\begin{aligned} a_l(k) &= \frac{k^{\gamma-2}}{(2\pi)^{1/2}} \frac{\Gamma((l-\gamma)/2 + 1)}{2^{\gamma-1/2}\Gamma((l+\gamma+1)/2)} \\ &\equiv k^{\gamma-2} f(l), \end{aligned}$$

where the reflection property

$$a_l(k) = (-1)^l a_l(-k), \tag{33}$$

follows directly from Eq. (31). Substitution for $a_l(p/r)$ in Eq. (32) then gives

$$\begin{aligned} g_l(r) &= r^{-\gamma} f(l) \int_{-\infty}^{\infty} p^\gamma\, dp\, h_l^{(1)}(p) \\ &= r^{-\gamma} \end{aligned}$$

for all r, as required.

C. Other completeness relationships

The relationship Eq. (16), can be written in terms of regular functions also if one treats the bound-state wavefunctions differently from the continuum wavefunctions. If the contour shown in Fig. 2 is brought down to the real axis, Eq. (16) becomes

$$\int_{-\infty}^{\infty} k\,dk\,\frac{\phi(k,r)f(k,r_0)}{L_+(k)} - 2\pi i \sum_{\text{Res}} \frac{k\phi(k,r)f(k,r_0)}{L_+(k)} = \pi i\delta(r-r_0).$$

The definition of $\phi(k,r)$, Eq. (15) can be used to rewrite this equation in the form

$$2\int_{-\infty}^{\infty} k^2\,dk\,\frac{\phi(k,r)\phi(k,r_0)}{L_+(k)L_-(k)} - 2\pi \sum_{\text{Res}} \frac{k\phi(k,r)f(k,r_0)}{L_+(k)} = \pi\delta(r-r_0). \tag{34}$$

where the residues occur because of the zero's

$L_+(k_n)=0.$

For these values of k_n

$$f(k_n,r_0) = \frac{2ik_n}{L_-(k_n)}\,\phi(k_n,r_0), \tag{35}$$

which completes the proof. This result is the standard form usually given for the completeness relationship.[9] For the spherical Bessel functions there are no bound states to consider so Eq. (34) reduces to

$$\frac{2}{\pi}\int_0^{\infty} k^2\,dk\,rj_l(kr)r_0\,j_l(kr_0) = \delta(r-r_0). \tag{36}$$

Finally, we note that if $f(-k,r)$ (and therefore L_-) is analytic in the upper-half of the complex k-plane, the defintion of $\phi(k,r)$ given in Eq. (15) can be used to show

$$(2\pi)^{-1}\int_{-\infty}^{\infty} dk\,f(-k,r)f(k,r_0) = \delta(r,r_0), \tag{37}$$

a result which might be expected considering the asymptotic properties given in Eq. (12). In particular for the Hankel functions,

$$(2\pi)^{-1}\int_{-\infty}^{\infty} k^2\,dk\,rh_l^{(1)}(kr)\,r_0\,h_l^{(2)}(kr_0) = \delta(r-r_0), \tag{38}$$

where the integration path must go either above or below the origin (the residue vanishes).

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Department of Energy, Division of Basic Sciences and in part by NRC Grant A-2228. One of us (TAW) wishes to thank the physics department of the University of British Columbia for their hospitality during which time this work was finished.

[1]T.A. Weber and C.L. Hammer, J. Math. Phys. **18**, 1562 (1977), see especially Sec. C. C.L. Hammer, T.A. Weber, and V.S. Zidell, Amer. Phys. **45**, 933 (1977).

[2]A.E. Taylor, *Introduction to Functional Analysis* (Wiley, New York, 1958), p. 362.

[3]Roger G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), pp. 329, 371.

[4]The function ϕ_λ given here is r times that given in Ref. 1.

[5]This is an illustration of a general theorem due to Poincaŕe. The theorem is stated by E. Hilb in *Encyklopödie der Mathematischen Wessenschaften*, (Teuber, Leipzig, 1915), Vol. 2, pt. 2, p. 501.

[6]See Ref. 3, p. 333.

[7]Note that $f(k,r)$ must be analytically continued in the upper half k-plane to the incoming wave solution $f(-k,r)$.

[8]*Handbook of Mathematical Functions*, edited by Milton Abramowitz and Irene A. Stegun (National Bureau of Standards, 1970), Section 19, p. 685.

[9]As defined in Ref. 3, $L_-(k)$ is not necessarily analytic for values of k in the upper half of the complex k plane. However $L_-(k_n)$ must always be well defined since for these values $\phi(k_n,r)$ is proportional to $f(k_n,r)$ and both of these functions are well defined everywhere in the upper half k plane including the real axis.

Construction of doubly orthogonal functions[a)]

John M. Jarem

Department of Electrical Engineering, University of Petroleum and Minerals, Dhahran, Saudi Arabia

(Received 25 May 1979; accepted for publication 30 August 1979)

It is shown that the extremal functions associated with a general functional form a set of biorthogonal and doubly orthogonal functions. The theory is applied to antenna theory to find the doubly orthogonal functions of the E and H plane strip source antennas.

I. INTRODUCTION

Sets of analytic functions $\{F_n\}$ which are orthogonal on the interval $-\infty < u < \infty$ with weight factor $h_2(u)$ and which are orthogonal on the interval $-1 \leqslant u \leqslant 1$ with weight factor $h_1(u)$ are called doubly orthogonal. These functions have proved to be of great value in the synthesis of continuous antenna sources (it is assumed that the far field radiation pattern of the antenna source is to approximate a desired, ideal radiation pattern as closely as possible) when the antenna source is subject to an energy or magnitude constraint (Ref. 1, pp. 154–634 and Ref. 2).

The usefulness of doubly orthogonal functions arises from the fact that the same functions $\{F_n\}$ may be used to describe both the far field radiation pattern which is defined only on the interval $-1 \leqslant u \leqslant 1$ and the energy constraint which is defined on the infinite interval $-\infty < u < \infty$.

A drawback to the applicability of these functions is that only a limited number of these functions are known to be doubly orthogonal. At the present time the only known doubly orthogonal functions are the spheroidal functions $\{\Psi_{\alpha,n}(c,u)\}$ [3] (which include as special cases the prolate spheroidal functions[4] and the even and odd Mathieu functions[3]) and the generalized spheroidal functions.[5]

The object of this paper will be to greatly expand the known class of doubly orthogonal functions and to also show how these functions may be constructed. The first part of the investigation will show that the extremal functions of certain types of linear functionals are doubly orthogonal. The second part of the paper will derive for the first time the doubly orthogonal functions associated with the real and reactive power of the E and H plane strip source antennas.

II. ANALYSIS

This part of the investigation will be concerned with showing that the set of extremal functions associated with a certain type of functional are doubly orthogonal. To formulate the problem let $f(t)$ be a function which is defined on the interval $-1 \leqslant t \leqslant 1$ and also let $F(cu)$ be the finite Fourier transform of $f(t)$, where c is a positive constant. (In the application to antenna theory c equals $\pi L/\lambda$, where L is the length of the antenna and λ is the wavelength of free space radiation.) The functional with which we are interested is defined by

$$[Q] = \frac{\int_{-\infty}^{\infty} h_2(u)\,|F(cu)|^2\,du}{\int_{-1}^{1} h_1(u)\,|F(cu)|^2\,du} = \frac{\langle f,\underline{H}_2 f\rangle}{\langle f,\underline{H}_1 f\rangle}, \quad (1)$$

where

$$F(cu) = \int_{-1}^{1} f(t)\,e^{jcut}\,dt \quad (2)$$

and where $f(t)$ behaves as

$$f(t) \sim (1-t^2)^{\beta}, \quad \beta > -1 \quad (3)$$

for t arbitrarily close to ± 1. Also

$$\langle f,g\rangle = \int_{-1}^{1} f^*(t)\,g(t)\,dt$$

is defined to be the inner product of f and g on $-1 \leqslant t \leqslant 1$. The integral operators $\underline{H}_1$ and $\underline{H}_2$ are obtained by an integration interchange in Eq. (1) and are given by

$$\underline{H}_1(c,t) = \int_{-1}^{1} dt' \left\{ \int_{-1}^{1} h_1(u) \exp[jcu(t-t')]\,du \right\}, \quad (4a)$$

$$\underline{H}_2(c,t) = \int_{-1}^{1} dt' \left\{ \int_{-\infty}^{\infty} h_2(u) \exp[jcu(t-t')]\,du \right\}. \quad (4b)$$

The weight factors $h_1(u)$ and $h_2(u)$ are also assumed to be even in u and nonnegative.

We have chosen this functional because the set of extremal functions associated with it for

$$\begin{aligned} h_1 &= (1-u^2)^{\alpha}, \quad |u| \leqslant 1, \\ h_2 &= (u^2-1)^{\alpha}, \quad |u| \geqslant 1, \\ h_2 &= 0, \quad |u| < 1, \quad \alpha > -1, \end{aligned} \quad (5)$$

turn out to be the spheroidal functions $\{\Psi_{\alpha,n}(c,u)\}$. We have also chosen this functional because it describes the real and reactive powers which are associated with a number of continuous line source antennas [Ref. 1, pp. 74–105] and is thus of practical value.

The second half of Eq. (1) has been expressed as an inner product of the operators $\underline{H}_1$ and $\underline{H}_2$ with the functions $f(t)$. The kernels of these operators are the Fourier transforms of $h_1(u)$ and $h_2(u)$ as defined in Eqs. (4a) and (4b). It is clear that $\underline{H}_1$ and $\underline{H}_2$ are Hermitian, positive definite, and real operators. It is assumed that $f(t)$ belongs to a class of functions such that the functional $[Q]$ is always convergent and that the value of β is consistent with this convergence.

To identify the extremals associated with the functional of Eq. (1) we vary $[Q]$ with respect to all functions which meet the conditions (3) and cause the functional to be convergent. This is a straightforward procedure which has been described in detail in Ref. 6 [pp. 1108–9, Eq. (9.4.6)]. The $[Q]$ functional after variation may be written in the form

[a)]Supported by a University of Petroleum and Minerals Research Grant.

$$\langle \delta f, (\underline{H}_2 - Q\underline{H}_1) f \rangle = 0, \tag{6}$$

where $Q = [Q]|_{f_{\text{EXT}}}$ is the value of the functional at the extremum and $\delta f(t)$ is the first variation of $f(t)$. The condition that $f(t)$ satisfy the boundary condition (3) is imposed by expanding $f(t)$ and $\delta f(t)$ in the series

$$f(t) = \sum_{k=0}^{\infty} x_k \phi_k(t), \tag{7a}$$

$$\delta f(t) = \sum_{k'=0}^{\infty} \delta x_{k'} \phi_{k'}(t), \tag{7b}$$

where it is assumed that the set of functions $\{\phi_k(t)\}$ are complete and orthonormal in the interval $-1 \leqslant t \leqslant 1$ and satisfies the condition of Eq. (3). Examples of functions $\{\phi_k\}$ which satisfy these properties are the Gegenbauer polynomials [Ref. 6, p. 783] with

$$\phi_k(t) = (1-t^2)^{\beta} T_k^{\beta}(t), \quad k = 0,1,\ldots, \quad \beta > -1 \tag{8}$$

and the spheroidal functions[3] with

$$\phi_k(t) = (1-t^2)^{\beta} \Psi_{\beta,n}(c,u), \quad k = 0,1,\ldots, \quad \beta > -1. \tag{9}$$

Another set which may be used when $\beta = 1$ are the functions

$$\phi_k(t) = \begin{cases} \cos \underline{k} t, & k = 0,2,4\cdots, \\ \sin \underline{k} t, & k = 1,3,5\cdots, \end{cases} \tag{10}$$

with $\underline{k} = [(k+1)/2]\pi$.

To proceed further with the analysis, we substitute (7a) and (7b) into Eq. (6) and carry out the inner products of the functions ϕ_k and $\phi_{k'}$ with operators $\underline{H}_1$ and $\underline{H}_2$. After setting the coefficients of the independent variation of $\delta x_{k'}$ to zero, the following matrix equation results:

$$[\underline{H}_2][x] = Q[\underline{H}_1][x], \tag{11}$$

where $[\underline{H}_1]$ and $[\underline{H}_2]$ are square matrices and $[x]$ is a column matrix of the coefficients of Eq. (7a). The matrix elements of this equation are defined by

$$[\underline{H}_1]_{k,k'} = \int_{-1}^{1} h_1(u) \Phi_{k'}^*(cu) \Phi_k(cu)\, du, \tag{12a}$$

$$[\underline{H}_2]_{k,k'} = \int_{-\infty}^{\infty} h_2(u) \Phi_{k'}^*(cu) \Phi_k(cu)\, du, \tag{12b}$$

where Φ_k is the finite Fourier transform of ϕ_k.

The finite Fourier transform of the $\{\phi_k\}$ in Eq. (8) [Ref. 6, p. 621] is

$$\Phi_k(cu) = \frac{j^k \sqrt{2\pi}\,\Gamma(k+2\beta+1)}{k!(cu)^{\beta+1/2}} J_{k+\beta+1/2}(cu), \quad k = 0,1,2, \tag{13}$$

and the finite Fourier transform of Eq. (9) [Ref. 3, p. 187] is

$$\Phi_k(cu) = j^k |\nu_{\beta,n}(c)| \Psi_{\beta,n}(c,u), \quad k = 0,1,2\cdots. \tag{14}$$

The finite Fourier transform of the $\{\Phi_k\}$ in Eq. (10) which will be used later in this paper is given (Ref. 1, p. 115)

$$\Phi_k(cu) = \int_{-1}^{1} \phi_k(t) e^{jcut}\, dt = \begin{cases} j^k \left[\dfrac{-2\underline{k}\cos cu}{(cu)^2 - \underline{k}^2} \right], & k = 0,2,\cdots, \\ j^k \left[\dfrac{-2\underline{k}\sin cu}{(cu)^2 - \underline{k}^2} \right], & k = 1,3,\cdots. \end{cases} \tag{15}$$

The Φ_k in all three examples given above turn out to be complete and orthonormal in the interval $-\infty < u < \infty$.

It can easily be seen that the matrix elements of Eqs. (12a) and (12b) are zero whenever k is even and k' is odd (or vice versa) since Φ_k and $\Phi_{k'}$ are even and odd, respectively (recall that h_1 and h_2 are even). Therefore it can be seen that the matrix equation (11) splits into two separate matrix equations one which describes the even coefficients and one which describes the odd.

The solution to the matrix equation (11) consists of an infinite set of eigenvectors which we denote

$$[x]_n^T = \begin{cases} [x_{0,n}, 0, x_{2,n}, 0\cdots]^T, & n = 0,2,4,\cdots, \\ [0, x_{1,n}, 0, x_{3,n}, \cdots]^T, & n = 1,3,5,\cdots, \end{cases} \tag{16}$$

corresponding to an infinite set of eigenvalues $0 < Q_0 < Q_1 \cdots$.

As is well known these eigenvectors satisfy the orthogonality relations

$$[x]_m^T [\underline{H}_2][x]_n = Q_n N_n \delta_{m,n}, \tag{17a}$$

$$[x]_m^T [\underline{H}_1][x]_n = N_n \delta_{m,n}, \tag{17b}$$

where N_n is a positive normalization constant and $\delta_{m,n}$ is the Kronecker delta.

The doubly orthogonal functions associated with the functional of Eq. (1) may be constructed from the matrix eigenvectors $[x]_n$. Let $f_n(t)$ and $F_n(cu)$ be the functions defined by the equations

$$f_n(t) = \sum_{k=0}^{\infty} x_{k,n} \phi_n(t) \tag{18}$$

and

$$F_n(cu) = \int_{-1}^{1} f_n(t) e^{jcut}\, dt = \sum_{k=0}^{\infty} x_{k,n} \Phi_n(cu). \tag{19}$$

The double orthogonality relations that these functions $\{F_n(cu)\}$ satisfy are given by the equations

$$\langle f_m, \underline{H}_1 f_n \rangle = \int_{-1}^{1} h_1 F_m^* F_n\, du = N_n \delta_{m,n}, \tag{20}$$

$$\langle f_m, \underline{H}_2 f_n \rangle = \int_{-\infty}^{\infty} h_2 F_m^* F_n\, du = Q_n N_n \delta_{m,n}. \tag{21}$$

These relations are a result of the orthogonality relations given in Eqs. (17a) and (17b) and are proved in the Appendix. The first part of Eqs. (20) and (21) shows the additional interesting property that the set of functions $\{f_n\}$ is biorthogonal to the set of functions $\{\underline{H}_1 f_n\}$. This property is needed in order to determine the coefficients when a known function is expanded in a series of the set $\{f_n\}$ on the interval $-1 \leqslant t \leqslant 1$.

We would like to point out that the construction method which has been presented here to produce doubly orthogonal functions has the additional degree of freedom that the boundary conditions on the set of functions $\{f_n\}$ from which the functions $\{F_n\}$ arise may be specified. This is extremely important at least in antenna theory where the sources (which would presumably be represented by the set of functions $\{f_n\}$) are expected to meet physical boundary conditions say in a slot or aperture.

In concluding this section we would like to note that the formalism presented here for constructing doubly orthogonal functions is numerically practical and easy to implement. The matrix elements require only a few hundred points of numerical integration for an accurate evaluation and also many of the currently developed eigenanalysis computer programs may be used to solve the matrix equations in a few seconds for reasonably small matrices (25×25).

III. STRIP SOURCE ANTENNAS

In this section we will illustrate the preceding method by finding for the first time the doubly orthogonal functions which are useful in the description of the real and reactive power of the E and H plane strip source antennas (Ref. 1, pp. 74–105). This set of doubly orthogonal functions will represent the real and reactive power of these antennas in their simplest possible form and thus will greatly facilitate antenna synthesis methods which seek to limit the ratio of reactive power to real power to a fixed value.

The ratio of the reactive power to the real power which is called the quality factor Q[7] has been derived by Rhodes (Ref. 1, pp. 74 – 105) for the E and H plane strip source antennas. The quality factor for the H-plane strip source antenna is given by Eq. (1) with

$$h_1(u) = (1-u^2)^{1/2}, \quad |u|\leqslant 1,$$
$$h_2(u) = 0, \quad |u|<1,$$
$$h_2(u) = u^2/(u^2-1)^{1/2}, \quad |u|\geqslant 1,$$

and the quality factor of the E plane strip source is given by Eq. (1) with

$$h_1(u) = (1-u^2)^{-1/2}, \quad |u|\leqslant 1,$$
$$h_2(u) = 0, \quad |u|<1,$$
$$h_2(u) = (u^2-1)^{-1/2}, \quad |u|\geqslant 1.$$

The function $f(t)$ represents the electric field in both antennas and the boundary conditions for the E-plane and H-plane strip sources are given by Eq. (3) with $\beta = 1$ and $\beta = 2$, respectively (Ref. 1, p. 89).

To obtain an approximate numerical solution for the doubly orthogonal functions of both antennas, the functions $\{\Phi_k\}$ of Eq. (15) have been used to calculate the matrix coefficients of $[\underline{H}_1]_{k,k'}$ and $[\underline{H}_2]_{k,k'}$ up to a value of $k=49$ and $k'=49$. The two resulting 25×25 matrix equations which correspond to the even and odd coefficients of Eqs. (12a) and (12b) have been solved on an IBM 360–158 computer using Harwell's eigenanalysis program EA12AD. The resulting doubly orthogonal functions for both antennas turned out to be well behaved and the double orthogonality conditions of Eqs. (20) and (21) were verified.

Table I lists the eigenvalues Q_n of the E- and H-plane strip source antennas for $c=6$ and also displays for comparison with the E-plane strip source the eigenvalues of the spheroidal functions $\{\Psi_{-1/2,n}(c,u)\}$ which were obtained from Ref. 3 (p. 208) with $Q_n = \gamma_n - 1$. As can be seen from Table I, the eigenvalues of the E-plane strip source antenna as obtained with $\beta = 1$ [which is the only physically correct value of β (Ref. 1, p. 89)] interestingly enough compare very closely to that found by Rhodes with $\beta = -\frac{1}{2}$. This result seems to imply that reactive and real power of an antenna is not greatly affected by the nature of the electromagnetic fields near the aperture edge. This statement also represents an answer to Rhodes' question (Ref. 1, p. 40) about what effect the edge behavior of an electromagnetic field has on its associated value of quality factor. We remark that the eigenvalues of the H-plane strip source antenna and the associated eigenfunctions have been obtained here for the first time.

IV. SUMMARY AND CONCLUSIONS

In this investigation it has been shown that the set of extremal functions $\{F_n(cu)\}$ associated with the functional of Eq. (1) are orthogonal on the interval $-1\leqslant u\leqslant 1$ and also on the infinite interval $-\infty<u<\infty$. Previously this had been shown to be true only for the spheroidal functions and generalized spheroidal functions. Now for the first time it is shown that the class of functions which are doubly orthogonal is a very large one and given by the extremals of Eq. (1). We also note that the construction method presented here offers the additional degree of freedom that the functions $f_n(t)$ from which the doubly orthogonal functions arise may be made to behave as Eq. (3) near the points $t=\pm 1$.

We conclude this paper by noting that the preceding theory of double orthogonality may be extended to functionals of vectors $F^i(c_1u_1, c_2u_2\cdots)$, $i=1,\ldots,n$ over multidimensional variables $j=1,\ldots,m$. The doubly orthogonal functions which are associated with real and reactive power of rectangular apertures (two dimensions) are an example of where functionals of vectors F^i over a multidimensional domain (the real power is described in a region outside a circle in two dimensions) would be useful.

TABLE I.

n	Q of E-plane for $c=6$ (from matrix eq.)	Q of H-plane for $c=6$	Q of $\Psi_{-\frac{1}{2},n}(c,u)$ for $c=6$ (from Ref. 3)
0	0.111174 − 03	0.438691 − 03	0.103114 − 03
1	0.454066 − 02	0.201505 − 01	0.424436 − 03
2	0.721500 − 01	0.384336 + 00	0.682299 − 01
3	0.473604 + 00	0.374175 + 01	0.456053 + 00
4	0.187972 + 01	0.269618 + 02	0.180946 + 01
5	0.110122 + 02	0.227449 + 03	0.103794 + 02
6	0.128359 + 03	0.328553 + 04	0.118398 + 03
7	0.232237 + 04	0.714732 + 05	0.210748 + 04

ACKNOWLEDGMENTS

I would like to thank Dr. Thaddeus Kobylarz and Dr. Preston Clement for their helpful advice towards the preparation of this paper.

APPENDIX

To prove Eq. (21) we substitute f_m and f_n or Eq. (18) into Eq. (21) and obtain

$$\langle f_m, \underline{H}_2 f_n \rangle = \int_{-1}^{1} \left\{ \left[\sum_{k'=0}^{\infty} x_{k,m} \phi_m \right]^* \int_{-1}^{1} dt' \times \left\{ \int_{-\infty}^{\infty} h_2 \exp[jcu(t-t')]\, du \right\} \cdot \left[\sum_{k=0}^{\infty} x_{k,n} \phi_n \right] \right\} dt. \tag{A1}$$

If we interchange u and t integration and interchange u and t' integration we obtain

$$\begin{aligned} \langle f_m, \underline{H}_2 f_n \rangle &= \int_{-\infty}^{\infty} du\, h_2(u) \left[\sum_{k'=0}^{\infty} x_{k',m} \times \int_{-1}^{1} \phi_m(t') e^{jcut'} dt' \right]^* \cdot \left[\sum_{k=0}^{\infty} x_{k,n} \int_{-1}^{1} \phi_n(t) e^{jcut} dt \right] \\ &= \int_{-\infty}^{\infty} h_2(u) F_m^* F_n\, du \end{aligned} \tag{A2}$$

$$= \int_{-\infty}^{\infty} du\, h_2(u) \left[\sum_{k'=0}^{\infty} x_{k',m} \Phi_m^*(cu) \right] \times \left[\sum_{k=0}^{\infty} x_{k,n} \Phi_n(cu) \right].$$

This proves the equality of the first part of Eq. (21). If we interchange summation and integral in the last part of (A2) we obtain

$$\begin{aligned} \langle f_m, \underline{H}_2 f_n \rangle &= \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} x_{k',m} x_{k,n} \int_{-\infty}^{\infty} h_2 \Phi_m^* \Phi_n\, du \\ &= [x]_m^T [\underline{H}_2][x]_n. \end{aligned} \tag{A3}$$

Since the u integral in Eq. (A3) is just the matrix element of Eq. (12b) and the double summation is just matrix multiplication, we have

$$\langle f_m, \underline{H}_2 f_n \rangle = \int_{-\infty}^{\infty} h_2 F_m^* F_n\, du = N_n Q_n \delta_{mn} \tag{A4}$$

by Eq. (17a). The proof of Eq. (20) is similar.

[1]D. Rhodes, *Synthesis of Planar Antenna Sources* (Clarendon, Oxford, 1974).
[2]R.L. Fante, IEEE Trans. Antennas and Propagation **AP-18**, 177 (1970).
[3]D. Rhodes, "On the Spheroidal Functions," J. Res. Nat. Bur. Stand. B, Math. Sci. **74B** (3), 187 (1970).
[4]D. Slepian and H.O. Pollak, Bell Syst. Tech. J. **40**, 43 (1961).
[5]D. Slepian, Bell Syst. Tech. J. **43**, 3009 (1964).
[6]P.M. Morse and H. Feshback, *Methods of Theoretical Physics-Part II* (McGraw-Hill, New York, 1953).
[7]D. Rhodes, J. Franklin Inst. **302**, 225 (1976).

Quadratic Hamiltonians: The four classes of quadratic invariants, their interrelations and symmetries

P. G. L. Leach

Department of Applied Mathematics, La Trobe University, Bundoora, 3083, Australia

(Received 21 September 1978; revised manuscript received 11 January 1979)

The quadratic invariants of the three basic quadratic Hamiltonian systems—attractive oscillator, repulsive oscillator, and free particle— are shown to be the same. These invariants are divided into two categories, useful and nonuseful. The definition of useful is in terms of contributing to the (quadratic invariant based) symmetry group of the appropriate Hamiltonian. Usefulness is not invariant under a time-dependent linear canonical transformation. Hence different classes of invariants produce the different symmetry groups for the three different types of quadratic Hamiltonian considered here. The paper concludes with a consideration of the useful transformation of arbitrary quadratic Hamiltonians.

I. INTRODUCTION

The importance of quadratic forms in the mathematical analysis of physical problems cannot be gainsaid. The reasons for this importance lie both in physics and in mathematics. It is in the nature of things for a physical context to produce, as a first approximation at least, a quadratic form of some type or another. Mathematically, quadratic forms are amenable to analysis without overmuch effort, are susceptible to elegant treatment and yield results which are often receptive to simple pictorial representation. The increase in effort required to analyze higher order systems and the lack of obvious significance in the results coupled with the reasonable accuracy of the quadratic approximation in many physical applications has served to concentrate attention on quadratic systems even further. It is therefore appropriate that we examine every aspect of such systems to extract the utmost understanding of their properties.

We are concerned here with a particular type of quadratic system, that of the quadratic Hamiltonian. In a general sense the Hamiltonian is that of an n-dimensional system although the examples used will be confined to three-dimensional systems. Quadratic Hamiltonians are characteristic of small motions about an equilibrium point and find application from celestial mechanics to quantum mechanics. Such Hamiltonians have attracted considerable attention in the literature, including some contributions from the present writer. In the course of our investigations of the invariants of time-dependent Hamiltonians, both quadratic and of higher order, it seemed wise to look again at the time-independent isotropic harmonic oscillator.[1] Of the four classes of quadratic invariant found for that system, only two, the angular momentum tensor and the Jauch–Hill–Fradkin tensor seemed to be of physical significance.[2,3] Certainly the other two classes offered no contribution to a discussion of the dynamical symmetry of the system..

When we turn to the other basic components of a quadratic Hamiltonian,[4] the free particle and the repulsive oscillator (in both forms), we see that these hitherto obscure quadratic constants do have a role to play. In this article we examine the relationships between the four classes of quadratic invariant, the four basic components of a quadratic Hamiltonian and the associated symmetry groups. The primary tool in these investigations is the time-dependent linear canonical transformation. This is not out of prejudice against nonlinear transformation,[5] but because of the natural relationship between quadratic Hamiltonians and linear transformations, both in classical and quantum mechanics.[6] Indeed, in part we are looking for the maximal invariance groups generated by a basis formed from invariants which are quadratic forms.[7]

2. THE LINEAR TRANSFORMATIONS

The theory of linear canonical transformations has been discussed elsewhere[8] and is summarized here. Writing the dynamical variables as

$$q_i = z^\mu, \quad i = 1,n, \quad \mu = 1,n,$$
$$p_i = z^\mu, \quad i = 1,n, \quad \mu = n+1,2n, \tag{2.1}$$

a (homogeneous) quadratic Hamiltonian for an n-dimensional system may be written as

$$H = \tfrac{1}{2}\mathbf{z}^T A \mathbf{z}, \tag{2.2}$$

where A is a real symmetric $2n\times 2n$ matrix.[9] The Hamiltonian (2.2) may be transformed by a linear canonical transformation to

$$\bar{H} = \tfrac{1}{2}\bar{\mathbf{z}}^T \bar{A}\bar{\mathbf{z}}, \tag{2.3}$$

where $\bar{A}$ is again a real symmetric $2n\times 2n$ matrix. Writing the transformation as

$$\bar{z} = Sz, \tag{2.4}$$

the transformation matrix S satisfies

$$\dot{S} = J\bar{A}S - SJA \tag{2.5a}$$

$$SJS^T = J \tag{2.5b}$$

where J is the $2n\times 2n$ symplectic matrix and the require-

 0022-2488/80/010032-06$01.00

ment (2.5b), that S be a member of the symplectic group of matrices is the condition that the transformation be canonical. To avoid any misunderstanding we emphasize three points. Firstly, (2.5a) has a real unique solution for a given initial condition $S(t_0)$; secondly, if $S(t_0)$ is canonical, then $S(t)$ is canonical, and thirdly, as S is time-dependent, H and $\bar{H}$ are not to be assumed to be scalars of equal value.

For our primary purpose we are interested in the following Hamiltonians[10]:

$$H_a = \tfrac{1}{2}\mathbf{z}^T\mathbf{z}, \quad H_r = \tfrac{1}{2}\mathbf{z}^T R\mathbf{z}, \qquad H_s = \tfrac{1}{2}\mathbf{z}^T S\mathbf{z}, \quad H_f = \tfrac{1}{2}\mathbf{z}^T F\mathbf{z}, \tag{2.6}$$

where

$$R = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}, \quad S = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}. \tag{2.7}$$

The subscripts a, r, s, and f are used to describe the four classes of Hamiltonian in which we are interested. For the sake of convenience and because of its familarity we take H_a as our starting point. The transformation matrices from H_a to the others are[8]

$$H_a \to H_r: \begin{bmatrix} -(\dot{\alpha}e^t + \dot{\beta}e^{-t}), & (\alpha e^t + \beta e^{-t}) \\ -(\dot{\alpha}e^t + \dot{\beta}e^{-t}), & (\alpha e^t + \beta e^{-t}) \end{bmatrix}, \tag{2.8}$$

where

$$\alpha = A \sin t + B \cos t, \quad \beta = C \sin t + D \cos t, \tag{2.9}$$

and A, B, C, and D are constant matrices, as will be the case for the transformations below. They are arbitrary subject to the constraint of (2.5b).

$$H_a \to H_s: \begin{bmatrix} -\dot{\alpha}e^t, & \alpha e^t \\ -\dot{\beta}e^{-t}, & \beta e^{-t} \end{bmatrix}, \tag{2.10}$$

where

$$\alpha = B \cos t - A \sin t, \quad \beta = D \cos t - C \sin t. \tag{2.11}$$

$$H_a \to H_f: \begin{bmatrix} -\dot{\alpha}t + \dot{\beta}, & \alpha t - \beta \\ \dot{\alpha}, & \alpha \end{bmatrix}, \tag{2.12}$$

where

$$\alpha = A \cos t + B \sin t, \quad \beta = C \cos t + D \sin t. \tag{2.13}$$

The transformation matrices (2.8), (2.10), and (2.12) [with the expressions given for α and β in (2.9), (2.11), and (2.13), respectively] are the most general solutions of (2.5a). Under the constraint (2.5b) each provides the most general homogeneous linear canonical transformation from the homogeneous quadratic form H_a to the homogeneous quadratic forms H_r, H_s, and H_f, respectively. Just as we restrict our attention to homogeneous quadratic forms to avoid unnecessary complication, so also we restrict the transformations to homogeneous ones, as our interest is in quadratic invariants.

To reduce the amount of computational algebra it is convenient to select suitable matrices for A, B, C, and D. This does not reduce the generality of the results which we obtain. In particular, for $H_a \to H_r$ we use

$$A = -\tfrac{1}{2}I, \quad B = \tfrac{1}{2}I, \quad C = -\tfrac{1}{2}I, \quad D = -\tfrac{1}{2}I; \tag{2.14}$$

for $H_a \to H_s$ we use

$$A = I, \quad B = 0, \quad C = 0, \quad D = I; \tag{2.15}$$

and for $H_a \to H_f$ we use

$$A = I, \quad B = 0, \quad C = 0, \quad D = I. \tag{2.16}$$

3. THE QUADRATIC INVARIANTS OF H_a

The scalar quadratic form

$$I = \tfrac{1}{2}\mathbf{z}^T M\mathbf{z}, \tag{3.1}$$

in which M is a real symmetric $2n \times 2n$ matrix, is an invariant of H_a (2.6) provided

$$\mathbf{z}^T(JM + \tfrac{1}{2}\dot{M})\mathbf{z} = 0. \tag{3.2}$$

Writing M in block form as

$$M = \begin{bmatrix} U & W \\ W^T & V \end{bmatrix}, \tag{3.3}$$

where U and V are symmetric, (3.2) is equivalent to

$$\mathbf{q}^T(\tfrac{1}{2}\dot{U} - W)\mathbf{q} = 0, \quad \mathbf{p}^T(\tfrac{1}{2}\dot{V} + W^T)\mathbf{p} = 0, \qquad \mathbf{p}^T(\tfrac{1}{2}\dot{W}^T - V)\mathbf{q} + \mathbf{q}^T(\tfrac{1}{2}\dot{W} + U)\mathbf{p} = 0. \tag{3.4}$$

It follows that the elements of the matrices U, V, and W are the solutions of the differential equations

$$\begin{aligned} &\dot{U}_{ii} - 2W_{ii} = 0, \quad \dot{V}_{ii} + 2W_{ii} = 0, \\ &\dot{U}_{ij} = W_{ij} + W_{ji}, \quad \dot{V}_{ij} = -(W_{ij} + W_{ji}), \\ &\dot{W}_{ii} = -U_{ii} + V_{ii}, \quad \dot{W}_{ij} + \dot{W}_{ji} = -2U_{ij} + 2V_{ij} \end{aligned} \tag{3.5}$$

(no summation on repeated indices).

The solution set of (3.5) is found readily and is

$$\begin{aligned} &U = C + A \sin 2t + B \cos 2t, \\ &V = C - A \sin 2t - B \cos 2t, \\ &W = D + A \cos 2t - B \sin 2t. \end{aligned} \tag{3.6}$$

The arbitrary constant matrices A, B, and C are symmetric while D is skew-symmetric. The number of distinct arbitrary constants is $n(2n+1)$. Using (3.6) we may write the general quadratic invariant for H_a as

$$I = \tfrac{1}{2}\mathbf{z}^T\left\{\begin{bmatrix} C & D \\ -D & C \end{bmatrix} + \begin{bmatrix} A & -B \\ -B & -A \end{bmatrix}\sin 2t + \begin{bmatrix} B & A \\ A & -B \end{bmatrix}\cos 2t\right\}\mathbf{z}. \tag{3.7}$$

Given the existence of four distinct arbitrary matrices it is evident that there are four classes of invariant, each of which could be written down in a variety of ways. As there is a standard form for the angular momentum tensor and for the Jauch–Hill–Fradkin tensor, we use them and write down the other two classes in similar forms. Thus, we have

$$2L_{ij} = q_i p_j - q_j p_i, \tag{3.8}$$

$$2A_{ij} = p_i p_j + q_i q_j, \tag{3.9}$$

$$2C_{ij} = (p_i p_j - q_i q_j)\sin 2t - (q_i p_j + p_i q_j)\cos 2t, \tag{3.10}$$

$$2D_{ij} = (p_i p_j - q_i q_j)\cos 2t + (q_i p_j + p_i q_j)\sin 2t,, \tag{3.11}$$

In the usual notation, L_{ij} is the angular momentum tensor and the components of the angular momentum vector are given by

$$L_k = \epsilon_{ijk}L_{ij}. \tag{3.12}$$

The constants A_{ij} are the components of the Jauch–Hill–Fradkin tensor, while the C_{ij} and D_{ij} do not appear to have received attention and have no specific name as yet.

The properties of the L_{ij} and A_{ij} have received adequate attention elsewhere,[2,3,8,11] especially in relation to their providing a basis for the generators of the symmetry group of H_a, SU(n). Unlike the L_{ij} and A_{ij}, the C_{ij} and D_{ij} do not have zero Poisson Bracket (commutator) with H_a. Consequently, they do not contribute to a description of the symmetry group of H_a. We do note that the two sets of Poisson brackets (commutator) $[L_{ij}, C_{kl}]$ and $[L_{ij}, D_{kl}]$, $i,j,k,l = 1, n$, are closed. For the three-dimensional problem there are three linearly independent L_{ij}, six linearly independent A_{ij}, six linearly independent C_{ij} and six linearly independent D_{ij}, in all, 21 linearly independent quantities. The generators of the symmetry group of H_a SU(3), come from the nine L_{ij} and A_{ij}, As H_a itself is one of the linear combinations of the L_{ij} and A_{ij}, there remain eight other linearly independent combinations to provide the generators of SU(3).

4. THE QUADRATIC INVARIANTS OF H_r

From (2.8), (2.9), and (2.14), the transformation matrix for the transformation $H_a \rightarrow H_r$ is

$$\begin{bmatrix} I(\cosh t\cos t + \sinh t\sin t), & I(\sinh t\cos t - \cosh t\sin t) \\ I(\cosh t\sin t + \sinh t\cos t), & I(\cosh t\cos t - \sinh t\sin t) \end{bmatrix} \tag{4.1}$$

$$\Leftrightarrow \frac{\partial}{\partial t}\begin{bmatrix} \cosh t\sin t, & \cosh t\cos t \\ \sinh t\sin t, & \sinh t\cos t \end{bmatrix} \otimes I. \tag{4.2}$$

While the order of I is arbitrary, we take the systems to be three-dimensional and so the order as three. Applying the transformation (4.2), in the new coordinate system the invariants are[12]

$$2L_{ij} = q_i p_j - q_j p_i, \tag{4.3}$$
$$2A_{ij} = (p_i p_j + q_i q_j)\cosh 2t - (q_i p_j + p_i q_j)\sinh 2t, \tag{4.4}$$
$$2C_{ij} = (p_i p_j + q_i q_j)\sinh 2t - (q_i p_j + p_i q_j)\cosh 2t, \tag{4.5}$$
$$2D_{ij} = p_i p_j - q_i q_j. \tag{4.6}$$

The first observation which we make is that each of these quantities is a constant of the motion described by H_r. The second is that it is now the L_{ij} and D_{ij} which have zero Poisson brackets (quantum mechanically, commutators) with the Hamiltonian H_r. From our knowledge of the attractive oscillator, the interesting invariants will be L_{ij} and D_{ij}. We describe some of their classical and quantal properties below. We distinguish between those invariants which have a zero Poisson bracket (commutator) with the Hamiltonian and those which do not by terming the former *useful*.[13]

5. CLASSICAL PROPERTIES OF THE USEFUL H_r INVARIANTS

In the three-dimensional case the D_{ij} have the following properties:

$$D_{ij} = D_{ji}, \tag{5.1}$$
$$d(D_{ij})/dt = 0, \tag{5.2}$$
$$T_r(D_{ij}) = H_r, \tag{5.3}$$
$$D_{ii}D_{jj} - D_{ij}^2 = -L_{ij}^2 = -\tfrac{1}{4}(\epsilon_{ijk}L_k)^2 \quad \text{(summation on } k \text{ only)}, \tag{5.4}$$
$$D_{ij}L_j = 0 = L_i D_{ij}, \tag{5.5}$$
$$D_{ij}D_{jk} = H_r D_{ik} - L_{ij}L_{jk} = H_r D_{ik} + \tfrac{1}{4}(\mathbf{L}^2\delta_{ik} - L_i L_k), \tag{5.6}$$
$$\operatorname{adj}(D_{ij}) = L_{ik}L_{kj} = \tfrac{1}{4}(L_i L_j - \mathbf{L}^2\delta_{ij}), \tag{5.7}$$
$$q_i\{H_r\delta_{ij} - D_{ij}\}q_j = L_{ij}L_{ij} = \tfrac{1}{4}\mathbf{L}^2, \tag{5.8}$$
$$q_i\{H_r\delta_{ij} - D_{ij}\}p_j = 0, \tag{5.9}$$
$$p_i\{H_r\delta_{ij} - D_{ij}\}p_j = L_{ij}L_{ij} = \tfrac{1}{4}\mathbf{L}^2. \tag{5.10}$$

The resemblance between the properties of D_{ij} and those of the Jauch–Hill–Fradkin tensor is very close, the only difference being in (5.4) and (5.6), in which the angular momentum terms are of opposite sign.

The invariants D_{ij} play a role for H_r similar to the roles played by A_{ij} for H_a and the Runge–Lenz vector for the Kepler problem. In particular, from (5.5) we observe that the angular momentum vector is an eigenvector of the matrix $[D_{ij}]$ corresponding to the eigenvalue zero. Equation (5.8) is the orbit equation whose features follow from the eigenvalues of $[D_{ij}]$. Apart from the zero noted above, the eigenvalues are found [using (5.3) and (5.4)] to be

$$\lambda_{\pm} = \tfrac{1}{2}\{H_r \pm (H_r^2 + \mathbf{L}^2)^{1/2}\}. \tag{5.11}$$

As expected, the orbit equation is an hyperbola with the ratio of the lengths of the semiaxes given by $(\lambda_+/\lambda_-)^{1/2}$. The plane of the orbit is normal to the direction of the angular momentum vector.

In his paper Fradkin[3] noted that a complete description of a motion with a conserved quantity was possible only because of the periodic (or closed) nature of the system. It would appear that the statement must be amended to embrace the present situation in which the orbit is certainly not closed. The important property of such a conserved quantity is that it admits a quadratic form which satisfies an orbit equation of the type (5.8), i.e., elliptic or hyperbolic.[14]

6. THE SYMMETRY GROUP OF THE USEFUL H_r INVARIANTS

The similarity of the algebraic properties of the H_r invariants D_{ij} to those of the Jauch–Hill–Fradkin tensor suggests that they, together with the L_{ij}, provide a basis for the generators of the dynamical symmetry group for H_r. In a quantum mechanical context we define the operators (following Fradkin)

$$\begin{aligned} D_0 &= (2D_{33} - D_{11} - D_{22}), \\ D_\epsilon &= -\epsilon(D_{13} + i\epsilon D_{23}), \quad \epsilon = \pm 1, \\ D_{2\epsilon} &= (D_{11} - D_{22} + 2i\epsilon D_{12}), \quad \epsilon = \pm 1, \end{aligned} \tag{6.1}$$

and use the angular momentum operators L_3 and

$$L_\epsilon = L_1 + i\epsilon L_2, \quad \epsilon = \pm 1. \tag{6.2}$$

All of these operators commute with H_r. Amongst each other the commutation relations are

$$\begin{aligned} &[L_3, D_0] = [D_0, D_{2\epsilon}] = [D_\epsilon, D_{2\epsilon}] = [L_\epsilon, D_{2\epsilon}] = 0, \\ &[L_\epsilon, L_{-\epsilon}] = [D_\epsilon, D_{-\epsilon}] = -\tfrac{1}{2}[D_{2\epsilon}, D_{-2\epsilon}] = 2\epsilon\hbar L_3, \\ &[L_\epsilon, D_{-\epsilon}] = \hbar D_0, \\ &\epsilon[L_3, L_\epsilon] = \tfrac{2}{3}[D_0, D_\epsilon] = -[D_{-\epsilon}, D_{2\epsilon}] = \hbar L_\epsilon, \end{aligned} \tag{6.3}$$

$$\epsilon[L_3,D_\epsilon] = \tfrac{1}{6}[L_\epsilon,D_0] = \tfrac{1}{4}[L_{-\epsilon},D_{2\epsilon}] = \hbar D_\epsilon,$$
$$\epsilon[L_3,D_{2\epsilon}] = 2[L_\epsilon,D_\epsilon] = 2\hbar D_{2\epsilon}.$$

These commutation relations are not those characteristic of the operators of the SU(3) group. In particular the results for $[D_\epsilon,D_{-\epsilon}]$, $[D_{2\epsilon},D_{-2\epsilon}]$, $[D_0,D_\epsilon]$ and $[D_{-\epsilon},D_{2\epsilon}]$ differ by a sign. This difference may be overcome by defining

$$D_0' = iD_0, \quad D_\epsilon' = iD_\epsilon, \quad D_{2\epsilon}' = iD_{2\epsilon}. \tag{6.4}$$

Replacing unprimed by primed operators in the commutators of (6.3), the relations become those characteristic of the operators which provide a suitable basis for SU(3). The change of basis in (6.4) is called the Weyl unitary trick.[15] In the new basis the symmetry group is SU(3) and so in the original basis it is SU(2,1), as would be expected for H_r. Thus what was a nonuseful invariant for H_a is a useful invariant for H_r.

7. THE QUADRATIC INVARIANTS OF H_s

From (2.10), (2.11), and (2.15) the matrix of the transformation $H_a \rightarrow H_s$ is

$$\begin{bmatrix} \cos te^t, & -\sin te^t \\ \sin te^{-t}, & \cos te^{-t} \end{bmatrix} \otimes I. \tag{7.1}$$

As before, although the system can be of arbitrary order, we restrict ourselves to a three-dimensional system. Under the transformation (7.1), the invariants become

$$2L_{ij} = q_i p_j - q_j p_i, \tag{7.2}$$
$$2A_{ij} = q_i q_j e^{-2t} + p_i p_j e^{2t}, \tag{7.3}$$
$$2C_{ij} = -(q_i p_j + p_i q_j), \tag{7.4}$$
$$2D_{ij} = p_i p_j e^{2t} - q_i q_j e^{-2t}. \tag{7.5}$$

Each of these quantities is an invariant of the system described by H_s. Those which have zero Poisson bracket with H_s and L_{ij} and C_{ij}, which contrasts with L_{ij} and A_{ij} for H_a and L_{ij}, and D_{ij} for H_r. It is this pair of invariant classes which provides the interesting information about the dynamical symmetry of H_s.

8. PROPERTIES OF THE USEFUL INVARIANTS OF H_s

As it is the tensor invariant C_{ij} which provides the additional information about the motion described by H_s, we concentrate on its properties. Classically, these are

$$C_{ij} = C_{ji}, \tag{8.1}$$
$$d(C_{ij})/dt = 0, \tag{8.2}$$
$$T_r(C_{ij}) = -H_s, \tag{8.3}$$
$$C_{ii}D_{jj} - C_{ij}^2 = -L_{ij}^2 = -\tfrac{1}{4}(\epsilon_{ijk}L_k)^2 \tag{8.4}$$

(summation on k only),

$$C_{ij}L_j = 0 = L_i C_{ij}, \tag{8.5}$$
$$C_{ij}C_{jk} = H_s C_{ik} - L_{ij}L_{jk} = H_s C_{ik} + \tfrac{1}{4}\{\mathbf{L}^2\delta_{ik} - L_i L_k\}, \tag{8.6}$$
$$\mathrm{adj}(C_{ij}) = L_{ik}L_{kj} = \tfrac{1}{4}\{L_i L_j - \mathbf{L}^2\delta_{ij}\}, \tag{8.7}$$
$$q_i(H_s\delta_{ij} + C_{ij})q_j = 0, \tag{8.8}$$
$$q_i(H_s\delta_{ij} + C_{ij})p_j = -\tfrac{1}{4}\mathbf{L}^2, \tag{8.9}$$
$$p_i(H_s\delta_{ij} + C_{ij})p_j = 0. \tag{8.10}$$

These properties, with the exception of the last three, are essentially those of the D_{ij} given in Sec. 5. The eigenvalues of the matrix $[C_{ij}]$ are

$$\lambda = 0, \quad \tfrac{1}{2}\{-H \pm (H^2 + \mathbf{L}^2)^{1/2}\}, \tag{8.11}$$

which, as would be expected, indicate hyperbolic motion in a plane normal to the direction of the angular momentum vector.

What then is the difference between the properties of C_{ij} vis á vis H_s and D_{ij} vis á vis H_r? Essentially none, as is easily seen from the properties of the quantum mechanical operators

$$C_0 = (2C_{33} - C_{11} - C_{22}),$$
$$C_\epsilon = -\epsilon(C_{13} + i\epsilon C_{23}), \quad \epsilon = \pm 1, \tag{8.12}$$
$$C_{2\epsilon} = (C_{11} - C_{22} + 2i\epsilon C_{12}), \quad \epsilon = \pm 1.$$

We find that the commutation relations amongst these C's and with the angular momentum operators L_3 and L_ϵ are exactly the same as those given for H_r in Eqs. (6.3). It follows immediately that the symmetry group of H_s is SU(2,1). The role played by the C_{ij} for H_s is the same as that played by the D_{ij} for H_r. We emphasize, however, that the C_{ij} and D_{ij} are not the same invariants.

9. THE QUADRATIC INVARIANTS OF H_f

From (2.12), (2.13), and (2.16), the matrix of the transformation $H_a \rightarrow H_f$ is

$$\begin{bmatrix} t\sin t + \cos t, & t\cos t - \sin t \\ \sin t, & \cos t \end{bmatrix} \otimes I. \tag{9.1}$$

Under the transformation (9.1) the invariants become

$$2L_{ij} = q_i p_j - q_j p_i, \tag{9.2}$$
$$2A_{ij} = q_i q_j + (1 + t^2)p_i p_j - t(q_i p_j + p_i q_j), \tag{9.3}$$
$$2C_{ij} = 2tp_i p_j - q_i p_j - p_i q_j, \tag{9.4}$$
$$2D_{ij} = (1 - t^2)p_i p_j - q_i q_j + t(q_i p_j + p_i q_j). \tag{9.5}$$

We observe that only L_{ij} has zero Poisson bracket with H_f. As they stand the other invariants are of no relevance in describing the invariance properties of H_f.[16] However, defining

$$2E_{ij} = A_{ij} + D_{ij} = p_i p_j, \tag{9.6}$$

we observe the following classical properties:

$$E_{ij} = E_{ji}, \quad \frac{d(E_{ij})}{dt} = 0,$$
$$T_r(E_{ij}) = H_f, \quad E_{ii}E_{jj} - E_{ij}E_{ji} = 0 \quad \text{(no summation)}, \tag{9.7}$$
$$E_{ij}L_j = 0 = L_i E_{ij}, \quad E_{ij}E_{jk} = H_f E_{ik}, \quad \mathrm{adj}(E_{ij}) = 0,$$
$$q_i(H_f\delta_{ij} - E_{ij})q_j = \tfrac{1}{4}\mathbf{L}^2, \quad q_i(H_f\delta_{ij} - E_{ij})p_j = 0,$$
$$p_i(H_f\delta_{ij} - E_{ij})p_j = 0,$$

which reflect some of those of the Jauch–Hill–Fradkin tensor. We note in particular that the eigenvalues of the matrix $[E_{ij}]$ are O, O, and H and that $\mathbf{L}$ is an eigenvector.

However, were we to define an E_0, E_ϵ, and $E_{2\epsilon}$ in the usual way, the commutation relations are not those of the generators of either SU(3) or SU(2,1). The invariant E_{ij} does not provide any increase in the symmetry of H_f.

10. COMMENT

Before making some remarks about the general quadratic Hamiltonian, we have a few comments to make on some aspects of the preceding sections especially with respect to quantum mechanics. For the three-dimensional case each of the Hamiltonians considered has 21 linearly independent quadratic constants of the motion. Only those which have been termed useful play a role in providing a basis for the generators of the appropriate symmetry group. It is only these which commute with the Hamiltonian under consideration. Each Hamiltonian is Hermitian. This is not the case for all of the invariants. As an example consider the C and D invariants of H_a. In the quantum mechanical context it is appropriate to replace sines and cosines by the imaginary exponential. We define two invariants as

$$\begin{aligned} 2U_{ij} = 2iC_{ij} + 2D_{ij} &= (p_i p_j - q_i q_j)e^{2it} - i(q_i p_j + p_i q_j)e^{2it} \\ &= (F_{ij} - iG_{ij})e^{2it}, \end{aligned} \tag{10.1}$$

$$2V_{ij} = -2iC_{ij} + 2D_{ij} = (F_{ij} + iG_{ij})e^{-2it}. \tag{10.2}$$

Since F_{ij} and G_{ij} are Hermitian, it is obvious that U_{ij} and V_{ij} are not, and in fact

$$U_{ij}^\dagger = V_{ji}, \quad V_{ij}^\dagger = U_{ji}. \tag{10.3}$$

This situation, however, does not apply generally. For H_r the explicitly time-dependent invariants, when written in matrix form, are Hermitian. For H_s they are symmetric and, for H_f, two classes form Hermitian matrices and the third a symmetric matrix. Those which do form a symmetric matrix are individually Hermitian.

If we examine the time development of the various quadratic invariants using Heisenberg's equation (c.f. Ref. 7, p. 501)

$$\tilde{O}(t) = \exp(-itH)\tilde{O}(0)\exp(itH), \tag{10.4}$$

we find that those invariants which do not contain t explicitly satisfy the equation trivially (since they commute with the appropriate H). For those which contain t explicitly, (10.4) is still satisfied in all cases. Only the algebra is more difficult. To illustrate the result we consider the algebraically simplest case, that for H_f. We have

$$\begin{aligned} &H_f = \tfrac{1}{2}p_k p_k, \\ &2A_{ij}(t) = q_i q_j + (1+t^2)p_i p_j - t(q_i p_j + p_i q_j), \\ &2C_{ij}(t) = 2tp_i p_j - (q_i p_j + p_i p_j), \\ &2D_{ij}(t) = (1-t^2)p_i p_j - q_i q_j + t(q_i p_j + p_i q_j), \end{aligned} \tag{10.5}$$

which, for the sake of brevity, we rewrite as

$$\begin{aligned} &H_f = \tfrac{1}{2}\beta_{kk} \quad 2A_{ij}(t) = \alpha_{ij} + (1+t^2)\beta_{ij} - t\gamma_{ij}, \\ &2C_{ij}(t) = 2t\beta_{ij} - \gamma_{ij}, \\ &2D_{ij}(t) = -\alpha_{ij} + (1-t^2)\beta_{ij} + t\gamma_{ij}. \end{aligned} \tag{10.6}$$

The commutation relations between H_f and α_{ij}, β_{ij}, and γ_{ij}, respectively, are

$$[H, \alpha_{ij}] = -i\gamma_{ij}, \quad [H, \beta_{ij}] = 0, \quad [H, \gamma_{ij}] = -2i\beta_{ij}. \tag{10.7}$$

It is relatively easy to show that

$$\begin{aligned} &H^n \alpha_{ij} = \alpha_{ij}H^n - in\gamma_{ij}H^{n-1} - n(n-1)\beta_{ij}H^{n-2}, \\ &H^n\beta_{ij} = \beta_{ij}H^n, \quad H^n\gamma_{ij} = \gamma_{ij}H^n - 2in\beta_{ij}H^{n-1}. \end{aligned} \tag{10.8}$$

Using (10.8), it is just a matter of algebra to verify that (10.4) is satisfied by each invariant in (10.6)

With these further considerations it is clear that the properties of the quadratic invariants as invariants are not affected by the presence or otherwise of explicit time dependence. It is only in relation to the Hamiltonian that there is any marked difference between explicitly time-dependent and time-independent quadratic invariants. The latter invariants, which we have termed useful, provide information about the invariance symmetry of the Hamiltonian. The former do not.

11. THE GENERAL QUADRATIC HAMILTONIAN

As was noted in Sec. 2, the homogeneous quadratic Hamiltonian

$$H = \tfrac{1}{2}\mathbf{z}^T A\mathbf{z} \tag{11.1}$$

may be transformed to

$$\bar{H} = \tfrac{1}{2}\bar{\mathbf{z}}^T\bar{A}\bar{\mathbf{z}} \tag{11.2}$$

by a linear canonical transformation. When the transformation is time-dependent, there need be no relationship between the natures of the matrices A and $\bar{A}$. The few simple examples given in Sec. 2 amply demonstrate this. The theory of systems of linear differential equations guarantees a continuous solution of (2.5) under minimal restraints on the elements of A and $\bar{A}$.

Given an arbitrary quadratic Hamiltonian the problem of determining the nature of the motion is essentially one of reducing it, by means of transformation, to a recognizable form. A collection of difficulties occurs to make this a nontrivial problem. Suppose that A is time independent. It does not follow that it can be diagonalized by a time-independent transformation. A symmetric matrix is diagonalized by an orthogonal transformation whereas a canonical transformation is symplectic. The two coincide for one-dimensional problems, but only accidentally for problems of higher dimension. The best which one can do with a time-independent transformation is to convert the Hamiltonian to normal form.[17] The normal form may contain combinations of H_a, H_r and H_f together with some other forms not discussed here. While we can still make progress with such combinations, it would be better if a simpler approach could be used.

When we turn to time-dependent quadratic Hamiltonians, the use of time-independent linear transformations is not likely to be of much value as the resulting Hamiltonian will still be time dependent. Hence, it is rather more useful to use a time-dependent transformation. However, two difficulties arise. Firstly, H and $\bar{H}$ are no longer equal and any discussion of the symmetry of $\bar{H}$ is in fact a discussion of the symmetry of the invariant I associated with the system described by H. The invariant I is simply $\bar{H}$ expressed in terms of the original coordinates. The second and more serious difficulty, which applies to both time-dependent and time-independent systems, is choosing the appropriate form for $\bar{H}$, since any choice of $\bar{H}$ may be attained classically under a time-dependent transformation. We saw in an earlier section what happened to the useful invariants of H_a under time-dependent transformations. Then it was easy enough to see which constants were useful. When H, I, and the associated invariants of I contain time explicitly, we shall distinguish between useful and nonuseful invariants by their having zero or nonzero Poisson brackets (or commutators in quantum mechanics) with $I\,(\Leftrightarrow\bar{H})$.

When the useful invariants of I have been identified, we have the further problem of deciding whether they and I are the appropriate invariants for H. This problem is closely related to that of the normal form. Classically, the choice of $\bar{H}$ is immaterial for solving the problem (in theory at least, some choices could be impractical). Quantum mechanically, the choice is not so free. In an earlier paper[11] we saw that the energy eigenvalues of the three-dimensional anisotropic oscillator could be obtained by integral transform from those of the corresponding isotropic oscillator. However, the same procedure applied to the free particle using a classically acceptable transformation produces nonsense. Clearly, $\bar{H}$ should describe a system which is qualitatively the same as that of H.

Were it possible to diagonalize A with a time-independent linear canonical transformation, there would be no problem. With A diagonalized, the classical and quantum treatment is trivial. We are faced with two tasks. The first is to identify the relationship between the normal and diagonal form of a given Hamiltonian. The second is to ensure that any time-dependent transformation does not alter the nature of the Hamiltonian. We hope to report on these matters shortly.

ACKNOWLEDGMENTS

Correspondence from K. Mariwaller (Madras) and K.B. Wolf (Mexico) prompted some of the thoughts behind this paper. I thank them for their interest and accept full responsibility for the outcome of their prompting. I also thank R. Rimmer (La Trobe) for useful discussions and the referee for several constructive suggestions.

[1] P.G.L. Leach, *The invariants of quadratic Hamiltonians: linear and quadratic invariants for the time-independent isotropic harmonic oscillator* (preprint, Department of Applied Mathematics, La Trobe University, 1978).

[2] J.M. Jauch and E.H. Hill, Phys. Rev. 57, 641–5 (1940).

[3] D.M. Fradkin, Am. J. Phys. **33**, 207–11 (1965).

[4] Partly prompted by private communications from K. Mariwaller (Madras) and K.B. Wolf (Mexico).

[5] Indeed we have made extensive use of them, see P.G.L. Leach, J. Math. Phys. **20**, 86–95 (1978); **20**, 96–100 (1979); **21**, 38–43 (1980).

[6] Cf. L. Van Hove, Acad. Roy. des Sci. des Lett. et des Beaux-Arts Belg. Classe des Sci. Bull., Ser. 5 **37**, 610–29 (1951).

[7] The use of nonlinear transformations and nonquadratic invariants can lead to a larger maximal kinematical group, cf., for the massive free Schrödinger equation, A.O. Barut, Helv. Phys. Acta **46**, 496–503 (1973).

[8] P.G.L. Leach, J. Math. Phys. **18**, 1608–11 (1977).

[9] Mathematically there is no need for this restriction, but, given the real origin of such Hamiltonians, the mathematical generalization seems superfluous.

[10] To obviate unnecessary symbols, mass and coupling constants are taken as unity. There is no essential simplification.

[11] P.G.L. Leach, J. Math. Phys. **19**, 446–51 (1978).

[12] No attempt is made to distinguish between differing coordinate systems as there will be four in all and the distinguishing markings would render the expressions less comprehensible.

[13] The use of a zero Poisson bracket with the Hamiltonian as a criterion of utility is sufficient for time-independent problems. For time-dependent Hamiltonians there would be two classes of useful invariants, those with zero Poisson bracket with the Hamiltonian and those with zero Poisson bracket with the invariant of the problem.

[14] The reader is referred to the discussion of a time-dependent problem in N.J. Günther and P.G.L. Leach, J. Math. Phys. **18**, 572–6 (1977). The orbit equations need not describe an ellipse or hyperbola of constant semiaxes lengths

[15] R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974), p. 43.

[16] Although they are relevant to the noninvariance symmetry associated with H_f, cf. Leach.[8]

[17] Cf. J. Williamson, Am. J. Math. **59**, 141 (1936), who gives a complete classification for two-dimensional systems; see also A. Ciampi, Indiana U. Math. J. **23**, 513–26 (1973).

A note on the Hénon–Heiles problem

P. G. L. Leach
Department of Applied Mathematics, La Trobe University, Bundoora, 3083, Australia

(Received 27 July 1978)

The Hamiltonian for the Hénon–Heiles problem, $H=(1/2)(p_1^2+p_2^2+q_1^2+q_2^2)+q_1^2q_2-(1/3)q_2^3$, is a particular example of time-independent Hamiltonians for two-dimensional oscillator systems with third degree anharmonicity. It has been used as a model for galactic motion. There has been much discussion of the possible existence of an integral other than the Hamiltonian. In this note we show that the Hénon–Heiles Hamiltonian in particular and the class in general does not possess an invariant series which is explicitly time-independent other than the Hamiltonian itself.

1. INTRODUCTION

In recent years there has been a considerable body of literature devoted to the constants of the motion described by quadratic Hamiltonians.[1] Much of the practical motivation for this work is found in the motion of a charged particle in an electromagnetic field. In the case of an axially symmetric field this problem, in the first approximation, has as Hamiltonian

$$H=\tfrac{1}{2}p^2+\tfrac{1}{2}w^2q^2, \tag{1.1}$$

where p and q are the canonically conjugate variables and w is usually time-dependent in the practical context. In this case, the Hamiltonian (1.1) has been successfully tackled from two viewpoints. The first, using Kruskal's method,[2] was the construction of an adiabatic invariant series which proved to be exact.[3] The very fact of this result promoted further work which led to the second approach. This was to tackle the problem from the viewpoint of time-dependent linear canonical transformations.[4]

A natural extension of this work has been to systems of higher dimension[5] with the Hamiltonian

$$H=\tfrac{1}{2}\mathbf{z}^TA\mathbf{x}+\mathbf{B}^Tz+c, \tag{1.2}$$

$$\begin{aligned} z^\nu &= q_i, \quad \nu=1,n,\ i=1,n,\\ z^\nu &= p_i, \quad \nu=n+1,2n,\ i=1,n \end{aligned} \tag{1.3}$$

and the $2n\times 2n$ real symmetric matrix A, the $2n$-vector $\mathbf{B}$, and the scalar C may be time-dependent. In particular it was seen that there existed n^2 linearly independent quadratic constants for H (1.2).[6] One of these is the constant which takes the place of the Hamiltonian and which may be interpreted as the Hamiltonian of a time-independent isotropic harmonic oscillator expressed in transformed coordinates, the actual transformation being that which transforms the isotropic Hamiltonian to H (1.2).[7]

In view of the success which attended the quadratic approximation, it is natural that investigation of anharmonic systems has been undertaken. Because the problem is founded in motion in a time-dependent electromagnetic field, the coefficients of the Hamiltonian have been taken as time-dependent.[8] For the particular case of

$$H=\tfrac{1}{2}p^2+\tfrac{1}{2}w^2(t)q^2+\lambda(t)q^3 \tag{1.4}$$

sufficient work[9] was done for a formal invariant series to be written down. The structure of that series was such that the method would apply to Hamiltonians with different degrees of anharmonicity. It should be emphasized that this series is not an adiabatic invariant series. Work on such a series has been done by Ottoy.[10]

A natural extension to the problems considered above is the multidimensional anharmonic problem. Immediately we come into conjunction with a different field of study, that of celestial mechanics, in which a model for galactic motion has been taken to be a two-dimensional oscillator with third-order anharmonic terms. Admittedly the model Hamiltonian is time-independent, but, from the work on quadratic Hamiltonians, it may be believed with some justification that an understanding of the time-independent problem will provide some insights for the understanding of the time-dependent problem.

One model which has been found to be useful is that of Hénon–Heiles.[11] Numerical work on this model[12] has suggested that, for low energies at least, there exists a third[13] isolating integral.[14] The numerical work also suggests that this integral either ceases to exist or ceases to be isolating for sufficiently high values of the energy. As far as we know, there has been no determination of an analytic form for the third integral, although there has been work on an asymptotic integral.[15] One of the difficulties with this type of problem is that there are opinions against the existence of a third integral[16] and opinions for its existence.[17] The case against such existence has been made in the generic case by Markus and Meyer.[18]

While neither we, nor any of the writers whom we quote, claim that the computer evidence can be regarded in any way as proof, it is suggestive. In particular the discussion of Hénon and Heiles does suggest an integral in series form with either a limited range of convergence or which is isolating for low energies and becomes nonisolating for higher energies.

The Hamiltonian of the Hénon–Heiles problem is time-independent. In this note we shall explore the possible existence of time-independent integrals which may be expressed as series.

 0022-2488/80/010038-06$01.00

2. FORMULATION OF THE INVARIANT SERIES

The Hamiltonian of the problem considered by Hénon and Heiles is

$$H = \tfrac{1}{2}(q_1^2 + q_2^2 + p_1^2 + p_2^2) + q_1^2 q_2 - \tfrac{1}{3}q_2^3, \tag{2.1}$$

where q_1 and q_2 are the canonical coordinates and p_1 and p_2 the canonically conjugate momenta. As our interest is more general, we shall consider

$$H = H^{(2)} + H^{(3)}, \tag{2.2}$$

where

$$H^{(2)} = \tfrac{1}{2}(q_1^2 + q_2^2 + p_1^2 + p_2^2), \tag{2.3}$$

$$H^{(3)} = \sum_{i=1}^{3} A_i q_1^i q_2^{3-i}, \tag{2.4}$$

the A_i being real, time-independent, scalars. There is no real reduction in generality in writing the coefficients of both q_1^2 and q_2^2 as unity.

We formally define an invariant series for H [(2.2)] as

$$I = I^{(2)} + I^{(3)} + \cdots, \tag{2.5}$$

where $I^{(j)}$ is a homogeneous polynomial of degree j in the canonical variables. In line with our comment above, I is taken as time-independent and so is an invariant or integral if

$$[I,H]_{\mathrm{PB}} = 0 \tag{2.6}$$

$$\Leftrightarrow \begin{cases} [I^{(2)},H^{(2)}]_{\mathrm{PB}} = 0, & (2.7\mathrm{a}) \\ [I^{(j)},H^{(2)}]_{\mathrm{PB}} = -[I^{(j-1)},H^{(3)}]_{\mathrm{PB}}, \quad j = 3,\cdots. & (2.7\mathrm{b}) \end{cases}$$

Equations (2.7) define a recursion method for obtaining the coefficients of the higher terms of the invariant series from those of $I^{(2)}$.

3. THE INVARIANT TERM $I^{(2)}$

Writing $I^{(2)}$ as

$$I^{(2)} = \sum_{r_1 + r_2 = 2} \sum_{j_1=0}^{r_1} \sum_{j_2=0}^{r_2} B_{j_1 j_2}^{r_1 r_2} q_1^{j_1} q_2^{j_2} p_1^{r_1 - j_1} p_2^{r_2 - j_2}, \tag{3.1}$$

(2.7a) is equivalent to the homogeneous system of equations

$$M^{(2)}\mathbf{u}^{(2)} = 0, \tag{3.2}$$

where the 8×8 matrix $M^{(2)}$ may be written in block form

$$M^{(2)} = \begin{bmatrix} K_0^2 & 0 & 0 \\ 0 & K_1^2 & 0 \\ 0 & 0 & K_2^2 \end{bmatrix}. \tag{3.3}$$

The matrices K_0^2, K_1^2, and K_2^2 are given by

$$K_0^2 = G_0^2 = G_2^2 = K_2^2, \tag{3.4}$$

$$K_1^2 = \begin{bmatrix} G_1^2 & I \\ -I & G_1^2 \end{bmatrix}, \tag{3.5}$$

where

$$G_0^2 = \begin{bmatrix} 0 & 1 & 0 \\ -2 & 0 & 2 \\ 0 & -1 & 0 \end{bmatrix}, \tag{3.6}$$

$$G_1^2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{3.7}$$

The G matrices are skew-centrosymmetric and have zero or pure imaginary eigenvalues (see Leach[8]; also Clement[19]). The eigenvalues of the K matrices are 0, $\pm 2i$ for K_0^2 and K_2^2 and 0(2), $\pm 2i$ for K_1^2, the (2) indicating a double root.

The vector $\mathbf{u}^{(2)}$ consists of the coefficients $B_{j_1 j_2}^{r_1 r_2}$ arranged in the following order: $r_1 = 2, j_1 = 0,1,2$; $r_1 = 1$, $r_2 = 1, j_2 = 0, j_1 = 0,1$; $r_1 = 1$, $r_2 = 1$, $j_1 = 0$, $j_2 = 0,1$; $r_2 = 2$, $j_2 = 0,1,2$. For subsequent invariants the coefficients will be arranged in similar order.

The homogeneous system (3.2) has rank four (= order 8-4 zero eigenvalues) and its solution may be expressed in terms of four arbitrary constants. Thus $I^{(2)}$ is a linear combination of four quadratic terms. There is no need to calculate them here as the time-independent quadratic invariants of $H^{(2)}$ are well known (cf. Jauch and Hill,[1] Fradkin,[1] and Leach[7]). Indeed the purpose of the development above was to indicate the flavor of the work below in the case of the higher order terms of I. We have

$$\begin{aligned} I^{(2)} = {} & B_1(q_1^2 + p_1^2) + B_2(q_1 q_2 + p_1 p_2) \\ & + B_3(q_2^2 + p_2^2) + B_4(q_1 p_2 - q_2 p_1), \end{aligned} \tag{3.8}$$

the first three being the components of the Jauch–Hill–Fradkin tensor and the fourth angular momentum.

For the present we make no selection of which combination of the four invariants we shall use to obtain the higher order terms. However, we note that the choice

$$B_1 = B_3, \quad B_2 = 0 = B_4 \tag{3.9}$$

results in $I^{(2)}$ being a scalar multiple of $H^{(2)}$ and $I^{(j)}, j > 3$, being zero. In this case I is simply a scalar multiple of H. As an integral of the motion must be independent of H for it to be isolating, this I is of no value in the search for such an integral.

4. THE INVARIANT TERM $I^{(3)}$

Proceeding in a similar fashion to that of Sec. 3, we write

$$I^{(3)} = \sum_{r_1 + r_2 = 3} \sum_{j_1=0}^{r_1} \sum_{j_2=0}^{r_2} C_{j_1 j_2}^{r_1 r_2} q_1^{j_1} q_2^{j_2} p_1^{r_1 - j_1} p_2^{r_2 - j_2}, \tag{4.1}$$

and, from (2.7b), with suitable rearrangement, obtain

$$M^{(3)}\mathbf{u}^{(3)} = \mathbf{v}^{(3)}, \tag{4.2}$$

where $M^{(3)}$ is the block diagonal matrix

$$M^{(3)} = [K_i^3], \quad i = 0,3, \tag{4.3}$$

with

$$K_0^3 = G_0^3 = G_3^3 = K_3^3, \tag{4.4}$$

$$K_1^3 = \begin{bmatrix} G_1^3 & I_3 \\ -I_3 & G_1^3 \end{bmatrix} = K_2^3, \tag{4.5}$$

$$G_0^3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -3 & 0 & 2 & 0 \\ 0 & -2 & 0 & 3 \\ 0 & 0 & -1 & 0 \end{bmatrix},$$

$$G_1^3 = \begin{bmatrix} 0 & 1 & 0 \\ -2 & 0 & 2 \\ 0 & -1 & 0 \end{bmatrix}. \tag{4.6}$$

The vector $\mathbf{u}^{(3)}$ consists of the coefficients $C^{r_1r_2}_{j_1j_2}$ in an ordering corresponding to that of the $B^{r_1r_2}_{j_1j_2}$ used above. The nonhomogeneous term $\mathbf{v}^{(3)}$ contains the coefficients of $-[I^{(2)},H^{(3)}]_{\mathrm{PB}}$ in the same ordering as that of the $C^{r_1r_2}_{j_1j_2}$. As the eigenvalues of $M^{(3)}$ are all nonzero [for K^3_0 and K^3_3, $\pm i$, $\pm 3i$ and for K^3_1 and K^3_2, $\pm i(2)$, $\pm 3i$], it has an inverse and

$$\mathbf{u}^{(3)} = (M^{(3)})^{-1}\mathbf{v}^{(3)}. \tag{4.7}$$

Since $M^{(3)}$ and so $M^{(3)-1}$ are block matrices, it is convenient to divide $\mathbf{u}^{(3)}$ and $\mathbf{v}^{(3)}$ into corresponding subvectors so that (4.7) becomes

$$\mathbf{u}^{(3)}_i = (K^3_i)^{-1}\mathbf{v}^{(3)}_i, \quad i = 0,3. \tag{4.8}$$

Calculation of the Poisson bracket of $I^{(2)}$ with $H^{(3)}$ gives

$$\mathbf{v}^{(3)}_0 = \begin{bmatrix} 0 \\ 0 \\ 6B_1A_3 + B_2A_2 \\ B_4A_2 \end{bmatrix}, \quad \mathbf{v}^{(3)}_3 = \begin{bmatrix} 0 \\ 0 \\ B_2A_1 + 6B_3A_0 \\ -B_4A_1 \end{bmatrix},$$

$$\mathbf{v}^{(3)}_1 = \begin{bmatrix} 0 \\ 0 \\ 3B_2A_3 + 2B_3A_2 \\ 0 \\ 4B_1A_2 + 2B_2A_1 \\ 2B_4A_1 - 3B_4A_3 \end{bmatrix},$$

$$\mathbf{v}^{(3)}_2 = \begin{bmatrix} 0 \\ 0 \\ 2B_1A_1 + 3B_2A_0 \\ 0 \\ 2B_2A_2 + 4B_3A_1 \\ 3B_4A_0 - 2B_4A_2 \end{bmatrix}. \tag{4.9}$$

Substituting (4.9) into (4.8), we obtain

$$\mathbf{u}^{(3)}_0 = \frac{1}{3}\begin{bmatrix} -2B_4A_2 \\ 0 \\ -3B_4A_2 \\ 6B_1A_3 + B_2A_2 \end{bmatrix}, \quad \mathbf{u}^{(3)}_3 = \frac{1}{3}\begin{bmatrix} 2B_4A_1 \\ 0 \\ 3B_4A_1 \\ B_2A_1 + 6B_3A_0 \end{bmatrix},$$

$$\mathbf{u}^{(3)}_1 = \frac{1}{3}\begin{bmatrix} -4B_4A_1 + 6B_4A_3 \\ -6B_2A_3 - 4B_3A_2 + 4B_1A_2 + 2B_2A_1 \\ -2B_4A_1 + 3B_4A_3 \\ 6B_2A_3 + 4B_3A_2 - 4B_1A_2 - 2B_2A_1 \\ -4B_4A_1 + 6B_4A_3 \\ 3B_2A_3 + 2B_3A_2 + 4B_1A_2 + 2B_2A_1 \end{bmatrix},$$

$$\mathbf{u}^{(3)}_2 = \frac{1}{3}\begin{bmatrix} -6B_4A_0 + 4B_4A_2 \\ -4B_1A_1 - 6B_2A_0 + 2B_2A_2 + 4B_3A_1 \\ -3B_4A_0 + 2B_4A_2 \\ 4B_1A_1 + 6B_2A_0 - 2B_2A_2 - 4B_3A_1 \\ -6B_4A_0 + 4B_4A_2 \\ 2B_1A_1 + 3B_2A_0 + 2B_2A_2 + 4B_3A_1 \end{bmatrix}. \tag{4.10}$$

5. THE INVARIANT TERM $I^{(4)}$

Again proceeding as before, but using $D^{r_1r_2}_{j_1j_2}$ instead of $C^{r_1r_2}_{j_1j_2}$, the coefficients of $I^{(4)}$ are required to satisfy the equation

$$M^{(4)}\mathbf{u}^{(4)} = \mathbf{v}^{(4)}, \tag{5.1}$$

where $M^{(4)}$ is the block diagonal matrix

$$M^{(4)} = [K^4_i], \quad i = 0,4, \tag{5.2}$$

and

$$K^4_0 = G^4_0 = G^4_4 = K^4_4, \tag{5.3}$$

$$K^4_1 = \begin{bmatrix} G^4_1 & I_3 \\ -I_3 & G^4_1 \end{bmatrix} = \begin{bmatrix} G^4_3 & I_3 \\ -I_3 & G^4_3 \end{bmatrix} = K^4_3, \tag{5.4}$$

$$K^4_2 = \begin{bmatrix} G^4_2 & I_2 & 0 \\ -2I_2 & G^4_2 & 2I_2 \\ 0 & -I_2 & G^4_2 \end{bmatrix}. \tag{5.5}$$

The G matrices are

$$G^4_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -4 & 0 & 2 & 0 & 0 \\ 0 & -3 & 0 & 3 & 0 \\ 0 & 0 & -2 & 0 & 4 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix},$$

$$G^4_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -3 & 0 & 2 & 0 \\ 0 & -2 & 0 & 3 \\ 0 & 0 & -1 & 0 \end{bmatrix},$$

$$G^4_2 = \begin{bmatrix} 0 & 1 & 0 \\ -2 & 0 & 2 \\ 0 & -1 & 0 \end{bmatrix}. \tag{5.6}$$

The eigenvalues of K^4_0 and K^4_4 are $\{0, \pm 2i, \pm 4i\}$, of K^4_1 and K^4_3 are $\{0(2), \pm 2i(2), \pm 4i\}$, and of K^4_2 are $\{0(3), \pm 2i(2), \pm 4i\}$.

As the matrix $M^{(4)}$ is singular, the solution of (5.1) will be a solution of the corresponding homogeneous equation (with nine arbitrary constants) plus a particular solution for the nonhomogeneous part. For such a solution for the nonhomogeneous term must be consistent with the homogeneous term must be consistent with the homogeneous part. Again we divide the matrix and vectors into their natural blocks. We write the subvectors of $\mathbf{u}^{(4)}$ as $\mathbf{u}^4_i$ and the components as u^i_j and similarly for $\mathbf{v}^{(4)}$,

$$\mathbf{u}^4_0 = \begin{bmatrix} \frac{1}{4}(2u^4_{02} - v^4_{01}) \\ v^4_{00} \\ u^4_{02} \\ -v^4_{04} \\ \frac{1}{4}(2u^4_{02} + v^4_{03}) \end{bmatrix} \tag{5.7}$$

with u^4_{02} arbitrary and

$$3v^4_{00} + v^4_{02} + 3v^4_{04} = 0, \tag{5.8}$$

$$\mathbf{u}_1^4 = \begin{bmatrix} u_{10}^4 \\ u_{11}^4 \\ u_{10}^4 + \frac{1}{2}(v_{11}^4 - v_{14}^4) \\ u_{11}^4 + \frac{1}{2}(v_{12}^4 + v_{17}^4) \\ -u_{11}^4 + v_{10}^4 \\ u_{10}^4 + v_{14}^4 \\ -u_{11}^4 - \frac{1}{2}(v_{12}^4 + 3v_{17}^4) \\ u_{10}^4 + \frac{1}{2}(v_{11}^4 + 2v_{13}^4 - v_{14}^4) \end{bmatrix} \tag{5.9}$$

with u_{10}^4 and u_{11}^4 arbitrary and

$$\begin{aligned} &3v_{10}^4 + v_{12}^4 + v_{15}^4 + 3v_{17}^4 = 0, \\ &v_{11}^4 - 3v_{13}^4 - 3v_{14}^4 + v_{16}^4 = 0, \end{aligned} \tag{5.10}$$

$$\mathbf{u}_2^4 = \begin{bmatrix} u_{20}^4 \\ u_{21}^4 \\ u_{20}^4 + \frac{1}{2}(v_{21}^4 - u_{24}^4) \\ -u_{21}^4 + v_{20}^4 \\ u_{24}^4 \\ u_{21}^4 + v_{22}^4 \\ u_{20}^4 + \frac{1}{2}(v_{23}^4 - u_{24}^4) \\ -u_{21}^4 + v_{20}^4 + v_{26}^4 \\ u_{20}^4 + \frac{1}{2}(v_{21}^4 + v_{25}^4) \end{bmatrix}, \tag{5.11}$$

with u_{20}^4, u_{21}^4, and u_{24}^4 arbitrary and

$$\begin{aligned} &v_{21}^4 - v_{23}^4 + v_{25}^4 - v_{27}^4 = 0, \\ &2v_{20}^4 + v_{24}^4 + 2v_{28}^4 = 0, \\ &v_{20}^4 + v_{22}^4 + v_{26}^4 + v_{28}^4 = 0. \end{aligned} \tag{5.12}$$

The components of $\mathbf{u}_3^4$ are those of $\mathbf{u}_1^4$ with the first subscript 1 replaced by 3 and those of $\mathbf{u}_4^4$ are those of $\mathbf{u}_0^4$ with the first subscript 0 replaced by 4.

The elements of $\mathbf{v}^{(4)}$ arise from $-[I^{(3)},H^{(3)}]$. Using the results in (4.10), we obtain

$$\mathbf{v}_0^4 = \tfrac{1}{3}\begin{bmatrix} 0 \\ 0 \\ B_4A_2\{-12A_3 - 4A_1\} \\ A_2\{-6B_2A_3 - 4B_3A_2 + 4B_1A_2 + 2B_2A_1\} \\ B_4A_2\{-6A_3 - 2A_1\} \end{bmatrix}, \tag{5.13}$$

$$\mathbf{v}_1^4 = \tfrac{1}{3}\begin{bmatrix} 0 \\ 0 \\ B_4\{A_2(8A_2 - 12A_0) + A_3(36A_3 - 24A_1)\} \\ A_3\{-18B_2A_3 - 12B_3A_2 + 12B_1A_2 + 6B_2A_1\} \\ +A_2\{8B_1A_1 + 12B_2A_0 - 8B_3A_1 - 4B_2A_2\} \\ 0 \\ B_4\{-12A_2A_2 - 8A_1A_1 + 12A_3A_1\} \\ -24B_2A_3A_1 - 4B_3A_2A_1 + 4B_1A_2A_1 + 4B_2A_1A_1 + 36B_2A_3A_3 \\ +24B_3A_2A_3 - 24B_1A_2B_3 - 6B_2A_0A_2 + 2B_2A_2A_2 \\ B_4\{-2A_2A_2 - 4A_1A_1 - 6A_1A_3 + 18A_3A_3 - 6A_0A_2\} \end{bmatrix}, \tag{5.14}$$

$$\mathbf{v}_2^4 = \tfrac{1}{3}\begin{bmatrix} 0 \\ 0 \\ B_4\{-18A_0A_3 + 12A_2A_3 + 6A_1A_2\} \\ 0 \\ 24B_4\{A_3A_2 - A_0A_1\} \\ B_3\{-8A_2A_2 + 12A_3A_1 - 16A_1A_1\} + B_2\{-4A_1A_2 - 6A_2A_3 \\ -18A_0A_3 + 24A_1A_0\} + B_1\{8A_2A_2 - 12A_3A_1 + 16A_1A_1\} \\ B_4\{-6A_1A_2 - 12A_1A_0 + 18A_3A_0\} \\ B_3\{-12A_0A_2 + 16A_2A_2 + 8A_1A_1\} + B_2\{-18A_0A_3 - 6A_0A_1 \\ +24A_2A_3 - 4A_1A_2\} + B_1\{12A_0A_2 - 16A_2A_2 - 8A_1A_1\} \\ B_4\{-24A_0A_1 + 9A_0A_3 + 18A_3A_2\} \end{bmatrix}, \tag{5.15}$$

$$\mathbf{v}_3^4 = \tfrac{1}{3}\begin{bmatrix} 0 \\ 0 \\ B_4\{A_1(-8A_1+12A_3)+A_0(-36A_0+24A_2)\} \\ A_1\{12B_2A_3+8B_3A_2-8B_1A_2-4B_2A_1\} \\ +A_0\{-12B_1A_1-18B_2A_0+6B_2A_2+12B_3A_1\} \\ 0 \\ B_4\{-12A_2A_0+8A_2A_2+12A_1A_1\} \\ -6B_2A_3A_1+2B_2A_1A_1-4B_1A_1A_2-24B_2A_0A_2+4B_2A_2A_2 \\ +4B_3A_1A_2+24B_1A_1A_0+36B_2A_0A_0-24B_3A_1A_0 \\ B_4\{2A_1A_1+4A_2A_2+6A_1A_3+6A_0A_2-18A_0A_0\} \end{bmatrix}, \tag{5.16}$$

$$\mathbf{v}_4^4 = \tfrac{1}{3}\begin{bmatrix} 0 \\ 0 \\ B_4A_1\{12A_0+4A_2\} \\ A_1\{-6B_2A_0+4B_3A_1-4B_1A_1+2B_2A_2\} \\ B_4A_1\{2A_2+6A_0\} \end{bmatrix}. \tag{5.17}$$

The requirements of consistency, (5.8), (5.9), and (5.10), (together with the corresponding ones for $\mathbf{u}_3^4$ and $\mathbf{u}_4^4$) are, in terms of the A 's and B 's,

$$\begin{aligned} &B_4A_2(3A_3+A_1)=0, \\ &B_4(9A_3A_3-3A_3A_1-3A_2A_0-2A_1A_2+A_2A_2)=0, \\ &B_3(30A_3A_2+10A_2A_1)+B_2(-21A_3A_1+45A_3A_3-21A_0A_2 \\ &\quad -7A_2A_2+2A_1A_1)+B_1(-30A_3A_2-10A_2A_1)=0, \\ &B_3(2A_1A_3-4A_1A_1+2A_0A_2-4A_2A_2)+B_2(5A_0A_1-5A_2A_3) \\ &\quad +B_1(-2A_1A_3+4A_1A_1-2A_2A_0+4A_2A_2)=0, \\ &B_4(10A_3A_2-12A_0A_1+3A_0A_3)=0, \\ &B_4(-9A_0A_0+3A_0A_2+3A_1A_3+2A_2A_2+A_1A_1)=0, \\ &B_3(-30A_1A_0-10A_1A_2)+B_2(-21A_3A_1+45A_0A_0-21A_2A_0 \\ &\quad +7A_1A_1+2A_2A_2)+B_1(30A_1A_0+10A_2A_1)=0, \\ &B_4A_1(3A_0+A_2)=0. \end{aligned} \tag{5.18}$$

Note that the number of independent conditions is reduced by one owing to the duplication of (5.17), row 5.

6. THE CONSISTENCY REQUIREMENTS AND THE HÉNON–HEILES PROBLEM

For the Hénon–Heiles Hamiltonian,

$$A_2=1,\quad A_0=-\tfrac{1}{3},\quad A_1=0=A_3. \tag{6.1}$$

Substituting these values into (5.17), we see that the requirement of consistency leads to

$$B_1=B_3,\quad B_2=0=B_4, \tag{6.2}$$

i.e., the quadratic invariant is a scalar multiple of $H^{(2)}$. Furthermore, $\mathbf{v}^4$ is then identically zero and $I^{(3)}$ is just the same multiple of $H^{(3)}$, i.e., only a scalar multiple of the original Hamiltonian satisfies the consistency requirements and then trivially.

Apart from the trivial case discussed above, we may well ask whether there is any set of values of the A_i and B_i for which the consistency conditions are satisfied. Examining the terms in B_4, the assumption that $B_4 \neq 0$ leads in all ways to each of the A 's being zero, and so we must take A as zero. Examining the other terms, we have the set of equations

$$\begin{bmatrix} a & b & -a \\ c & d & -c \\ e & f & -e \end{bmatrix}\begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \mathbf{0}, \tag{6.3}$$

where $a, b, c, d, e,$ and f are the coefficients of the appropriate equations of (5.17). Clearly this has the solution

$$B_1=B_3,\quad B_2=0. \tag{6.4}$$

For any nontrivial solution we require that the rank of the coefficient matrix be one, i.e.,

$$ad-bc=0,\quad cf-de=0,\quad fa-eb=0, \tag{6.5}$$

or, equivalently, that there exist scalars α, β, γ, not all zero such that

$$\begin{aligned} &\alpha a+\beta c+\gamma e=0, \\ &\alpha b+\beta d+\gamma f=0. \end{aligned} \tag{6.6}$$

A little elementary algebra shows that this is not the case unless the A 's are all zero. The only other possibility is that

$$b = 0 = d = f. \tag{6.7}$$

In this case the solution to (6.3) is

$$B_1 = B_3, \quad B_2 \text{ arbitrary}. \tag{6.8}$$

The consequence of (6.7) as far as the A 's are concerned is that the following solutions are possible.

$$\begin{aligned} &A_0 = 0 = A_2, \quad A_1 = 3A_3 \\ &A_1 = 0 = A_3, \quad A_2 = 3A_0 \\ &3A_0 = A_1 = A_2 = 3A_3 \\ &3A_0 = -A_1 = A_2 = -3A_3. \end{aligned} \tag{6.9}$$

Each one of these represents an exactly integrable case as is readily seen by use of the transformation

$$\begin{aligned} &q_1 = (Q_1 + Q_2)/\sqrt{2}, \quad q_2 = (-Q_1 + Q_2)/\sqrt{2} \\ &p_1 = (P_1 + P_2)/\sqrt{2}, \quad p_2 = (-P_1 + P_2)/\sqrt{2}. \end{aligned} \tag{6.10}$$

7. COMMENT

We have seen that a time-independent invariant series does not exist for Hamiltonians of the Hénon–Heiles type due to inconsistency in the determination of the coefficients of the fourth order terms. Even had they been consistent, the same question of consistency would arise at the determination of all even order terms. In one sense we may be regarded as fortunate in not having to pursue the question of consistency to the higher order terms.

The existence of the third integral is still an open question. We have merely precluded one possible form. Clearly the next step in the search is to introduce time dependence into the invariant series to overcome the problem of inconsistency which arose with the time-independent approach. The results of that work will be announced in a further note.

ACKNOWLEDGMENT

We wish to thank Professor J. Ford of the Georgia Institute of Technology for some useful suggestions regarding Sec. 6.

[1]Cf. C.J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463–68 (1976); D.M. Fradkin, Amer. J. Phys. **33**, 207–211 (1965); J.M. Jauch and E.L. Hill, Phys. Rev. **57**, 641–45 (1940); G.H. Katzin and J. Levine, J. Math. Phys. **18**, 1267–74 (1977); A. Leung and K. Meyer, J. Diff. Equations **17**, 32–43 (1975); H.R. Lewis, Jr., Phys. Rev. Lett. **18**, 510–12 (1967); H.R. Lewis, Jr., Phys. Rev. **172**, 1313–15 (1968); H.R. Lewis, Jr., J. Math. Phys. **9**, 1976–86 (1968); J.E. Littlewood, Ann. Phys. (N.Y.) **29**, 13–18 (1964); J.W. Macki, J. Math. Phys. **18**, 1256–58 (1977); W. Sarlet, Proc. ICNO1975 (to appear); W. Sarlet, preprint, University of Ghent, March 1977; K.R. Symon, J. Math. Phys. **11**, 1320–30 (1970); L.S. Wollenberg, J. Math. Phys. **16**, 1352–57 (1975).

[2]M. Kruskal, J. Math. Phys. **3**, 806–28 (1962); see also W. Sarlet, Ann. Phys. (N.Y.) **92**, 232–47, 248–61 (1975).

[3]See H.R. Lewis, references in Ref. 1.

[4]P.G.L. Leach, SIAM J. Appl. Math. **34**, no. 3 (1978); N.J. Günther and P.G.L. Leach, J. Math. Phys. **18**, 572–76 (1977).

[5]P.G.L. Leach, J. Math. Phys. **18**, 1608–11, 1902–07 (1977); **19**, 446–51 (1978).

[6]I.e., the exact counterpart to the time-independent problem; cf. D.M. Fradkin, Ref. 1.

[7]The reason for using the isotropic oscillator Hamiltonian is that it provides a simple formulation for the constants of the motion; see Leach, J. Math. Phys. **19**, 446–51 (1978).

[8]P.G.L. Leach, "On nonlinear transformations for time-dependent polynomial Hamiltonians," J. Math. Phys. (to appear).

[9]P.G.L. Leach, "Towards an invariant for the time-dependent anharmonic oscillator," J. Math. Phys. (to appear).

[10]J.P. Ottoy, PhD thesis, University of Ghent, Belgium, 1976 (unpublished).

[11]M Hénon and C. Heiles, Astron. J. **69**, 73–79 (1964).

[12]Cf. F.G. Gustavson, Astron. J. **71**, 670–85 (1966).

[13]In the model under consideration, energy is the first integral and angular momentum the second; see Hénon and Heiles, Ref. 11.

[14]For the meaning of isolating integral see A. Wintner, *The Analytical Foundations of Celestial Mechanics* (Princeton U.P., Princeton, N.J., 1947); G. Contropoulos, Astron. J. **68**, 1–14 (1963).

[15]A.H.P. van der Burgh, J. Sound and Vibration, **42**, 463–75 (1975); **49**, 93–103 (1976); F. Verhulst, INCO Konferenz 1975, Bd. I/2, edited by B. Schmidt (Akademie Verlag, Berlin, 1977).

[16]Cf. J.H. Jeans, Mon. Not. Roy. Astron. Soc. **76**, 70–84 (1915).

[17]Cf. Contropoulos, Ref. 14; also T.Y. Thomas, Proc. Nat. Acad. Sci. U.S.A. **32**, 10–15 (1946).

[18]L. Markus and K.R. Meyer, Mem. Amer. Math. Soc., no. 144 (1974).

[19]P.A. Clement, SIAM Rev. **1**, 50–52 (1959).

Mobility of nonlinear systems[a)]

Bogdan Mielnik

Department of Mathematics, King's College, London, England[b)]

(Received 19 September 1978)

Global mobility is defined and is found to be decisive for the structure of a physical system. The structure of some simple nonlinear systems is elucidated by studying their dynamical mobility. It is shown that due to an excessive mobility a nonlinear system may acquire some classical features.

1. INTRODUCTION

In the usual formation of a physical theory the elements of geometry, such as the structure of states, the geometry of the phase space, and the "logic" are introduced *a priori*. In the quantum theories these structures have a traditional form borrowed from the theory of Hilbert spaces. The Hilbert space formalism has also been used for the quantization of nonlinear fields including Einstein's gravity. However, objections were raised, that the quantum theory cannot be consistently extended to this new domain without questioning its traditional structure (Penrose). In fact, it has been shown that even in their proper domain the quantum theories cannot be applied without worrying about consistency (Haag). A development of this line of thought exhibited a consistency link between the dynamics and the structure of any theory.[1] As it turns out, the basic structural elements such as the quantum logic and the "conditional probabilities" are not exclusive to quantum theories but exist also though in a generalized form, in any statistical theory. Moreover, they are of dynamical origin. This suggests that the usual formulation of a theory might be reversed, by starting from the dynamics and then determining the structure.

Below, this idea is applied to determine the geometric structure of some simple nonlinear systems. However, in order to fix geometry, the dynamics must be understood in a "global" sense. This leads to the idea of *dynamical mobility* which is developed in Sec. 2, following ideas published earlier in Refs. 1 and 2. The structural consequences of the mobility are reviewed in Sec. 3. It is then shown that there is an essential difference in mobility between linear and nonlinear wave mechanics. For the nonlinear models considered here the phase space is so flexible under dynamical transformations that the system acquires some classical properties. The consequences of this fact for the physics of nonlinear models are discussed in Sec. 6. The possible significance of mobility for other domains of physics is considered in Sec. 7.

2. CONCEPT OF MOBILITY

In many physical theories the dynamics is represented by a one-parameter group of transformations of a physical structure (dynamical group). However, if one is interested in the entire dynamical nature of a system, some wider algebraic structures become essential.

Suppose that one knows the set Φ of all pure states (denoted $\phi, \psi, \ldots$) of a hypothetical physical system. Below, only minimal assumptions defining the structure of Φ will be adopted. It will not be assumed that Φ must be a Hilbert space. It seems reasonable to assume that Φ is a topological space with some physically meaningful topology. In what follows, it will be assumed also that Φ possesses the structure of a generalized differential manifold (possibly of infinite dimension). This fact cannot be immediately motivated by physical arguments, but happens to be true for the phase spaces of existing physical theories (e.g., for symplectic manifolds and for unit spheres in Hilbert spaces). For the moment, no more assumptions concerning the structure of Φ will be made: it will be left open whether Φ will give rise to a classical phase space, or to a quantum mechanical phase space, or, perhaps, to a new type of phase space corresponding to a new physical theory.

Suppose further that the system with the pure states $\phi\in\Phi$ does not exist alone. It exists in a certain external world whose possible states denoted $\xi,\eta,\ldots$ (called "external conditions") form a certain set Ξ. Below, the external conditions $\xi\in\Xi$ will be assumed of the "ideal" type, i.e., yielding a well posed problem of evolution and not introducing any dissipation into the behavior of the system (the same idealization is represented by external fields in quantum theories). Consistently, it will be assumed that every $\xi\in\Xi$ induces a unique family of diffeomorphisms $g(t,t';\xi):\Phi\rightarrow\Phi$ $(t\leqslant t')$ mapping the pure states onto the pure states and representing the evolution process which the system undergoes under the influence of the external conditions ξ. The difference between the two possible views of dynamics can be now put as follows: According to the standard approach, dynamics deals with a concrete evolution process corresponding to some given external condition and represented by a unique family of evolution operations $g(t,t') = g(t,t';\xi_0)$ $(\xi_0\in\Xi)$. According to the point of view which might be called global, this description is too narrow. The dynamics is a "plural" concept: it stands not for one but for an infinity of evolution processes $g(t,t';\xi)$ corresponding to all possible conditions $\xi\in\Xi$. One of the mathematical structures reflecting the above global view is the set of all operations g: $\Phi\rightarrow\Phi$, which can be achieved by all possible evolution processes.

Definition 1: An invertible operation g: $\Phi\rightarrow\Phi$ is called

[a]Partly supported by Polish Grant MR7 and by the Norman Foundation.
[b]On leave of absence from the Institute of Theoretical Physics, Warsaw University.

 0022-2488/80/010044-11$01.00

achievable if it belongs to the closure of the set of all operations $\{g(t,t';\xi): t\leqslant t', \xi \in \Xi\}$ in the topology corresponding to the following definition of convergence. *Definition* 2: A sequence of mappings $g_n: \Phi \to \mathscr{T}$, where $\mathscr{T}$ is a topological space, is called *convergent* to a mapping $g: \Phi \to \mathscr{T}$ if for every $\phi \in \Phi$ and every $\phi_n \in \Phi$ $(n = 1,2,...,)$, $\phi_n \to \phi$, there is $g_n(\phi_n) \to g(\phi)$. In what follows, all the limiting transitions will be understood in the sense of Definition 2.[3]

Now let G denote the class of all dynamically achievable differentiable transformations of Φ. The class G represents one of global features of the dynamical theory. Since it may be assumed that the set of external conditions Ξ does not distinguish between parts of the time axis, the achievable operations can be repeated with a time delay. Hence, they can be superposed. Moreover, the limiting transition conserves the product (superposition). Hence, the class G has the structure of a topological semigroup. This semigroup is wider than the traditionally studied one-parameter (dynamical) semigroup. It represents the whole ability of the system to be transformed: it has thererfore been called the *semigroup of global mobility*.[2] The semigroup G is present in many dynamical schemes which differ from the quantum theory. Thus, for example, for the nonlinear Schrödinger waves obeying the evolution equation with the "external potentials"

$$i\frac{\partial}{\partial t}\psi = -\Delta\psi + \epsilon f(|\psi|^2)\psi + V(x,t)\,\psi, \qquad (2.1)$$

G would be the semigroup of all nonlinear transformation of the manifold of ψ waves generated by Eqs. (2.1) for all possible potentials $V(x,t)$.

Besides the semigroup aspect, the global dynamics admits also a certain group-theoretical description. Since the external conditions $\xi \in \Xi$ are "ideal," it may be assumed that operations $g(t,t';\xi)$ are invertible, and their inverse are again diffeomorphisms of Φ. Henceforth the achievable operations span a certain transformation group of Φ. *Definition* 3: The smallest topologically closed group of transformations $\Phi \to \Phi$ containing the semigroup G will be called the *dynamical mobility group* of the system and denoted M. The information contained in M is more general and vague than that represented by G: it only tells which transformations $\Phi \to \Phi$ are not "against the nature of the system" without granting, however, that they can be effectively achieved. The above idea of mobility is nonrelativistic. It acquires a relativistic meaning if the system exists in Galileo or Minkowski space–time and if the evolution equations are covariant.

In spite of the domination of the "fixed Hamiltonian approach," the global description of dynamics is gradually making its way in recent developments. It is close to the idea that the implications of the dynamics are best seen in the theory of open systems (Havas and Plebanski). It is also related to the idea that the physical theory might be conveniently stated in terms of "impotence principles" (Bergmann and Sudarshan). In quantum field theory some typical mobility problems were posed by Schwinger by considering S matrices depending on the external fields.[4] The global description is one of the main points in the C^*-algebraic approach due to Haag and Kastler[5] and subsequent approaches dealing with infinities of "operations".[6–9] The problem of mobility in quantum mechanics was raised by Lubkin.[10] It has been subsequently realized that the significance of the global structures extends beyond the Hilbert space formalism of quantum theories. The global mobility semigroup is a flexible structural element which can exist in any theory and is not limited by regularities required from * algebras and quantum logic. Inversely, it is precisely the semigroup G which explains the origin of any possible regularities of a dynamical theory.

3. IMPLICATIONS OF MOBILITY

(1) *Invariance*: As recently found, in quantum mechanics the semigroups G and the group M both coincide with the unitary group[2,11] This indicates that in the general theory M is the right substitute of the missing unitary group.

(2) *Geometry of the phase space*: An idea arises that the geometry of the phase space Φ should not be *a priori* assumed. It should be found by applying the Klein program to the group M acting on the manifold Φ.

(3) *Functional observables*: One of the most essential implications of the semigroup G concerns the statistical theories. A statistical theory arises if the pure states $\phi \in \Phi$ are not observed directly but only via some secondary statistical effects called "observables." In quantum mechanics the observables are quadratic forms of the state vector and theorefore are represented by linear operators. This need not necesarily be so in a general statistical theory based on arbitrary dynamics.[12,1,13] Here one deals with functional observables.[1] *Definition* 5: A *functional observable* is any continuous real function $f: \Phi \to \mathbb{R}$ whose values $f(\phi)$ are interpretable as averages over the pure states ϕ of a certain statistical experiment. Given the manifold of pure states Φ, the statistical theory is defined by specifying the class of these real functions on Φ which are observables[1]. This class will be denoted F. For known reasons, F should be a linear class, closed in the topology of Definition 2. As has already been found, the functional observables contain no less information than the algebras of observables[14,15,1]. In particular, they determine the geometry elements of the statistical theory such as the convex set S of all pure and mixed states[1] and the quantum logic *et al.*[16,17,13,18] Given the manifold Φ, the contents of F may serve to classify the theories: the richer F is the more precise perception of the statistical ensembles and the "more classical" the theory.[19]

Now, it turns out that the class of observables is conditioned by the mobility. Indeed, having a method of measuring an observable $f \in F$ and having an operation $g \in G$, one can modify the measurement by letting the system undergo first the transformation $g: \Phi \to \Phi$ and measuring f afterwards: the resulting new observable is the superposition $f \circ g: \Phi \to \Phi \to \mathbb{R}$. Hence, F must be closed under the mobility semigroup $F \circ G \subset F$. If one wants the theory to have a group-theoretical invariance similar to the unitary invariance of quantum theories, it is natural to assume that the class F is invariant under the whole mobility group $F \circ M = F$. The theory with that property will be called group invariant.[20]

If the invariance of the theory is assumed, the class of observables F becomes a strictly group-theoretical concept. The number of possibilities of constructing invariant statistical theories for a given dynamics depends on the number of invariant function spaces ("special functions") which the group M allows to exist on the manifold Φ. To choose one of these invariant spaces as the class F one needs only a minimal statistical information. As a consequence, the traditional geometric elements of the theory (quantum logic), previously studied by the axiomatic approaches, find their most natural explanation in the dynamical mobility of the system. What one has here is a certain flexible mathematical mechanism interrelating the motion and the geometry of the theory. Below, this mechanism will be used to examine the structure of simplified models of the nonlinear Schrödinger dynamics (2.1).

4. TWO-COMPONENT MODEL

A simple spin model of the nonlinear Schrödinger dynamics (2.1) is obtained by replacing the space continuum by the discrete space composed of two points 1 and 2 and by replacing the wave ψ by a two-component complex vector $\psi = ||\psi_j||, j = 1,2$:

$$i\frac{\partial}{\partial t}\begin{Vmatrix}\psi_1\\ \psi_2\end{Vmatrix} = -\begin{Vmatrix}0 & 1\\ 1 & 0\end{Vmatrix}\begin{Vmatrix}\psi_1\\ \psi_2\end{Vmatrix} + \epsilon\begin{Vmatrix}f(|\psi_1|^2) & 0\\ 0 & f(|\psi_2|^2)\end{Vmatrix}\begin{Vmatrix}\psi_1\\ \psi_2\end{Vmatrix} + \begin{Vmatrix}V_1 & 0\\ 0 & V_2\end{Vmatrix}\begin{Vmatrix}\psi_1\\ \psi_2\end{Vmatrix}. \tag{4.1}$$

Here, the matrix $\|{}^0_1\ {}^1_0\|$ appears instead of the Laplace operator, the real numbers V_1, V_2 represent an "external potential," and the real function $f(\zeta)$ continuous for $\zeta \geqslant 0$, defines the nonlinearity. The quantity $|\psi_1|^2 + |\psi_2|^2 = |\psi|^2$ is conserved for all V_j's ("superconservation law"), therefore, further considerations will be restricted to ψ vectors for which $|\psi|^2 = R =$ const. Vectors differing only by a phase remain so after the evolution (4.1); therefore, one can consider a hypothetical system whose states are the rays of the vectors ψ. The geometric representation of the resulting manifold Φ is obtained after the standard transition to real coordinates:

$$\begin{aligned} x &= \psi_1\bar{\psi}_2 + \psi_2\bar{\psi}_1 = (\psi,\sigma_1\psi),\\ y &= i^{-1}(\psi_1\bar{\psi}_2 - \psi_2\bar{\psi}_1) = (\psi,\sigma_2\psi),\\ z &= |\psi_1|^2 - |\psi_2|^2 = (\psi,\sigma_3\psi), \end{aligned} \tag{4.2}$$

where σ_j's are the Pauli matrices. Here, $x^2 + y^2 + z^2 = (|\psi_1|^2 + |\psi_2|^2)^2 = R^2$ and so, the manifold of pure states is the sphere of radius R in $\mathbb{R}^3$.

The equations (4.1) in terms of real coordinates read

$$\frac{\partial}{\partial t}x = \epsilon\omega(z)\,y + \lambda y,$$

$$\frac{\partial}{\partial t}y = -2z - \epsilon\omega(z)\,x - \lambda x,$$

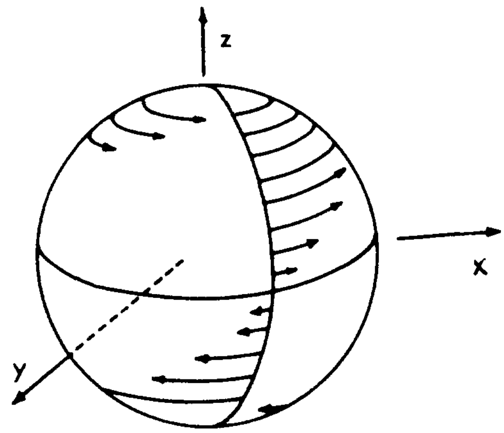


FIG. 1.

$$\frac{\partial}{\partial t}z = 2y, \tag{4.3}$$

where $\omega(z) = \omega(z,R) = f[(R+z)/2] - f[(R-z)/2]$ is an odd function defining the nonlinearity and $\lambda = V_1 - V_2 \in \mathbb{R}$ is an arbitrary constant representing the external force. In what follows it will be assumed for simplicity that the function $\omega(z,R)$ is either strictly increasing or strictly decreasing for $-R \leqslant z \leqslant R$. To ensure that the further considerations are valid for every R sphere, this will be assumed true for every $R \geqslant 0$. This imposes conditions on f which are fulfilled, for example, by a wide class of polynomials: $f(\zeta) = \zeta$, $\zeta^2, \zeta^3, \zeta^5, \zeta^7$,, etc. Below, the evolutions (4.3) will be represented by vector fields (generators)

$$\frac{\partial}{\partial t} = X = -2I_x + \epsilon\omega(z)I_z + \lambda I_z, \tag{4.4}$$

where $I_x = z(\partial/\partial y) - y(\partial/\partial z)$, $I_y = x(\partial/\partial z)$, $-z(\partial/\partial x)$, $I_z = y(\partial/\partial x) - x(\partial/\partial y)$ are the generators of rotations on the sphere Φ. In the expression (4.4) the part $X_0 = -2I_x + \epsilon\omega(z)I_x = A + \epsilon N$ stands for the free evolution. $A = -2I_x$ generates the rigid rotation about the x axis, while $N = \omega(z)I_z$ stands for the nonlinear part of the motion. On every circle $z =$ const., N produces just a rotation around z. However, for different z's the angular velocities of these rotations are differnt. The resulting transformations are "torsions" of the sphere Φ (Fig. 1). The part λI_z in Eq. (4.4) contributes with a rigid rotation around the z axis with any desired velocity $\lambda = V_1 - V_2$. It turns out that the dynamical mobility defined by the vector fields (3.4) is much richer than that of the linear system ($\epsilon = 0$).

Achievable operations

The simplest dynamical operations are of the form $\exp\tau X$ ($\tau \geqslant 0$), where X are the vector fields (4.4). By superposing these operations and making limiting transitions one can generate some new exponential operations. Thus, by taking a strong external force and letting it act for a short time ($\lambda = \gamma/\epsilon$, $\tau = \epsilon$, $\epsilon \to 0$), one obtains in the limit the transformation $\exp\gamma I_z = \lim_{\epsilon\to 0}\exp[\epsilon(X_0 + \gamma I_z/\epsilon)]$, an arbitrary rigid rotation around z, interpreted as a "shock transformation" which the system undergoes under the influence of a δ-like pulse of the external potential: $V_j(t) = \Gamma_j\,\delta(t - t_0)$;

$\Gamma_1 - \Gamma_2 = \gamma$. Having that transformation, one can now generate the product of operations $\exp \gamma I_z \exp X \times \exp(-\gamma I_z) = \exp\tilde{X}$, where X is given by Eq. (4.4), $\gamma\in R$, and the modified generator $\tilde{X}$ is

$$\tilde{X} = -2(\cos\gamma\, I_x + \sin\gamma\, I_y) + \epsilon\omega(z)\, I_z + \lambda\, I_z. \qquad (4.5)$$

By varying γ and λ one obtains a family of vector fields on Φ generating achievable operations. It contains in particular

$$\begin{aligned} X_0 &= -2I_x + \epsilon\omega(z)\, I_z, \\ V &= \gamma I_z, \\ X_1 &= 2I_x + \epsilon\omega(z)\, I_z. \end{aligned} \qquad (4.6)$$

The class of dynamically achievable operations can be further extended by applying the Trotter formula:

$$\exp(\tau_1\, Y_1 + \tau_2\, Y_2) = \lim_{n\to\infty}\left(\exp\frac{\tau_1 Y_1}{n}\exp\frac{\tau_2 Y_2}{n}\right)^n. \qquad (4.7)$$

Given two continuous vector fields on Φ both generating families of achievable operators $\exp\tau_1\, Y_1$ and $\exp\tau_2\, Y_2$ $(\tau_1, \tau_2 \geqslant 0)$, one can also generate (4.7). Hence, the list of vector fields generating the achievable operations contains all positive combinations of the fields (4.5). In particular, by taking $\frac{1}{2}X_1 + \frac{1}{2}X_0$ one ends up with

$$\epsilon N = \epsilon\omega(x)\, I_z. \qquad (4.8)$$

This signifies that by using adequate external forces one can maneuver the system to perform an operation dictated by the purely nonlinear part of Eq. (4.1), the linear part producing no final effect. Taking more general positive combinations of the fields (4.5), one sees that the operations $\exp Y$ for the following vector fields Y are also achievable:

$$\begin{aligned} Y = &-2(\tau_1\cos\gamma_1 + \tau_2\cos\gamma_2)I_x \\ &+ 2(\tau_1\sin\gamma_1 + \tau_2\sin\gamma_2)\, I_y \\ &+ \epsilon(\tau_1 + \tau_2)\omega(z)I_z + \lambda\, I_z \\ &(\lambda, \gamma_1, \gamma_2\in\mathbb{R};\ \tau_1, \tau_2 \geqslant 0). \end{aligned} \qquad (4.9)$$

In particular, taking $\cos\gamma_1 = -\cos\gamma_2 = 1$, one obtains the following two-parameter family of vector fields:

$$\begin{aligned} Z = &-2(\tau_1 - \tau_2)\, I_x + \epsilon(\tau_1 + \tau_2)\omega(z)\, I_z + \lambda I_z, \\ &(\lambda\in\mathbb{R}, \tau_1, \tau_2 \geqslant 0). \end{aligned} \qquad (4.10)$$

To read the information contained in Eqs. (4.8)–(4.10), some differential geometric representation of dynamical mobility is needed.

5. MOBILITY CONES

Even in the simplest dynamical theories the effective knowledge of G is limited by the impossibility of resolving the evolution equations explicitly. However, there is essential information which can be obtained without integrating the evolution equations. Below, Φ is assumed to be a differentiable submanifold of a real Banach space and G denotes any topologically closed semigroup of transformations $\Phi \to \Phi$. *Definition* 6: A continuous vector field X on Φ generating a one-parameter group of diffeomorphisms $\exp\tau X : \Phi\to\Phi$ is called a *generator field* of G if $\exp\tau X\in G$ for every $\tau \geqslant 0$. Here, it is not assumed that G is the mobility semigroup. If so, it is still not assumed that the operations

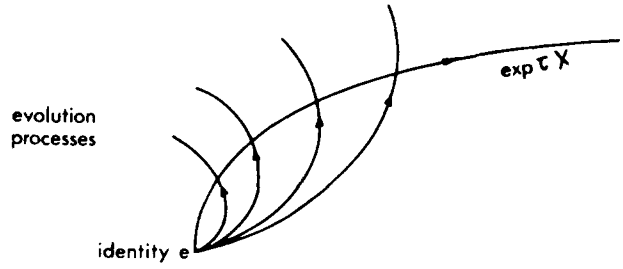


FIG. 2.

$\exp\tau X$ $(\tau\geqslant 0)$ are achievable in turn by one single evolution process or that $\tau \geqslant 0$ is a measure of time. Indeed, it may happen that the transformations $\exp\tau X$ are achievable separately each by its own process and that there is no common evolution process which would accomplish all of them in succession (Fig. 2).

Among the generator fields of G of special interest are those which are bounded and fulfil the Lipschitz condition; they will be called fields of BL class. The fields of BL class are integrable; moreover, they make the Trotter formula (4.7) convergent in the sense of Definition 2. Now, if the fields Y_1 and Y_2 are generators of G, of BL class, and if $\tau_1, \tau_2 \geqslant 0$, it follows from Eq. (4.7) that the field $\tau_1 Y_1 + \tau_2\, Y_2$ is a generator field too. Hence, the set of all BL-class generators of G has the structure of a *convex cone* of vector fields on Φ. In practice, it is also of interest to consider smaller cones of the fields. *Definition* 7: Any topologically closed convex cone of BL-class generators of G is a *cone of generators*. Note that every cone of generators covers Φ by a field of convex *vector cones* in the tangent spaces. *Definition* 8: Given a semigroup of diffeomorphisms of Φ and given a cone of generators K, the *mobility cone* K_ϕ at a point $\phi\in\Phi$ is the cone of vectors obtained by taking the values at ϕ of the vector fields $X\in K$. Intuitively, the "mobility cone" at a point ϕ encompasses all the "allowed directions" in which the point ϕ may be displaced by the small semigroup operations $\exp\tau X$ $(\tau\geqslant 0)$ with generators $X\in K$. A particular importance must be assigned to the mobility cones defined by all BL-classes generators of G: they will be called the *cones of the semigroup G on the manifold* Φ. One has the following lemma:

Lemma: Let Φ be a real Banach manifold covered by the field of mobility cones of the semigroup G. Let $\tau \to \phi(\tau)$ $(\tau\geqslant 0)$ be a differentiable trajectory on Φ whose tangent vector at every point is contained in the local mobility cone (integral trajectory of the cone field). Then every point $\phi(\tau)$ on the trajectory can be achieved (at least approximately) from the point $\phi(0)$ by means of motions generated by the semigroup G: $\phi(\tau)\in\overline{G\,\phi(0)}$ (Fig. 3.). It is thus seen that, besides defining the directions, the mobility cones also provide information concerning the finite displacements of any point of Φ: this information is obtained by constructing the integral trajectories of the field of cones. It turns out to be obvious that the mobility cones play a role in the global dynamics similar to that of characteristics surfaces in the theory of hyperbolic partial differential equations.

The mobility of a single point does not yet exhaust the dynamical information. What one would like to know is,

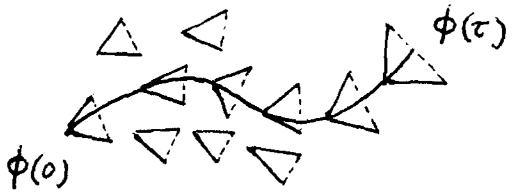


FIG. 3.

also, how one can transform the manifold Φ as a whole. This involves the following question: how can one simultaneously displace any number of points on Φ? For a finite set of n points this information is again represented by mobility cones. As a semigroup of transformations of Φ, G acts also as a semigroup of transformations on every Cartesian product

$$\underbrace{\Phi \times \cdots \times \Phi}_{n}$$

according to

$$g(\phi_1 \times \cdots \times \phi_n) = g(\phi_1) \times \cdots \times g(\phi_n)$$
$$(g \in G, \phi_1, \ldots, \phi_n \in \Phi).$$

Consequently, G determines on $\Phi \times \ldots \times \Phi$ a field of mobility cones: this field represents the possible simultaneous displacements of n points on Φ by the transformations of the semigroup G. A part of that information can be represented geometrically on the initial manifold Φ without the need for constructing Cartesian products. Let ϕ and ϕ_0 be two points on Φ and suppose that one is interested in the possibilities of moving ϕ while ϕ_0 is kept fixed. This corresponds to the subsemigroup $G_{\phi_0} = \{g \in G : g\phi_0 = \phi_0\}$ ("little semigroup"). The semigroup G_{ϕ_0} has again some mobility cones on the manifold Φ: they represent the freedom to move ϕ while ϕ_0 is fixed. More generally, one can consider any subset $\Phi_0 \subset \Phi$ and the corresponding subsemigroup $G_{\Phi_0} = \{g \in G : g\phi = \phi$ for $\phi \in \Phi_0\}$. The mobility cones of G_{Φ_0} on Φ represents the freedom of moving any point ϕ while all the points of Φ_0 are fixed, They might be called little cones of G. As is seen, the problems of mobility always lead to a similar type of structure.

Definition 8: A *mobility manifold* is a (generalized) differentiable manifold Φ covered by a field of convex cones defined in the tangent spaces to Φ, which are interpreted as the sets of allowed directions in which the points of the manifold are free to move. The cones can be of arbitrary dimension and can degenerate at some points. (The vanishing of the mobility cone at any point $\phi_0 \in \Phi$ means that ϕ_0 is "stationary.") An example of a mobility manifold with no degeneracy is relativistic space–time, if the "future cone" at every point is distinguished. Another example of that structure arises in the theory of the two-component model of Sec. 4.

4. TWO-COMPONENT MODEL (CONTINUATION)

If Φ is the sphere with dynamics defined by Eqs. (4.3), a class of generators of G is given by Eq. (4.5). Since γ and λ in Eq. (4.5) are arbitrary, the values of the fields X at any point cover all the direction on the local tangent plane. Hence, the mobility cones of G on Φ are just the tangent spaces. As a consequence, given any point $\phi \in \Phi$, there are no points on Φ which could not be, at least approximately, achieved from ϕ. In fact, one can even show that every point of Φ can be achieved exactly ($G\phi = \Phi$), which makes Φ a homogeneous space of G. In this respect nonlinear dynamics does not yet differ from linear dynamics. The difference emerges when one considers the simultaneous displacements of many points of Φ.

Let $\phi_0 \in \Phi$ and consider the motions of any point $\phi \in \Phi$ arising from those transformations which keep ϕ_0 invariant. The generators of the corresponding subsemigroup G_{ϕ_0} can be found by distinguishing among the vector fields (4.4)–(4.10) those which vanish at ϕ_0. Since G contains rigid notations around the z axis (shock transformations), it can be assumed without loss of generality that ϕ_0 lies on the x,z plane: $\phi_0 = (x_0, 0, z) = (R\cos\alpha, 0, R\sin\alpha)$. A family of generators vanishing at ϕ_0 can be now constructed from vector fields (4.10). At $\phi = \phi_0$ the field (4.10) coincides with $I = -2(\tau_1 - \tau_1) I_x + \epsilon(\tau_1 + \tau_2)\omega(z_0) I_z + \lambda I_z$. To vanish at ϕ_0 this field must be proportional to the generator I_0 of rigid rotations around ϕ_0 axis: $I_0 = \cos\alpha I_x + \sin\alpha I_z$. The condition $I = \Lambda I_0 (\Lambda \in R) \Longleftrightarrow ((\tau_1 - \tau_2) I_x + [\epsilon(\tau_1 + \tau_2)\omega(z_0) + \lambda] I_z = \Lambda \cos\alpha I_x + \Lambda \sin\alpha I_z$ leads to

$$-\frac{1}{2}\frac{\epsilon(\tau_1 + \tau_2)\omega(z_0)\tau\lambda}{\tau_1 - \tau_2} = \tan\alpha \Rightarrow \lambda = -2(\tau_1 - \tau_2) \times \tan\alpha - \epsilon(\tau_1 - \tau_2)\omega(z_0).$$

Assuming $\cos\alpha > 0$ and putting $\tau = (\tau_1 + \tau_2)/\cos\alpha$ and $\epsilon = (\tau_2 - \tau_1)/(\tau_2 + \tau_1)$, one obtains the following cone of vector fields vanishing at ϕ_0:

$$Y = \tau\{2\epsilon I_0 + \epsilon \cos\alpha[\omega(z_0) - \omega(z)] I_z\},$$
$$\tau \geqslant 0, \; -1 \leqslant \epsilon \leqslant 1 \tag{4.11}$$

They cover the sphere Φ with a field of mobility cones whose boundaries are marked by two vector fields:

$$Y_+ = 2I_0 + \epsilon\cos\alpha\,[\omega(z_0) - \omega(z)]\, I_z$$

and

$$Y_- = -2I_0 + \epsilon\cos\alpha\,[\omega(z_0) - \omega(z)]\, I_z.$$

For $\epsilon > 0$ and $\omega(z)$ decreasing, this leads to the following differential picture of mobility (Fig 4).

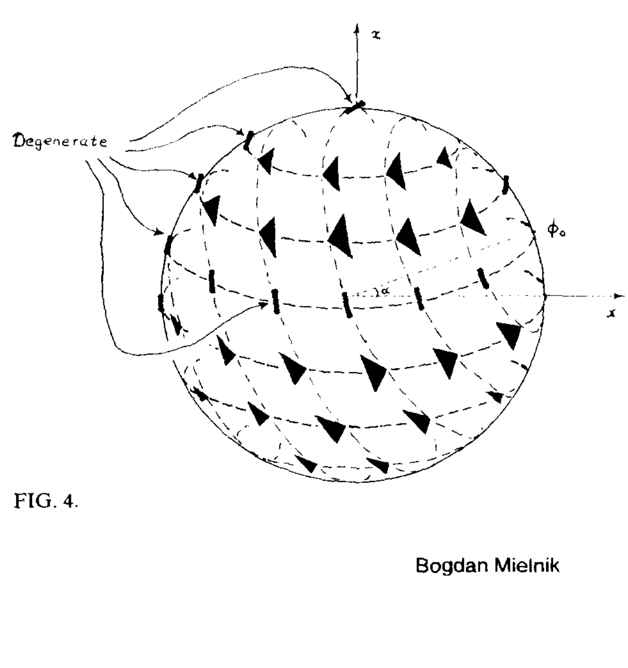


FIG. 4.

Figure 4 exhibits an essential difference between the linear and the nonlinear models. In the linear case the evolution transformations are the isometries of the sphere and so, having fixed any point $\phi_0 \in \Phi$, one can move the other points only along the circles around ϕ_0. The corresponding mobility cones are one dimensional (straight lines). This is no longer so in the nonlinear model ($\epsilon \neq 0$). Here, for any fixed ϕ_0 the second point ϕ acquires a higher degree of mobility: it is not bound to move along the circles but can be displaced in the directions forming the cones on Fig. 4. By constructing integral trajectories of this field of cones one can see that in the nonlinear case no forbidden domains for the point ϕ exist on the sphere Φ: keeping ϕ_0 fixed, one can still maneuver the other point ϕ into any desired vicinity on Φ. In particular, ϕ can tend to either "north pole" ($z = R$) or "south pole" ($z = -R$) of Φ (this maneuver involves an integral trajectory spiraling around the corresponding pole). Now, by applying a sequence of dynamical transformations (first moving ϕ and leaving ϕ_0 invariant, then moving ϕ_0 and keeping ϕ fixed), one can see that the dynamical system (4.3) enjoys a curious property. Any two distinct points on Φ can be simultaneously displaced to two arbitrary neighborhoods by an adequate dynamical transformation. In particular, every two points can be moved arbitrarily close to the north and south poles of Φ. In terms of complex vectors in Eq. (4.1) this means that for every two vectors $\psi = ||\psi_1, \psi_2||$ and $\eta = ||\eta_1, \eta_2||$ there exists a dynamical transformation which simultaneously reduces almost to zero the first component of ψ and the second component of η.

The increased mobility of the nonlinear system (4.3) has consequences for its physics. This is seen by comparing the two component vectors ψ in Eqs. (4.1) with the quantum mechanical state vectors (as for example, that of spin or polarization). It is an essential property of quantum states that they cannot be arbitrarily maneuvered. This is due to an impossibility law of quantum mechanics which forbids the precise separation of pure states ("second impossibility").[1] Given two state vectors ψ and, η, one cannot, in general, produce a filtering process which would accept all the systems in the pure state ψ while rejecting all systems in the state η. Every filter which transmits all ψ particles must unavoidably transmit at least the average fraction $|(\psi, \eta)|^2$ of the η particles.[1] The above law finds consistent expression in the quantum mechanical evolution equations. The dynamical operations generated by these equations are unitary and conserve the scalar products. As a result, one cannot displace two state vectors arbitrarily: where one goes the other follows at a constant distance. The above separation impossibly holds no longer for the nonlinear dynamics (4.3). As seen from the mobility cones in Fig. 4 every two states, however close at the beginning, can be arbitrarily separated at the end. This makes possible the construction of infinitely selective measuring devices, contrary to the quantum mechanical impossibility principle. Indeed, if at least one pair of states on Φ can be distinguished sharply, without an element of statistical uncertainty (e.g., the states $||1,0||$ and $||0,1||$ at the poles of Φ), then there is also a sharp distinction between any two states: the method consists of first displacing them toward two opposite poles of Φ and then applying the selective measurement. The existence of infinitely selective filters is one of the features of classical theories. This indicates that because of an excessive dynamical mobility the nonlinear system (4.3) acquires some classical properties.

This statement can be given a more exact sense if it is assumed that Eqs. (4.1) define the dynamics in an invariant statistical theory. Then, in agreement with the considerations of Sec. 3, the class of observables F depends on the mobility group M defined by Eqs. (4.1). The group M can be described in terms of generators which now form a Lie algebra of vector fields spanned by Eqs. (4.4)–(4.10). Since among the generators are $X_0 = -2I_x + \epsilon N$ and $\epsilon N = \epsilon\omega(z)I_z$, the algebra must contain also $\frac{1}{2}(\epsilon N - X_0) = I_x$. On the other hand, it contains I_z. Hence, it must also contain $I_y = [I_x, I_z]$. Thus, the Lie algebra of generators of M is spanned by the three rotation generators I_x, I_y, I_z and the "nonlinear generator" $\epsilon N = \epsilon\omega(z)I_z$. The first three vector fields generate the rotation group SO(3). Consistently, M is the group spanned by SO(3) and by the nonlinear rotations represented in Fig. 1. Now, the problem of the special function spaces of M on Φ becomes simple. Since $M \supset$ SO(3), every invariant function space of M must be also invariant under SO(3) and so it must be composed of some number of the spherical function subspaces. By examining the operator $\omega(z)I_z$, one proves that it yields transitions (either direct or indirect) between any two subspaces of spherical functions (except for the subspace of constants). The class F contains constants and is invariant under M. Moreover, F must contain some functions besides constants (otherwise the states would not be distinguishable and the theory would become trivial). Hence, F contains all the spherical functions subspaces. Since F is closed, this implies that F coincides with the class of all continuous real functions on Φ [$F = C(\Phi)$]. As a consequence, all the quantum mechanical imposibility principles are broken (see Ref. 1), and the construction of the convex set S described in Ref. 7 leads to a generalized simplex with the resulting quantum logic distributive. Hence, every group invariant statistical theory built up upon the nonlinear evolution equations (3.1) is "structurally classical" in the sense of the classification of Sec. 3.

6. NONLINEAR LATTICE MODEL

The considerations of Sec. 4 also permit one to determine the mobility of a more general lattice model of the nonlinear Schrödinger dynamics. Here, the wave function ψ is replaced by an infinite-component complex vector of $l^2 \psi = ||\psi_j||, j = ..., -1,0,1,\cdots$. Instead of the Schrödinger wave equation (3.1), one has a differential-difference equation with a finite difference operator playing the role of Laplacian $(\Delta\psi)_j = \psi_{j+1} - 2\psi_j + \psi_{j-1}$. As the term $-2\psi_j$ is inessential and can be absorbed by redefining the potential, the resulting equation is

$$i\frac{\partial\psi_j}{\partial t} = -(\psi_{j+1} + \psi_{j-1}) + \epsilon f(|\psi_j|^2)\psi_j + V_j\,\psi_j, \qquad (6.1)$$

where the numbers $V_j (j = ..., -1,0,1,...)$ form an analog of the external potential and the real function $f: R^+ \to R$ defines the nonlinearity. Below, f is assumed continuous with its first derivative. The quantity $|\psi|^2 = \Sigma_{j=-\infty}^{+\infty} |\psi_j|^2$ is conserved by the evolution (6.1), and further considerations will be restricted to the subset of unit vectors $\Psi = \{\psi \in l^2 : |\psi|^2 = 1\}$. Since any two vectors of Ψ which are proportional remain proportional after the evolution (6.1), it may be assumed that they label the same pure state. With this assumption the manifold of pure states Φ becomes the set of unit rays in l^2. Below, it will be convenient to use Ψ instead of Φ to label the pure states, remembering, however, that the unit vectors $\psi \in \Psi$ are redundant. Before any elements of a statistical theory are introduced it is worthwhile to examine the dynamical mobility of the ψ vectors obeying Eq. (6.1). As before, the semigroup of mobility will be described in terms of the generator fields now acting as differential operators on the functions $\phi \in C^\infty(\Psi)$ of an infinity of coordinates: $\phi(\psi) = \phi(\ldots, \psi_{-1}, \bar{\psi}_{-1}, \psi_0, \bar{\psi}_0, \ldots)$. To write down these generators explicitly it is convenient to rewrite Eq. (6.1) as

$$\frac{\partial}{\partial t}\psi_j = i(\psi_{j+1} + \psi_{j-1}) - i\epsilon f(|\psi_j|^2)\,\psi_j - iV_j\,\psi_j, \tag{6.2a}$$

$$\frac{\partial}{\partial t}\bar{\psi}_j = -i(\bar{\psi}_{j+1} + \bar{\psi}_{j-1}) + i\epsilon f(|\psi_j|^2)\,\bar{\psi}_j + iV_j\,\bar{\psi}_j. \tag{6.2b}$$

Now, the form of the vector field X generating Eq. (6.1) is found from the condition $(\partial/\partial t)\,\phi(\psi) = X\phi(\psi)$ $[\psi \in \Psi, \phi \in C^\infty(\Psi)]$:

$$X = \sum_{j=-\infty}^{+\infty}\Big[i(\psi_{j+1} + \psi_{j-1})\frac{\partial}{\partial\psi_j} - i(\bar{\psi}_{j+1} + \bar{\psi}_{j-1})\frac{\partial}{\partial\bar{\psi}_j} - i\epsilon f(|\psi_j|^2)\Big(\psi_j\frac{\partial}{\partial\psi_j} - \bar{\psi}_j\frac{\partial}{\partial\bar{\psi}_j}\Big) - iV_j\Big(\psi_j\frac{\partial}{\partial\psi_j} - \bar{\psi}_j\frac{\partial}{\partial\bar{\psi}_j}\Big)\Big]. \tag{6.3}$$

The field (6.2) is an analog of the Hamilton operator in the nonlinear theory. In particular, the choice $V_j = 0$ $(j = \ldots, -1, 0, 1, \ldots)$ gives the generator of the free evolution X_0. By regrouping terms in Eq. (6.3) one obtains

$$X_0 = \sum_{j=-\infty}^{+\infty}(D_j^+ + D_j^-) + \sum_{j=-\infty}^{+\infty} f(|\psi_j|^2)\,D_j, \tag{6.4}$$

where

$$D_j^+ = i\Big(\psi_{j+1}\frac{\partial}{\partial\psi_j} - \bar{\psi}_{j+1}\frac{\partial}{\partial\bar{\psi}_j}\Big), \tag{6.5}$$

$$D_j^- = i\Big(\psi_{j-1}\frac{\partial}{\partial\psi_j} - \bar{\psi}_{j-1}\frac{\partial}{\partial\bar{\psi}_j}\Big), \tag{6.6}$$

$$D_j = -i\Big(\psi_j\frac{\partial}{\partial\psi_j} - \bar{\psi}_j\frac{\partial}{\partial\bar{\psi}_j}\Big). \tag{6.7}$$

The operators (6.6)–(6.7) are the elementary "tendencies" in the evolution processes (6.1); they act on the components of ψ according to the simple rules

$$\begin{array}{ccc}
D_j^+: & D_j^-: & D_j: \\
\psi_j \to i\,\psi_{j+1} & \psi_j \to i\,\psi_{j-1} & \psi_j \to -i\,\psi_j \\
\bar{\psi}_j \to -i\,\bar{\psi}_{j+1} & \bar{\psi}_j \to -i\,\bar{\psi}_{j-1} & \bar{\psi}_j \to i\,\bar{\psi}_j \\
\left.\begin{array}{c}\psi_k\\ \bar{\psi}_k\end{array}\right\}\to 0 & \left.\begin{array}{c}\psi_k\\ \bar{\psi}_k\end{array}\right\}\to 0 & \left.\begin{array}{c}\psi_k\\ \bar{\psi}_k\end{array}\right\}\to 0 \\
(k \neq j) & (k \neq j) & (k \neq j)
\end{array} \tag{6.8}$$

Since the summation in Eq. (6.4) runs from $-\infty$ to $+\infty$, the generator X_0 may equivalently be written

$$X_0 = \sum_{j=-\infty}^{+\infty}(D_j^+ + D_{j+1}^-) + \sum_{j=-\infty}^{+\infty}\epsilon f(|\psi_j|^2)\,D_j. \tag{6.9}$$

The transition from Eqs. (6.4) to (6.9) involves only a simple rearrangement of terms, and so does not affect the convergence of the infinite sum (6.4). The operators $A_j = D_j^+ + D_{j+1}^-$ produce the "circular operations" in the pair of components ψ_j and ψ_{j+1}, and vanish elsewhere:

$$A_j = D_j^+ + D_{j+1}^- = i\Big(\psi_{j+1}\frac{\partial}{\partial\psi_j} + \psi_j\frac{\partial}{\partial\psi_{j+1}} - \bar{\psi}_{j+1}\frac{\partial}{\partial\bar{\psi}_j} - \bar{\psi}_j\frac{\partial}{\partial\bar{\psi}_{j+1}}\Big):$$

$$\psi_j \underset{i}{\overset{i}{\rightleftarrows}} \psi_{j+1}, \qquad \bar{\psi}_j \underset{-i}{\overset{-i}{\rightleftarrows}} \bar{\psi}_{j+1},$$

$$\psi_k \to 0, \qquad \bar{\psi}_k \to 0. \tag{6.10}$$

Their natural counterparts are

$$B_j = \psi_j\frac{\partial}{\partial\psi_{j+1}} - \psi_{j+1}\frac{\partial}{\partial\psi_j} + \bar{\psi}_j\frac{\partial}{\partial\bar{\psi}_{j+1}} - \bar{\psi}_{j+1}\frac{\partial}{\partial\bar{\psi}_j},$$

$$\psi_j \underset{+1}{\overset{-1}{\rightleftarrows}} \psi_{j+1}, \qquad \bar{\psi}_j \underset{+1}{\overset{-1}{\rightleftarrows}} \bar{\psi}_{j+1},$$

$$\psi_k \to 0, \qquad \bar{\psi}_k \to 0, \tag{6.11}$$

$$C_j = D_{j+1} - D_j = i\Big(\psi_j\frac{\partial}{\partial\psi_j} + \psi_{j+1}\frac{\partial}{\partial\psi_{j+1}} - \bar{\psi}_j\frac{\partial}{\partial\bar{\psi}_j} + \bar{\psi}_{j+1}\frac{\partial}{\partial\bar{\psi}_{j+1}}\Big):$$

$$\begin{array}{l}
\psi_j \to i\,\psi_j, \quad \bar{\psi}_j \to -i\,\bar{\psi}_j, \\
\psi_{j+1} \to -i\,\psi_{j+1}, \quad \bar{\psi}_{j+1} \to i\,\bar{\psi}_{j+1}, \\
\psi_k \to 0, \quad \bar{\psi}_k \to 0 \quad (k \neq j).
\end{array} \tag{6.12}$$

The differential operators A_j, B_j, C_j, when acting on the pair of components ψ_j, and ψ_{j+1}, behave as Pauli matrices multiplied by the imaginary unit i. The evolution generator (6.2) can now be written

$$X = \sum_{j=-\infty}^{+\infty} A_j + \sum_{j=-\infty}^{+\infty}\epsilon f(|\psi_j|^2)\,D_j + \sum_{j=-\infty}^{+\infty} V_j\,D_j = A + \epsilon N + V, \tag{6.13}$$

the meaning of A, N, and V self-evident. Notice that the "nonlinear" and potential parts N and V can be equivalently written in terms of C_j by applying an Abel transformation. One has

$$V = \sum_{j=-\infty}^{+\infty} w_j C_j, \tag{6.14}$$

where the numbers $w_j\,(j = ..., -1,0,1,...)$ represent a superpotential: $w_{j-1} - w_j = V_j$. If the sum $\Sigma_{j=-\infty}^{+\infty} f(|\psi_j|^2)$ is convergent, N can be written explicitly in any of the forms

$$N = \sum_{j=-\infty}^{+\infty} \left\{\left[\sum_{k=j+1}^{+\infty} f(|\psi_k|^2)\right] C_j\right\} + ...$$
$$= \sum_{j=-\infty}^{+\infty} \left\{\left[\sum_{k=-\infty}^{j} f(|\psi_k|^2)\right] C_j\right\} + ... \tag{6.15}$$

The terms dropped in Eqs. (6.14) and (6.15) generate only trivial rotations in ψ.

Achievable operations

The vector fields (6.13) are the basic generators of G, the simplest achievable operations being of the form $\exp \tau X$, where $\tau \geqslant 0$ and X is given by Eq. (6.13). More general operations can be now obtained by taking products and limits of the operations $\exp \tau X$. In particular, by taking a strong external potential $V_j = (1/\epsilon)\,\alpha_j$ and letting it act during a short time interval $\tau = \epsilon \to 0$, one achieves the dynamical operation $\exp \alpha$, where the generator α has the form of the pure potential part of Eq. (6.13) (shock transformation)

$$\alpha = \sum_{j=-\infty}^{+\infty} \alpha_j D_j \quad (\alpha_j \in \mathbb{R}). \tag{6.16}$$

By now superposing the "natural evolution" $\exp\tau X$ with two opposite shock transformations generated by Eq. (6.16), one can produce a new class of dynamical transformations of the form $\exp\alpha \exp\tau X \exp(-\alpha) = \exp\tau\widetilde{X}$, where $\widetilde{X} = \exp\alpha X \exp(-\alpha)$. Since $e^{\alpha}\psi_j e^{-\alpha} = e^{-i\alpha_j}\psi_j$,

$$e^{\alpha}\bar{\psi}_j e^{-\alpha} = e^{i\alpha_j}\bar{\psi}_j, e^{\alpha}(\partial/\partial\psi_j)e^{-\alpha} = e^{i\alpha_j}(\partial/\partial\psi_j),$$
$$e^{\alpha}(\partial/\partial\bar{\psi}_j)e^{-\alpha} = e^{-i\alpha_j}(\partial/\partial\bar{\psi}_j),$$

and

$$e^{\alpha} D_j e^{-\alpha} = D_j,$$

hence the modified generator becomes

$$\widetilde{X} = i \sum_{j=-\infty}^{+\infty} \left(e^{i\gamma_j}\psi_{j+1}\frac{\partial}{\partial\psi_j} + e^{-i\gamma_j}\psi_j\frac{\partial}{\partial\psi_{j+1}} - e^{-i\gamma_j}\bar{\psi}_{j+1}\frac{\partial}{\partial\bar{\psi}_j} + e^{i\gamma_j}\bar{\psi}_j\frac{\partial}{\partial\bar{\psi}_{j+1}}\right) + \epsilon N + V, \tag{6.17}$$

where $\gamma_j = \alpha_j - \alpha_{j+1}$. After an obvious decomposition

$$\widetilde{X} = \sum_{j=-\infty}^{+\infty} (\cos\gamma_j A_j + \sin\gamma_j B_j) + \epsilon N + V. \tag{6.18}$$

The vector fields (6.18) now induce a wide class of achievable operations. In particular, taking $\cos\gamma_j = -1$ $(j = ..., -1,0,1,...)$, $V = 0$, one can obtain an analog of the free evolution generators with the sign of the "linear part" reversed

$$X_1 = -A + \epsilon N. \tag{6.19}$$

More generally, taking $\cos\gamma_j = \epsilon_j = \pm 1$ and $V \neq 0$, one obtains an analog of the generator (6.13) with only some of the signs in A reversed:

$$X_\epsilon = \sum_{j=-\infty}^{+\infty} \epsilon_j A_j + \epsilon N + V. \tag{6.20}$$

The unit sphere Φ admits a natural embedding in a real Banach space. If the potentials $V = ||V_j||, j = ..., -1,0,1,...$ are bounded, the vector fields (6.2)–(6.20) are bounded and fulfil the Lipschitz condition which makes possible the application of the Trotter formula (4.7). This formula allows one to approximate the evolution operations corresponding to positive linear combinations of the vector fields (6.20). In particular, one can approximate

$$\exp\tau\epsilon N = \exp(\tfrac{1}{2}\tau X_0 + \tfrac{1}{2}\tau X_1) \quad (\tau \geqslant 0). \tag{6.21}$$

This means that in a single evolution act one can annihilate with any desired accuracy the effect of the linear part of the free evolution, forcing the system to perform an operation dictated by the "pure nonlinearity." The Trotter mechanism provides also an analog of the evolution generator (6.13) with only a part of the linear terms cancelled and part of them still present. This is done by taking an average of the vector fields (6.13) and (6.15). The resulting new generator is

$$X_0 = \tfrac{1}{2}(X + X_\epsilon) = \sum_{j=-\infty}^{+\infty} \theta_j A_j + \epsilon N + V, \tag{6.22}$$

where $\theta_j = \frac{1}{2}(\epsilon_j + 1)$ takes the values 0 and 1 for $j = ..., -1,0,1,....$ The existence of the vector fields (6.22) among the generators of the semigroup G means that by applying adequate external forces one can influence the system to perform a selective operation, where some of the linear terms of A do not contribute but some are essential (for $\theta_j = 1$). Obviously, if $\theta_k = 0$ for a certain k, then the evolution operation leaves the subsets of components $\{\psi_j : j \leqslant k\}$ and $\{\psi_j : j > k\}$ decoupled. If $\theta_k = \theta_l = 0$ for two integers $k,l,k < l$, then the subset of components $\{\psi_{k+1},...,\psi_l\}$ becomes decoupled from all the rest [meaning that the components $\phi_{k+1},...,\phi_l$ of the vector $\phi = (\exp\tau X_0)(\psi)$ depend only upon the coordinates $\psi_{k+1},...,\psi_l$ of the vector ψ]. The simplest "decoupled operations" are obtained by taking $\theta_j = \delta_{jk}$ in Eq. (6.22). The corresponding generators are

$$X_{(k,k+1)} = A_k + \epsilon N + V. \tag{6.23}$$

The operations $\exp\tau X_{(k,k+1)}$ transform the ψ vectors very simply. Every component ψ_j with $j \neq k, k+1$ is multiplied by a phase factor. The only two components transformed nontrivially are ψ_k and ψ_{k+1}. In the two dimensional subspace of these components the generator (6.23) induces the transformations correspondingly exactly to the ones of Sec. 4. Thus, employing adequate external forces, one can operate on every two components ψ_k and ψ_{k+1}, making them imitate the model of Sec. 4. This fact has important consequences for the mobility of the lattice model (6.1). Indeed, one can see that by operating on successive pairs of components of ψ one can produce evolution operations separating two arbitrarily close state vectors. As an example, consider two "wave packets" $\psi, \eta \in \Psi$ both with finite num-

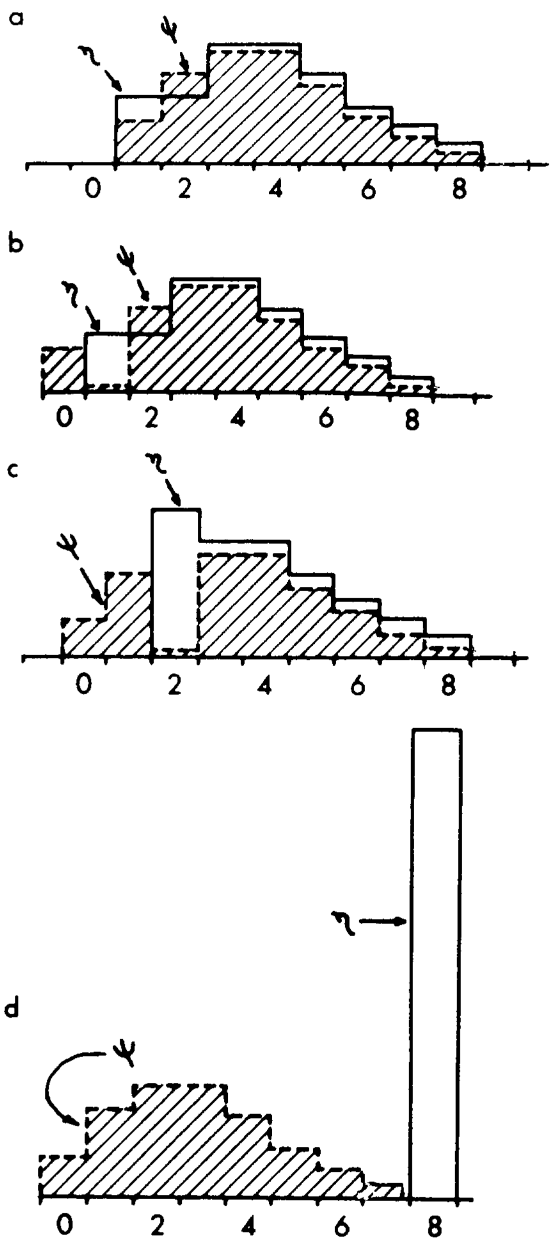


FIG. 5.

bers of nonvanishing components, differing slightly one from the other: $|\psi_1|^2 = |\eta_1|^2 + \delta$, $|\psi_2|^2 = |\eta_2|^2 - \delta$, $\psi_3 = \eta_3,...,\psi_n = \eta_n$; $\psi_k = \eta_k = 0 (k \neq 1,...,n)$. (The absolute values of ψ and η are drawn in Fig. 5). Now, by using the operations $\exp\tau X_{(0,1)}$, one can make the pairs ψ_0, ψ_1 and η_0, η_1 to behave like replicas of the system of Sec. 4. In particular, one can perform an operation which almost compresses both components of ψ into the single cell $j = 0$ while simultaneously shifting the components of η to the cell $j = 1$ (north and south poles of the sphere in Fig. 4). Having accomplished this manuever with a sufficient accuracy, the next transformation must be of the form $\exp\tau X_{(1,2)}$ and operate in the cell $j = 1,2$. As before, the contents of ψ in this cell are shifted to the left (to subcell $j = 1$) whereas the contents of η are shifted to the right (subcell $j = 2$). Proceeding so, one defines a sequence of separation operations in the successive cells (0,1), (1,2), (2,3),..., $(n-1, n)$ always shifting the local contents of ψ to the left and that of η to the right. The successive stages of these operations can approximate, with any desired accuracy, the sequence of picture in Fig. 5.

As a final result, the waves packets ψ and η, though almost identical at the beginning, are almost completely separated at the end: the packet η is compressed to the single cell $j = n$ whereas the packet ψ is filtered to the left. The existence of the above separation operation allows the construction of infinitely selective measuring devices able to accept ψ and simultaneously reject η with any desired accuracy: for if there is a method of distinguishing precisely the space-separated packets (which is likely to be assumed), then there is also a method of filtering unmistakably any two packets, however little they differ from one another. As in Sec. 4, the existence of the infinitely selective measurements violates the quantum mechanical impossibility principles and shows that the nonlinear wave packet (6.1) acquires some classical features.

Assume now that an invariant statistical theory is built up on the dynamical equations (6.1). Then one deals with the mobility group M generated by Eqs. (6.1). The contents of M may be estimated according to the Lie algebra of its generators fields of BL class. Because of the arbitrary γ_j's in Eq. (6.13), this algebra contains every pair of vector fields A_j, B_j $(j = ..., -1,0,1,...)$. Consistently, it also contains every C_j $(j = ..., -1,0,1,...)$. For any fixed j, the triple of vector fields A_j, B_j and C_j generates all possible unitary transformations of the pair of components ψ_j and ψ_{j+1}. By superposing many unitary transformations in the successive two-dimensional cells of components ψ_j and ψ_{j+1}, one can generate, in the limit, any unitary transformation in the whole of l^2. Hence, M contains the unitary group in l^2. The other dynamical operations are due to the vector field N. Hence, M is the smallest topologically closed group of transformations of Ψ which contains the unitary transformations and the nonlinear operations $\exp\tau\epsilon N$. The corresponding transformation group of the set of rays Φ is so rich that the theory of special functions on Φ becomes trivial: the only admissible class of functions F invariant under the group M and essentially wider than the subspace of constants is the whole class of real continuous function on Φ $[F = C(\Phi)]$. As a consequence, every group invariant statistical theory built up on Eqs. (6.1) has a classical structure of states (a simplex). As already noticed in Sec. 3, this does not mean, however, the existence of hidden parameters or the possibility of returning to a causal scheme (see also Ref. 12).

7. CORRESPONDENCE

The results of Sec. 6 exhibit certain special features of nonlinear wave dynamics. The nonlinear system (6.1) is characterized by an excessive mobility. If only a nonlinear term is present in Eqs. (6.1), the mobility group "blows up." The class of observables becomes so rich that the convex set S becomes a generalized simplex (classical geometry). The same phenomenon has been already observed, for a different class of nonlinear equations, by Haag and Bannier.[12] A question arises whether the return to the simplex geometry is a general tendency in nonlinear theories? Another problem concerns the correspondence principle: it looks as if there is a structural discontinuity between the linear theory and the nonlinear theory with nonlinearity tending to zero.

This last phenomenon arises from the fact that the scheme of Sec. 3 has a global character. The structures introduced in Sec. 2 contain a limiting transition which does not commute with vanishing of nonlinearity ($\epsilon \to 0$). Indeed, the semigroup G is the family of all transformations $\Phi \to \Phi$ which are dynamically achievable, no matter how long a time may they require to be accomplished. Similarly, F was defined as the class of functions representing all possible experiments, no matter how much time and accuracy they may require to be performed. Hence, the class F represents a kind of absolute knowledge.

The above global description cannot be consistently avoided if one is interested in the complete nature of the system. However, for some practical purposes it may become overloaded with information. The operations of separation of the wave packets represented in Fig. 5 always exist: however, if the nonlinerity is too small, it may take too long a time to accomplish them and therefore they are of no practical importance. This suggests that in some circumstances the global scheme might be ignored in favor of partial descriptions.

Instead of representing the system in terms of the whole class of observables F, it might be convenient to introduce some subclasses $F' \subset F$ corresponding to restricted types of experiments. An interesting example of "partial perception" arises if the measurements are confined to a finite time interval $[t,t']$ and limited by a finite accuracy. Let $F_{[t,t']}$ denote the corresponding class of observables. For $t = t'$, $F_{[t,t']}$ contains only the "immediately measurable" observables. For the system of Sec. 6 they may be assumed identical with the primitive observables of the "Born interpretation": $p_j(\psi) = |\psi_j|^2$; $p(\psi) = \Sigma_i \mu_j p_j(\psi)$. If $\tau = t' - t > 0$, but ϵ is very small, so that $\epsilon\tau$ is small, the higher order contributions to the quadratic probabilities are still beyond the threshold of detection. An observer confined to the interval $[t,t']$ therefore will not register the existence of the higher order observables: the class $F_{[t,t']}$ reflecting *his* approximate experience will coincide with the class of quadratic forms as in the linear theory. The situation changes if the interval $[t,t']$ increases. The terms proportional to $(\epsilon\tau)^n$ $(n = 1,2,...)$ then successively cross the detection barrier and $F_{[t,t']}$ is gradually extended to include the higher order forms of ψ. This produces a sequence of increasingly precise schemes with expanding observables and regressing impossibility principles. [1] Finally, the class of observables extends to include all the real forms of ψ, thus providing an infinitely precise perception of the pure states (rays) characteristic for classical theories. This process is conditioned by the value of ϵ: the smaller the nonlinearity, the longer the time necessary to enrich the observables. It is thus seen that the nonlinear theory with $\epsilon \to 0$ tends to the linear one in the local sense, simultaneously conserving its different global shape.

The description in terms of partial observables is known in the $*$-algebra approach to quantum theories. In particular, the class $F_{[t,t']}$ is an inexact analog of the Haag algebra of observables localized in space–time. However, there is an essential structural difference. In present day quantum theories the partial pictures of a theory can differ only in size but not in shape: they all reproduce the same type of structure based on $*$ algebras and Hilbert spaces. This is not necessarily so in a generalized statistical theory. Here the system may reveal different shapes to different observers. An example is precisely the nonlinear system of Sec. 6 which exhibits a whole variety of geometric structures from the orthodox quantum one up to the orthodox classical. [21] It thus looks as though Haag's idea of partial observables, when translated to the domain of nonlinear theories, might reveal a new mechanism of consistency between many simultaneous structures of the same physical system. If a finite accuracy is assumed, it can also show how the nonlinearity can provide an asymptotic return to a classical geometry, which does not imply the return to a causal scheme. The nonlinear system (6.1) becomes classical when it grows old. The system for which this happens might be called "asymptotically classical."

8. OPEN PROBLEMS

The considerations of Secs. 6 and 7 illustrate the somewhat special place of global mobility. This concept is likely to appear in any dynamical theory of open systems. Once it exists it starts to play a decisive role: for because of the consistency links it conditions the geometry of the theory. This may be of interest for some domains where the dynamics has been already postulated but the physical interpretation is still missing.

(1) Recently, certain c-number fields have been studied because of the existence of solitons. The most consistent quantization of these fields seemingly leads via Feynman path integrals. [22] One might also think, however, that the soliton fields are the soliton fields, are already the proper nonlinear representations for the physical quanta (as in de Broglie's proposals) and therefore should not be further quantized. Instead, they should be investigated "as they are". This would involve the study of the soliton geometry and prolongation structures. [23] In the case of an open system this would lead to the problem of mobility. An interesting example of an open soliton system of form (2.1) with logarithmic nonlinearity $f(\zeta) = \ln\zeta$ and with external potentials was discussed by Białynicki-Birula and Mycielski. [24] Though the function $\omega(z) = \frac{1}{2}\ln[(R+z)/2] - \ln[(R-z)/2]$ has singularities, the results of Sec. 4 suggest that the gaussons of Bialynicki-Birula–Mycielski are structurally classical, even though they may appear in a probabilistic theory.

(2) One of the gaps in the present day theory lies in between the quantum theories and general relativity. The attempts to quantize gravity by means of the standard formalism have not been conclusive. One of the most intriguing problems concerns the hypothetical graviton. There are opinions that the single graviton in vacuum (if any such entity exists) should be described like any other particle with spin, in the spirit of the linear laws of quantum mechanics, and that the nonlinearity of the macroscopic gravitational

field is simply due to the interactions in a cloud of many gravitons. A completely different theory of a nonlinear graviton has been proposed by Penrose,[25] who postulates that the graviton is described by a left-flat complex Einsteinian field and carries its own portion of the curvature. An algebraic theory of complex Einsteinian fields has been developed recently by Plebanski.[26] The theory of the nonlinear graviton is not yet complete as the graviton exists "in itself" and the statistical interpretation is missing. The general statistical theory of Sec. 3 is not directly applicable here, as it was formulated in a flat space–time, and moreover there is not yet a theory describing the graviton as an open system Still the scheme of Sec. 3 is at least devoid of some handicaps: for it seems that the general mobility group rather than the unitary group and the functional rather than the operator observables are general enough to describe the hypothetical entity.

ACKNOWLEDGMENTS

The author is grateful to Professor B. Nagel for his hospitality at the Royal Institute of Technology in Stockholm and to Professor F.A.E. Pirani for his hospitality at King's College, London, where parts of this investigation were carried. The author would like to thank also Professor J. Plebanski, Professor G. Massieu, Professor H. Moreno, and Professor A. Garcia for the opportunity of presenting these ideas on the Symposium on Mathematical Physics in Mexico (1976) and to Professor R. Ingarden for a similar opportunity at the Symposium on Mathematical Physics in Torun (1976). Thanks are also due to Professor I. Bialynicki-Birula, Professor A. Trautman, Dr. J. Waniewski, Professor S. Woronowicz, and other colleagues for their interest in this work and helpful discussions.

[1]B. Mielnik, Commun. Math. Phys. **37**, 221 (1974).
[2]B. Mielnik, Rep. Math. Phys. **12**, 331 (1977).
[3]The above definition arises from the demand that the convergence of physically interpretable functions should make sense even if the function arguments are not measurable with infinite accuracy.
[4]J. Schwinger, Phys. Rev. **91**, 713 (1953).
[5]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[6]E.G. Beltrametti and G. Casinelli, Commun. Math. Phys. **40**, 7 (1975).
[7]E.B. Davies, Commun. Math. Phys. **49**, 113 (1976).
[8]C.M. Edwards, Commum. Math. Phys. **24**, 260 (1971).
[9]J.C.T. Pool, Commun. Math. Phys. **9**, 118 (1968); **9**, 212 (1968).
[10]E. Lubkin, J. Math. Phys. **15**, 673 (1974).
[11]J. Waniewski, Rep. Math. Phys. **11**, 331 (1977).
[12]R. Haag and U. Bannier, Commun. Math. Phys. **60**, 1 (1978).
[13]S.L. Woronowicz, Commun. Math. Phys. **51**, 243 (1976).
[14]E.M. Alfsen and W. Schultz, Report of the Institute of Mathematics, University of Oslo, ISBN 82-553-0206g (1975).
[15]G. Ludwig, Z. Naturforsh. Teil **22A**, 1303 (1967); **22**, 1324 (1967); see also *Foundations of Quantum Mechanics and Ordered Linear Space*, Lecture Notes in Physics, edited by A. Hartkämper and H. Neumann, Marburg 1973, (Springer, Berlin, 1973), Vol. 29.
[16]G. Birkhoff, and J. von Neumann, Ann. Math. **37**, 823 (1936).
[17]C. Piron, Helv. Phys. Acta **42**, 842 (1969).
[18]B. Mielnik, in *Quantum Mechanics, Determinism, Causality and Particles*, edited by Flato *et al.*, (Reidel, Dordrecht, 1976), p. 117.
[19]This classification concerns only a part of the properties of the statistical theory. It coincides with the one based on the distributive property of quantum logic. However, as pointed out by Haag and Bannier, it does not imply the existence of hidden parameters. Thus, there might be schemes structurally classical but still different from a classical theory.
[20]In some cases the group invariance follows from the semigroup invariance [J. Waniewski (private discussion)].
[21]An equivalent observation has already been made by Haag and Bannier, who noticed that the reduction of the wave packet for the nonlinear waves would lead to the pure states with a diminished norm. As a consequence, the information which we possess could be essentially affected by the information which has been lost. The significance at the Haag–Bannier paradox for the applicability of the nonlinear models is still to be discussed.
[22]L.D. Faddeev, Lecture at the Scandynavian Elementary Particle Symposium, Stockholm, May 1977.
[23]F.A.E. Pirani, Lecture at the Symposium on Mathematical Physics, Torun, December 1976).
[24]I. Bialynicki-Birula and J. Mycielski, Ann Phys. (Leipzig) **100**, 62 (1976).
[25]R. Penrose, Gen. Rel Grav: **7**, 31 (1976).
[26]J.F. Plebanski, J. Math. Phys. **16**, 2403 (1975); **18**, 2511 (1977).

A phase space approach to generalized Hamilton–Jacobi theory

J. Samuel
Department of Physics, Indian Institute of Science, Bangalore, 560012, India

(Received 5 December 1978)

Generalizations of H–J theory have been discussed before in the literature. The present approach differs from others in that it employs geometrical ideas on phase space and classical transformation theory to derive the basic equations. The relation between constants of motion and symmetries of the generalized H–J equations is then clarified.

INTRODUCTION

The Hamilton–Jacobi form of dynamics derives much of its importance today from the analogies it has with quantum mechanics. At a purely classical level, it has application in celestial mechanics where it provides the framework for canonical perturbation theory.

In the usual form of the theory, the momenta turn out to be the gradient of a function on configuration space called Hamilton's principal function. Rund[1] suggested that one may give the momenta alternative representation and arrive at a generalization of the theory. In Rund's theory the momenta are expressed in terms of n functions on configuration space (n being the number of degrees of freedom). It was pointed out by Baumeister[2] that the classical problem of Pfaff was relevant in choosing the representation for the momenta. The number of functions in the representation here lies between 1 and n. Baumeister's form of the theory contains as special cases both Rund's theory and traditional H–J theory.

Mukunda[3] has discussed Hamilton–Jacobi theory using geometrical ideas on phase space. As pointed out in Ref. 3 working in phase space rather than configuration space is advantageous because the action of canonical transformations is easy to visualize. The present approach to generalized Hamilton–Jacobi theory relies largely on geometrical ideas developed in Ref. 3. It is not clear at the moment whether the generalized theory has any relevance to quantum mechanics. It does seem safe to say however that analogies with quantum mechanics, if they exist, are not straightforward and will require some unearthing.

The material to be presented is arranged as follows. In Sec. I, surfaces in phase space and their behavior under canonical transformations of phase space are discussed. In Sec. II, the surfaces are given a representation (following Ref. 2) in terms of so called Clebsch potentials. The behavior of Clebsch potentials under the action of canonical transformations is also worked out. In Sec. III, the fact that the time evolution of a system is a canonical transformation is used to write down the equations for time evolution of the Clebsch potentials. These constitute the system of generalized H–J equations. In the last section, it is shown that a constant of motion may be used to generate new solutions of these equations from a given one.

I. GEOMETRY OF PHASE SPACE SURFACES

Consider the $2n$-dimensional phase space of a dynamical system: (p_r, q^r), $r = 1,\dots,n$. Let M be an m-dimensional surface embedded therein. The points of M may be parametrized by u_α, $\alpha = 1,\dots,m$ so that the surface now appears as $(p_r(u), q^r(u))$. Given a phase space function $A(p,q)$ one may restrict its arguments to M, thereby obtaining a function of u,

$$(A(p,q))_M = A(p(u),q(u)). \tag{1.1}$$

The symbol on the left denotes restriction to M.

The Lagrange bracket of two parameters u_α, u_β is defined as

$$[u_\alpha,u_\beta] = \left(\frac{\partial p_r(u)}{\partial u_\beta}\frac{\partial q^r(u)}{\partial u_\alpha} - \frac{\partial p_r(u)}{\partial u_\alpha}\frac{\partial q^r(u)}{\partial u_\beta} \right). \tag{1.2}$$

These form an $m\times m$, real, antisymmetric matrix. The rank of this matrix is assumed constant over M and is called the symplectic rank of the surface M. It is even and independent of the choice of parameters u_α. The rank of the augmented $m\times m+1$ matrix

$$\left([u_\alpha,u_\beta], p_r(u)\frac{\partial q^r(u)}{\partial u_\alpha} \right) \tag{1.3}$$

is called the character of M. From its definition it is clear that it is either equal to the symplectic rank or exceeds it by one.

Canonical transformations of phase space will be thought of in the active sense as a one-to-one mapping of phase space onto itself. Only infinitesimal transformations are considered and in what follows, higher order terms than we require will be dropped without comment. Under the action of a canonical transformation generated by $A(p,q)$, a surface M parametrized by u_α will go over to a surface M' which we once again parametrize by u_α. Thus a point $(p_r(u),q^r(u))$ on M will go over to a point $(p_r'(u),q'^r(u))$ on M'. The relation between the new and old surfaces is given by

$$\begin{aligned} p_r'(u) &= p_r(u) + \epsilon\{A, p_r\}_M, \\ q'^r(u) &= q^r(u) + \epsilon\{A,q^r\}_M. \end{aligned} \tag{1.4}$$

ϵ is an infinitesimal parameter and the braces are restrictions of Poisson brackets to M.

Canonical transformations have the property

$$p_r'(u)\frac{\partial q'^r(u)}{\partial u_\alpha} = p_r(u)\frac{\partial q^r(u)}{\partial u_\alpha} + \frac{\partial w(u)}{\partial u_\alpha}, \tag{1.5}$$

where $w(u) = \epsilon(A + p_r\{A,q^r\})_M$.

The well known invariance of Lagrange brackets under canonical transformations follows easily from Eqs. (1.5):

 0022-2488/80/010055-05$01.00

$$[u_\alpha, u_\beta]_{q'p'} = [u_\alpha, u_\beta]_{qp}. \tag{1.6}$$

Thus the matrix of Lagrange brackets and therefore the symplectic rank are canonical invariants. Under a canonical transformation the last column of the augmented matrix (1.3) will increase by a gradient while the other columns remain the same. Since at a given point, by choosing $A\,(p,q)$, one can make this term have any value, it is clear that the character in general may change under a canonical transformation. For this reason the character of a surface will not play an important role in our considerations.

Under a canonical transformation generated by $A\,(p,q)$, if the points of M remain in M, then the transformation is said to leave M invariant; or more briefly, M is an eigensurface of $A\,(p,q)$. If $(p_r'(u), q'^r(u))$ and $(p_r(u), q^r(u))$ describe the same surface M, it is clear that for some u',

$$p_r'(u) = p_r(u'), \quad q'^r(u) = q^r(u'). \tag{1.7}$$

This leads, via a Taylor expansion for $(p_r(u'), q^r(u'))$, to the $2n$ equations

$$\begin{aligned} \frac{\partial p_r(u)}{\partial u_\alpha}\,\delta u_\alpha &= \epsilon\{A, p_r\}_M, \\ \frac{\partial q^r(u)}{\partial u_\alpha}\,\delta u_\alpha &= \epsilon\{A, q^r\}_M, \end{aligned} \tag{1.8}$$

where $\delta u_\alpha = u_\alpha' - u_\alpha$ are m suitable quantities. Thus if M is to be an eigensurface of $A\,(p,q)$, the $2n$-m conditions (1.8) must be satisfied by M and the phase space function $A\,(p,q)$. By means of laborious but quite straightforward manipulations with Eqs. (1.8) one can show that if M is an eigensurface of two phase space functions A and B, it is also an eigensurface of their Poisson bracket $\{A,B\}$. One might have expected this from the Baker–Campbell–Hausdorff formula[4] for the composition of canonical transformations. Its verification, though tedious in our present description of surfaces, is easy in an alternative one which we now discuss.

An m-dimensional surface M may be characterized by $2n-m$ independent phase space functions, $\varphi_\sigma(p,q)$, $\sigma = 1,\ldots,2n-m$, whose restrictions to M vanish. This description is suggested by the theory of constrained systems[5] and we will freely use theorems and vocabulary developed in that context. A phase space function $A\,(p,q)$ is said to vanish weakly on M if its restriction to M vanishes.

This is written

$$A \approx 0. \tag{1.9}$$

If the restrictions of its first derivatives with respect to q and p also vanish on M, then A is said to vanish strongly. This is written

$$A \equiv 0. \tag{1.10}$$

We refer to Ref. 5 for the result that a function which vanishes weakly on M is strongly equal to a linear combination of the φ's which describe M.

Under a canonical transformation, if M goes over to M', one may describe M' by the vanishing of $\varphi_\sigma'(p,q)$. The relation between the two sets of φ's is

$$\varphi_\sigma'(p,q) = \varphi_\sigma(p,q) + \epsilon\{\varphi_\sigma, A\}, \tag{1.11}$$

A being the generator of the transformation. The condition for a canonical transformation generated by A to leave M invariant is

$$\{\varphi_\sigma, A\} \approx 0. \tag{1.12}$$

With use of the Jacobi identity and the result following Eq. (1.10), it is easily seen that for two phase space functions A and B, $\{\varphi_\sigma, A\} \approx 0$ and $\{\varphi_\sigma, B\} \approx 0$ imply $\{\varphi_\sigma, \{A, B\}\} \approx 0$. Thus, an eigensurface of A and B is also an eigensurface of their Poisson bracket. The corresponding result in the theory of constrained systems is proved in Ref. 5. Its statement in that context is that the Poisson bracket of two first class functions is also first class.

We now specialize to the case where $m = n$. Further let us assume that the q's themselves may be made to serve as parameters on M. We thus restrict the surfaces under consideration and this restriction will hold through the rest of this article. The points of M now appear as $p_r(q)$. The Lagrange brackets take the form

$$[q_i, q_j] = \frac{\partial p_i(q)}{\partial q^j} - \frac{\partial p_j(q)}{\partial q^i}. \tag{1.13}$$

Under a canonical transformation a surface M described by $p_r(q)$ will go over to M' which we describe by $p_r'(q)$. We use the same q's to describe both surfaces. This is at variance with an earlier convention whereby the original point on M and the transformed point on M' shared the same parameter. Taking this into account in Eq. (1.4) we have

$$p_r'(q) = p_r(q) + \epsilon\left(\frac{\partial A}{\partial q^r}\right)_M + \epsilon\,\frac{\partial p_r(q)}{\partial q^j}\left(\frac{\partial A}{\partial p_j}\right)_M, \tag{1.14}$$

or rewriting,

$$p_r'(q) = p_r(q) + \epsilon\,\frac{\partial A_M}{\partial q^r} + \epsilon\,[q_j, q_r]\{A, q^j\}_M. \tag{1.15}$$

This completes the geometrical framework which will be used in subsequent sections.

II. THE CLEBSCH REPRESENTATION

With every n-dimensional surface M we associate a covariant vector field $p_r(q)$. From $p_r(q)$ let us form $\omega = p_r(q)dq^r$ and think of it as a Pfaffian form. The character and symplectic rank of the form are defined as the character and symplectic rank of the associated surface.

Given a Pfaffian form ω with symplectic rank $2m$ the theory of Pfaffian forms[6] gives a systematic reduction procedure for arriving at an equivalent form

$$P_\alpha(q)\,dQ^\alpha(q) + d\psi(q), \tag{2.1}$$

where $\alpha = 1,\ldots,m$ and P_α, Q^α, and ψ are $2m+1$ functions on configuration space. The P's and Q's are functionally independent of one another. If the character of ω is $2m+1$, then ψ is functionally independent of the P's and Q's and Eq. (2.1) is in fact the most economical (in the sense that it employs the smallest number of functions) representation of ω. If the character is $2m$, then ψ is functionally dependent on (P_α, Q^α) and may be got rid of by a further Pfaffian reduction. Since the possibility of this last reduction depends on the character, which is not a canonical invariant concept, we will not make use of it. We content ourselves with the representation (2.1) for ω hereafter called the minimal one. The P's and Q's are $2m$ independent functions and ψ may or may not depend functionally on the P's and Q's. This leads to what is called the Clebsch representation for $p_r(q)$

$$p_r(q) = P_\alpha \frac{\partial Q^\alpha(q)}{\partial q^r} + \frac{\partial \psi(q)}{\partial q^r}. \tag{2.2}$$

The $2m+1$ functions $(P_\alpha, Q^\alpha, \psi)$ will be referred to as Clebsch potentials. The Clebsch potentials may be thought of as providing a third description of surfaces in phase space. Of course, only n-dimensional surfaces parametrizable by q are thus described.

The representation (2.2) is not unique. Two sets of Clebsch potentials are said to be related by a Clebsch gauge transformation if they describe the same surface $p_r(q)$. The relation between the two sets is now clarified.

$$P_\alpha\, dQ^\alpha + d\psi = P'_\alpha\, dQ'^\alpha + d\psi' \tag{2.3}$$

implies

$$P_\alpha\, dQ^\alpha - P'_\alpha\, dQ'^\alpha = d(\psi' - \psi) = d\chi. \tag{2.4}$$

As shown in Ref. 6, P'_α, Q'^α, and χ are functions of P_α and Q^α only. Equation (2.4) implies that the two sets (P'_α, Q'^α) and (P_α, Q^α) are related by a canonical transformation in the $2m$-dimensional (P_α, Q^α) space. All minimal representations are thus related by such transformations. Conversely, any such transformation will provide a new minimal representation starting from a given one. If the transformation is infinitesimal, the relation between the two sets of potentials may be conveniently expressed with the help of the infinitesimal generator of the transformation $F(P,Q)$

$$\begin{aligned}
\delta P_\alpha &= P_\alpha' - P_\alpha = \epsilon \frac{\partial F(P,Q)}{\partial Q^\alpha}, \\
\delta Q^\alpha &= Q'^\alpha - Q^\alpha = -\epsilon \frac{\partial F(P,Q)}{\partial P_\alpha}, \\
\delta \psi &= \psi' - \psi = \epsilon\Big(P_\alpha \frac{\partial F(P,Q)}{\partial P_\alpha} - F(P,Q)\Big).
\end{aligned} \tag{2.5}$$

F is an arbitrary function of its arguments and ϵ is an infinitesimal parameter. There is a risk of confusion between canonical transformations in the $2m$-dimensional (P_α, Q^α) space and phase space. To avoid this, in what follows the former will only be referred to as Clebsch gauge transformations.

Let us introduce the shorthand notation $\pi(q)$ to stand for the $2m+1$ Clebsch potentials $(P_\alpha(q), Q^\alpha(q), \psi(q))$. $\pi(q)$ defines a surface $p_r(q)$ via Eq. (2.2). Under a canonical transformation (in phase space!) the surface $p_r(q)$ will go over to $p'_r(q)$ which we may represent by new Clebsch potentials $\pi'(q)$. It follows from the canonical invariance of symplectic rank that the new set $\pi'(q)$ has the same number of functions $(2m+1)$ as $\pi(q)$. It is also clear, because of the Clebsch gauge degree of freedom, that $\pi'(q)$ cannot be determined by $\pi(q)$ and the canonical transformation. We expect arbitrariness in $\pi'(q)$ to the extent of a Clebsch gauge transformation. Use of the Clebsch representation (2.2) in Eq. (1.15) gives

$$P'_\alpha\, dQ'^\alpha + d\psi' = P_\alpha\, dQ^\alpha + d\psi + \epsilon\, dA_M + \epsilon\Big[dP_\alpha \frac{\partial Q^\alpha}{\partial q^j} - \frac{\partial P_\alpha}{\partial q^j} dQ^\alpha\Big]\{A, q^j\}_M. \tag{2.6}$$

A is the generator of the transformation and M the surface $p_r(q)$.

$$\begin{aligned}
P'_\alpha\, dQ'^\alpha + d\psi' &= \Big(P_\alpha - \epsilon\{A,q^j\}_M \frac{\partial P_\alpha}{\partial q^j}\Big) d\Big(Q^\alpha - \epsilon\{A,q^j\}_M \frac{\partial Q^\alpha}{\partial q^j}\Big) \\
&\quad + d\Big(\psi + \epsilon A_M + \epsilon P_\alpha \frac{\partial Q^\alpha}{\partial q^j}\{A,q^j\}_M\Big).
\end{aligned} \tag{2.7}$$

Comparing the two sides of this equation with the use of Eq. (2.5) yields

$$\begin{aligned}
\delta P_\alpha &= -\epsilon\{A,q^j\}_M \frac{\partial P_\alpha}{\partial q^j} + \epsilon \frac{\partial F(P,Q)}{\partial Q^\alpha}, \\
\delta Q^\alpha &= -\epsilon\{A,q^j\}_M \frac{\partial Q^\alpha}{\partial q^j} + \epsilon \frac{\partial F(P,Q)}{\partial P_\alpha}, \\
\delta\psi &= \epsilon A_M + \epsilon P_\alpha \frac{\partial Q^\alpha}{\partial q^j}\{A,q^j\}_M \\
&\quad + \epsilon\Big(P_\alpha \frac{\partial F(P,Q)}{\partial P_\alpha} - F(P,Q)\Big).
\end{aligned} \tag{2.8}$$

These equations give the change in the Clebsch potentials as a result of a canonical transformation of phase space. The appearance of the arbitrary function $F(P,Q)$ on the right is as anticipated.

III. THE GENERALIZED H–J EQUATIONS

Consider a system with the $2n$-dimensional phase space (p_r, q^r) and Hamiltonian $H(p,q,t)$. We can at $t=0$ think of an n-dimensional surface in phase space whose points appear as $p_r(q,0)$. Let the points of this surface move with time obeying Hamilton's equations. The points, as they move, fill out an $(n+1)$-dimensional region of phase space which can be viewed as consisting of an n parameter family of classical trajectories. At any time t, the points define a surface M whose points appear as $p_r(q,t)$. The correspondence between the one parameter family of surfaces and the n parameter family of trajectories is to be noted. The surface M may be given a Clebsch representation $\pi(q,t)$. The time evolution of a system is a canonical transformation generated by its Hamiltonian. To obtain the equations that determine the time evolution of Clebsch potentials, we replace A by H and ϵ by $-\delta t$ in Eqs. (2.8) and take the limit. The results are:

$$\begin{aligned}
\frac{\partial P_\alpha}{\partial t} &= \{H,q^j\}_M \frac{\partial P_\alpha}{\partial q^j} - \frac{\partial F(P,Q,t)}{\partial Q^\alpha}, \\
\frac{\partial Q^\alpha}{\partial t} &= \{H,q^j\}_M \frac{\partial Q^\alpha}{\partial q^j} - \frac{\partial F(P,Q,t)}{\partial P_\alpha}, \\
\frac{\partial\psi}{\partial t} &= -H_M - \{H,q^j\}_M P_\alpha \frac{\partial Q^\alpha}{\partial q^j} + F(P,Q,t) \\
&\quad - P_\alpha \frac{\partial F(P,Q,t)}{\partial P_\alpha}.
\end{aligned} \tag{3.1}$$

M is the time dependent surface whose Clebsch potentials are $\pi(q,t)$. Equations (3.1) constitute the system of generalized Hamilton–Jacobi equations. They are coupled first order partial differential equations for (P_α,Q^α,ψ) with independent variables q^r and t. If some specific choice is made for the arbitrary function $F(P,Q,t)$, then a well-defined initial value problem results.

From the remarks in Sec. II it is clear that the number of Clebsch potentials required to describe the surface does not change with time. Neither is the functional independence of the P's and Q's affected by time development. The functional independence of ψ however may change with time. A specific choice of the arbitrary function $F(P,Q,t)$ will be called a choice of gauge. The choice of gauge $F=0$ seems to enjoy a certain importance over others. In this gauge, P_α and Q^α turn out to be constants along the classical trajectories, while the rate of change of ψ is equal to the Lagrangian. That is to say, if $p_r(t)$ and $q^r(t)$ are solutions to Hamilton's equations, then

$$\begin{aligned} &P_\alpha(q(t),t)=P_\alpha(q(0),0),\\ &Q^\alpha(q(t),t)=Q^\alpha(q(0),0),\\ &\frac{d}{dt}[\psi(q(t),t)]=L(q(t),\dot q(t),t). \end{aligned} \tag{3.2}$$

Although P_α and Q^α are constant along the trajectories, they must not be thought of as "constants of motion" in the usual sense of the term as they are not phase space functions but configuration space functions.

IV. SYMMETRIES AND CONSTANTS OF MOTION

Given a solution $\pi(q,t)$ of Eqs. (3.1) one can effect on it a canonical transformation (as described in Sec. II) to give $\pi'(q,t)$. Here we discuss under what conditions $\pi'(q,t)$ is also a solution. If $G(p,q,t)$ is the phase space function which generates the transformation, a sufficient condition is that the effect of two successive transformations generated by G and H is the same regardless of the order in which they are applied. This diagram makes matters clearer:

$$\begin{tikzcd} \pi(t+\delta t) \arrow[r, "\epsilon G(t+\delta t)"] & \pi'(t+\delta t) \\ \pi(t) \arrow[u, "-\delta t H"] \arrow[r, "\epsilon G(t)."] & \pi'(t) \arrow[u, "-\delta t H"'] \end{tikzcd} \tag{4.1}$$

With each of the π's occurring in (4.1) one can associate an n-dimensional surface in phase space which may be described by the vanishing of n independent phase space functions $\varphi_\sigma(p,q)$, $\sigma=1\cdots n$. The effect of canonical transformations on such functions has been discussed in Sec. 1 to first order in infinitesimals. Here it is necessary to keep terms up to the second order in infinitesimals ϵ and δt jointly. Computing $\varphi'_\sigma(p,q,t+\delta t)$ [corresponding to $\pi'(t+\delta t)$] by the clockwise route in (4.1) gives

$$\begin{aligned}\varphi'(t+\delta t)=\varphi(t+\delta t)&-\epsilon\{G(t+\delta t),\varphi(t+\delta t)\}\\ &+\frac{\epsilon^2}{2}\{G(t+\delta t),\{G(t+\delta t),\varphi(t+\delta t)\}\},\end{aligned} \tag{4.2}$$

where the index σ has been dropped and

$$\varphi(t+\delta t)=\varphi(t)+\delta t\{H,\varphi(t)\}+\frac{\delta t^2}{2}\{H,\{H,\varphi(t)\}\}. \tag{4.3}$$

This gives

$$\begin{aligned}\varphi'_c(t+\delta t)=\varphi(t)&-\epsilon\{G,\varphi\}+\delta t\{H,\varphi\}\\ &+\frac{\epsilon^2}{2}\{G,\{G,\varphi\}\}+\frac{\delta t^2}{2}\{H,\{H,\varphi\}\}\\ &-\epsilon\delta t\{G,\{H,\varphi\}\}-\epsilon\delta t\left\{\frac{\partial G}{\partial t},\varphi\right\},\end{aligned} \tag{4.4}$$

where c denotes that this has been evaluated following the clockwise route. All time arguments on the r.h.s. of 4.4 are t. Likewise, following the anticlockwise route,

$$\begin{aligned}&\varphi'_a(t+\delta t)\\ &\quad=\varphi(t)+\delta t\{H,\varphi\}-\epsilon\{G,\varphi\}+\frac{\delta t^2}{2}\{H,\{H,\varphi\}\}\\ &\qquad+\frac{\epsilon^2}{2}\{G,\{G,\varphi\}\}-\epsilon\delta t\{H,\{G,\varphi\}\}.\end{aligned} \tag{4.5}$$

The difference between the two is given by

$$\begin{aligned}&\varphi'_c(t+\delta t)-\varphi'_a(t+\delta t)\\ &\quad=-\epsilon\delta t\left[\{G,\{H,\varphi\}\}+\{H,\{\varphi,G\}\}+\left\{\frac{\partial G}{\partial t},\varphi\right\}\right].\end{aligned} \tag{4.6}$$

Use of Jacobi's identity in (4.6) gives

$$\varphi'_c(t+\delta t)-\varphi'_a(t+\delta t)=-\epsilon\delta t\left\{\{G,H\}+\frac{\partial G}{\partial t},\varphi\right\}, \tag{4.7}$$

which can be written with the aid of the phase space function

$$A=\{G,H\}+\frac{\partial G}{\partial t} \tag{4.8}$$

as

$$\varphi'_c(t+\delta t)-\varphi'_a(t+\delta t)=-\epsilon\delta t\{A,\varphi\}, \tag{4.9}$$

which is the change in $\varphi_\sigma(p,q,t)$ due to a canonical transformation generated by $A(p,q,t)$. This might have been expected, save for the explicit time dependence of $G(p,q,t)$, from the BCH formula[4] for the composition of canonical transformations. In all of the above we have dealt with the effect of canonical transformations not on the Clebsch potentials themselves but on the underlying phase space structures. We can of course apply the transformations to the Clebsch potentials directly. This involves some algebra and yields the same result. In terms of Clebsch potentials, (4.9) may be written (save for a pure gauge) as

$$\pi'_c(t+\delta t)-\pi'_a(t+\delta t)=-\epsilon\delta t\tilde{A}\pi(t), \tag{4.10}$$

where the rhs is a symbolic notation for the change in $\pi(t)$ due to a canonical transformation generated by $A(p,q,t)$ [see Eqs. (2.8)].

If G is a constant of motion, A from its definition is zero. It then follows from (4.10) that the route followed in evaluating $\pi'(t+\delta t)$ is immaterial and so $\pi'(q,t)$ is also a solution of the GHJ equations. Thus a constant of motion can be used to map the manifold of solutions onto itself. If the surface described by the vanishing of $\varphi_\sigma(p,q,t)$ is an eigensurface of G then the new solution is Clebsch gauge related to the old one. Note in this context that since G is a constant of motion its eigensurfaces remain eigensurfaces under time evolution. Thus the two solutions are "the same" or distinct for all times.

Conversely let $G(p,q,t)$ be a phase space function which generates a transformation which takes every solution of Eqs. (3.1) onto another such solution. Then the difference $\pi_c'(t+\delta t) - \pi_a'(t+\delta t)$ in (4.10) must be a pure gauge. This means that the surface described by the vanishing of $\varphi_\sigma(p,q,t)$ must be an eigensurface of A. Thus every n-dimensional surface parametrizeable by q must be invariant under the transformation generated by A. Since any point in phase space can be thought of as the intersection of two such surfaces, the transformation generated by A must take every point into itself. A must therefore be a pure function of time which can in fact be set to zero after a harmless redefinition of G. Thus, G is a constant of motion.

V. CONCLUSION

A distinctive feature of the Hamilton–Jacobi form of dynamics is its grouping together of separate phase space trajectories into families.[7] Thus one solution of the H–J equation corresponds to an entire family of solutions of Hamilton's equations constructed in a special way. It is enlightening to view the generalizations of Rund and Baumeister in this context. The generalizations result in greater freedom in the manner in which families are to be constructed. Thus, solutions of the generalized Hamilton–Jacobi equations correspond to a larger set of families of phase space trajectories than do solutions of the Hamilton–Jacobi equation.

The form of the theory as put forward here is essentially the same as Baumeister's. However the considerations that lead to the generalized H–J equations in Ref. 2 are different, being analytic and mainly concerned with configuration space. The Eqs. (3.1) are identical in content to the corresponding ones of Ref. 2 but they differ in viewpoint and in emphasis.

This discussion of generalized H–J theory parallels closely Mukunda's discussion of H–J theory. Geometrical ideas like the association of phase space surfaces with solutions of the H–J equations have come over whole into the generalized theory. However the surfaces under consideration in Ref. 3 are a more restricted class since it was assumed there that they have zero symplectic rank. This assumption has been dropped in the passage from the usual to the generalized theory. As had better be true, the considerations in this article agree with those of Ref. 3 for the case of surfaces with vanishing symplectic rank. It is remarked in Ref. 3 that the discussion has been implicitly restricted to local properties of phase space. This remark would seem to apply here too. While understanding such limitations of the method, we feel that the phase space approach to generalized H–J theory is a useful and enlightening one.

ACKNOWLEDGMENTS

I am indebted to Professor N. Mukunda for suggesting this line of work and for guidance, criticism, and help through its progress. I have profited from discussions with Professor A.P. Balachandran (who suggested the alternative description in Sec. I) and Dr. R. Nityananda. Thanks also to the N.C.E.R.T., New Delhi for financial aid.

[1]H. Rund, Arch. Rat. Mech. Anal. **65**, 305 (1977).
[2]R. Baumeister, "Generalized Hamilton–Jacobi theories," Arizona State University preprint, 1978.
[3]N. Mukunda, Proc. Indian Acad. Sci. A **87**, 85 (1978).
[4]E.C.G. Sudarshan and N. Mukunda, *Classical Dynamics—A Modern Perspective* (Wiley, New York, 1974), Chap. 14.
[5]P.A.M. Dirac, Can. J. Math. **2**, 129 (1950).
[6]C. Caratheodory, "Calculus of variations and partial differential equations of the first order," Part I (Holden–Day, San Francisco, 1967).
[7]P.A.M. Dirac, Can. J. Math. **3**, 1 (1951).

On electromagnetic multipole fields in a finite, spherically symmetric region

Arno D. Steiger
Lawrence Livermore Laboratory, University of California, Livermore, California 94550

(Received 9 October 1978; accepted for publication 12 June 1979)

The electromagnetic eigenfields for the region bounded by two concentric spheres are discussed and compared with the corresponding eigenfields of a spherical cavity. These characteristic fields are the solenoidal and irrotational multipole solutions of the vector Helmholtz equation that satisfy the source-free boundary conditions. They constitute a complete set for the expansion of an arbitrary, square-integrable electromagnetic field, which may be generated by surface and volume sources. The frequencies of the solenoidal and irrotational eigenfields for the annular region are analyzed as functions of the radius ratio, $\alpha = r_1/r_2$ ($r_1 < r_2$ = constant), of the two concentric spheres. The results are illustrated by graphs and tables. Two relations obtained by applying the implicit function theorem to the transcendental eigenfrequency equations are also derived by calculating the work performed against the radiation pressure as the electromagnetic field is compressed adiabatically. The multipole fields are expressed in terms of vector spherical harmonics and vector spherical multipoles. Two formulas for the reduction of vector products of multipole fields to sums of vector spherical harmonics are derived.

INTRODUCTION AND SUMMARY

In a finite space region with well-defined, but not necessarily perfectly conducting boundary surfaces, a square-integrable electromagnetic field can be expanded in terms of the (solenoidal and irrotational) eigenvector fields of the bounded region.[1-5] These eigenvector fields, referred to as cavity modes, are the solutions of the vector Helmholtz equation that satisfy certain boundary conditions. In this communication we consider two types of finite regions, a cavity bounded by a single sphere and a cavity bounded by two concentric spheres, with emphasis on the latter type of domain. As the eigenvector fields of these (spherically symmetric) domains we take the solenoidal and irrotational *multipole* solutions[6,7] of the boundary value problems consisting of the vector Helmholtz equation and the requirements that the tangential components of an electric multipole vector and the normal component of a magnetic multipole vector vanish at the boundary. These boundary conditions are satisfied by an electromagnetic field on a perfectly conducting surface. Although the multipole eigenvectors satisfy these ideal boundary conditions, they constitute a complete basis for electromagnetic field expansions even if the actual boundary surfaces are not perfectly conducting.[1-5] This property of the eigenvectors is essential, since in a bounded domain an electromagnetic field may be generated not only by charge and current distributions within the domain, but also by surface sources on the boundary (e.g., waveguide openings); furthermore, an electromagnetic field may be attenuated through imperfectly conducting boundary surfaces. The solenoidal modes are also called normal modes because they represent the characteristic electromagnetic oscillations of an isolated (no surface or volume sources) cavity.

[a]Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore Laboratory under contract number W-7405-ENG-48.

Three physical systems which may be idealized as electromagnetic field domains bounded by two concentric spheres are (1) a spherical microwave cavity which is partially filled with a central plasma,[8-10] (2) the cavity bounded by the surface of the earth and by the ionosphere (in this cavity the so-called Schuhmann resonances[11] may be excited), (3) pulsar cavity radiation in the region between the surface of a rotating magnetic neutron star and the radiation-plasma interface.[12]

Normal mode frequencies for the cavity domain of the first system were calculated and measured by Boyen *et al.*[13] In the part of their work dealing with a cold uniform central plasma, these authors assume that the cavity shell is a perfect conductor and that the tangential components of the electric and magnetic field vectors are continuous across the plasma boundary surface. Their results are displayed as graphs. The same boundary conditions were used in Ref. 14 in order to compute tables of normal mode frequencies for the region bounded by a spherical cavity shell and the surface of a central, nontransparent plasma core, characterized by a negative dielectric constant.

In Sec. 1 the solenoidal and irrotational multipole solutions of the vector Helmholtz equation are derived from the multipole solutions of the scalar Helmholtz equation. The vector multipole solutions are expressed in terms of vector spherical harmonics and vector spherical multipoles.[15] Section 2 contains the construction of the solenoidal and irrotational characteristic multipole fields (modes) for the cavity Ω_{r_0}, bounded by the sphere $r = r_0$, and for the cavity $\Omega(r_1, r_2)$, bounded by the concentric spheres $r = r_1$ and $r = r_2 (r_1 < r_2)$. As shown in Sec. 1, there are no solenoidal fields of multipole order zero. Irrotational (oscillatory) modes of multipole order zero can exist in Ω_{r_0} as an electric or as a magnetic field, in $\Omega(r_1, r_2)$, however, only as a magnetic field. On the other hand, an electrostatic field can exist only in $\Omega(r_1, r_2)$, since only for this cavity is the boundary not singly connected. A magnetostatic field can exist neither in Ω_{r_0} nor in $\Omega(r_1, r_2)$, since both domains are simply connected. For either cavity, the frequency spectrum of the transverse electric solenoidal modes is identical with the frequency spectrum of the irrotational electric modes whose multipole orders are not zero. Furthermore, for either cavity, the frequencies of the irrotational magnetic mode of multipole order zero are equal to the frequencies of the irrotational electric mode (and therefore also of the transverse electric solenoidal mode) of multipole order one. The frequencies of the solenoidal and irrotational modes of $\Omega(r_1,r_2)$ are functions of the ratio parameter $\alpha = r_1/r_2$. A discussion of these functional dependencies for r_2 = constant and $0<\alpha<1$ is the subject of Sec. 3. The following features emerge: (1) All the eigenfrequencies of the transverse electric (solenoidal) and electric irrotational modes increase monotonically with α. (2) For each multipole order l ($l = 1,2,\cdots$), the smallest frequency of a transverse magnetic (solenoidal) mode and the smallest frequency of a magnetic irrotational mode decrease monotonically as α increases in the open interval (0,1), both circular eigenfrequencies tending towards the minimum value $(c/r_2)\,[l(l+1)]^{1/2}$ as α approaches unity. (3) All the other frequencies of the transverse magnetic (solenoidal) modes, and of the magnetic irrotational modes, each have a minimum for some value of α in the open interval (0,1). (4) For particular values of α certain frequencies that belong to different multipole eigenvectors coincide, thus giving rise to mode-crossing degeneracies. Transitions between degenerate modes, referred to as mode instabilities may be induced by surface currents. Numerical values of roots of the transcendental equations which determine the eigenfrequencies of the domain $\Omega(r_1,r_2)$, the minima of some of these roots, and values of certain degenerate roots are tabulated in Refs. 16–18. Section 3 also contains graphs displaying mode-crossing degeneracies as intersections of curves that represent these roots as functions of α. When multiplied by c/r_2, these roots are equal to the circular eigenfrequencies. Some of the analytic results of Sec. 3, whose derivation is based on the implicit function theorem, are derived in Sec. 4 by calculating the work performed against the radiation pressure if the electromagnetic field in $\Omega(r_1,r_2)$ is compressed adiabatically. The force exerted on the boundary by the field of a normal mode is positive (in the direction of the outward normal) only when the corresponding frequency increases with increasing α. The Poynting vector, the momentum density, and the angular momentum density of an electromagnetic radiation field can be expressed as sums of vector spherical harmonics by means of two formulas that are derived in the Appendix.

Electromagnetic field expansions in terms of the solenoidal and irrotational multipole eigenvectors, constructed in Sec. 2, are not uniformly convergent on a boundary which carries surface sources. Term-by-term differentiation of the field expansions are then not permissible.[3,4]

For the expansion of an electromagnetic field generated by boundary surface sources alone, it may not be necessary to invoke irrotational modes, provided that the normal modes satisfy the boundary conditions that conform to these surface sources (rather than the boundary conditions for a perfectly conducting surface). This was demonstrated by Schelkunoff[19] for a rectangular cavity.

1. THE MULTIPOLE SOLUTIONS OF THE VECTOR HELMHOLTZ EQUATION

The homogeneous wave equation for a vector field $\mathscr{F}(t,\mathbf{x})$ with harmonic time dependence,

$$\mathscr{F}(t,\mathbf{x}) = re[\mathbf{F}(\mathbf{x})e^{-i\omega t}] = \tfrac{1}{2}[\mathbf{F}(\mathbf{x})e^{-i\omega t} + \mathbf{F}^*(\mathbf{x})e^{i\omega t}]\,, \tag{1.1}$$

reduces to the vector Helmholtz equation

$$(\nabla^2 + k^2)\mathbf{F}(\mathbf{x}) = 0\,, \tag{1.2}$$

where the wave number k is assumed to be constant. In spherical coordinates, $\mathbf{x} = (r,\theta,\phi)$, three independent solutions of (1.2) are[7]

$$\mathbf{F}^{(0)}(r,\theta,\phi) = (1/k)\nabla\psi(r,\theta,\phi)\,, \tag{1.3a}$$

$$\mathbf{F}^{(1)}(r,\theta,\phi) = \mathbf{r}\times\nabla\psi(r,\theta,\phi) = -\nabla\times[\mathbf{r}\psi(r,\theta,\phi)]\,, \tag{1.3b}$$

$$\mathbf{F}^{(2)}(r,\theta,\phi) = -(1/k)\nabla\times\nabla\times[\mathbf{r}\psi(r,\theta,\phi)]\,, \tag{1.3c}$$

where ψ is a solution of the scalar Helmholtz equation

$$(\nabla^2 + k^2)\psi(r,\theta,\phi) = 0\,. \tag{1.4}$$

The vector field (1.3a) is irrotational,

$$\nabla\times\mathbf{F}^{(0)} = 0\,, \tag{1.5}$$

while the vector fields (1.3b) and 1.3c) are solenoidal,

$$\nabla\cdot\mathbf{F}^{(1)} = 0,\quad \nabla\cdot\mathbf{F}^{(2)} = 0\,. \tag{1.6}$$

The vectors (1.3a) and (1.3b) are perpendicular,

$$\mathbf{F}^{(0)}\cdot\mathbf{F}^{(1)} = 0\,. \tag{1.7}$$

From (1.4) it is immediated that (1.3a) is a solution of (1.2), and that

$$\nabla\cdot\mathbf{F}^{(0)} = -k\psi\,. \tag{1.8}$$

The vector fields (1.3b) and (1.3c) are related by

$$\nabla\times\mathbf{F}^{(1)} = k\,\mathbf{F}^{(2)}\,. \tag{1.9a}$$

$$\nabla\times\mathbf{F}^{(2)} = k\,\mathbf{F}^{(1)}\,. \tag{1.9b}$$

From (1.9) follows

$$-\nabla\times\nabla\times\mathbf{F}^{(\sigma)} + k^2\mathbf{F}^{(\sigma)} = 0,\quad \sigma = 1,2\,. \tag{1.10}$$

By virtue of (1.6), the two equations (1.10) imply that (1.3b) and (1.3c) are solutions of (1.2). The substitutions

$$\left.\begin{aligned} \mathbf{F}^{(\sigma)} &= (\sqrt{\epsilon})\mathbf{E} \quad (1.11a)\\ \mathbf{F}^{(\sigma')} &= i(\sqrt{\mu})\mathbf{H} \quad (1.11b)\end{aligned}\right\}\,,\quad \sigma,\sigma' = 1,2,\quad \sigma'\neq\sigma$$

$$k = (\sqrt{\epsilon\mu})(\omega/c) \tag{1.11c}$$

transform the reduced Maxwell equations (1.9) into the ordinary Maxwell equations for the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ in a source-free domain characterized by the dielectric and magnetic permeabilities ϵ and μ. The *multipole* solution to (1.4) of order (l,m),

$$l = 0,1,2,\cdots,\quad m = -l,-l+1,\ldots,l\,. \tag{1.12}$$

which is regular on the unit sphere,

$$\mathscr{S}^2 = \{0\leqslant\theta\leqslant\pi, 0\leqslant\phi<2\pi\}\,,$$

can be expressed as

$$\psi_{lm}(r,\theta,\phi) = \zeta_l(kr)Y_{lm}(\theta,\phi)\,, \tag{1.13}$$

where

$$\zeta_l(kr) = C_l j_l(kr) + D_l n_l(kr)\,, \tag{1.14}$$

with j_l and n_l denoting spherical Bessel functions of the first and second kind. The constants C_l and D_l are determined by normalization and boundary conditions; $D_l = 0$, if the radial domain contains the point $r = 0$, where the function n_l is singular. The spherical harmonics, $Y_{lm}(\theta,\phi)$, are normalized such that[11]

$$\int d(\theta,\phi) Y^*_{l'm'}(\theta,\phi) Y_{lm}(\theta,\phi) = \delta_{l'l}\delta_{m'm} , \tag{1.15}$$

where

$$\int d(\theta,\phi)\cdots = \int_0^\pi \sin\theta\, d\theta \int_0^{2\pi} d\phi \cdots . \tag{1.16}$$

In accordance with the definitions (1.3), the scalar multipole solutions (1.13) generate the vector multipole fields

$$\mathbf{F}^{(0)}_{lm}(r,\theta,\phi) = (1/k)\nabla\psi_{lm}(r,\theta,\phi) , \tag{1.17a}$$

$$\mathbf{F}^{(1)}_{lm}(r,\theta,\phi) = \mathbf{r}\times\nabla\psi_{lm}(r,\theta,\phi) , \tag{1.17b}$$

$$\mathbf{F}^{(2)}_{lm}(r,\theta,\phi) = (1/k)\nabla\times[\mathbf{r}\times\nabla\psi_{lm}(r,\theta,\phi)] , \tag{1.17c}$$

which are solutions of (1.2). As the functions (1.13) constitute a complete set for the expansion of a solution to (1.4), the three types of vector multipole fields (1.17) are three independent complete sets for the expansions of the corresponding three linearly independent vector fields (1.3). With the normalization (1.15) of the spherical harmonics, it follows from (1.13), (1.14), and (1.17a) that (with $\mathbf{e}_r = \mathbf{r}/r$)

$$\begin{aligned}\mathbf{F}^{(0)}_{00} &= \frac{1}{\sqrt{4\pi}}\left(\frac{d}{d(kr)}\zeta_0(kr)\right)\mathbf{e}_r \\ &= -\frac{1}{\sqrt{4\pi}}[C_0 j_1(kr) + D_0 n_1(kr)]\mathbf{e}_r .\end{aligned} \tag{1.18a}$$

Since the operator $\mathbf{r}\times\nabla$ does not act on the radial coordinate, (1.17b) and (1.17c) imply that

$$\mathbf{F}^{(1)}_{00} = 0 \tag{1.18b}$$

and

$$\mathbf{F}^{(2)}_{00} = 0 . \tag{1.18c}$$

The expressions (1.17) involve complex vector fields on the unit sphere $\mathscr{S}^2$. In the Hilbert space, $\mathscr{H}(\mathscr{S}^2)$, of these vector fields, the inner product is defined by

$$\langle \boldsymbol{\Phi},\boldsymbol{\psi}\rangle = \int d(\theta,\phi)\boldsymbol{\Phi}^*(\theta,\phi)\cdot\boldsymbol{\psi}(\theta,\phi) . \tag{1.19}$$

A set of orthonormal basis vectors in $\mathscr{H}(\mathscr{S}^2)$ which transform according to the irreducible unitary representations of the three-dimensional rotation group are the vector spherical harmonics

$$\mathscr{Y}^m_{lj}(\theta,\phi) = \sum_{\mu,\tau} \langle j\mu 1\tau|lm\rangle Y_{j\mu}(\theta,\phi)\boldsymbol{\epsilon}_\tau . \tag{1.20}$$

The spherical basis vectors, $\boldsymbol{\epsilon}_\tau$, are related to the Cartesian unit vectors $\mathbf{e}_x,\mathbf{e}_y,\mathbf{e}_z$ by the equations

$$\boldsymbol{\epsilon}_\tau = \begin{cases} \mp(1/\sqrt{2})(\mathbf{e}_x \pm i\mathbf{e}_y), & \text{if } \tau = \pm 1 , \\ \mathbf{e}_z , & \text{if } \tau = 0 . \end{cases} \tag{1.21}$$

Therefore,

$$\boldsymbol{\epsilon}^*_\tau = (-1)^\tau \boldsymbol{\epsilon}_{-\tau} , \tag{1.22}$$

and

$$\boldsymbol{\epsilon}^*_{\tau'}\cdot\boldsymbol{\epsilon}_\tau = \delta_{\tau'\tau} . \tag{1.23}$$

From (1.15), (1.23), and the orthogonality of the Clebsch–Gordan (CG) transformation it follows that

$$\langle \mathscr{Y}^{m'}_{l'j'},\mathscr{Y}^m_{lj}\rangle = \delta_{l'l}\delta_{j'j}\delta_{m'm} . \tag{1.24}$$

On account of (1.22),

$$Y^*_{j\mu}(\theta,\phi) = (-1)^\mu Y_{j,-\mu}(\theta,\phi)$$

and

$$\langle j\mu 1\tau|lm\rangle = (-1)^{-j-1+l}\langle j,-\mu 1,-\tau|l,-m\rangle ,$$

the following relation holds for the complex conjugate vector spherical harmonics:

$$\mathscr{Y}^{*m}_{lj}(\theta,\phi) = (-1)^{l+m-j-1}\mathscr{Y}^{-m}_{lj}(\theta,\phi) . \tag{1.25}$$

The scalar and the vector components of expression (1.20) normal to the unit sphere are given by

$$(\mathbf{e}_r\cdot\mathscr{Y}^m_{lj}) = -\langle l010|j0\rangle Y_{lm} . \tag{1.26}$$

The relation (1.26) can be derived by expressing the unit vector $\mathbf{e}_r = \mathbf{r}/r$ in terms of the spherical basis vectors (1.21),

$$\mathbf{e}_r = \left(\frac{4\pi}{3}\right)^{1/2}\sum_\rho Y_{1\rho}(\theta,\phi)\boldsymbol{\epsilon}_\rho , \tag{1.27}$$

and by reducing products of the spherical harmonics $Y_{1\rho}$ and $Y_{j\mu}$ by means of the CG transformation. From (1.26) it is immediate that $\mathscr{Y}^m_{ll}$ is tangent to the unit sphere,

$$\mathbf{e}_r\cdot\mathscr{Y}^m_{ll} = 0 , \tag{1.28}$$

whereas $\mathscr{Y}^m_{l,l+1}$ and $\mathscr{Y}^m_{l,l-1}$ have radial as well as tangential components. The representations in terms of vector spherical harmonics of the multipole fields (1.17) are [Eqs. (B11), (A14), (B5), and (B.14) of Ref. 8]

$$\begin{aligned}&\mathbf{F}^{(0)}_{lm}(r,\theta,\phi) \\ &= [1/(2l+1)^{1/2}][(l+1)^{1/2}\zeta_{l+1}(kr)\mathscr{Y}^m_{l,l+1}(\theta,\phi) \\ &\quad + (\sqrt{l})\zeta_{l-1}(kr)\mathscr{Y}^m_{l,l-1}(\theta,\phi)] ,\end{aligned} \tag{1.29a}$$

$$\mathbf{F}^{(1)}_{lm}(r,\theta,\phi) = i[l(l+1)]^{1/2}\zeta_l(kr)\mathscr{Y}^m_{l,l}(\theta,\phi) , \tag{1.29b}$$

and

$$\begin{aligned}&\mathbf{F}^{(2)}_{lm}(r,\theta,\phi) \\ &= [1/(2l+1)^{1/2}][l(l+1)^{1/2}\zeta_{l+1}(kr)\mathscr{Y}^m_{l,l+1}(\theta,\phi) \\ &\quad - (l+1)(\sqrt{l})\zeta_{l-1}(kr)\mathscr{Y}^m_{l,l-1}(\theta,\phi)] ,\end{aligned} \tag{1.29c}$$

where

$$\zeta_{l\pm 1} \equiv C_l j_{l\pm 1} + D_l n_{l\pm 1} . \tag{1.29d}$$

The Poynting vector of the radiation field (1.11) can be decomposed into multipole vector fields, i.e., into components transforming according to unitary irreducible representations of the three-dimensional rotation group, by means of formula (A4). This formula reduces the vector product of two vector spherical harmonics to a sum of vector spherical harmonics. A multipole decomposition of the angular momentum density of the radiation field (1.11) can be obtained by means of the triple vector product reduction (A7). In the Hilbert space $\mathscr{H}(\mathscr{S}^2)$ with the inner product (1.19) the vector spherical harmonics (1.20) are orthonormal, as is implied by (1.24). In the three-dimensional (complex) vector space V^3, spanned by the spherical unit vectors (1.21), however, the three linearly independent vector spherical harmonics $\mathscr{Y}^m_{l,l}, \mathscr{Y}^m_{l,l+1}, \mathscr{Y}^m_{l,l-1}$ do not constitute an orthogonal set, since [Eq. (A2) of Ref. 15 implies that]

$$\mathscr{Y}^m_{ll}\cdot\mathscr{Y}^m_{l,l\pm1}=0,\quad \mathscr{Y}^m_{l,l+1}\cdot\mathscr{Y}^m_{l,l-1}\neq 0\,. \tag{1.30}$$

On the other hand, the vector spherical multipoles

$$\mathbf{X}^{(L)}_{lm}(\theta,\phi)=\mathbf{e}_r Y_{lm}(\theta,\phi)\,, \tag{1.31a}$$

$$\mathbf{X}^{(M)}_{lm}(\theta,\phi)=[l(l+1)]^{-1/2}(1/i)\mathbf{r}\times\nabla Y_{lm}(\theta,\phi)\,, \tag{1.31b}$$

$$\begin{aligned}\mathbf{X}^{(E)}_{lm}(\theta,\phi)&=-i\mathbf{e}_r\times\mathbf{X}^{(M)}_{lm}(\theta,\phi)\\&=[l(l+1)]^{-1/2}r\nabla Y_{lm}(\theta,\phi)\,,\end{aligned} \tag{1.31c}$$

which are referred to as longitudinal, transverse magnetic, and transverse electric, obviously constitute an orthogonal vector set in V^3. The connections between the vector spherical multipoles and the vector spherical harmonics are[15]

$$\begin{aligned}\mathbf{X}^{(L)}_{lm}(\theta,\phi)=(2l+1)^{-1/2}[(l)^{1/2}\mathscr{Y}^m_{l,l-1}(\theta,\phi)\\-(l+1)^{1/2}\mathscr{Y}^m_{l,l+1}(\theta,\phi)\,,]\end{aligned} \tag{1.32a}$$

$$\mathbf{X}^{(M)}_{lm}(\theta,\phi)=\mathscr{Y}^m_{ll}(\theta,\phi)\,, \tag{1.32b}$$

$$\begin{aligned}\mathbf{X}^{(E)}_{lm}(\theta,\phi)=(2l+1)^{-1/2}[(l+1)^{1/2}\mathscr{Y}^m_{l,l-1}(\theta,\phi)\\+(l)^{1/2}\mathscr{Y}^m_{l,l+1}(\theta,\phi)]\,.\end{aligned} \tag{1.32c}$$

The relation (1.32b) is an immediate consequence of (1.17b), (1.29b), and (1.31b). The relations (1.32a) and (1.32c) can be derived from the definitions (1.31a) and (1.31c) by using Eqs. (B10) and (B11) of Ref. 8. From (1.24) and (1.32) it is evident that the vector spherical multipoles are orthonormal with respect to the inner product, Eq. (1.19), in $\mathscr{H}(\mathscr{S}^2)$, namely

$$\langle\mathbf{X}^{(\sigma')}_{l'm'},\mathbf{X}^{(\sigma)}_{lm}\rangle=\delta_{l'm}\delta_{m'm}\delta_{\sigma',\sigma},\quad \sigma',\sigma=L,M,E\,. \tag{1.33}$$

The corresponding scalar product in V^3 is given by Eq. (26) of Ref. 15. By replacing in (1.29) the vector spherical harmonics by the vector spherical multipoles, in accordance with the relations (1.32), one obtains a decomposition of the multipole solutions (1.17) into radial and tangential components (namely, into parts normal and tangent to a sphere),

$$\begin{aligned}(\mathbf{F}^{(0)}_{lm})_\parallel&=(2l+1)^{-1}[l\zeta_{l-1}(kr)\\&\quad-(l+1)\zeta_{l+1}(kr)]\mathbf{X}^{(L)}_{lm}(\theta,\phi)\,,\\(\mathbf{F}^{(0)}_{lm})_\perp&=[l(l+1)]^{1/2}(kr)^{-1}\zeta_l(kr)\mathbf{X}^{(E)}_{lm}(\theta,\phi)\,,\end{aligned} \tag{1.34a}$$

$$\begin{aligned}(\mathbf{F}^{(1)}_{lm})_\parallel&=0\,,\\(\mathbf{F}^{(1)}_{lm})_\perp&=\mathrm{i}[l(l+1)]^{1/2}\zeta_l(kr)\mathbf{X}^{(M)}_{lm}(\theta,\phi)\,,\end{aligned} \tag{1.34b}$$

$$\begin{aligned}(\mathbf{F}^{(2)}_{lm})_\parallel&=-l(l+1)(kr)^{-1}\zeta_l(kr)\mathbf{X}^{(L)}_{lm}(\theta,\phi)\,,\\(\mathbf{F}^{(2)}_{lm})_\perp&=(2l+1)^{-1}[l(l+1)]^{1/2}[l\zeta_{l+1}(kr)\\&\quad-(l+1)\zeta_{l-1}(kr)]\mathbf{X}^{(E)}_{lm}(\theta,\phi)\,.\end{aligned} \tag{1.34c}$$

These relations facilitate the evaluation of the Maxwell stresses exerted by the fields (1.17) on a spherical surface. From (1.34a) and (1.34b) it is obvious that the orthogonality (1.7) is ensured by the orthogonality of the vector spherical multipoles.

2. THE MULTIPOLE FIELDS IN A BOUNDED DOMAIN

Each of the solutions (1.3) of the vector Helmholtz equation, (1.2), are now assumed to exist in a bounded domain, Ω, and to satisfy one of the two homogeneous boundary conditions[1-5]

$$\mathbf{n}\times\mathbf{F}^{(i)}=0\quad(i=0,1,2)\ \text{on}\ \Sigma \tag{2.1}$$

or

$$\mathbf{n}\cdot\mathbf{F}^{(i)}=0\quad(i=0,1,2)\ \text{on}\ \Sigma\,, \tag{2.2}$$

where the boundary Σ consists of one or several closed surfaces, and where $\mathbf{n}$ denotes the unit normal to Σ. Together, (2.1) and (2.2) imply that there are no surface sources on Σ. On a perfectly conducting boundary (ideal cavity shell), the tangential components of the electric field, $\mathbf{F}^{(i)}_{(e)}$, vanishes in accordance with (2.1), while the normal component of the magnetic field, $\mathbf{F}^{(i)}_{(h)}$, vanishes in accordance with (2.2). The irrotational field (1.3a) satisfies (2.1) if its generating function ψ, which is a solution of (1.4), satisfies the boundary condition

$$\psi=\text{const}\quad\text{on}\ \Sigma\,. \tag{2.3}$$

If the wave number k in (1.4) is different from zero, the constant in (2.3) may be chosen as zero:

$$\psi=0\quad\text{on}\ \Sigma,\quad\text{if}\ k\neq0\,. \tag{2.4}$$

On account of (1.8), the boundary condition (2.4) implies that (1.3a) satisfies

$$\nabla\cdot\mathbf{F}^{(0)}=0\quad\text{on}\ \Sigma\,, \tag{2.5}$$

in addition to (2.1). The boundary condition (2.2) for (1.3a) is equivalent to

$$\mathbf{n}\cdot\nabla\psi=0\quad\text{on}\ \Sigma\,. \tag{2.6}$$

The relations (1.9) imply that the solenoidal fields (1.3b) and (1.3c) cannot independently satisfy the boundary conditions (2.1) and (2.2), i.e., if (1.3b) [(1.3c)] is subject to (2.1), then (1.3c) [(1.3b)] satisfies (2.2). From the spherical coordinate representation of the curl operator appearing in (1.3b) and in (1.3c) it follows that on a spherical boundary surface of radius $r_0,\mathscr{S}^2_{r_0}$, the conditions (2.1) and (2.2) are satisfied by (1.3b) and (1.3c), respectively, when

$$\psi=0\quad\text{on}\ \mathscr{S}^2_{r_0}\,, \tag{2.7}$$

and by (1.3c) and (1.3b), respectively, when

$$\frac{\partial(r\psi)}{\partial r}=0\quad\text{on}\ \mathscr{S}^2_{r_0}\,. \tag{2.8}$$

On $\mathscr{S}^2_{r_0}$, the condition (2.4) is obviously identical with (2.7), and (2.6) becomes

$$\frac{\partial\psi}{\partial r}=0\quad\text{on}\ \mathscr{S}^2_{r_0}\,. \tag{2.9}$$

If the generating function ψ is identified with the multipole solution (1.13) of the scalar Helmholtz equation, (1.4), the boundary conditions (2.7), (2.8), and (2.9) reduce to

$$\zeta_l(kr_0)=0\,, \tag{2.10}$$

$$[(d/dr)(r\zeta_l(kr))]_{r=r_0}=0\,, \tag{2.11}$$

and

$$[(d/dr)\zeta_l(kr)]_{r=r_0}=0\,, \tag{2.12}$$

respectively, where ζ_l is defined by (1.14). For the multipole fields (1.17), the boundary conditions (2.1) and (2.2) are implied by (2.10), (2.11), and (2.12). In particular, the boundary conditions

$$\mathbf{n}\times\mathbf{F}^{(0)}_{lm}=0\quad\text{on}\ \mathscr{S}^2_{r_0} \tag{2.13}$$

and

$$\left.\begin{array}{l}\mathbf{n}\times\mathbf{F}^{(1)}_{lm}=0\\ \mathbf{n}\cdot\mathbf{F}^{(2)}_{lm}=0\end{array}\right\}\quad\text{on }\mathscr{S}^2_{r_0}\tag{2.14}$$

are consequences of (2.10), while it follows from (2.11) that

$$\left.\begin{array}{l}\mathbf{n}\times\mathbf{F}^{(2)}_{lm}=0\\ \mathbf{n}\cdot\mathbf{F}^{(1)}_{lm}=0\end{array}\right\}\quad\text{on }\mathscr{S}^2_{r_0}\tag{2.15}$$

and from (2.12) that

$$\mathbf{n}\cdot\mathbf{F}^{(0)}_{lm}=0\quad\text{on }\mathscr{S}^2_{r_0}\tag{2.16}$$

If the field domain $\Omega=\Omega_{r_0}$ is the cavity bounded by $\mathscr{S}^2_{r_0}$, the boundary conditions (2.10), (2.11), and (2.12) hold for

$$\zeta_l(kr)=C_l j_l(kr),\tag{2.17}$$

in accordance with the comment following the definition (1.14). These three boundary conditions are then the eigenvalue equations whose roots determine three denumerably infinite sequences of characteristic wavenumbers [and by (1.11c) of characteristic frequencies] of the cavity domain Ω_{r_0}. Since

$$j_l(ze^{m\pi i})=e^{ml\pi i}j_l(z),\quad (l,m=0,1,2,\cdots)\tag{2.18}$$

all three eigenvalue equations are invariant under the transformation $k\to-k$, which means that only positive roots ($k>0$) need be considered. With (2.17), one obtains from (2.10)–(2.12)

$$j_l(\lambda)=0:\tag{2.19a}$$

$$\lambda_{ln}=r_0k_{ln},\quad \lambda_{l,n+1}>\lambda_{l,n},\quad n=1,2,\cdots,\tag{2.19b}$$

$$lj_{l+1}(\eta)=(l+1)j_{l-1}(\eta):\tag{2.20a}$$

$$\eta_{ln}=r_0k'_{ln},\quad \eta_{l,n+1}>n_{l,n},\quad n=1,2,\cdots,\tag{2.20b}$$

and

$$(l+1)j_{l+1}(\mu)=lj_{l-1}(\mu):\tag{2.21a}$$

$$\mu_{ln}=r_0k''_{ln},\quad \mu_{l,n+1}>\mu_{l,n},\quad n=1,2\cdots.\tag{2.21b}$$

If the solenoidal multipole fields (1.17b) and (1.17c) are subject to the boundary conditions (2.14) [(2.15)], they constitute the transverse electric (TE) [transverse magnetic (TM)] normal modes of the ideal cavity resonator bounded by the (perfectly conducting) spherical shell of radius r_0, provided that the radial dependence of the generating multipole function (1.13) is given by (2.17). In accordance with (1.11), the electric and the magnetic vector are

$$\mathbf{E}^{\mathrm{TE}}_{nlm}=(\sqrt{\epsilon})\mathbf{F}^{(1)}_{(e)nlm},\tag{2.22a}$$

$$\mathbf{H}^{\mathrm{TE}}_{nlm}=i(\sqrt{\mu})\mathbf{F}^{(2)}_{(e)nlm},\tag{2.22b}$$

for the TE_{nlm} mode, and

$$\mathbf{E}^{\mathrm{TM}}_{nlm}=(\sqrt{\epsilon})\mathbf{F}^{(2)}_{(h)nlm},\tag{2.23a}$$

$$\mathbf{H}^{\mathrm{TM}}_{nlm}=i(\sqrt{\mu})\mathbf{F}^{(1)}_{(h)nlm},\tag{2.23b}$$

for the TM_{nlm} mode.

Here,

$$\mathbf{F}^{(1)}_{(g)nlm}=\mathbf{r}\times\nabla\psi_{(g)nlm}\quad (g=e,h),\tag{2.24a}$$

$$\mathbf{F}^{(2)}_{(g)nlm}=(k^{(g)}_{ln})^{-1}\nabla\times\mathbf{F}^{(1)}_{(g)nlm},\tag{2.24b}$$

and

$$\psi_{(g)nlm}=C^{(g)}_{ln}j_l(k^{(g)}_{ln}r)Y_{lm}(\theta,\phi),\tag{2.25}$$

where $k^{(e)}_{ln}=k_{ln}$ is defined by (2.19), and $k^{(h)}_{ln}=k'_{ln}$ by (2.20).

The normal modes (2.22) and (2.23) are the *free* electromagnetic oscillations of an isolated spherical cavity. These oscillations exist for the multipole orders specified by (1.12) with the restriction $l\neq0$, as is evident from (1.18b) and (1.18c). For the description of *forced* electromagnetic oscillations in the cavity it is necessary to add to the normal modes the irrotational eigensolutions of the boundary value problems consisting of (1.2) and (2.13) or of (1.2) and (2.16). The boundary condition (2.13) defines for the cavity bounded by the perfectly conducting shell $\mathscr{S}^2_{r_0}$ the irrotational electric eigenvectors

$$\mathbf{F}^{(0)}_{(e)nlm}=(1/k_{ln})\nabla\psi_{(e)nlm},\tag{2.26}$$

while (2.16) defines the irrotational magnetic eigenvectors

$$\mathbf{F}^{(0)}_{(h)nlm}=(1/k''_{ln})\nabla\psi^{(0)}_{(h)nlm},\tag{2.27}$$

with

$$\psi^{(0)}_{(h)nlm}=C^{(h,0)}_{ln}j_l(k''_{ln}r)Y_{lm}(\theta,\phi),\tag{2.28}$$

where k''_{ln} is determined by (2.21).

If the boundary Σ of the field domain Ω consists of two concentric spheres,

$$\Sigma(r_1,r_2)=\mathscr{S}^2_{r_1}+\mathscr{S}^2_{r_2}\quad (r_1<r_2),$$

$$\Omega(r_1,r_2)=\Omega_{r_2}-\Omega_{r_1},$$

the boundary conditions (2.10), (2.11), and (2.12) can be satisfied on $\Sigma(r_1,r_2)$ by the radial function (1.14) with $C_l\neq0$ and $D_l\neq0$,

$$\zeta_l(kr_1)=\zeta_l(kr_2)=0,\tag{2.29}$$

$$[(d/dr)(r\zeta_l(kr))]_{r=r_1,r_2}=0,\tag{2.30}$$

and

$$[(d/dr)\zeta_l(kr)]_{r=r_1,r_2}=0.\tag{2.31}$$

Substitution of (1.14) in (2.29)–(2.31) yields the following eigenvalue equations, whose roots λ, η, and μ can be considered as functions of the parameter:

$$\alpha\equiv r_1/r_2\quad (0<\alpha<1),\tag{2.32}$$

where

$$\lambda,\eta,\mu=kr_2:\tag{2.33}$$

$$j_l(\alpha\lambda)n_l(\lambda)=j_l(\lambda)n_l(\alpha\lambda),\tag{2.34}$$

$$s_l(\alpha\eta)t_l(\eta)=s_l(\eta)t_l(\alpha\eta),\tag{2.35a}$$

$$s_l(z)\equiv(d/dz)[zj_l(z)],\quad t_l(z)\equiv(d/dz)[zn_l(z)],\tag{2.35b}$$

and

$$v_l(\alpha\mu)w_l(\mu)=v_l(\mu)w_l(\alpha\mu),\tag{2.36a}$$

$$v_l(z)=(d/dz)j_l(z),\quad w_l(z)=(d/dz)n_l(z).\tag{2.36b}$$

The equations (2.34), (2.35a), and (2.36a) are invariant under the transformations $\lambda\to-\lambda$, $\eta\to-\eta$, and $\mu\to-\mu$, respectively; this follows from (2.18) and from

$$n_l(ze^{m\pi i})=(-1)^m e^{ml\pi i}n_l(z)\quad (l,m=0,1,2,\cdots).$$

Therefore, only positive roots need be considered in a discussion of the eigenvalue spectra defined by (2.34), (2.35), and (2.36),

$$\lambda_{ln}(\alpha) = r_2 k_{ln}(\alpha), \quad n = 1,2\cdots, \\ \lambda_{l,n+1}(\alpha) > \lambda_{l,n}(\alpha), \tag{2.37}$$

$$\eta_{ln}(\alpha) = r_2 k'_{ln}(\alpha), \quad n = 1,2,\cdots, \\ \eta_{l,n+1}(\alpha) > \eta_{ln}(\alpha), \tag{2.38}$$

and

$$\mu_{ln}(\alpha) = r_2 k''_{ln}(\alpha), \quad n = 1,2,\cdots, \\ \mu_{l,n+1}(\alpha) > \mu_{l,n}(\alpha). \tag{2.39}$$

The normal modes of a cavity bounded by two concentric spherical shells of infinite conductivity are described by the fields (2.22) and (2.23) if the generating multipole function in (2.24) is given by

$$\psi_{(g)nlm} = C_{ln}^{(g)}[j_l(k_{ln}^{(g)}r) + \gamma_{ln}^{(g)} n_l(k_{ln}^{(g)}r)]Y_{lm}(\theta,\phi) \quad (g = e,h), \tag{2.40}$$

where with (2.34) and (2.37)

$$k_{ln}^{(e)} = k_{ln}(\alpha) = \lambda_{ln}(\alpha)/r_2, \tag{2.41a}$$

$$\gamma_{ln}^{(e)} = -j_l(\lambda_{ln})/n_l(\lambda_{ln}) \tag{2.41b}$$

and where with (2.35) and (2.38)

$$k_{ln}^{(h)} = k'_{ln}(\alpha) = \eta_{ln}(\alpha)/r_2, \tag{2.42a}$$

$$\gamma_{ln}^{(h)} = -s_l(\eta_{ln})/t_l(\eta_{ln}). \tag{2.42b}$$

With the generating multipole function (2.40) the solenoidal eigenvectors (2.22) and (2.23) satisfy on $\Sigma(r_1,r_2)$ the boundary conditions (2.14) and (2.15), respectively. The irrotational electric eigenvector (2.26) satisfies on $\Sigma(r_1,r_2)$ the boundary condition (2.13) with the eigenvalue k_{ln} and the generating multipole function $\psi_{(e)nlm}$ defined by (2.40) and (2.41). The irrotational magnetic eigenvector (2.27) satisfies on $\Sigma(r_1,r_2)$ the boundary condition (2.16) if its generating multipole function, $\psi_{(h)nlm}^{(0)}$, is given by

$$\psi_{(h)nlm}^{(0)} = C_{ln}^{(h,0)}[j_l(k''_{ln}r) + \gamma_{ln}^{(h,0)} n_l(k''_{ln}r)]Y_{lm}(\theta,\phi), \tag{2.43}$$

where with (2.36) and (2.39)

$$k''_{ln}(\alpha) = \mu_{ln}(\alpha)/r_2, \tag{2.44a}$$

$$\gamma_{ln}^{(h,0)} = -v_l(\mu_{ln})/w_l(\mu_{ln}). \tag{2.44b}$$

Since the boundary $\Sigma(r_1,r_2)$ consists of two separate surfaces (spherical capacitor), a zero frequency ($k = 0$) electric field,

$$\mathbf{E}^{(0)} = -\nabla\Phi = -[d\Phi(r)/dr]\mathbf{e}_r \tag{2.45}$$

may exist which by (1.2) is a solution to the Laplace equation for the domain $\Omega(r_1,r_2)$. On $\Sigma(r_1,r_2)$ the irrotational field (2.45) satisfies the boundary condition (2.1) [but not the boundary condition (2.5)] if its generating function, Φ, is determined by the boundary value problem

$$\nabla^2\Phi = 0 \quad \text{in } \Omega(r_1,r_2), \\ \Phi = \Phi_1 = \text{const} \quad \text{on } \mathscr{S}_{r_1}^2, \\ \Phi = \Phi_2 = \text{const} \quad \text{on } \mathscr{S}_{r_2}^2 \\ \Phi_1 \neq \Phi_2, \tag{2.46}$$

as is evident from (1.4) and (2.3). The solution to (2.46) is

$$\Phi = [1/(r_2 - r_1)][(\Phi_1 - \Phi_2)(r_1 r_2/r) + r_2\Phi_2 - r_1\Phi_1]. \tag{2.47}$$

The stationary electric field (2.45) can be expressed as a zero-order multipole field by virtue of the representations

$$\mathbf{e}_r = -(\sqrt{4\pi})\mathscr{Y}_{01}^{0}(\theta,\phi) \tag{2.48a}$$

or

$$\mathbf{e}_r = (\sqrt{4\pi})\mathbf{X}_{00}^{(L)}(\theta,\phi). \tag{2.48b}$$

Equation (2.48a) follows from (1.20) and (1.27), while (2.48b) is an immediate consequence of (2.48a) and (1.32a).

In the domain Ω_{r_0} the zero-order multipole solution (1.18a) of the vector Helmholtz equation ($k \neq 0$) is realizable as the irrotational electric field given by (2.26), (2.19), and (2.25) for $l = 0$, and as the irrotational magnetic field given by (2.27), (2.21), and (2.28) for $l = 0$, since the eigenvalue equations (2.19a) and (2.21a) have both nontrivial solutions when $l = 0$. In the domain $\Omega(r_1,r_2)$, however, (1.18a) can be realized only as the irrotational magnetic field given by (2.27), (2.36), (2.39), and (2.43) for $l = 0$ since the eigenvalue equation (2.34), which also determines the characteristic frequencies of the irrotational electric fields (2.26), has *no* nontrivial solutions when $l = 0$.[20] The eigenvalue equation (2.21a) [(2.36)] for $l = 0$ is identical with the eigenvalue equation (2.19a) [(2.34)] for $l = 1$, so that $\mu_{0n} = \lambda_{1n}$ ($n = 1,2,\cdots$). While the realizable fields of multipole order zero are oscillatory electric and oscillatory magnetic in Ω_{r_0}, they are stationary electric [Eqs. (2.45) and (2.48)] and oscillatory magnetic in $\Omega(r_1,r_2)$. A stationary magnetic field can exist neither in Ω_{r_0}, nor in $\Omega(r_1,r_2)$, since both domains are simply connected.[1-5]

If the domain Ω_{r_0}, $[\Omega(r_1,r_2)]$ contains neither charge nor current distributions and if the boundary $\mathscr{S}_{r_0}^2$ $[\Sigma(r_1,r_2)]$ is free of surface sources [and if in (2.47) $\Phi_1 = \Phi_2$], the solenoidal eigenfields (2.22) and (2.23) with the generating functions (2.25) [(2.40)] are a complete set of basis vectors for the expansion of an arbitrary electromagnetic field in Ω_{r_0}, $[\Omega(r_1,r_2)]$. However, if surface or volume sources are present, the solenoidal eigenfields are in general no longer a complete set of basis vectors. It is then necessary to add to the solenoidal eigenvectors the irrotational electric and magnetic eigenfields (2.26) and (2.27) with the generating functions (2.25) for $g = e$ and (2.28), in order to obtain a complete set of basis vectors in the domain Ω_{r_0}. In the domain $\Omega(r_1,r_2)$ a complete set of basis vectors is obtained by supplementing the solenoidal eigenfields with the irrotational vectors (2.26) and (2.27), whose generating functions are given by (2.40) for $g = e$ and by (2.43), and with the stationary electric field (2.45) if $\Phi_1 \neq \Phi_2$. An electric field **E** and a magnetic field **H** in a domain Ω whose boundary Σ carries surface sources cannot satisfy the boundary conditions (2.1) and (2.2) everywhere on Σ. Therefore, with surface sources on Σ the solenoidal eigenfields (2.22) and (2.23) (i.e., the normal modes) do not constitute a complete set of basis vectors for the expansions of **E** and **H** even if there are no charge or current distributions in Ω (i.e., even if div**E** $= 0$ and div**H** $= 0$ in Ω). Waveguide openings and absorbing boundary surfaces (imperfectly conducting cavity walls) are possible realizations of surface sources (or sinks). The domain $\Omega(r_1,r_2)$ with imperfectly conducting inner boundary surface $\mathscr{S}^2_{r_1}$ is an idealization of the system consisting of an over-dense energy absorbing plasma core of radius r_1 at the center of a spherical microwave cavity of radius r_2. An eigenmode expansion of the microwave field containing the plasma core then involves the solenoidal as well as the irrotational eigenfields.

Since the roots of the eigenvalue equations (2.19a), (2.20a), and (2.21a) [(2.34), (2.35), and (2.36)] do not depend on the multipole index $m(m = -l, -l+1,...,l)$, the wavenumbers [and, by (1.11c), the frequencies] of the solenoidal and irrotational eigenfields (2.22), (2.23), (2.26), and (2.27) of Ω_{r_0}, $[\Omega(r_1,r_2)]$ are at least $(2l+1)$-fold. This multiplicity is referred to as azimuthal degeneracy of the eigenfrequencies. The coincidence of the frequencies of the multipole fields (2.22) and (2.26), both determined by the roots of (2.19a) in Ω_{r_0}, and by the roots of (2.34) in $\Omega(r_1,r_2)$, constitutes another type of eigenfrequency degeneracy. A third type of eigenfrequency degeneracy, caused by mode crossing, occurs in the domain $\Omega(r_1,r_2)$ for certain values of the ratio parameter $\alpha = r_1/r_2$. This degeneracy is discussed in the following section.

3. THE EIGENFREQUENCIES OF THE CAVITY $\Omega(r_1,r_2)$ AS FUNCTIONS OF THE RATIO r_1/r_2

We now analyze the dependence of the roots λ_{ln} of (2.34), η_{ln} of (2.35), and μ_{ln} of (2.36) on the continuous real parameter α whose domain is the open interval $(0,1) = \{\alpha:0<\alpha<1\}$ [see Ref. 17].

If Eq. (2.34) is written as

$$F_l(\alpha,\lambda) = 0, \tag{3.1a}$$

where

$$F_l(\alpha,\lambda) = j_l(\alpha\lambda)n_l(\lambda) - j_l(\lambda)n_l(\alpha\lambda), \tag{3.1b}$$

then

$$\frac{d\lambda}{d\alpha} = -\frac{\partial F_l/\partial\alpha}{\partial F_l/\partial\lambda}. \tag{3.2}$$

For $\lambda = \lambda_{ln}$ the partial derivatives in (3.2) are

$$\left.\frac{\partial F_l}{\partial\alpha}\right|_{\lambda=\lambda_{ln}} = \lambda_{ln}[j_{l-1}(\alpha\lambda_{ln})n_l(\lambda_{ln}) - j_l(\lambda_{ln})n_{l-1}(\alpha\lambda_{ln})] \tag{3.3a}$$

and

$$\left.\frac{\partial F_l}{\partial\lambda}\right|_{\lambda=\lambda_{ln}} = \alpha[j_{l-1}(\alpha\lambda_{ln})n_l(\lambda_{ln}) - j_l(\lambda_{ln})n_{l-1}(\alpha\lambda_{ln})] + j_l(\alpha\lambda_{ln})n_{l-1}(\lambda_{ln}) - j_{l-1}(\lambda_{ln})n_l(\alpha\lambda_{ln}). \tag{3.3b}$$

Since $\lambda_{ln}\neq 0$ $(l,n = 1,2,\cdots)$ and since the relation

$$j_{l-1}(\alpha\lambda)n_l(\lambda) - j_l(\lambda)n_{l-1}(\alpha\lambda) = 0$$

is incompatible with (3.1), it follows from (3.3a) that

$$\left.\frac{\partial F_l}{\partial\alpha}\right|_{\lambda=\lambda_{ln}} \neq 0 \quad (0<\alpha<1). \tag{3.4}$$

The right-hand side of Eq. (3.3b) is different from zero for $0<\alpha<1$, and vanishes if and only if $\alpha = 1$. Therefore,

$$d\lambda_{ln}(\alpha)/d\alpha \neq 0 \quad (0<\alpha<1), \tag{3.5}$$

which means that λ_{ln} is a monotonic function of α. This implies that if,

$$\lambda_{ln}(\alpha_2) > \lambda_{ln}(\alpha_1) \quad \text{for } \alpha_2>\alpha_1, \tag{3.6}$$

where α_1 and α_2 are in (0,1), then $\lambda_{ln}(\alpha)$ is monotonically increasing for $0<\alpha<1$. The roots λ_{ln} given in Ref. 16 for $l = 1(1)15$ and $n = 1(1)30$ satisfy condition (3.6). Since

$$\lim_{\alpha\to 1}\frac{\partial F_l}{\partial\alpha} \neq 0, \quad \lim_{\alpha\to 1}\frac{\partial F_l}{\partial\lambda} = 0,$$

one has with condition (3.6)

$$\lim_{\alpha\to 1}\frac{d\lambda_{ln}}{d\alpha} = +\infty. \tag{3.7}$$

By virtue of (3.1) and the cross-product relation

$$j_{l-1}(z)n_l(z) - j_l(z)n_{l-1}(z) = -z^2, \tag{3.8}$$

it is possible to rewrite the expressions (3.3a) and (3.3b) as

$$\left.\frac{\partial F_l}{\partial\alpha}\right|_{\lambda=\lambda_{ln}} = -(\alpha^2\lambda_{ln})^{-1}\tau_l^{-1}(\alpha,\lambda_{ln}), \tag{3.9a}$$

$$\left.\frac{\partial F_l}{\partial\lambda}\right|_{\lambda=\lambda_{ln}} = \lambda_{ln}^{-2}[\tau_l(\alpha,\lambda_{ln}) - \alpha^{-1}\tau_l^{-1}(\alpha,\lambda_{ln})], \tag{3.9b}$$

where

$$\tau_l(\alpha,\lambda) = j_l(\alpha\lambda)/j_l(\lambda) = n_l(\alpha\lambda)/n_l(\lambda). \tag{3.9c}$$

From (3.2) and (3.9) it is immediate that

$$\frac{d\lambda_{ln}}{d\alpha} = \frac{\lambda_{ln}/\alpha}{\alpha\tau_l^2(\alpha,\lambda_{ln}) - 1}. \tag{3.10}$$

If (2.35) is written as

$$G_l(\alpha,\eta) = 0, \tag{3.11a}$$

where

$$G_l(\alpha,\eta) = s_l(\alpha\eta)t_l(\eta) - s_l(\eta)t_l(\alpha\eta), \tag{3.11b}$$

then

$$\frac{d\eta}{d\alpha} = -\frac{\partial G_l/\partial\alpha}{\partial G_l/\partial\eta}. \tag{3.12}$$

For $\eta = \eta_{ln}$ the partial derivatives in (3.12) are

$$\left.\frac{\partial G_l}{\partial\alpha}\right|_{\eta=\eta_{ln}} = \alpha^{-2}\eta_{ln}^{-1}[l(l+1) - (\alpha\eta_{ln})^2]\rho_l^{-1}(\alpha,\eta_{ln}), \tag{3.13a}$$

$$\left.\frac{\partial G_l}{\partial\eta}\right|_{\eta=\eta_{ln}} = \alpha^{-1}\eta_{ln}^{-2}[l(l+1) - (\alpha\eta_{ln})^2]\rho_l^{-1}(\alpha,\eta_{ln}) - \eta_{ln}^{-2}[l(l+1) - \eta_{ln}^2]\rho_l(\alpha,\eta_{ln}), \tag{3.13b}$$

where

$$\rho_l(\alpha,\eta) = s_l(\alpha,\eta)/s_l(\eta) = t_l(\alpha,\eta)/t_l(\eta). \tag{3.13c}$$

In the derivation of (3.13) the cross product relation (3.8) was used. Clearly, $\rho_l(\alpha,\eta)\neq 0$ and $\rho_l^{-1}(\alpha,\eta)\neq 0$. Since $\eta_{ln}\neq 0$ $(l,n = 1,2,\cdots)$, it follows from (3.12) and (3.13) that

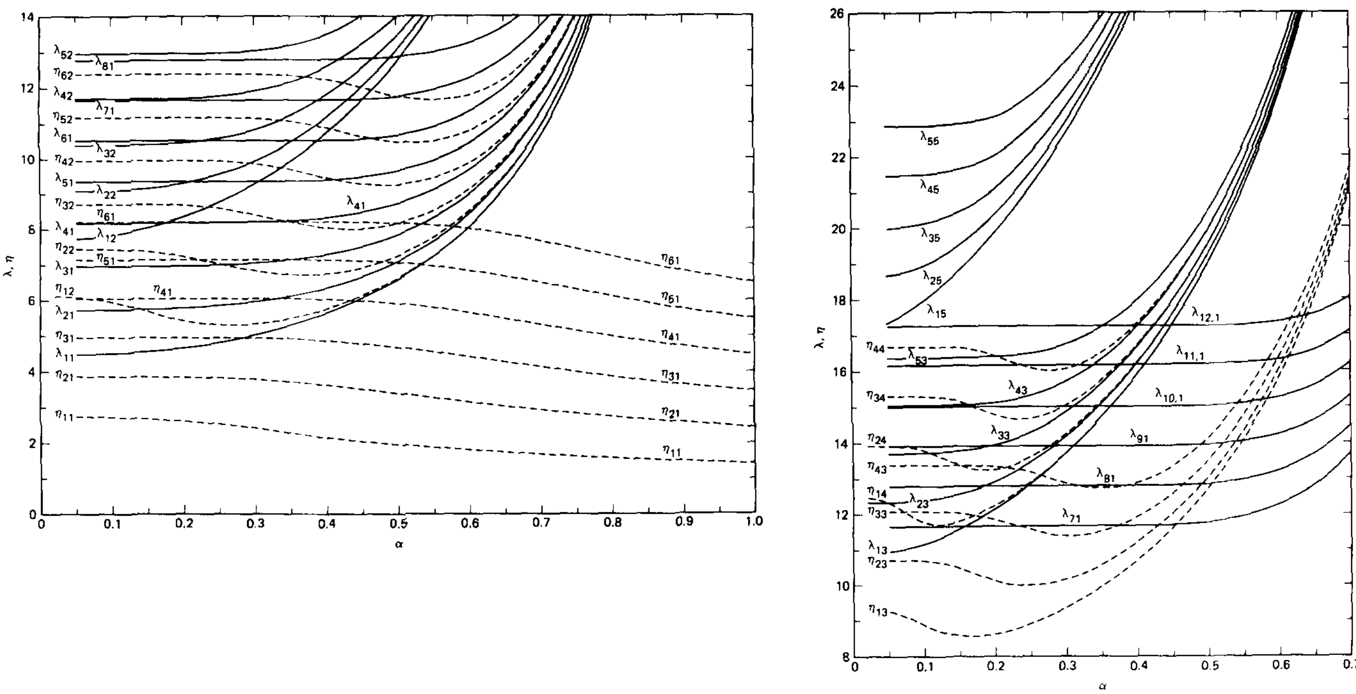
$$d\eta_{ln}/d\alpha = 0, \quad 0<\alpha<1 \tag{3.14}$$

if and only if

$$\alpha\eta_{ln} = [l(l+1)]^{1/2}. \tag{3.15}$$

Furthermore, (3.12) and (3.13) imply that

$$\lim_{\alpha\to 1} d\eta_{l1}/d\alpha = -[l(l+1)]^{1/2} \tag{3.16}$$



FIGS. 1 and 2. Roots λ_{ln} of Eq. (2.34) and η_{ln} of Eq. (2.35) represented as functions of $\alpha = r_1/r_2$ by solid and dashed curves, respectively. The circular eigenfrequencies of the transverse electric or the irrotational electric and of the transverse magnetic multipole fields in $\Omega(r_1,r_2)$ are $\omega_{ln} = (c/r_2)\lambda_{ln}$ and $\omega_{ln} = (c/r_2)\eta_{ln}$, respectively.

if

$$\lim_{\alpha\to 1} \eta_{l1} = [l(l+1)]^{1/2}, \tag{3.17}$$

and

$$\lim_{\alpha\to 1} d\eta_{ln}/d\alpha = \infty \quad (n>1) \tag{3.18}$$

if (3.17) does not hold. The relation (3.16) was derived by means of l'Hospital's rule. The index $n = 1$ is assigned to the root satisfying (3.16) and (3.17), since it is the lowest root for a given value of l.[16,17] The roots η_{l1} $(l = 1,2,\cdots)$ are monotonically decreasing functions of α. In accordance with (3.14), all the roots η_{ln} with $n > 1$ are not monotonic functions of α. The minima of these roots and the corresponding values of

TABLE I. Minima of the roots η_{ln} in Figs. 1 and 2.

l	n	$\alpha_{\min}$	$\eta_{\min}$
1	2	0.26590	5.3186
2	2	0.36541	6.7034
3	2	0.43306	7.9992
4	2	0.48354	9.2488
5	2	0.52322	10.4683
6	2	0.55553	11.6659
1	3	0.16571	8.5345
2	3	0.24494	10.0004
3	3	0.30448	11.3770
4	3	0.35200	12.7048
1	4	0.12079	11.7085
2	4	0.18533	13.2167
3	4	0.23663	14.6396
4	4	0.27925	16.0150

α, which are determined by (3.11) and (3.15), are presented in Ref. 17 for $l = 1(1)15$ and $n = 1(1)30$. By substituting in (3.12) for the partial derivatives the expressions (3.13a) and (3.13b) one obtains

$$\frac{d\eta_{ln}}{d\alpha} = \frac{\eta_{ln/\alpha}}{\alpha\sigma_l(\alpha,\eta_{ln})\rho_l^2(\alpha,\eta_{ln}) - 1}, \tag{3.19}$$

where

$$\sigma_l(\alpha,z) = [l(l+1) - z^2]/[l(l+1) - (\alpha z)^2]. \tag{3.20}$$

Graphical representations of some of the roots, λ_{ln} and η_{ln}, of the transcendental equations (3.1) and (3.11) as functions of the parameter α are given in Figs. 1 and 2. In accordance with (1.11c), (2.37), and (2.38), the intersections of the curves $\lambda_{ln}(\alpha)$ and $\eta_{ln}(\alpha)$ indicate that α-dependent degeneracies of the characteristic frequencies of the fields (2.22), (2.23), and (2.26) occur in the cavity $\Omega(r_1,r_2)$. These degeneracies are referred to as mode crossing. The values of the minima of the curves $\eta_{ln}(\alpha)$ which appear in Figs. 1 and 2 are listed in Table I together with the values of the parameter α for which these minima occur.

If (2.36) is expressed as

$$K_l(\alpha,\mu) = 0, \tag{3.21a}$$

where

$$K_l(\alpha,\mu) = v_l(\alpha\mu)w_l(\mu) - v_l(\mu)w_l(\alpha\mu), \tag{3.21b}$$

then

$$\frac{d\mu}{d\alpha} = -\frac{\partial K_l/\partial\alpha}{\partial K_l/\partial\mu}. \tag{3.22}$$

For $\mu = \mu_{ln}$ the partial derivatives in (3.22) are

$$\left.\frac{\partial K_l}{\partial \alpha}\right|_{\mu=\mu_{ln}} = \alpha^{-4}\mu_{ln}^{-3}[l(l+1)-(\alpha\mu_{ln})^2]\chi_l^{-1}(\alpha,\mu_{ln})\,, \tag{3.23a}$$

$$\left.\frac{\partial K_l}{\partial \mu}\right|_{\mu=\mu_{ln}} = \alpha^{-3}\mu_{ln}^{-4}[l(l+1)-(\alpha\mu_{ln})^2]\chi_l^{-1}(\alpha,\mu_{ln})\,, \\ -\mu_{ln}^{-4}[l(l+1)-\mu_{ln}^2]\chi_l(\alpha,\mu_{ln})\,, \tag{3.23b}$$

where

$$\chi_l(\alpha,\mu) = v_l(\alpha\mu)/v_l(\mu) = w_l(\alpha\mu)/w_l(\mu)\,. \tag{3.23c}$$

By substituting in (3.22) for the partial derivatives the expressions (3.23a) and (3.23b) one obtains

$$\frac{d\mu_{ln}}{d\alpha} = \frac{\mu_{ln}/\alpha}{\alpha^3\sigma_l(\alpha,\mu_{ln})\chi_l^2(\alpha,\mu_{ln})-1}\,, \tag{3.24}$$

where $\sigma_l(\alpha,\mu_{ln})$ is defined by (3.20). From (3.22) and (3.23) it is evident that the functional dependence on the parameter α of the roots μ_{ln} is similar to that of the roots η_{ln}. In particular,

$$d\mu_{ln}/d\alpha = 0, \quad 0<\alpha<1 \tag{3.25}$$

if and only if

$$\alpha\mu_{ln} = [l(l+1)]^{1/2}\,; \tag{3.26}$$

furthermore,

$$\lim_{\alpha\to 1} d\mu_{l1}/d\alpha = [l(l+1)]^{1/2} \tag{3.27}$$

if

$$\lim_{\alpha\to 1}\mu_{l1} = [l(l+1)]^{1/2}\,, \tag{3.28}$$

and

$$\lim_{\alpha\to 1} d\mu_{ln}/d\alpha = \infty \quad (n>1) \tag{3.29}$$

if (3.28) does not hold. Figure 3 shows the dependence on the parameter α of some of the roots, μ_{ln}, of the transcendental equation (3.21); the minima of these roots and the corresponding values of α are exhibited in Table II.

From Figs. 1, 2, and 3 it is then evident that the following types of crossing degeneracies occur for the roots $\lambda_{ln}(\alpha)$, $\eta_{ln}(\alpha)$, and $\mu_{ln}(\alpha)$:

$$\lambda_{ln}(\alpha)=\lambda_{l'n'}(\alpha),\quad \eta_{ln}(\alpha)=\eta_{l'n'}(\alpha),\quad \mu_{ln}(\alpha)=\mu_{l'n'}(\alpha)\,, \\ \lambda_{ln}(\alpha)=\eta_{l'n'}(\alpha),\quad \lambda_{ln}(\alpha)=\mu_{l'n'}(\alpha),\quad \eta_{ln}(\alpha)=\mu_{l'n'}(\alpha)\,. \tag{3.30}$$

Crossing degeneracies exist therefore in the domain $\Omega(r_1,r_2)$ not only for the frequencies of the normal modes (2.22) and (2.23), but also for the frequencies of the irrotational electric and magnetic eigenvectors (2.26) and (2.27).

TABLE II. Minima of the roots μ_{ln} in Fig. 3.

l	n	$\alpha_{\min}$	$\mu_{\min}$
1	2	0.24495	5.7735
2	2	0.34837	7.0314
3	2	0.41909	8.2657
4	2	0.47183	9.4782
5	2	0.51321	10.6725
6	2	0.54682	11.8517
1	3	0.15635	9.0453
2	3	0.23626	10.3680
3	3	0.29675	11.6736
4	3	0.34512	12.9582

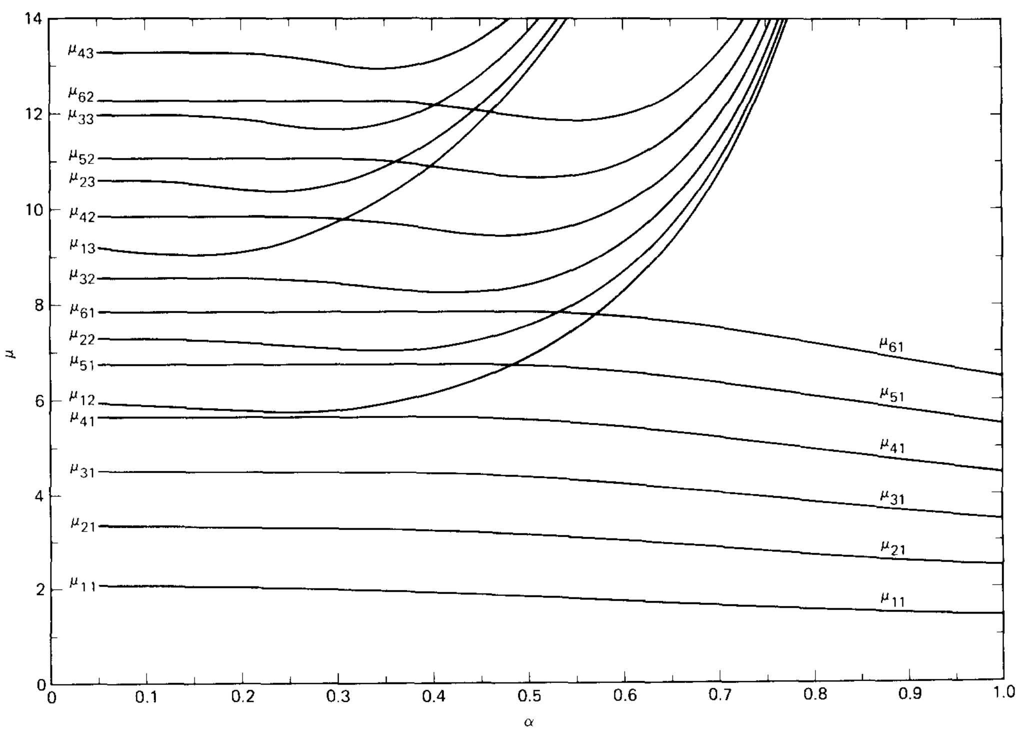


FIG. 3. Roots μ_{ln} of Eq. (2.36) as functions of $\alpha = r_1/r_2$. The circular eigenfrequencies of the irrotational magnetic multipole fields in $\Omega(r_1,r_2)$ are $\omega_{ln} = (c/r_2)\mu_{ln}$.

The roots $\eta_{11}(\alpha)$, $\eta_{21}(\alpha)$, $\mu_{11}(\alpha)$, and $\mu_{21}(\alpha)$ are obviously free of mode-crossing degeneracies. Values of the degenerate roots (3.30), together with the values of the parameter α for which those degeneracies occur, are presented in Ref. 18.

4. ADIABATIC COMPRESSION OF THE CHARACTERISTIC ELECTROMAGNETIC FIELDS IN $\Omega(r_1,r_2)$

The relations (3.10) and (3.19) can be derived by calculating the adiabatic change of the time-average electromagnetic energy of the normal modes in $\Omega(r_1,r_2)$ which is caused by a spherically symmetric expansion or contraction of the boundary surfaces. For the compression of the electric and magnetic fields in $\Omega(r_1,r_2)$ to be adiabatic it is necessary that the boundary surfaces $\mathscr{S}^2_{r_1}$ and $\mathscr{S}^2_{r_2}$ be perfect conductors and that their displacements be sufficiently slow so that transitions between adjacent modes cannot be induced by the Doppler effect.

If the radius of the inner sphere, $\mathscr{S}^2_{r_1}$, is increased from $r_1 = \alpha r_2$ to $r_1 + \delta r_1 = (\alpha + \delta\alpha) r_2$, the time-average electromagnetic field energy changes by the amount

$$\delta U = -r_1^2 \delta r_1 \int d(\theta,\phi) \mathbf{F}_p \cdot \mathbf{e}_r = -r_2^3 \alpha^2 \delta\alpha \int d(\theta,\phi) \mathbf{F}_p \cdot \mathbf{e}_r , \tag{4.1}$$

where $d(\theta,\phi)$ is defined by (1.16) and where for $\epsilon = \mu = 1$ [Eq. (3.5) of Ref. 8]

$$\begin{aligned}\mathbf{F}_p &= -\mathbf{e}_r (1/8\pi)[\mathbf{H}_\perp^2 - \mathbf{E}_\parallel^2]_{(r=r_1)} \\ &= -\mathbf{e}_r(1/16\pi)[(\mathbf{H}_{nlm})_\perp^* \cdot (\mathbf{H}_{nlm})_\perp \\ &\quad - (\mathbf{E}_{nlm})_\parallel^* \cdot (\mathbf{E}_{nlm})_\parallel]_{(r=r_1)}\end{aligned} \tag{4.2}$$

is the time-average electromagnetic pressure on the (perfectly conducting) sphere $\mathscr{S}^2_{r_1}$.[21] In the second equation (4.2), the time-averaging was performed in accordance with (1.1). To evaluate (4.1) for the transverse electric normal modes, only the part of the magnetic field (2.22b) that is tangent to $\mathscr{S}^2_{r_1}$ is needed. From (1.34c), (2.32), (2.33), (2.34), (2.40), (2.41), and (3.8) one obtains

$$\begin{aligned}(\mathbf{H}^{\mathrm{TE}}_{nlm})_{\perp(r=r_1)} &= iC^{(e)}_{ln}[l(l+1)]^{1/2}(\alpha\lambda_{ln})^{-2} \\ &\quad \times [n_l(\alpha\lambda_{ln})]^{-1} \mathbf{X}^{(E)}_{lm}(\theta,\phi) .\end{aligned} \tag{4.3}$$

If (4.3) is substituted in (4.2), and if in (4.1) the solid-angle integration is performed by means of (1.33), one finds

$$\begin{aligned}\delta U^{\mathrm{TE}}_{ln} &= |C^{(e)}_{ln}|^2 r_2^3 \alpha^2 \delta\alpha (16\pi)^{-1} l(l+1)(\alpha\lambda_{ln})^{-4} \\ &\quad \times [n_l(\alpha\lambda_{ln})]^{-2} .\end{aligned} \tag{4.4}$$

In order to derive the corresponding expression for the transverse magnetic normal modes, the radial part of the electric field (2.23a) and the tangential part of the magnetic field (2.23b) are needed on $\mathscr{S}^2_{r_1}$. From (2.23a), (1.34c), (2.32), (2.33), (2.35), (2.40), (2.42), and (3.8) follows

$$\begin{aligned}(\mathbf{E}^{\mathrm{TM}}_{nlm})_{\parallel(r=r_1)} &= -C^{(h)}_{ln} l(l+1)(\alpha\lambda_{ln})^{-2} \\ &\quad \times [t_l(\alpha\eta_{ln})]^{-1} \mathbf{X}^{(L)}_{lm}(\theta,\phi) .\end{aligned} \tag{4.5}$$

From (2.23b), (1.34b), (2.32), (2.33), (2.35), (2.40), (2.42), and (3.8) follows

$$\begin{aligned}(\mathbf{H}^{\mathrm{TM}}_{nlm})_{\perp(r=r_1)} &= iC^{(h)}_{ln}[l(l+1)]^{1/2}[\alpha n_{ln} t_l(\alpha\eta_{ln})]^{-1} \\ &\quad \times \mathbf{X}^{(M)}_{lm}(\theta,\phi) .\end{aligned} \tag{4.6}$$

By combining (4.1), (4.2), (4.5), and (4.6), and by utilizing (1.33) one arrives at

$$\begin{aligned}\delta U^{\mathrm{TM}}_{ln} &= |C^{(h)}_{ln}|^2 r_2^3 \alpha^2 \delta\alpha (16\pi)^{-1} l(l+1)(\alpha\eta_{ln})^{-4} \\ &\quad \times [t_l(\alpha\eta_{ln})]^{-2}[(\alpha\eta_{ln})^2 - l(l+1)] .\end{aligned} \tag{4.7}$$

For the transverse electric and the transverse magnetic normal modes (2.22) and (2.23) the total time-average electromagnetic energy in the domain $\Omega(r_1,r_2)$ [which also equals the total instantaneous electromagnetic energy, since the boundary $\Sigma(r_1,r_2)$ is assumed to be perfectly conducting] is for $\epsilon = \mu = 1$ given by

$$\begin{aligned}U_{ln} &= \frac{1}{16\pi}\int_{r_1}^{r_2} dr\, r^2 \int d(\theta,\phi)[|\mathbf{E}_{nlm}|^2 + |H_{nlm}|^2] \\ &= \frac{1}{8\pi}\int_{r_1}^{r_2} dr\, r^2 \int d(\theta,\phi)|\mathbf{E}_{nlm}|^2 \\ &= \frac{1}{8\pi}\int_{r_1}^{r_2} dr\, r^2 \int d(\theta,\phi)|\mathbf{H}_{nlm}|^2 .\end{aligned} \tag{4.8}$$

The radial integrations can be carried out by usig the relation [Ref. 8, Eqs. (A1) and (A2)]

$$\begin{aligned}\int_{r_1}^{r_2} r^2\, dr [j_l(kr) + \gamma n_l(kr)]^2 &= \tfrac{1}{2} r^3 \{ [j_l(kr) + \gamma n_l(kr)]^2 \\ &\quad - [j_{l+1}(kr) + \gamma n_{l+1}(kr)][j_{l-1}(kr) + \gamma n_{l-1}(kr)]\}\Big|_{r_1}^{r_2} .\end{aligned} \tag{4.9}$$

If the normal modes (2.22) and (2.23) are represented by (1.29b, 1.29c), the angular integrations are performed by means of (1.24). By combining the formulas (1.29b, 1.29c) with the generating functions (2.40) one then arrives at the results

$$\begin{aligned}U^{\mathrm{TE}}_{ln} &= |C^{(e)}_{ln}|^2 r_2^3 \alpha^3 (16\pi)^{-1} l(l+1)(\alpha\lambda_{ln})^{-4} \\ &\quad \times [n_l(\alpha\lambda_{ln})]^{-2}[\alpha\tau_l^2(\alpha,\lambda_{ln}) - 1] ,\end{aligned} \tag{4.10}$$

where $\tau_l(\alpha,\lambda)$ is defined by (3.9c), and

$$\begin{aligned}U^{\mathrm{TM}}_{ln} &= |C^{(h)}_{ln}|^2 r_2^3 \alpha^3 (16\pi)^{-1} l(l+1)(\alpha\eta_{ln})^{-4}[t_l(\alpha\eta_{ln})]^{-2} \\ &\quad \times [(\alpha\eta_{ln})^2 - l(l+1)][\alpha\sigma_l(\alpha,\eta_{ln})\rho_l^2(\alpha,\eta_{ln}) - 1] ,\end{aligned} \tag{4.11}$$

where $t_l(\alpha\eta)$, $\rho_l(\alpha\eta)$, and $\sigma_l(\alpha\eta)$ are defined by (2.35b), (3.13c), and (3.20).

From (4.4) and (4.10) it is immediate that

$$\frac{dU^{\mathrm{TE}}_{ln}}{d\alpha} = \frac{U^{\mathrm{TE}}_{ln}/\alpha}{\alpha\tau_l^2(\alpha,\lambda_{ln}) - 1} , \tag{4.12}$$

and from (4.7) and (4.11) it is immediate that

$$\frac{dU^{\mathrm{TM}}_{ln}}{d\alpha} = \frac{U^{\mathrm{TM}}_{ln}/\alpha}{\alpha\sigma_l(\alpha,\eta_{ln})\rho_l^2(\alpha,\eta_{ln}) - 1} . \tag{4.13}$$

Since the elementary excitations of the electromagnetic fields in $\Omega(r_1,r_2)$ are photons of energy $\hbar\omega_{ln}$ (and angular momentum $l\hbar$), one may write

$$U^{\mathrm{TE}}_{ln} = N^{\mathrm{TE}}_{ln}\hbar\omega^{\mathrm{TE}}_{ln} = N^{\mathrm{TE}}_{ln} c\hbar k_{ln} \tag{4.14}$$

and

$$U^{\mathrm{TM}}_{ln} = N^{\mathrm{TM}}_{ln}\hbar\omega^{\mathrm{TM}}_{ln} = N^{\mathrm{TM}}_{ln} c\hbar k'_{ln} . \tag{4.15}$$

The conservation of the photon numbers N^{TE}_{ln} and N^{TM}_{ln} in $\Omega(r_1,r_2)$ is ensured by the assumption of perfect conductivity of the boundary $\Sigma(r_1,r_2)$. The relations (4.14) and (4.15) together with (2.37) and (2.38) imply that (4.12) and (4.13) are equivalent with (3.10) and (3.19), respectively.

APPENDIX

The vector product of two spherical basis vectors (1.21) can be expressed as

$$\boldsymbol{\epsilon}_\rho \times \boldsymbol{\epsilon}_\sigma = i(\sqrt{2})\langle 1\rho 1\sigma|1\tau\rangle \boldsymbol{\epsilon}_\tau . \qquad \text{(A1)}$$

From the definition (1.20) we obtain after reducing each product $Y_{j'\mu'} Y_{j\mu}$ and using Eq. (A1) the following relation for the vector product of two vector spherical harmonics:

$$\begin{aligned}&\mathcal{Y}^{m'}_{l'j'}(\theta,\phi)\times\mathcal{Y}^{m}_{lj}(\theta,\phi)\\ &= i\sqrt{2}\left(\frac{(2j'+1)(2j+1)}{4\pi}\right)^{1/2}\sum_{\bar j}(2\bar j+1)^{-1/2}\langle j'0j0|\bar j0\rangle\\ &\quad\times\sum_{\mu'\tau'\mu\tau\bar\mu\bar\tau}\langle j'\mu'1\tau'|l'm'\rangle\,\langle j\mu 1\tau|lm\rangle\,\langle j'\mu' j\mu|\bar j\bar\mu\rangle\\ &\quad\times\langle 1\tau' 1\tau|1\bar\tau\rangle\, Y_{\bar j\bar\mu}(\theta,\phi)\boldsymbol{\epsilon}_{\bar\tau} . \qquad \text{(A2)}\end{aligned}$$

The summation with respect to the indices μ',τ',μ,τ may be performed in accordance with the recoupling transformation[6]

$$\begin{aligned}&\sum_{\mu'\tau'\mu\tau}\langle j'\mu'1\tau'|l'm'\rangle\,\langle j\mu1\tau|lm\rangle\langle j'\mu'j\mu|\bar j\bar\mu\rangle\,\langle 1\tau'1\tau|1\bar\tau\rangle\\ &\quad= [3(2\bar j+1)(2l'+1)(2l+1)]^{1/2}\sum_{\bar l}\langle\bar j\bar\mu 1\bar\tau|\bar l\bar m\rangle\\ &\qquad\times\langle l'm'lm|\bar l\bar m\rangle\begin{Bmatrix} j' & 1 & l'\\ j & 1 & l\\ \bar j & 1 & \bar l\end{Bmatrix}. \qquad \text{(A3)}\end{aligned}$$

By substituting (A3) and (A2) and taking account of the definition (1.20) we arrive at the formula for the decomposition of the vector product of two vector spherical harmonics,

$$\begin{aligned}&\mathcal{Y}^{m'}_{l'j'}(\theta,\phi)\times\mathcal{Y}^{m}_{lj}(\theta,\phi)\\ &= i\left(\frac{3}{2\pi}\right)^{1/2}[2j'+1)(2j+1)(2l'+1)(2l+1)]^{1/2}\\ &\quad\times\sum_{\bar j}\langle j'0j0|\bar j0\rangle\sum_{\bar l}\langle l'm'lm|\bar l,m'+m\rangle\begin{Bmatrix} j' & 1 & l'\\ j & 1 & l\\ \bar j & 1 & \bar l\end{Bmatrix}\\ &\quad\times\mathcal{Y}^{m'+m}_{\bar l\bar j}(\theta,\phi) . \qquad \text{(A4)}\end{aligned}$$

The relation (A4) enables one to represent the Poynting vector $\mathbf{S}=(c/4\pi)\mathbf{E}\times\mathbf{B}$ and the momentum density $c^{-2}\mathbf{S}$ of the electromagnetic radiation field (1.11) by irreducible spherical tensors, namely the vector spherical harmonics (1.20). In order to obtain such an irreducible tensor representation for the angular momentum density $c^{-2}\mathbf{r}\times\mathbf{S}$ it is necessary to decompose the triple vector product $\mathbf{e}_r\times(\mathcal{Y}^{m'}_{l'j'}\times\mathcal{Y}^{m}_{lj})$ into vector spherical harmonics. On account of (1.26) one may write

$$\begin{aligned}&\mathbf{e}_r\times(\mathcal{Y}^{m'}_{l'j'}\times\mathcal{Y}^{m}_{lj})\\ &\quad= -\langle l010|j0\rangle Y_{lm}\mathcal{Y}^{m'}_{l'j'} + \langle l'010|j'0\rangle Y_{l'm'}\mathcal{Y}^{m}_{lj} . \qquad \text{(A5)}\end{aligned}$$

The relation

$$\begin{aligned}&Y_{lm}\mathcal{Y}^{m'}_{l'j'} = [(2l+1)(2j'+1)/4\pi]^{1/2}\\ &\quad\times\sum_{\mu'\tau'}\langle j'\mu'1\tau'|l'm'\rangle\boldsymbol{\epsilon}_{\tau'}\sum_{\bar j\bar\mu}(2\bar j+1)^{-1/2}\\ &\quad\times\langle l0j'0|\bar j0\rangle\,\langle lmj'\mu'|\bar j\bar\mu\rangle Y_{\bar j\bar\mu} , \qquad \text{(A6)}\end{aligned}$$

can be obtained by using (1.20) and by reducing products of spherical harmonics. By substituting in (A6) the recoupling transformation[6]

$$\begin{aligned}&\langle lmj'\mu'|\bar j\bar\mu\rangle\,\langle j'\mu'1\tau'|l'm'\rangle\\ &\quad= -(-1)^{j'+l'}[(2l'+1)(2\bar j+1)]^{1/2}\\ &\qquad\times\sum_{\bar l}\langle l'm'lm|\bar l,m'+m\rangle\\ &\qquad\times\langle\bar j\bar\mu 1\tau'|\bar l,m'+m\rangle\begin{Bmatrix} l & l' & \bar l\\ 1 & \bar j & j'\end{Bmatrix},\end{aligned}$$

one finds after invoking the definition (1.20)

$$\begin{aligned}&Y_{lm}(\theta,\phi)\mathcal{Y}^{m'}_{l'j'}(\theta,\phi)\\ &\quad= -(-1)^{j'+l'}[(2l+1)(2l'+1)(2j'+1)/4\pi]^{1/2}\\ &\qquad\times\sum_{\bar l}\langle l'm'lm|\bar l,m'+m\rangle\\ &\qquad\times\sum_{\bar j}\langle l0j'0|\bar j0\rangle\begin{Bmatrix} l & l' & \bar l\\ 1 & \bar j & j'\end{Bmatrix}\mathcal{Y}^{m'+m}_{\bar l\bar j}(\theta,\phi) .\end{aligned}$$

In view of this result, Eq. (A5) becomes

$$\begin{aligned}&\mathbf{e}_r\times(\mathcal{Y}^{m'}_{l'j'}\times\mathcal{Y}^{m}_{lj}) = [(2l+1)(2l'+1)/4\pi]^{1/2}(-1)^{l'}\\ &\times\sum_{\bar l,\bar j}\Big[(-1)^{j}(2j'+1)^{1/2}\langle l010|j0\rangle\,\langle l0j'0|\bar j0\rangle\begin{Bmatrix} l & l' & \bar l\\ 1 & \bar j & j'\end{Bmatrix}\\ &-(-)^{j-\bar l}(2j+1)^{1/2}\langle l'010|j'0\rangle\,\langle l'0j0|\bar j0\rangle\begin{Bmatrix} l' & l & \bar l\\ 1 & \bar j & j\end{Bmatrix}\Big]\\ &\cdot\langle l'm'lm|\bar l,m'+m\rangle\mathcal{Y}^{m'+m}_{\bar l,\bar j} . \qquad \text{(A7)}\end{aligned}$$

[1] R. Müller, "Theory of Cavity Resonators," in G. Goubau, *Electromagnetic Waveguides and Cavities* (Pergamon, New York, 1961), Chap. II.
[2] K. Kurokawa, IRE Trans. **MIT-6**, 178 (1958).
[3] K. Kurokawa, *An Introduction to the Theory of Microwave Circuits* (Academic, New York, 1969).
[4] R.E. Collin, *Foundations for Microwave Engineering* (McGraw–Hill, New York, 1966).
[5] F.E. Borgnis and C.H. Papas, *Encyclopedia of Physics* (Springer–Verlag, Berlin, 1958), Vol. XVI, Sec. 33.
[6] M.E. Rose, *Multipole Fields* (Wiley, New York, 1955).
[7] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), Chap. 13.
[8] A.D. Steiger, J. Math. Phys. **18**, 312 (1977).
[9] C.H. Woods, Phys. Fluids **20**, 252 (1976).
[10] E. Garelis, Phys. Fluids **17**, 2002 (1974).
[11] J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed.
[12] J.J. Monaghan, Astron. Astrophys. **29**, 383 (1973).
[13] H.L. Boyen, A.M. Messiaen, and P.E. Vandenplas, J. Appl. Phys. **40**, 2296 (1969).
[14] R.L. Pexton and A.D. Steiger, Math. Comp. **31**, 1000 (1977).
[15] M. Daumens and P. Minnaert, J. Math. Phys. **17**, 1903 (1976).
[16] R.L. Pexton and A.D. Steiger, Math. Comp. **31**, 752 (1977).
[17] R.L. Pexton and A.D. Steiger, Math. Comp. **32**, 511 (1978).
[18] R.L. Pexton and A.D. Steiger, Math. Comp. **33**, 1041 (1979).
[19] S.A. Schelkunoff, J. Appl. Phys. **26**, 1231 (1955).

[20] For $l=0$, the eigenvalue equations (2.34) and (2.35) both reduce to $\tan(\alpha x)=\tan x$ (with $x=\lambda$ and $x=\eta$), which equation obviously has no nontrivial solution when $0<\alpha<1$.
[21] Since on the surface of a perfect conductor the tangential components of the electric field and the normal component of the magnetic field vanish, the electromagnetic pressure is always normal to such a surface.

Relativistic Brownian motion and the space–time approach to quantum mechanics

Sisir Roy

Indian Statistical Institute, Calcutta-35, India

(Received 12 May 1978; revised manuscript received 30 October 1978)

An attempt has been made to extend the stochastic quantization procedures introduced by Nelson in the nonrelativistic case to the relativistic case in the four-dimensional Finsler space. The space–time in the microdomain is considered to be quantized and a more general concept of probability is needed to have a consistent and complete theory of quantum mechanics.

I. INTRODUCTION

The probability theoretical approach to nonrelativistic quantum mechanics has been introduced by Nelson [1] and developed by Guerra and Ruggiero [2] to deepen our understanding of space–time with microlocal structure, the microscopic domain of less than 10^{-13} cm in length. Recently, Caubet [3] has treated the stochastic quantization procedure in the relativistic case and tried to show that the closeness of ω_2 $[\omega_2 = d(\hbar s)]$ is enough to define the relativistic Brownian motion. However, then we should have trouble with the fifth dimension in the forms ω_1 and ω_2 as well as the vanishing nature of the diffusion coefficient in the relativistic limit as pointed out by Hakim. [4] It is worth mentioning that Yasue [5] has derived the relativistic wave equation along Nelson's approach within the context of the theory of elementary domains as proposed by Yukawa. However, this approach remains futile from the physicists' point of view unless one can explain the spectrum of elementary particles consistently within the context of the theory of elementary domains of the space–time.

The recent development of the theory of nonlocal fields for extended particles [6] and the concept of space–time quantization have created a lot of interest and rethinking of the stochastic quantization with the relativistic treatment in a more realistic fashion. The aim of this paper is to derive the relativistic wave equations in the theory of elementary domains in four-dimensional Finsler space so as to avoid the concept of the fifth dimension. For our convenience, let us briefly recapitulate the theory of space–time quantization and the model of leptons.

II. THE THEORY OF SPACE-TIME QUANTIZATION

The theory of nonlocal fields and the concept of elementary domains as proposed by Yukawa created a lot of interest in explaining the characteristic properties of elementary particles from their internal structures as well as to avoid the divergence difficulty inherent in the usual local field theory. However, unfortunately Yukawa and his collaborators did not succeed in producing any consistent and complete theory. Recently, a new theory of nonlocal field has been proposed by Bandyopadhyay, [7] where the charge and mass of an electron (as well as muon) can be taken to occur as a result of n-photon–neutrino weak interactions, when photons and neutrinos are represented by the relation $e = ng$, g being the photon–neutrino weak coupling constant ($g \simeq 10^{-10}e$). In this model electron and muon are depicted as $(\nu_e S)$ and $(\nu_\mu S)$, respectively, where s represents the system of photons interacting weakly at n space–time points with the extended structure of a two component neutrino. The two other components corresponding to the positive and negative energy states are formed when the form factor associated with the interaction changes its sign, implying that particles and antiparticles are mirror reflections of each other. This procedure helps us to unify weak and electromagnetic interactions and the accompanied violation of symmetry generates the photon as a Goldstone boson. If we take that the charge of a hadron is also due to the presence of a lepton in its structure then the charge spectrum of all hadrons can be interpreted on the basis of this concept of dynamical origin of charge.

In this picture, it is possible to show that the quantization of charge in units of e is related to the quantization of space–time, where each quantized space–time domain is determined by the region accommodating the specific n number of weak interactions involving extended structures of n photons and one neutrino, also considered to be of extended structure. In this context it is worthwhile to mention that a model of the lepton has been constructed out of n photons weakly interacting with the neutrino in the framework of nonlocal field theory. [8] In this picture, the mass and charge of the lepton arise due to the system of nonlocal interactions at different space–time points and the massive and charged lepton occupies the seat of a quantized space–time domain so that from the observational point of view the lepton will appear as a point particle. Since the interactions spinor is a two-component neutrino whose spin direction is either parallel or antiparallel to the momentum, the momentum direction of a neutrino becomes fixed. Thus, this model suggests a preferential direction in space within the fundamental domain, i.e., the internal space should be such as to violate Lorentz invariance. Taking this into consideration, it can be shown that the nonlocal fields representing the extended particles can be described in a consistent way.

From the above picture it is evident that each elementary domain is the seat of a lepton consisting of n number of photons and a single neutrino interacting weakly with the

 0022-2488/80/010071-06$01.00

photons. However, this lepton which derives mass (and charge) from the system of nonlocal interactions moves with a velocity less than that of its massless constituents (which always move with the speed of light). Thus, one may consider some sort of random motion of a particle inside the quantized domain so as to measure only the "mean" path which relates to the center of gravity of a large number of similar particles starting simultaneously from the same point and always confined within the quantized domain. These particles can be thought of as moving with the velocity of light but the mean velocity can be evaluated, generally, with a magnitude less than c and the deviations from the mean path can be regarded as a sort of Brownian motion. This splitting of the motion into a systematic part and a fluctuation is similar to the done in Langevin's equation, with its difference that the fluctuation can no longer be taken to be independent of the mean velocity, because of the constraint imposed by the fixed C, so that our problem reduces to the derivation of the Dirac equation from the relativistic Brownian motion in a locally anisotropic space–time.

III. RELATIVISTIC STOCHASTIC EQUATIONS

It seems rather attractive and natural to formulate the physical laws in the framework of space–time, relating them with the structure of space–time as has been shown by Einstein in his general theory of relativity with the large geometrical structure of space–time. In such a theory as that of Einstein, any local part of the world is assumed as the (Minkowski) flat space and all fundamental processes have been used to be represented in this frame, disregarding the microscopic scale structure. However, if we take it for granted that the Minkowski space was introduced with the calssical notion of light, the microlocal structure of physical space–time may not be the same as its macroscopic Minkowski structure.

In this section, we show that the usual relativistic quantum law is naturally related with the microlocal structure of space–time regarding the totality of four-dimensional elementary domains and that the Minkowski structure appears naturally in the macroscopic scale. Again in the above picture of space–time quantization, as the space–time in the microlevel is anisotropic, effects equivalent to the above, however, can be expected by the introduction of directional dependence into the space–time or of solving in Finsler space. The gravitational field is geometrized in terms of Riemann spaces, and Finsler spaces are, however, expected to geometrize the internal freedom of fields.

Roughly speaking, Finsler spaces are spaces of which the metric tensor $g_{\mu\nu}$ depends on the directional variable $\dot{x}^{\lambda}$ as well as the coordinate x^{λ}. The directional variables $\dot{x}^{\lambda}$ transform like a contravariant vector under a coordinate transformation $x^{\lambda\prime} = x^{\lambda\prime}(x^{\lambda})$. The combination of the coordiantes and the directional variable $(x^{\lambda},\dot{x}^{\lambda})$ is called the element of support. Finsler geometry however, is, formulated in such a way that only the direction of the vector $\dot{x}^{\lambda}$ has a meaning, but not its absolute value. This directional variable can be linked up to the anisotropy of space–time.

Now, this four-dimensional directivity of the locally anisotropic space enables us to retain the relativistic extension of the stochastic quantization which posses the hyperbolicity in its basic equation.

Let F be the four-dimensional Finsler space $q^0 = ict, q^i, v_0, v^i$, $i = 1,2,3$, $q,v\in F$, with x a diffusion process in F, and x_t the position of x at time t where t is the parameter time.

We denote the left and right derivatives of x by

$$(Lf)(q,v) = \lim_{s\uparrow t} E\left[\frac{f(x_t)-f(x_s)}{t-s}\right]\Bigg|_{x'=(q,v)} = \langle V_-,\nabla f\rangle + \langle V_-,\nabla_1 f\rangle - D\,\Box f$$

and

$$(Rf)(q,v) = \lim_{u\uparrow t} E\left[\frac{f(x_u)-f(x_t)}{u-t}\right]\Bigg|_{x_t=(q,v)} = \langle V_+,\nabla f\rangle + \langle V_+,\nabla_1 f\rangle + D^*\Box f,$$

where $\langle\,,\,\rangle$ denotes the inner product, and ∇, ∇_1 = grad and $\Box$ = div grad denote four-dimensional gradients and Laplacian operators, respectively, acting on differentiable functions f on F with respect to coordiantes and directional variables. $E[\,|\,]$ denotes the conditional expectation and D and D^* are diffusion coefficients. In Minkowski space, the d'Alembertion will be reduced to the simple form

$$\Box = g^{\lambda}(v)\frac{\partial}{\partial q\mu}\frac{\partial}{\partial q^{\lambda}}$$

and

$$(Rf)(q,v) = \langle V_+\nabla_f\rangle + D^*\Box f,$$
$$(Lf)(q,v) = \langle V_-\nabla_f\rangle - D\,\Box f.$$

For any particle, we have

$$V_0 = ic,$$

with V_0 the speed of the particle, so that

$$\langle \mathbf{V},\mathbf{V}\rangle = V_j V^j - c^2.$$

This concept of 4-velocity creates trouble in deriving the relativistic wave equation as the parameter time vanishes in the forms ω_1 and ω_2 (and so in the wave equation also. In fact as one looks upon the microworld in its four-dimensional entirely, time itself loses sense as the indicator of the development of phenomena contrary to the nonrelativistic case. For that reason Caubet has introduced a fifth term to both these forms ω_1 and ω_2 (and so in the wave equation also). However, in the locally anisotropic space–time, though $\langle\mathbf{V},\mathbf{V}\rangle$ does not contain time explicitly, it contains some directional variable v from which we have the stochastic quantization. As a result, the question of a fifth dimension as introduced by Caubet does not arise at all.

Now we assume the stationary property

$$P\{x_t^j\in dq^j | T_t = s\} = p\{x_s^j \in dq^j | T_s = s\},$$

whatever the parameter time t may be and $t\in T$. Here T denotes an open interval of the real line R.

A. Continuity of the space–time diffusion

Under classical assumptions of regularity concerning V_+, D^*, and the initial distribution of x, the operators $-L$

and R are adjoined to each other with respect to the measure $\rho\, dq$; in other words, whatever the functions with compact support $f,g,F\to R$ are, we have the following relation:

$$\int f(Lg)\rho\, dq + \int (Rf)g\rho\, dq = 0,$$

where $\rho{:}F\to R$ denotes the density of the conditional distribution of the process at time t such that

$$\rho(q,\tau)dq = \underset{[v]}{E}\,[f(q,v,\tau)]$$
$$= \mathrm{Prob}\{\lambda\,(\tau)\epsilon dq\},$$

with the measure $\rho(q,\tau)dqd\tau$. Here ρ is assumed to be independent of v because otherwise we will not get the corresponding probability density $\bar{\psi}(q,v)\psi(q,v) = \rho(q)$ for the ψ function of quantum mechanics which will be discussed later. As ρ is assumed to be independent of the directional variable v, then the probability density $\rho(q,\tau)$ satisfies, of course, the usual Kolmogorov-Fokker-Planck eqn.

$$\frac{\partial\rho}{\partial\tau}(q,\tau) = -\frac{\partial}{\partial q_\mu}\left\{b^{\mu}(q,v,\tau)\rho(q,\tau) + Dg^{\mu\lambda}\frac{\partial^2}{\partial q\mu\partial q\lambda}\rho(q,\tau)\right\},$$

where

$$\frac{\partial G}{\partial v}\cdot\frac{\partial\rho(q,\tau)}{\partial v} + \frac{F\,\partial\rho(q,\tau)}{\partial v} = \left(\frac{\partial G}{\partial v} + F\right)\frac{\partial\rho}{\partial v} = 0,$$

which can be verified immediately from the identity

$$\frac{d}{d\tau}E\,[f(x(\tau),v)] = E\,[Lf(x(\tau),v)] = E\,[Rf(x(\tau),v)]$$

in regard to the stochastic differential equation.

Let us introduce the current four-velocity field $u(q,v,\tau)$ and the osmotic four-velocity field $u(q,v,\tau)$ as, respectively,

$$v(q,v,\tau) = \tfrac{1}{2}\{b\,(q,v,\tau) + b_*(q,v,\tau)\},$$
$$u(q,v,\tau) = \tfrac{1}{2}\{b\,(q,v,\tau) - b_*(q,v,\tau)\},$$

where

$$b = (b^0,b^i),\quad i = 1,2,3$$

and

$$b^* = (b_0^*,b^{i*}),\quad b_0^* = -ic,\quad i = 1,2,3.$$

Then the equation of continuity becomes

$$\frac{\partial}{\partial\tau}\rho(w,\tau) + \frac{\partial}{\partial q\mu}\{v^{\mu}(q,v,\tau)\rho(q,\tau)\} = 0.$$

Now, considering the systematic four momentum and stochastic four momentum as p_μ and ϕ_μ, respectively, we can write

$$mD_c = p_\mu\left(\partial_\mu - \frac{\partial G^l}{\partial v^\mu}\frac{\partial}{\partial v^l} + F\frac{\partial}{\partial v^l}\right),$$

and

$$mD_s = \phi_\mu\left(\partial_\mu - \frac{\partial G^l}{\partial v^\mu}\frac{\partial}{\partial v^l}F\frac{\partial}{\partial v^l}\right),$$
$$+\frac{mD}{\sqrt{-g}}\left(\partial_\mu - \frac{\partial G^l}{\partial v_\mu}\frac{\partial}{\partial v^l} + F\frac{\partial}{\partial v^l}\right),$$
$$\times\sqrt{-g}g\,\mu\lambda\left(\partial_\lambda - \frac{\partial G^k}{\partial v_\lambda}\frac{\partial}{\partial v^k} + F\frac{\partial}{\partial v^k}\right),$$

where μ takes the values 1 to 4 with

$$q_\mu = (\mathbf{q},ict),\quad v_\mu = (v_0,\mathbf{v}).$$

D_c and D_s are the systematic and stochastic derivative operators, respectively. Here D_c and D_s are written in terms of c.m. coordinates and the directional variables v_μ. Here D_c and D_s the covariant differential operator $(\partial/\partial q_\mu) - (\partial G^l/\partial v^\mu)(\partial/\partial v^l)$ and $F(\partial/\partial v^\mu)$ are used. F is the metric fundamental function defined by $ds = F(q,dq)$ and G is defined as

$$2G(q,v) = \sigma_{hk}(q,v)v^hv^k.$$

B. Diffusion coefficient

Hakim has shown that, if the limit $\Delta t\to 0$ is used to calculate conditional probability densities, the only value for the diffusion constant D compatible with relativistic invariance is zero. To circumvent this difficult, if we discretize the time variable in the stochastic description, so that in the sequence of events which define trajectories, adjacent events have nonzero minimum temporal separation τ_0. In fact, in the microlocal structure the space–time is assumed to be Finsler space. Again it is well known in Finsler geometry that isotropic Finsler space is equivalent to the Reimann space of constant curvature and this constant curvature can be related to the fundamental length l_0 of the quantized space. Then taking the quantized nature of space–time, we have

$$\langle(dx)^2\rangle \simeq 2D\,dt$$

or

$$D = \frac{1}{2}\frac{\langle(dx)^2\rangle}{dt} = \lim_{\Delta t\to\tau_0}\frac{1}{2}\frac{\langle(dx)^2\rangle}{\Delta t}$$
$$= \frac{1}{2}c^2\tau_0 \text{ as } |\Delta x_i|/\tau_0 = C,\quad \forall\Delta x_i,$$
$$\dot{u},D = \frac{1}{2}c^2\frac{l_0}{c} = \frac{\hbar}{2m},$$

as $\tau_0 \sim (\hbar/mc)$, and m is the mass of the leptons (like electrons or muons).

So the diffusion coefficient in the relativistic limit becomes

$$D = \hbar/2m.$$

It is interesting to note that the finiteness of the diffusion coefficient is closely connected to the discrete nature of space–time.

C. Wave equations

Now the equation of motion may be written as [9]

$$D_cp_\mu - D_s\phi_\mu = f_{0\mu}^{(+)},$$

and

$$D_sp_\mu + D_c\phi_\mu = f_{0\mu}^{(-)},$$

where $f_{0\mu}^{(+)}$ and $f_{0\mu}^{(-)}$ stand for the external force acting on the particle, the plus or minus sign referring to its behavior under time reversal. To all order, the results for the electromagnetic case are

$$f_{0\mu}^{(+)} = \frac{e}{mc}F_{\mu\lambda}p_\lambda - \frac{e}{c}D_c[A_\mu + \partial_\lambda] + \frac{e}{mc}p_\lambda\partial_\lambda A_\mu,$$

where

$$\partial_\lambda = -\frac{\partial}{\partial v^l}\frac{\partial G^l}{\partial v_\lambda} + F\frac{\partial}{\partial v_\lambda}$$

and

$$f_{0\mu}^{(-)} = \frac{e}{mc}F_{\mu\lambda}\phi_\lambda - \frac{e}{c}D_c[A_\mu + \partial_\lambda] + \frac{e}{mc}q_\lambda\partial_\lambda A_\mu,$$

where $F_{\mu\lambda}$ stands for the electromagnetic tensor. Since the operators D_c and D_x have been written above only to second order, we can write the above equation to the same order, thus getting

$$f_{0\mu}^{(+)} = \frac{e}{mc}F_{\mu\lambda}p_\lambda \text{ and } f_{0\mu}^{(-)} = \frac{e}{mc}F_{\mu\lambda}\phi_\lambda,$$

where the gauge $\partial_\mu A_\mu = 0$ has been considered.

Now consider the Brownian particles as some sort of spinning rigid body in Finsler space and write down the derivatives in terms of relativistic Eulerian angles. Here the directional variables can be also expressed in terms of a set of angular coordinates considering the projections of the vector (v^λ) in the normal to the surface of the rigid body, the meridian, and the parallel which depend on the time and location of the body. Hence, taking into account the spin of the particle we write, as our fundamental systems of relativistic equations, the following set:

$$D_c p_\mu - D_s\phi_\mu = \frac{e}{mc}F_{\mu\lambda}p_\lambda + \left(\frac{ge}{4mc}\right)\partial_\mu S^c_{\rho\lambda}F_{\rho\lambda},$$

$$D_c\phi_\mu + D_s p_\mu = \frac{e}{mc}F_{\mu\lambda}\phi_\lambda + \left(\frac{eD}{c}\right)\partial_\lambda F_{\mu\lambda} + \left(\frac{ge}{4mc}\right)\partial_\mu S^s_{\rho\lambda}F_{\rho\lambda},$$

where S^c_λ and S^s_λ are the stochastic and systematic components of spin angular momentum, respectively.

By introducing the complex variables

$$P_\mu = p_\mu - i\phi_\lambda,$$

$$S^x_{\mu\lambda} = S^c_{\mu\lambda} - iS^s_{\mu\lambda},$$

$$f^x_\mu = f_{0\mu}^{(+)} - if_{0\mu}^{(-)} + \left(\frac{ge}{4mc}\right)\partial_\mu S^x_{\rho\lambda}F_{\rho\lambda},$$

we can write the above equations in the form

$$D_x P_\mu = f^x_\mu$$

and it complex conjugate. To integrate the above equations note the relation

$$[mD_x,\partial_\lambda] = -(\partial_x p_\mu)\partial_\lambda$$

and then write it in the form

$$p_\mu = \hbar\left[\frac{\partial}{\partial q^\mu}s(q,v,\tau) + \frac{\partial}{\partial v^\mu}s(q,v,\tau)\right] - eA^{(q)},$$

i.e., under the assumption that the form

$$\omega_2 = \langle MV_+A,dq\rangle_1 + \langle Is + B,dq\rangle_2$$

is closed, where I is the moment of inertia, and s is the spin; there exists a function

$S{:}F\to R$ such that $\omega_2 = d(\hbar S)$.

It is worth mentioning the Caubet has assumed the closeness of one form ω_2 and so exact as well, in analogy with the Jacobi equation. However, it is well known that, according to Bohr–Sommerfield quantization, the one form $\omega = p\,dq - H\,dt$ is not exact, i.e., there exists some topological constraint such that it can not be shrunk to a point. Here also there exists a space–time quantized domain such that $\oint\omega_2 = n\hbar$, where $\hbar$ is the quantum of action related to the fundamental length, so that the space–time may be thought of as a multiply connected region of stochastic fields with period $\hbar$ embedded in a smiply connected space–time.

Again the form

$$\omega_1 = \langle M\delta v + A,dq\rangle_1 + \langle I\delta v + B,qq\rangle_2,$$

which $\delta V = 2^{-1}(b - b_*)$ is closed by continuity of the process but not necessarily exact. If we define the scalar function by the relation

$$R(q,\tau) = \tfrac{1}{2}\ln P(q,\tau)$$

then, we have

$$\omega_1 = d(\hbar R).$$

Now putting the relativistic probability amplitude of the particle as

$$\psi(q,v,\tau) = \exp\{R(q,\tau) + is(q,v,\tau)\},$$

where

$$\bar\psi(q,v,\tau)\psi(q,v,\tau) = p(q,\tau),$$

we get the equations of motion as

$$\left[-i\hbar\left\{\partial_\mu - \frac{\partial G^l}{\partial v_\mu}\frac{\partial}{\partial v^l} + F\frac{\partial}{\partial v^l}\right\} - \frac{e}{c}A_\mu\right]$$

$$g_{\mu\lambda}\left[-i\hbar\left\{\partial_\lambda - \frac{\partial G^k}{\partial v_\lambda}\frac{\partial}{\partial v^\mu} + F\frac{\partial}{\partial v^\mu}\right\} - \frac{e}{c}A_\lambda\right]\psi + m^2c^2\psi = \left(\frac{ge}{2c}\right)s^2_{\mu\lambda}F_{\mu\lambda}\psi.$$

Again we know that in the case of a nonrelativistic spinning rigid body, the spin operator and its components satisfy the usual commutation relations

$$[\hat s_i,\hat s_j] = i\hbar\epsilon_{ijk}\hat s_k \quad\text{and}\quad [\hat s^2,\hat s_\mu] = 0,$$

i.e., we may construct simultaneous eigenfunctions of $\hat s^2$, $\hat s_3$, $\hat s_3'$, $3'$ referring to the body Z axis with eigenvalues $\hbar^v$, $(i+1)$, $\hbar m$, and $\hbar k$, respectively, where $j = 0, \frac{1}{2}, i, \ldots$ and $m,k = -i, -i+1\ldots$ for any given j.

Now like the covariant derivative

$$D_\mu = \partial_\mu - igA^i_\mu J^i,$$

here we have

$$\bar D_\mu = \partial_\mu - igA^i_\mu(J^i + k^i),$$

where k^i is a set of differential operator in $\partial/\partial v_i$ which satisfy the commutation relations of the

$$[k^i,k^j] = ic^{ijk}k^k,$$

and since J^i is v independent, $[k^i,J^j] = 0$. However, while the orbital angular momentum can have only half of the representation of the rotation group, i.e., integral spin only, k^i have in general all the representation of G.

Then if we start with a local symmetry group G, say $U(1)$ with fields $\psi(q)$ then, generalizing to the case were we have ν-dependent fields $\psi(q,\nu)$, and enlarged group $G\otimes G'$ results with $G'\simeq G$. The gauge parameters for G and G' are equal and, in other words, we have universality. Now, in our case, ψ is a function of q and ν; and if $\rho(q,\nu)=\bar{\psi}(q,\nu)\psi(q,\nu)$, i.e., if the additional gauge-dependent variable ν occurs in ρ, then we have the enlarged group but of course the concept of probability has to be replaced by a more general concept, namely, that of a gauge-dependent function $\rho(q,\nu)$. Then the enlarged group can be thought of as due to the manifestation of the anisotropy of the space–time where the particles and antiparticles may be considered as the mirror reflections of each other. Now under certain approximations that $A_0=0$ and hence $\partial_t\mathbf{A}=0$ and write

$$s^{x}_{\mu\lambda}F_{\mu\lambda}\simeq 2\mathbf{s}\times\mathbf{H},$$

we have

$$\Big(-i\hbar\partial_\mu - i\hbar F\frac{\partial}{\partial v^l} - i\hbar\frac{\partial e^l}{\partial v_\mu}\frac{\partial}{\partial v^l} - \frac{e}{c}A_\mu\Big)^2\phi + m^2c^2\phi = \Big(\frac{ge\hbar}{2c}\Big)\sigma\cdot\mathbf{H}\phi,$$

for $s=\frac{1}{2}$ and $g=\frac{1}{2}$ with

$$\phi=\begin{pmatrix}\psi_+\\ \psi_-\end{pmatrix}.$$

A known algebraic transformation may be used to cast the second order equation for a two component amplitude into a first order equation for a four-component amplitude. In face, defining ϕ and x by

$$\Phi=\frac{1}{2}(\phi-x),$$

$$\Big[i\hbar\partial_0-\sigma\cdot\Big(ih\nabla+i\hbar\nabla+\frac{e}{c}\mathbf{A}\Big)\Big]\phi=\tfrac{1}{2}mc(\phi+x),$$

we have

$$-i\hbar\partial_0\phi-\sigma\cdot\Big(i\hbar\nabla+i\hbar\nabla+\frac{e}{c}\mathbf{A}\Big)x=-mc\phi$$

and

$$\sigma\Big(i\hbar\nabla+i\hbar\nabla+\frac{e}{c}\mathbf{A}\Big)\phi=i\hbar\partial_0x.$$

Hence, if we introduce the matrices

$$\gamma_k=\begin{pmatrix}0 & -i\sigma_k\\ i\sigma_k & 0\end{pmatrix},\quad \gamma_4=\begin{pmatrix}I & 0\\ 0 & I\end{pmatrix},$$

$$\Sigma_k=\begin{pmatrix}\sigma_k & 0\\ 0 & \sigma_k\end{pmatrix},\quad \gamma_5=-\begin{pmatrix}0 & I\\ I & 0\end{pmatrix},$$

and the four-component amplitude $\psi=\binom{\phi}{\chi}$. Then we have the Dirac equation for electron or muon as

$$i\gamma_\mu\Big\{-i\hbar\partial_\mu-i\hbar\Big(F\frac{\partial}{\partial v^l}+\frac{\partial G^l}{\partial v_\mu}\frac{\partial}{\partial v^l}\Big)-\frac{e}{c}A_\mu\Big\}\psi + mc\psi=0.$$

Now in the Lorentz spaces

$$\psi=\psi(q)\psi(\nu),$$

so that the above equation can be reduced to the ordinary Dirac equation

$$i\gamma_\mu\Big(-i\hbar\partial_\mu-\frac{e}{c}A_\mu\Big)\psi(q)+mc\psi(q)=0$$

in the external space if

$$i\gamma_\mu\Big[-i\hbar F\frac{\partial}{\partial v^l}-i\hbar\frac{\partial G^l}{\partial v_\mu}\frac{\partial}{\partial v^l}\Big]\psi(\nu)=0$$

for internal variables.

IV. GLOBAL CONCEPT OF QUANTIZATION

The single valuedness and square integrability of the wave function ψ suffices for the global connotation of quantization. However, the stochastic derivation of the relativistic wave equation indicates the local validity of R and S as ω_1 and ω_2 are not necessarily exact, so that according to De Rham's theorem there does not exist any zero form, i.e., scalar functions S and R over the whole space–time. So the stochastic fields through the properties with which they are endowed many well define multiply connected manifolds of integration embedded or immersed in the simply connected space–time in the microphysical domain. It is interesting to note that Mandelstam and more recently Wu and Yang [10] studied, for example, the electromagnetic effects on matter which are described by a quantum mechanical wave function $\psi(n)$ as being equivalent to the presence of a path dependant phase factor $s(Cyn)=\exp[(-ie/\hbar c)\int_n^{y}A_\mu(n)d_n\mu]$ associated with the path Cyn appearing at the level of the wave function. Actually, the famous Bohm–Aharonov experiment showed that, in a multiply connected region where $f_{\mu\nu}=0$ (field strength), everywhere there are physical experiments for which the outcome depends on the loop integral

$$\frac{e}{\hbar c}\oint A_n d_n\mu$$

around an unshrinkable loop. Again an examination of the Bohm–Aharanov experiment indicates that, in fact, only the phase factor

$$\exp\Big(\frac{ie}{\hbar c}\oint A_\mu d_n\,\mu\Big)$$

and not the phase $(e/\hbar c)A_\mu d_n\,\mu$ is physically meaningful. In other words, the phase contains more information than the phase factor. However, the additional information is not measurable. They describe this as the nonintegrable (i.e., path-dependent) phase factor. This phase factor is the path-dependent element of the group $U(1)$, being the relevant group related in the quantum mechanical description of atomic phenomena to the concept of a probability density $\rho(n)=\psi^*(n)\psi(n)$ derived from the complex wave function $\psi(n)$. The structural group $U(1)$ of the bundle in Feynman's treatment, being identical with the associated electromagnetic gauge group $U(1)$ in Mandelstam's treatment, implies the operator of a covariant derivative on the bundle to be given by $D_\mu = \partial^\mu + i\,eA_\mu(n)$, with $A_\mu(n)$ denoting the electromagnetic potentials.

With this situation in mind let us consider the generalized wave function $\psi(n,v)$, depending on n and directional variable v, which is a cross section on a fiber bundle (considering the diffusing particles as bundles of timelike fibers in the quantized domain) constructed over space–time possessing a gauge or structural group more general than the group $U(1)$. Clearly, $\psi(n,v)$ will be at each point $n \in M4$ a representation of $\widetilde{G}$. Again it was shown that the enlarged group $\widetilde{G} \simeq U(1) \otimes U(1)$, i.e., if we start with a local symmetry group $U(1)$, we have higher order unitary group. Now, in the Finsler space, a bundle of carton type constructed over a flat or a curved space–time has a structural group which is necessarily noncompact. Correspondingly, it will be a nonunitary representation of G. We thus come to the conclusion that leaving the description of relativistic stochastic quantization in Finsler space on a wave function associated with a generalized (possibly noncompact) non-Abelian gauge group instead of the group $U(1)$ of quantum mechanics will replace at the same time the concept of probability by a more general concept, namely, that of a gauge-dependent scalar density function.

V. DISCUSSION

The stochastic interpretation of quantum mechanics emphasizes that the fundamental processes of nature are stochastic processes defined in a multiply connected space–time where the quantum of action ($\hbar$) is the one-dimensional period of one form ω_2. So whereas in the Copenhagen interpretation Newtonian equations are obtained in the limit $h \to 0$, here it corresponds to a stochastic process with zero stochastic force in a simply connected space–time where the period of ω_2 vanishes so that ω_2 can be written as the gradient of a scalar function, everywhere in the space–time.

Again the relativistic concept of stochastic quantization may be related to the concept of local anisotropy of the space–time so that the probability concept of quantum mechanics has to be replaced by a more general gauge dependant scalar function $\rho(n,v)$ in the microdomain of space–time. Then it raises a new possibility for v with the directional variables as hidden variables, which will be considered in the subsequent paper.

[1]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[2]F. Guerra and P. Ruggiero, Phys. Rev. Lett. **31**, 1022 (1973).
[3]J.P. Caubet, *Lecture notes in Mathematics*, edited by A. Dold and B. Eckmann (Springer, New York, 1975), No. 451, p. 113.
[4]R. Hakim, J. Math. Phys. **9**, 1905 (1968).
[5]Kunio Yasue, Prog. Theor. Phys. **57**, 318 (1977).
[6]P. Bandyopadhyay, Int. J. Theor. Phys. **10**, 175 (1974).
[7]P. Bandyopadhyay, Int. J. Theor. Phys. **8**, 323 (1973).
[8]P. Bandyopadhyay, and S. Roy, Int. J. Theor. Phys. **15**, 323 (1976).
[9]L. De la Pena-Aucrbach, J. Math. Phys. **12**, 452 (1971).
[10]Tai Tsun Wu and C.N. Yang, Phys. Rev. D **12**, 3845 (1975).

Generalized Stern–Gerlach experiments and the observability of arbitrary spin operators

Arthur R. Swift[a)]
Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01003
Ron Wright
Department of Mathematics, Kansas State University, Manhattan, Kansas 66506

(Received 6 March 1979; accepted for publication 8 June 1979)

Under the assumption that we can create in the laboratory any electromagnetic field consistent with Maxwell's equations, it is shown that an arbitrary Hermitian operator on a spin system can be measured using a suitable generalization of the Stern–Gerlach experiment. In particular, it is shown that every proposition about the spin-1 system is verifiable, answering the challenge of Hultgren and Shimony. The analysis also reveals complications in the standard Stern–Gerlach experiment of which many physicists are apparently not aware.

I. INTRODUCTION

The question of which operators in the Hilbert space representing a quantum mechanical system, actually correspond to quantities we can observe and, as a consequence, which projections actually correspond to propositions we can test, has been of perennial concern to those of us interested in the foundations of quantum mechanics. In the early stages of the development of the thoery it was generally assumed that every Hermitian operator (with perhaps some technical restrictions in the infinite-dimensional case) was observable. As late as 1958 we find the following from Dirac[1]:

> In practice it may be very awkward or perhaps even beyond the ingenuity of the experimenter, to devise an apparatus which could measure some particular operator, but the theory always allows one to imagine that the measurement can be made.

Nevertheless, with the discovery in the 1950's of superselection rules[2,3] it became clear that this assumption, along with the unrestricted superposition principle, was no longer tenable. Although it was still possible to maintain that all Hermitian operators not specifically ruled out by superselection rules were, in fact, observable, the alluring simplicity of the initial theory had been indelibly marred and there seems to have been an increasing uneasiness on the part of many workers. Wigner has given this disquiet its perhaps most eloquent expression[4]:

> We learn and teach, respectively, in courses on quantum mechanics that the measurable quantities, or in the words of Dirac, the observables, are hermitian operators. It can indeed be proved by means of the theory of measurement, that only hermitian operators can represent measurable quantities. Some books, and some lecturers, go further and claim that *all* hermitian (or more precisely, all self-adjoint) operators can be observables. However, if we ask how the measurement of a given self-adjoint operator should be carried out, the books and lecturers remain most secretive. One has, of course, no idea how a quantity such as $p + q$ or $pq + qp$ or pqp could be measured—in fact, clearly, most operators cannot. Still, many can···.
>
> There is, however, no rule which would tell us which self-adjoint operators are truly observables, nor is there any prescription known how the measurements are to be carried out, what apparatus to use, etc. In a theory with a positivistic undertone, this is a serious gap.

It is the main purpose of this paper to give, for spin systems, a rule for telling which self-adjoint operators are truly measurable—namely they all are—and to provide a prescription for how the experiments are to be carried out.

An intimately related question has recently been raised by Hultgren and Shimony,[5] with specific reference to spin-1 systems. Since the spectrum of a projection is contained in the set $\{0,1\}$, the projection can be thought of as a yes–no question, or proposition, about the system.[5,6,7] A projection represents a verifiable proposition exactly when the projection, as a Hermitian operator, is measurable. Hultgren and Shimony have proposed that the only verifiable propositions about a spin-1 system are the ones in the collection L_v of propositions corresponding to the standard Stern–Gerlach experiments, and have put forth a challenge to demonstrate the verifiability of the remaining propositions:

> If the standard formulation of quantum mechanics is applied to the spin-1 system, there is a projection operator corresponding to each proposition in L_v, but there are also projection operators which do not correspond to members of L_v and which in fact do not seem to correspond in any natural way to testable propositions...it is a good working hypothesis that those opertors corresponding to members of L_v have definite physical status which the others do not have (p. 381).
>
> ···An obligation is placed on the advocates of such programs (to recover the Hilbert space axioms via empirically justifiable axioms) to exhibit

a)Supported in part by the National Science Foundation

 0022-2488/80/010077-06$01.00

> the physical significance of propositions concerning the spin-1 system which are not elements of L_v ···(p. 390).

By showing that every Hermitian operator is observable, the "obligation" of Hultgren and Shimony is clearly fulfilled.

A third important and often-asked question is: Which rays represent realizable states of the system? It is shown below that for spin systems every ray does in fact represent a physically realizable state of the system.

II. THE OPERATIONAL MEANING OF THE QUESTIONS

In the last section we raised three questions which can be asked about the Hilbert space representing a quantum mechanical system. In one way or another each question asks which mathematical objects actually correspond to physical entities. The three questions were:

I. Which Hermitian operators are actually measurable (and how do we measure them)?

II. Which projections represent verifiable propositions?

III. Which rays represent physically realizable states?

Before attempting to answer any of the above questions for spin systems, as we propose to do, we wish to be very explicit about what we understand the questions to mean. We begin with question I.

If A is a Hermitian operator acting on a complex Hilbert space of dimension $n < \infty$, then the spectral theorem tells us that there is a unique set $\{a_1, a_2, ..., a_r\}$ of $r \leqslant n$ distinct real numbers (called the *eigenvalues* of A) and a unique set $\{P_1, P_2, ..., P_r\}$ of mutually orthogonal, nonzero projections (called the *eigenprojections* of A), such that $A = \Sigma_{i=1}^{r} a_i P_i$. The projection P_i is the projection onto the subspace of all eigenvectors of A corresponding to the eigenvalue a_i; if the eigenvalue a_i is degenerate, then the dimension of the projection P_i will be strictly greater than one.

We say that the Hermitian operator A, with spectral decomposition $A = \Sigma_{i=1}^{r} a_i P_i$, is *measurable* or *observable* exactly when there exists a physical operation E with outcomes $e_1, e_2, ..., e_r$, such that when the system is in the state represented by a density operator ρ, the outcome e_i is obtained with probability $\mathrm{Tr}\,\rho P_i$.[8] The term "physical operation" is to be understood in the sense of Randall and Foulis[9]:

> By a physical operation, we shall mean instructions that describe a well-defined, physically realizable, reproducible procedure and furthermore, that specify what must be observed and recorded. In particular a physical operation must require that, as a consquence of each execution of the instructions, one and only one symbol from a specified set R be recorded as the *result* of that realization of the operation.

In regards to the second question we have already said that a projection represents a *verifiable* and *testable* proposition if and only if the projection, as a Hermitian operator, is observable in the sense just discussed. It is clear that if every Hermitian operator is observable, then every projection represents a verifiable proposition. Without further assumptions (e.g., postulate C of Kharatyan[10] or axiom (iv) of Foulis and Randall[9]) the converse statement is false.

The final question is probably the most often misunderstood of the three. Kharatyan[10] and Streater and Wightman,[3] for example, identify it with question II. To us, a ray represents a *(physically) realizable* state if systems can be found such that the theoretical probabilities of experimental outcomes, or, equivalently the expectation values of various observables, as calculated from the ray in the usual way, agree with the observed values. There is no particular reason to believe that any physically realizable state has an "indicator outcome" which occurs with probability 1 when the system is in that state. Although it does not seem logically necessary, the only indisputable way which occurs to us for demonstrating that a ray corresponds to a physically realizable state is to give a repeatable procedure whereby systems of the proper type are prepared. It is in this way that we show below that for spin systems every ray corresponds to a physically realizable state.

We point out that all of these definitions require us to measure probabilities. From the standpoint of statistics this might seem like a nonoperational definition, but from the standpoint of physics, where measuring probabilities is equivalent to measuring intensities, this definition is sufficiently operational. For our purposes here, however, we can use the established principles of quantum mechanics to *prove* that our experimental outcomes will have the required probabilities.

In closing this section we note that we will have answered all three questions for spin systems in the advertised way—to wit, that every Hermitian operator is measurable, that every projection represents a physically verifiable proposition, and that every ray represents a physically realizable state—if given a Hermitian operator A in spin space, we can describe a physical apparatus by which particles in eigenstates of A corresponding to distinct eigenvalues are separated physically. By placing detectors judiciously and sending a single particle into the apparatus, we have a physical operation whose outcomes correspond to detection in the various counters; according to quantum theory these outcomes then occur with exactly the prescribed probabilities. Furthermore, given a ray, the apparatus corresponding to the projection onto that ray will produce a beam of particles in the desired state.

III. THE STANDARD STERN–GERLACH EXPERIMENT

In order to make the discussion in the next section of the generalized Stern–Gerlach experiments easier to understand, we first review quickly the standard Stern–Gerlach experiment. This apparatus, designed during the 1920's by Otto Stern and Walther Gerlach[11] has become the traditional means of demonstrating the quantization of angular momentum. In this experiment a beam of particles is passed through an inhomogeneous magnetic field $\mathbf{B}(r)$ whose direction is constant, but whose magnitude is dependent upon position:

$$\mathbf{B}(\mathbf{r}) = \lambda(\mathbf{r})\mathbf{B}_0 . \quad (1)$$

An essentially correct understanding of the experiment can be obtained by reasoning classically as follows: The magnetic energy associated with a particle with magnetic dipole moment $\boldsymbol{\mu}$ in the magnetic field $\mathbf{B}$ is

$$E_{\mathrm{mag}}(\mathbf{r}) = -\boldsymbol{\mu}\cdot\mathbf{B}(\mathbf{r}) = -\lambda(\mathbf{r})\mu_{B_0}B_0 ,$$

where μ_{B_0} is the component of the magnetic moment in the $\mathbf{B}_0$ direction and $B_0 = |\mathbf{B}_0|$; the particle experiences a force

$$\mathbf{F}(\mathbf{r}) = -\nabla E_{\mathrm{mag}}(\mathbf{r}) = +\mu_{B_0}B_0\nabla\lambda(\mathbf{r}).$$

Consequently particles with differing dipole components experience different forces and are separated spatially. Quantum mechanics enters the picture here only in that the magnetic dipole moment is proportional to the spin and the values of the spin component are found to be quantized, rather than continuously distributed, as the classical reasoning would lead us to predict.

The Stern–Gerlach experiments provide a complete set of experiments for spin-$\frac{1}{2}$ systems in the sense that for any Hermitian operator on a spin-$\frac{1}{2}$ system, there is a Stern–Gerlach experiment which measures it. For suppose that A is a Hermitian operator on the two-dimensional Hilbert space describing a spin-$\frac{1}{2}$ particle. The measurement of the spin component of the particle in the x_i direction is represented by an operator S_i on this Hilbert space. (We use units in which $\hbar = 1$.) Since the identity operator I together with the three spin operators S_1, S_2, and S_3 form a basis for the four-dimensional real vector space of all Hermitian operators, the given operator A can be uniquely written in the form

$$A = aI + b_iS_i , \tag{2}$$

where a and b_i are real numbers. The summation convention for repeated indices is used here and elsewhere in this paper.

Let μ_i be the operator corresponding to the measurement of the component of the magnetic moment in the x_i direction. By the Wigner–Eckart theorem, (μ_1,μ_2,μ_3) is proportional to (S_1,S_2,S_3) and the constant of proportionality is the total magnetic moment of the particle $(\mu_1,\mu_2,\mu_3) = \mu_0(S_1,S_2,S_3)$. In a magnetic field $\mathbf{B} = (B_1,B_2,B_3)$, the behavior of a spin-$\frac{1}{2}$ particle is described by the Hamiltonian

$$H = H_0 + \mu_0 B_iS_i ,$$

where the spin independent term H_0—which as an operator in spin space is just a scalar multiple of the identity—contains the kinetic energy. If the vector $\mathbf{B}_0$ in (1) is chosen parallel to the vector $\mathbf{b} = (b_1,b_2,b_3)$ in (2), then the eigenstates of H will be the same as those of A. With the magnetic field $\mathbf{B}(\mathbf{r}) = \lambda(\mathbf{r})\mathbf{B}_0$, the two eigenstates have energies

$$E(\mathbf{r}) = E_0 \pm \tfrac{1}{2}\mu_0 B_0\lambda(\mathbf{r}),$$

in the approximation that the spatial dependence of $\lambda(\mathbf{r})$ is constant over the dimensions of the particle. The force on the particle is then

$$\mathbf{F}(\mathbf{r}) = -\nabla E(\mathbf{r}) = \mp\tfrac{1}{2}\mu_0 B_0 \nabla\lambda(\mathbf{r}).$$

As a result, particles in different eigenstates experience different forces and are separated physically.

The question still remains as to whether there exists a function $\lambda(\mathbf{r})$ with a nonzero gradient, such that the magnetic field $\mathbf{B}(\mathbf{r}) = \lambda(\mathbf{r})\mathbf{B}_0$ satisfies Maxwell's equations for a static field, i.e., $\nabla\cdot\mathbf{B} = \nabla\times\mathbf{B} = 0$. The most naive choice for $\lambda(\mathbf{r})$ might be $\lambda(\mathbf{r}) = 1 + \mathbf{k}\cdot\mathbf{r}$ for $\mathbf{k}\neq 0$, since this implies $\nabla\lambda(\mathbf{r}) = \mathbf{k}\neq 0$. It turns out, however, that neither this, nor any other chocie for $\lambda(\mathbf{r})$ satisfies Maxwell's equations, because $\mathbf{B}(\mathbf{r}) = \lambda(\mathbf{r})\mathbf{B}_0$ implies that $0 = \nabla\times\mathbf{B} = \nabla\lambda(\mathbf{r})\times\mathbf{B}_0$ and $0 = \nabla\cdot\mathbf{B} = \nabla\lambda(\mathbf{r})\cdot\mathbf{B}_0$, whence $\nabla\lambda(\mathbf{r}) = 0$.

This incompatibility can be resolved by adding another term to the magnetic field and applying first-order perturbation theory. For example, let

$$\mathbf{B}(\mathbf{r}) = \frac{1+\mathbf{k}\cdot\mathbf{r}}{\mu_0}\mathbf{b} + \frac{\mathbf{b}\cdot\mathbf{r}}{\mu_0}\mathbf{k},$$

where $\mathbf{k}$ is any vector perpendicular to $\mathbf{b}$. In this case

$$\nabla\cdot\mathbf{B}(\mathbf{r}) = (2/\mu_0)\mathbf{k}\cdot\mathbf{b} = 0$$

and

$$\nabla\times\mathbf{B}(\mathbf{r})\equiv 0.$$

Maxwell's equations are satisfied. The Hamiltonian is

$$H = H_0 + (1+\mathbf{k}\cdot\mathbf{r})b_iS_i + (\mathbf{b}\cdot\mathbf{r})k_iS_i .$$

The term, $(1+\mathbf{k}\cdot\mathbf{r})b_iS_i$, has exactly the same eigenvectors as the operator A which we desire to measure. The first-order corrections to the energy due to the perturbation term $(\mathbf{b}\cdot\mathbf{r})k_iS_i$ [12] are of the form

$$(\mathbf{b}\cdot\mathbf{r})\langle\psi|k_iS_i|\psi\rangle ,$$

where ψ is an eigenvector for $(1+\mathbf{k}\cdot\mathbf{r})b_iS_i$, or equivalently, for A. It follows from the commutation relations $[S_i,S_j] = i\epsilon_{ijk}S_k$ and the orthogonality of $\mathbf{k}$ and $\mathbf{b}$, that the diagonal matrix elements $\langle\psi|k_iS_i|\psi\rangle$ are all zero. The energies to first order in are

$$L = \pm(1+\mathbf{k}\cdot\mathbf{r})(b/2),$$

and the forces on the particle in the two different eigenstates of A are

$$\mathbf{F}(\mathbf{r}) = -\nabla E(\mathbf{r}) = \mp(b/2)\mathbf{k}.$$

if $\nabla|\mathbf{B}|/|\mathbf{B}|$ is made sufficiently small, the second-order effects can be safely ignored, and the two states are separated spatially.

In the next section we discuss a generalized Stern–Gerlach experiment. As was the case here, we will express a given operator A in a standard basis and then show that by appropriate choice of electromagnetic fields the Hamiltonian can be made to have the same first-order eigenstates as A. In contrast to this section, the accordance with Maxwell's equations is made from the beginning by starting with scalar potentials satisfying Laplace's equation.

IV. GENERALIZED STERN–GERLACH EXPERIMENTS

Perhaps suprisingly, the standard Stern–Gerlach experiments do not form a complete set of experiments for particles with spin greater than $\frac{1}{2}$. Even for the case of spin-1 particles, there are many states which are not eigenstates of the operator b_iS_i, which measures the spin component in the $\mathbf{b} = (b_1,b_2,b_3)$ direction, for any value of $\mathbf{b}$.[5] (An example of such a state is $[1,1,1]/3^{1/2}$, in the basis of eigenstates of S_3.) Since eigenstates of b_iS_i are the only states that the standard Stern–Gerlach apparatus can filter, it is clear that some modification is necessary if all spin-1 operators are to be observable in this way.

That generalizations of the Stern–Gerlach experiment are possible, has been long suggested by other authors, notably Feynman[13] and more specifically, Lamb[8] who writes: "In some cases inhomogeneous electric fields could be used instead of these inhomogeneous magnetic fields." By using electric fields in the place of magnetic fields we can, in fact, observe certain spin operators not observable via the ordinary Stern–Gerlach experiments. For example, the electric field can interact with the electric quadrupole moment of the particle. (A spin- $\frac{1}{2}$ particle has only a magnetic dipole moment.)

The central result of this paper is that by using both electric and magnetic fields one can observe every spin operator. We turn our attention to the proof of this assertion.

The Hilbert space representing the spin state of a spin-s particle has dimension $2s+1$. As is apparently well known to some physicists, an arbitrary Hermitian operator A on this Hilbert space can be uniquely represented in the form

$$A = \sum_{k=0}^{2s} a^{(k)}{}_{i_1\cdots i_k} T^{(k)}{}_{i_1\cdots i_k}, \tag{3}$$

where $(a^{(k)}{}_{i_1\cdots i_k})$ are the components of a traceless, symmetric tensor of rank, k, $T^{(0)} = I$, $T^{(1)}{}_i = S_i$, and $T_{i_1\cdots i_n}$ is obtained from the product $S_{i_1}S_{i_2}\cdots S_{i_n}$ by symmetrizing and subtracting off the trace. (A tensor is (totally) symmetric if, for every pair of indices, interchanging them leaves the component invariant. A tensor is (totally) traceless, if, for every pair of indices, setting them equal and adding over all possible values for the index, gives a sum of zero.] For example, if $n=2$, $T_{ij} = 1/2(S_iS_j + S_jS_i) - \delta_{ij}S_kS_k/3$. There is again here, as always, an implied sum over the repeated indices. As in the case of spin- $\frac{1}{2}$, we will construct a Hamiltonian with the same eigenstates as the operator A (to first order in the sense of perturbation theory), but whose spatial variation induces different forces on particles in eigenstates corresponding to different eigenvalues. The operators $T^{(k)}{}_{i_1\cdots i_k}$ will be shown to be proportional to the operators measuring the various k-pole electromagnetic moments of the particle.

In order to write down the most general Hamiltonian for a spin-s particle in an electromagnetic field, we first develop an expression in terms of multipole moments, for the classical energy of an extended particle in an electromagnetic field described by scalar potentials Φ^E and Φ^M for the electric and magnetic fields, respectively. The potentials Φ^E and Φ^M are assumed to be analytic within the chamber of the experiment, and to satisfy Laplace's equation: $\nabla^2\Phi^E(\mathbf{r}) = \nabla^2\Phi^M(\mathbf{r}) = 0$. When the electric and magnetic fields are determined from the potentials via $\mathbf{E}(\mathbf{r}) = -\nabla\Phi^E(\mathbf{r})$ and $\mathbf{B}(\mathbf{r}) = -\nabla\Phi^M(\mathbf{r})$, Maxwell's equations for static fields are automatically satisfied. The classical energy for a particle with electric charge density ρ^E and magnetic charge density ρ^M is given by

$$W = \int_C (\rho^E(\mathbf{r})\Phi^E(\mathbf{r}) + \rho^M(\mathbf{r})\Phi^M(\mathbf{r}))\, d^3r, \tag{4}$$

where the coordinate $\mathbf{r}$ is the displacement from the center of the chamber and C is the volume occupied by the chamber. [Since there are, as best we know, no magnetic monopoles, the magnetic charge density is understood to be the negative of the divergence of the magnetic dipole density $\mathbf{M}(\mathbf{r}) : \rho^M(\mathbf{r}) = -\nabla\cdot\mathbf{M}(\mathbf{r})$.]

Suppose that the center of the particle is located at $\mathbf{z}$, near the center of the chamber. Relative to the center of the particle, an arbitrary point in the chamber has coordinate $\mathbf{y} = \mathbf{r} - \mathbf{z}$. We first expand Φ^E and Φ^M in a Taylor series about $\mathbf{z}$:

$$\Phi^E(\mathbf{r}) = \sum_{k=0}^{\infty} \frac{1}{k!}\Phi^E{}_{i_1\cdots i_k}(\mathbf{z}) y_{i_1}\cdots y_{i_k}$$

and

$$\Phi^M(\mathbf{r}) = \sum_{k=0}^{\infty} \frac{1}{k!}\Phi^M{}_{i_1\cdots i_k}(\mathbf{z}) y_{i_1}\cdots y_{i_k},$$

where

$$\Phi_{i_1 i_2\cdots i_k}(\mathbf{z}) = \frac{\partial}{\partial x_{i_k}}\frac{\partial}{\partial x_{i_{k-1}}}\cdots\frac{\partial}{\partial x_{i_1}}\Phi(\mathbf{r})\Big|_{\mathbf{r}=\mathbf{z}}.$$

If we set $\bar\rho^E(\mathbf{y}) = \rho^E(\mathbf{r})$ and $\bar\rho^M(\mathbf{y}) = \rho^M(\mathbf{r})$, then

$$\begin{aligned} W &= \int_C \Big[\bar\rho^E(\mathbf{y})\Big(\sum_{k=0}^{\infty}\frac{1}{k!}\Phi^E{}_{i_1\cdots i_k}(\mathbf{z})y_{i_1}\cdots y_{i_k}\Big) \\ &\quad + \bar\rho^M(\mathbf{y})\Big(\sum_{k=0}^{\infty}\frac{1}{k!}\Phi^M{}_{i_1\cdots i_k}(\mathbf{z})y_{i_1}\cdots y_{i_k}\Big)\Big]\, d^3y \\ &= \sum_{k=0}^{\infty}\frac{1}{k!}\Phi^E{}_{i_1\cdots i_k}(\mathbf{z})\int_C \bar\rho^E(\mathbf{y})y_{i_1}\cdots y_{i_k}\, d^3y \\ &\quad + \sum_{k=0}^{\infty}\frac{1}{k!}\Phi^M{}_{i_1\cdots i_k}(\mathbf{z})\int_C \bar\rho^M(\mathbf{y})y_{i_1}\cdots y_{i_k}\, d^3y. \end{aligned}$$

Let

$$q^{(k)}{}_{i_1}\cdots y_{i_k} = \int_C \bar\rho^E(\mathbf{y})y_{i_1}\cdots y_{i_k}\, d^3y,$$

the electric k-pole moment of the particle about its center, and let

$$m^{(k)}{}_{i_1\cdots i_k} = \int_C \bar\rho^M(\mathbf{y})y_{i_1}\cdots y_{i_k}\, d^3y,$$

the magnetic k-pole moment of the particle about its center. Since the k-pole moment tensors $q^{(k)}{}_{i_1\cdots i_k}$ and $m^{(k)}{}_{i_1\cdots i_k}$ are totally contracted with the traceless, symmetric tensors $\Phi^E{}_{i_1\cdots i_k}$ and $\Phi^M{}_{i_1\cdots i_k}$, respectively—the tensors are traceless because the potentials satisfy Laplace's equation—the multipole moment tensors may themselves be considered to be symmetric and traceless. Parity invariance for electromagnetic interactions implies for simple systems where the center of charge coincides with the center of mass that $\bar\rho^E(\mathbf{y}) = \bar\rho^E(-\mathbf{y})$ and $\bar\rho^M(\mathbf{y}) = -\bar\rho^M(-\mathbf{y})$. Hence, $q^{(k)}{}_{i_1\cdots i_k}$ is zero for k odd and $m^{(k)}{}_{i_1\cdots i_k}$ is zero for k even. Thus a particle has a magnetic dipole, octupole,$\cdots 2^{2n+1}$-pole ($n = 0,1,2,3,\cdots$), and an electric monopole, quadrupole, $\cdots 2^{2n+1}$-pole ($n = 0,1,2,\cdots$). Expressed with this new notation, the classical energy is

$$W = \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty}\frac{1}{k!}\Phi^E{}_{i_1\cdots i_k}(\mathbf{z})q^{(k)}{}_{i_1\cdots i_k} + \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty}\frac{1}{k!}\Phi^M{}_{i_1\cdots i_k}(\mathbf{z})m^{(k)}{}_{i_1\cdots i_k}.$$

The correspondence principle tells us that in converting this classical expression for the energy into a quantum mechanical Hamiltonian the measurable quantities $q^{(k)}{}_{i_1\cdots i_k}$ and $m^{(k)}{}_{i_1\cdots i_k}$ are replaced by Hermitian operators $Q^{(k)}{}_{i_1\cdots i_k}$ and $M^{(k)}{}_{i_1\cdots i_k}$. For a spin-$s$ particle the highest possible moment is of order $k=2s$, since for $k>2s$ both $Q^{(k)}{}_{i_1\cdots i_k}$ and $M^{(k)}{}_{i_1\cdots i_k}$ are zero.

A second and deeper aspect of the correspondence principle which is less often stated explicitly (but which is the basis for the importance of tensor operators in the theory of angular momentum), states that the $Q^{(k)}{}_{i_1\cdots i_k}$ and $M^{(k)}{}_{i_1\cdots i_k}$ transform under the rotations of the coordinate system exactly as their classical counterparts do; that is, $Q^{(k)}{}_{i_1\cdots i_k}$ and $M^{(k)}{}_{i_1\cdots i_k}$ are the components of traceless, symmetric, and hence irreducible, Cartesian tensor operators of rank k. This result may also be inferred from the fact that the Hamiltonian (see below) is, of course, invariant under rotations of the coordinate system; since the classical field moments $\Phi^{E}{}_{i_1\cdots i_k}(\mathbf{z})$ and $\Phi^{M}{}_{i_1\cdots i_k}(\mathbf{z})$ are tensors under rotation, the $Q^{(k)}{}_{i_1\cdots i_k}$ and $M^{(k)}{}_{i_1\cdots i_k}$ must transform oppositely in order that the Hamiltonian remain invariant. It follows from the Wigner–Eckart theorem that any two irreducible tensor operators of rank k acting on the Hilbert space of a particle of spin s, are proportional to one another. In the present case this means that there exist, for each allowed value of k, numbers Q_k and M_k such that $Q^{(k)}{}_{i_1\cdots i_k}=Q_k T^{(k)}{}_{i_1\cdots i_k}$ and $M^{(k)}{}_{i_1\cdots i_k}=M_k T^{(k)}{}_{i_1\cdots i_k}$, where the $T^{(k)}{}_{i_1\cdots i_k}$'s are the components of the tensor operator defined at the beginning of this section. The $Q^{(k)}{}_{i_1\cdots i_k}$ and $M^{(k)}{}_{i_1\cdots i_k}$ are the components of the multipole tensor operator and the Q_k's and M_k's are known to physicists as the electric and magnetic multipole moments, respectively. In light of the parity considerations above, $Q_k=0$ for k odd and $M_k=0$ for k even. It is observed physically that for values of $k\leqslant 2s$ which are not ruled out by symmetry arguments, the 2^k-pole moments are all nonzero.

We are now in a position to write down the general Hamiltonian for a spin s particle in terms of the multipole moments:

$$
\begin{aligned}
H(\mathbf{z}) = & \sum_{\substack{k=0\\k\ \text{even}}}^{2s} \frac{1}{k!}\Phi^{E}{}_{i_1\cdots i_k}(\mathbf{z})Q^{(k)}{}_{i_1\cdots i_k} + \sum_{\substack{k=0\\k\ \text{odd}}}^{2s} \frac{1}{k!}\Phi^{M}{}_{i_1\cdots i_k}(\mathbf{z})M^{(k)}{}_{i_1\cdots i_k}\\
= & \sum_{\substack{k=0\\k\ \text{even}}}^{2s} \frac{Q_k}{k!}\Phi^{E}{}_{i_1\cdots i_k}(\mathbf{z})T^{(k)}{}_{i_1\cdots i_k} + \sum_{\substack{k=0\\k\ \text{odd}}}^{2s} \frac{M_k}{k!}\Phi^{M}{}_{i_1\cdots i_k}(\mathbf{z})T^{(k)}{}_{i_1\cdots i_k}.
\end{aligned}
$$

If we define

$$
\Phi^{(k)}{}_{i_1\cdots i_k}(\mathbf{z}) = \begin{cases} (Q_k/k!)\Phi^{E}{}_{i_1\cdots i_k}(\mathbf{z}), & k \text{ even},\\ (M_k/k!)\Phi^{M}{}_{i_1\cdots i_k}(\mathbf{z}), & k \text{ odd},\end{cases}
$$

then $\Phi^{(k)}{}_{i_1 i_2\cdots i_k}(\mathbf{z})$ is a traceless symmetric tensor and

$$
H(\mathbf{z}) = \sum_{k=0}^{2s} \Phi^{(k)}{}_{i_1\cdots i_k}(z)T^{(k)}{}_{i_1\cdots i_k}.
$$

We now expand each of the $\Phi^{(k)}{}_{i_1\cdots i_k}$'s in a MacLauren series obtaining:

$$
\begin{aligned}
H(\mathbf{z}) &= \sum_{k=0}^{2s}\Big(\sum_{n=0}^{\infty}\frac{1}{n!}\Phi^{(k)}{}_{i_1\cdots i_k j_1\cdots j_n}(0)z_{j_1}\cdots z_{j_n}\Big)T^{(k)}{}_{i_1\cdots i_k}\\
&= \sum_{n=0}^{\infty}\frac{1}{n!}z_{j_1}\cdots z_{j_n}\sum_{k=0}^{2s}\Phi^{(k)}{}_{i_1\cdots i_k j_1\cdots j_n}(0)T^{(k)}{}_{i_1\cdots i_k}\\
&= \sum_{k=0}^{2s}\Phi^{(k)}{}_{i_1\cdots i_k}(0)T^{(k)}{}_{i_1\cdots i_k}\\
&\quad + z_j\sum_{k=0}^{2s}\Phi^{(k)}{}_{i_1\cdots i_k j}(0)T^{(k)}{}_{i_1\cdots i_k}\\
&\quad + \text{terms of second and higher order in } z.
\end{aligned}
$$

Returning to the operator $A=\Sigma_{k=0}^{2s}a^{(k)}{}_{i_1\cdots i_k}T^{(k)}{}_{i_1\cdots i_k}$ which is to be measured, we see that by choosing $\Phi^{(k)}{}_{i_1\cdots i_k}(0)=a^{(k)}{}_{i_1\cdots i_k}$ $(0\leqslant k\leqslant 2s)$—which we are free to do since this merely amounts to specifying a finite number of even-order coefficients in the Taylor expansion of the electric potential Φ^{E} and a finite number of odd-order coefficients in the Taylor expansion of the magnetic potential Φ^{M}—we can make the zero-order part of the Hamiltonian equal to A. When the particle is located at the center of the chamber, the energy eigenstates are exactly the eigenstates of A. By choosing $\Phi^{(k)}{}_{i_1\cdots i_k 1}(0)=a^{(k)}{}_{i_1\cdots i_k}$ $(0\leqslant k\leqslant 2s)$—which we are also free to do, independently of how $\Phi^{(k)}{}_{i_1\cdots i_k}$ was chosen, since this amounts to specifying a finite number of *odd-order* coefficients for Φ^{E} and *even-order* coefficients for Φ^{M}—we can make the z_1 "coefficient" in H equal to A as well. Although the coefficients $\Phi^{(k)}{}_{i_1\cdots i_k j}$ are not completely specified by this process, the symmetry requirements make it impossible, in general, to choose the coefficients A_2 and A_3 of z_2 and z_3, respectively, equal to 0.

The Hamiltonian, to first order in $\mathbf{z}$, is

$$
H_1 = A + z_1 A + z_2 A_2 + z_3 A_3 .
$$

Let us assume for a moment that A has no degenerate eigenvalues. If we denote the eigenvalues by a_i and the corresponding eigenstates by $|a_i\rangle$, then, according to perturbation theory, the energy E_i of a particle in the eigenstate $|a_i\rangle$ is, to first order in $\mathbf{z}$, the diagonal matrix element of the Hamiltonian H_1

$$
\begin{aligned}
E_i &= \langle a_i|H_1|a_i\rangle\\
&= a_i + z_1 a_i + z_2\langle a_i|A_2|a_i\rangle + z_3\langle a_i|A_3|a_i\rangle .
\end{aligned}
$$

The force experienced by particles in the eigenstate a_i is determined from the gradient of the energy. In particular, the force in the x_1 direction, at the center of the chamber, is

$$
F_1(0) = -\frac{\partial}{\partial z_1}E_i = -a_i .
$$

Consequently, particles in the eigenstates corresponding to different eigenvalues experience different forces in the x_1 direction and are physically separated in that direction. The presence of additional underdetermined forces in the x_2 and x_3 directions does not affect this separation to first order. By using additional homogeneous electric fields, for example, an experimenter could compensate for the x_2 and x_3 displacements and produce displacement only in the x_1 direction.

Actually, we have been lavish in choosing the coefficient A_1 of z_1 equal to A, for all that is needed is to have the *matrix elements* $\langle a_i | A_1 | a_i \rangle = a_i$. This requires specification of only $n = 2s + 1$ real parameters rather than $n^2 = 4s^2 + 4s + 1$. Satisfying the further condition that $\langle a_i | A_2 | a_i \rangle = \langle a_i | A_3 | a_i \rangle = 0$ requires an additional $2n$ real parameter. Thus only $3n$ real parameters are needed to produce a force in the x_1 direction proportional to the eigenvalue and zero forces in the x_2 and x_3 directions. On the other hand, we have available $\Sigma_{k=0}^{2s} 2(k+1) + 1 = 4s^2 + 8s + 3$ real parameters, which exceeds $3n = 6s + 3$ by $4s^2 + 2s$, a rapidly increasing positive number. Therefore it is probably possible in most actual situations to arrange for the apparatus to produce a force only in the x_1 direction.

If some of the eigenvalues of A are degenerate, we replace A by an operator A' all of whose eigenstates are also eigenstates of A, but whose eigenvalues are nondegenerate. The apparatus is then set up to measure A' and subsequently the subbeams of A are recombined with phase relations preserved.

We have reached our goal of designing a generalized Stern–Gerlach apparatus which, given an arbitrary Hermitian operator A acting in the Hilbert space of a spin-s particle, separates a beam of particles according to their eigenvalues with respect to the operator A, and we have shown that the electric and magnetic fields needed for this experiment are consistent with Maxwell's equations.

V. CONCLUSION

Modulo the ability to create in the laboratory any electromagnetic field consistent with Maxwell's equations, we have shown that, using a generalized Stern–Gerlach apparatus, every Hermitian operator acting on the Hilbert space of a spin-s particle can be measured and a beam of particles can be produced in the state corresponding to any given ray in the Hilbert space. If counters can be aranged which detect the presence of a particle (or measure the intensity of the beam) without changing its spin state, then the measurement we have proposed is a measurement of the first kind in the sense of Pauli;[7,14] that is, it obeys the projection postulate of von Neumann[6]: If the measurement is repeated immediately the same result will be obtained with certainty. It is also a minimal measurement of the operator in the sense of Herbut[15] because particles in different states corresponding to the same (degenerate) eigenvalue are not separated and the phase relation between them is not altered.

The experiments which we have proposed are admittedly "in principle" experiments and might be exceedingly difficult to carry out in the laboratory. We have assumed that we can create any electromagnetic field concordant with Maxwell's equations, whereas in the real world it is very difficult to produce electric fields with a significant gradient over the size of an atom, and furthermore, the quadrupole splitting of energy levels is a very small effect. These technological points do not have any bearing, however, on whether the operators in question are in principle observable, in the sense discussed in the introduction.

We have, of course, not answered the question in general of which Hermitian operators are measurable. It should be possible to adapt our methods to certain other finite-dimensional situations if some parameter could be found that plays a role analogous to the electromagnetic fields used here. For operators on infinite dimensional Hilbert spaces, for example, the operators $p + q$, $pq + qp$, or pqp suggested by Wigner, even our definition of measurability would have to be extended. We believe, however, that the conclusive demonstration of the measurability of all Hermitian operators in the nontrivial case of spin systems, gives significantly more credibility to the assumption, so prevalent among physicists, that all nonsuperselected Hermitian operators are observable.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to the participants in the joint Seminar on the Foundations of Quantum Mechanics at the University of Massachusetts/Amherst for providing a forum in which these ideas were first discussed.

The second author also expresses his gratitude to James Walker and Lawrence Weaver for many helpful discussions, and to the Department of Mathematics at the University of Massachusetts/Amherst for their hospitality during much of the time in which this research was carried out.

[1]P.A.M. Dirac, *The Principles of Quantum Mechanics* (Oxford Univ., Oxford, 1958), 4th ed., p. 37.
[2]G.C. Wick, A.S. Wightman, and E.P. Wigner, Phys. Rev. **88**, 101 (1952).
[3]R.F. Streater and A.S. Wightman, *PCT, Spin, Statistics and All That* (Benjamin, New York, 1964), p. 5.
[4]E.P. Wigner, "Epistomological Perspectives on Quantum Theory," in *Contemporary Research in the Foundations and Philosophy of Quantum Mechanics*, edited by C. Hooker (Reidel, Dordrecht, Holland, 1975), p. 369.
[5]F.O. Hultgren and A. Shimony, J. Math. Phys. **19**, 381 (1977).
[6]J. Von Neuman, *Mathematical Foundations of Quantum Mechanics* (Princeton Univ., Princeton, N.J., 1955).
[7]J.M. Jauch, *Foundations of Quamtum Mechanics* (Addison–Wesley, Reading, Mass., 1968), Chaps. 5, 8.
[8]W.E. Lamb, Jr., Phys. Today **22**, 23 (1969).
[9]D.J. Foulis and C.H. Randall, J. Math. Phys. **13**, 1667 (1972).
[10]S.C. Kharatyan, Sov. J. Nucl. Phys. **9**, 761 (1969).
[11]W. Gerlach and O. Stearn, Ann. Physik **74**, 673 (1924); **76**, 163 (1925).
[12]E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), 2nd ed., p. 414.
[13]R.P. Feynman, R.B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison–Wesley, Reading, Mass., 1965), Vol. III, pp. 5–11 and footnotes.
[14]W. Pauli, *Handbuch d. Physik*, edited by S. Flugge (Springer-Verlag, Berlin, 1958), Vol. V, part 1.
[15]F. Herbut, Ann. Phys. **55**, 271 (1969).

A kernel of Gel'fand–Levitan type for the three-dimensional Schrödinger equation

H. E. Moses[a)]
University of Lowell, Center for Atmospheric Research, College of Pure and Applied Science, Lowell, Massachusetts 01854

(Received 8 March 1979; accepted for publication 12 June 1979)

In a previous paper we introduced a Green's function for the three-dimensional Schrödinger equation analogous to the Green's function used to obtain the integral equation for the Jost wave functions in one dimension. The three-dimensional Green's function was used to define Jost wave functions for the three-dimensional problem and the completeness relations for these wave functions were obtained. In the present paper we use the three-dimensional Green's function to construct influence functions for the 3 + 3 ultrahyperbolic partial differential equation which have analogs to the causal properties of the corresponding influence functions for the 1 + 1 hyperbolic partial differential equation. Just as the 1 + 1 influence function can be used to obtain an integral equation for the one-dimensional Gel'fand–Levitan kernel in terms of the scattering potential, we use the 3 + 3 influence function to obtain an analogous integral equation for our proposed Gel'fand–Levitan kernel for the three-dimensional problem. Though much of the formalism for finding the properties of the kernel for the three-dimensional problem can be carried out in a straightforward manner, the interpretation of the triangularity properties is more difficult than in the one-dimensional case because of the complicated geometrical picture associated with the notion of causality. In addition to its use in obtaining a Gel'fand–Levitan kernel, the 3 + 3 influence function can be used to simplify the second term in an expansion of the potential in terms of the minimal scattering data. This simplification is also given. In the Appendix the asymptotic form of the three-dimensional Jost wave function is given in a form which is analogous to the asymptotic form for the one-dimensional Jost wave function and which is compatible with our notion of triangularity for the Gel'fand–Levitan kernel.

1. DERIVATION OF THE INTEGRAL EQUATION FOR THE GEL'FAND–LEVITAN KERNEL IN TERMS OF THE SCATTERING POTENTIAL

The present paper will be written as a direct extension of Ref. 1. When equations in Ref. 1 are referred to, a prime will be placed next to the equation number.

In the present paper it is our objective to give an equation for a proposed Gel'fand–Levitan kernel in terms of the scattering potential for the three-dimensional inverse problem. The analagous equation for the one-dimensional problem can be used to give the triangularity properties of the Gel'fand–Levitan kernel and to show the relation of the kernel to the potential (see for example, Ref. 2). As mentioned in Ref. 1, we shall use the Green's function (11′) in its three-dimensional interpretation to construct the integral equation for the Gel'fand–Levitan kernel, just as it is used in its one-dimensional interpretation for the integral equation for the one-dimensional kernel as in Ref. 2. To review the treatment for the one-dimensional problem, we repeat some of the results of Ref. 2.

The Jost wave function $f(x|p)$ which satisfies the integral equation (7′) using the Green's function (11′) is given in terms of the Gel'fand–Levitan kernel

$$f(x|p) = e^{ipx} + \int_{-\infty}^{\infty} K(x|x')\, e^{ipx'}\, dx'. \tag{1}$$

We now equate the two expressions for $f(x|p)$ given by Eqs. (1) and (7′):

$$\begin{aligned}\int_{-\infty}^{+\infty} & K(x|x')\, e^{ipx'}\, dx' \\ &= \int_{-\infty}^{+\infty} G_{pJ}(x-x')\, V(x') f(x'|p)\, dx' \\ &= \int_{-\infty}^{+\infty} G_{pJ}(x-x')\, V(x')\, e^{ipx'}\, dx' \\ &+ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} G_{pJ}(x-x')\, V(x')\, K(x'|x'')\, e^{ipx''}\, dx'\, dx''.\end{aligned} \tag{2}$$

On multiplying through by $(1/2\pi)\, e^{-ipy}$ and integrating with respect to y and using $(2\pi)^{-1}\int_{-\infty}^{+\infty} e^{ipx}\, dp = \delta(x)$, we have

$$\begin{aligned}K(x|y) = \int_{-\infty}^{+\infty} & R(x-x'|y-x')\, V(x')\, dx' \\ &+ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} R(x-x'|y-x'')\, V(x') \\ &\times K(x'|x'')\, dx'\, dx'',\end{aligned} \tag{3}$$

where

$$\begin{aligned}R(w|z) &= (2\pi)^{-1}\int_{-\infty}^{+\infty} G_{pJ}(w)\, e^{-izp}\, dp \\ &= (2\pi)^{-2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{ikw}\, e^{-izp}\, \gamma(p,k)\, dp\, dk.\end{aligned} \tag{4}$$

a)Research sponsored by the U.S. Army Research Office under Grant No. DAAG 29-78-G-0003 P-14919-M.

The second of Eqs. (4) follows from Eqs. (10′) and (11′).

Our choice of Eq. (11′) for $\gamma(p,k)$ is motivated by the fact that $R(w|z)$ and $K(x|y)$ have appropriate triangularity properties for the one-dimensional case.

For *any* choice of $\gamma(p,k)$ such that $(p^2-k^2)\gamma(p,k)=1$, it follows that

$$\frac{\partial^2}{\partial w^2}R(w|z)-\frac{\partial^2}{\partial z^2}R(w|z)=-\delta(w)\,\delta(z), \tag{5}$$

(i.e., $R(w|z)$ is an influence function for the 1 + 1 hyperbolic differential equation. As we have shown in Ref. 2 and shall show again below, the choice of (11′) for $\gamma(p,k)$ makes $R(w|z)$ causal or, equivalently, triangular. In turn, the triangularity in $R(w|z)$ assures us that $K(x|y)$ will have appropriate triangular properties. Let us define

$$H_0^x=-\frac{\partial^2}{\partial x^2},\quad H_0^y=-\frac{\partial^2}{\partial y^2}. \tag{6}$$

H_0 is thus the kinetic energy operator. From Eqs. (3)–(5)

$$(H_0^x-H_0^y)K(x|y)=\delta(x-y)V(x)+V(x)K(x|y), \tag{7}$$

or on defining K to be the integral operator whose kernel is $K(x|y)$ and I to be the identity operator which, when expressed as an integral operator, has as its kernel $\delta(x-y)$, Eq. (7) is

$$HU=UH_0, \tag{8}$$

where U is the wave operator given by $U=I+K$. Equation (8), in fact, is one of the starting points of the theory of inverse scattering as discussed in Ref. 3.

From Eq. (4) using Eq. (11′) (see also Ref. 2),

$$R(w|z)=\tfrac{1}{2}\eta(w-|z|), \tag{9}$$

where $\eta(x)$ is the Heaviside function

$$\eta(x)=0,\quad \text{for } x<0;\ \eta(x)=1,\quad \text{for } x\geqslant 0. \tag{10}$$

The expression Eq. (9) for $R(w|z)$ is a double Fourier transform of $\gamma(p,k)$ and for the one-dimensional problem is most conveniently obtained by first evaluating $G_{pJ}(w)$ and then using the first of Eq. (4). We make a point of this manner of evaluating $R(w|z)$ because in the three-dimensional case G_{pJ} cannot be evaluated. Nevertheless, it is still possible to obtain the three-dimensional analog of $R(w|z)$.

From Eq. (3), one obtains after the methods used in Ref. 2, in which the causal character of $R(w|z)$ as given by Eq. (9) plays an essential role,

$$K(x|y)=0,\quad \text{if } x<y, \tag{11}$$

$$\begin{aligned}K(x|y)=\tfrac{1}{2}\int_{-\infty}^{(x+y)/2}V(x')\,dx'+\tfrac{1}{2}\int_{-\infty}^{(x+y)/2}V(x')\\ \times\int_{y-x+x'}^{x'}K(x'|z)\,dx'\,dz+\tfrac{1}{2}\int_{(x+y)/2}^{x}V(x')\\ \times\int_{y-x+x'}^{y+x-x'}K(x'|z)\,dz\,dx',\quad \text{if } x\geqslant y.\end{aligned} \tag{12}$$

From Eq. (12) we obtain the familiar result $V(x)=2(d/dx)K(x|x)$.

We shall now repeat the arguments which lead to the three-dimensional analog of Eq. (3). The three-dimensional analog of $R(w|z)$, the "double Fourier transform" of $\gamma(p,k)$, is an influence function for the 3 + 3 ultrahyperbolic partial differential equation. This analog can be found explicitly. It is a distribution with "causal" or "triangular" properties. However, because it is difficult to visualize these concepts in the six-dimensional space, we are as yet unable to write the analog of Eq. (11) or (12). The bulk of the present paper is devoted to deriving the analog of $R(w|z)$.

We now parallel for three dimensions what we have done for one dimension. We shall represent the Jost function of Eq. (20′) as

$$f(\mathbf{x}|\mathbf{p})=e^{i\mathbf{p}\cdot\mathbf{x}}+\int K(\mathbf{x}|\mathbf{x}')\,e^{i\mathbf{p}\cdot\mathbf{x}'}\,d\mathbf{x}', \tag{13}$$

where the integration is taken over all space. On equating the expression for the Jost function by Eq. (13) with that in terms of the Jost Green's function given by Eq. (20′) in a manner entirely analogous to the derivation of Eq. (3), we obtain an equation for the proposed three-dimensional Gelfand–Levitan kernel $K(\mathbf{x}|\mathbf{x}')$:

$$\begin{aligned}K(\mathbf{x}|\mathbf{y})=\int R(\mathbf{x}-\mathbf{x}'|\,\mathbf{y}-\mathbf{x}')\,V(\mathbf{x}')\,d\mathbf{x}'\\ +\iint R(\mathbf{x}-\mathbf{x}'|\,\mathbf{y}-\mathbf{x}'')\\ \times V(\mathbf{x}')\,K(\mathbf{x}'|\mathbf{x}'')\,d\mathbf{x}'\,d\mathbf{x}'',\end{aligned} \tag{14}$$

where the influence function $R(\mathbf{w}|\mathbf{z})$, which satisfies the inhomogeneous 3 + 3 ultrahyperbolic equation

$$\nabla_{\mathbf{w}}^2R(\mathbf{w}|\mathbf{z})-\nabla_{\mathbf{z}}^2R(\mathbf{w}|\mathbf{z})=-\delta(\mathbf{w})\,\delta(\mathbf{z}), \tag{15}$$

is given by

$$R(\mathbf{w}|\mathbf{z})=(2\pi)^{-3}\int G_{pJ}(\mathbf{w})\,e^{-i\mathbf{p}\cdot\mathbf{z}'}\,d\mathbf{p}, \tag{16}$$

$$R(\mathbf{w}|\mathbf{z})=(2\pi)^{-6}\iint e^{i\mathbf{k}\cdot\mathbf{w}}\,e^{-i\mathbf{z}\cdot\mathbf{p}}\,\gamma(p,k)\,d\mathbf{p}\,d\mathbf{k}. \tag{17}$$

The three-dimensional analog of Eq. (7) is

$$(H_0^x-H_0^y)K(\mathbf{x}|\mathbf{y})=\delta(\mathbf{x}-\mathbf{y})V(\mathbf{x})+V(\mathbf{x})K(\mathbf{x}|\mathbf{y}), \tag{18}$$

where

$$H_0^x=-\nabla_{\mathbf{x}}^2,\quad H_0^y=-\nabla_{\mathbf{y}}^2. \tag{19}$$

The analog of Eq. (8) also follows from Eq. (19) as did Eq. (8) itself from Eq. (7).

Equation (1) for the one-dimensional Jost wave function in terms of the triangular Gelfand–Levitan kernel as well as the integral equation (7′) for the wave function contains the boundary condition $\lim_{x\to-\infty}f(x|p)=e^{ipx}$. In the Appendix we shall derive the analogous boundary condition for the three-dimensional Jost wave function, even though we cannot evaluate the Jost Green's function of Eq. (20′) or give the triangularity properties for the three-dimensional Gel'fand–Levitan kernel in a completely explicit form.

2. DERIVATION OF THE EXPRESSION FOR THE 3 + 3-DIMENSIONAL INFLUENCE FUNCTION

It will be useful to write

$$R(\mathbf{w}|\mathbf{z})=\frac{1}{4\pi|z|}\nabla_{\mathbf{w}}^2H(\mathbf{w}|\mathbf{z}), \tag{20}$$

where z is the optical radius of $\mathbf{z}$ and $H(\mathbf{w}|\mathbf{z})$ is given by

$$H(\mathbf{w}|\mathbf{z}) = -\frac{|z|}{(2\pi)^4\pi}\int\int e^{i\mathbf{k}\cdot\mathbf{w}}\, e^{-i\mathbf{z}\cdot\mathbf{p}}\gamma(p,k)\, d\mathbf{p}\frac{d\mathbf{k}}{k^2}. \quad (21)$$

Our major effort will be to obtain $H(\mathbf{w}|\mathbf{z})$ which will be a perfectly well behaved function exhibiting causality properties. By contrast, $R(\mathbf{w}|\mathbf{z})$ will be a symbolic function because of the operation $\nabla^2_{\mathbf{w}}$ in Eq. (20). We shall first evaluate the integral with respect to $\mathbf{p}$ in Eq. (21). One has

$$\int e^{-i\mathbf{z}\cdot\mathbf{p}}\gamma(p,k)\, d\mathbf{p} = \int e^{-i\mathbf{z}\cdot\mathbf{p}}[\eta(k)\gamma_-(p^2-k^2) + \eta(-k)\gamma_+(p^2-k^2)]\, d\mathbf{p}. \quad (22)$$

However, as is well known,

$$\int e^{-i\mathbf{z}\cdot\mathbf{p}}\gamma_\pm(p^2-k^2)\, d\mathbf{p} = \frac{(2\pi)^2}{2}|z|^{-1}\exp(\pm i|z||k|), \quad (23)$$

so that Eq. (22) becomes

$$\begin{aligned}\int e^{-i\mathbf{z}\cdot\mathbf{p}}\gamma(p,k)\, d\mathbf{p} &= \frac{(2\pi)^2}{2}|z|^{-1}[\eta(k)\exp(-i|z||k|) + \eta(-k)\exp(i|z||k|)]\\ &= \frac{(2\pi)^2}{2}|z|^{-1}e^{-i|z|k}. \quad (24)\end{aligned}$$

Thus, from Eq. (21), on using optical coordinates for the integration over $\mathbf{k}$ [$\mathbf{k} = k(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$] and optical coordinates for $\mathbf{w}$ [$\mathbf{w} = w(\sin\theta'\cos\phi', \sin\theta'\sin\phi', \cos\theta')$],

$$H(\mathbf{w}|\mathbf{z}) = -\frac{1}{(2\pi)^3}\int_0^{2\pi}d\phi \times \int_0^{\pi/2}\sin\theta\, d\theta\int_{-\infty}^{+\infty}e^{ik(\lambda-|z|)}\, dk, \quad (25)$$

where

$$\lambda = w[\sin\theta\sin\theta'\cos(\phi-\phi') + \cos\theta\cos\theta']. \quad (26)$$

Since only the absolute value of z appears, it will be useful to define

$$q = |z|, \quad (27)$$

and

$$H(\mathbf{w}|q) \equiv H(\mathbf{w}|\mathbf{z}) \quad (28)$$

In Eq. (25) we introduce the variable

$$\rho = \phi - \phi'. \quad (29)$$

We note that

$$\int_0^{2\pi}d\rho\cdots = 2\int_0^{\pi}d\rho\cdots. \quad (30)$$

Thus, from Eq. (25),

$$H(\mathbf{w}|q) = -\frac{1}{2\pi^2}\int_0^{\pi}d\rho\int_0^{\pi/2}\sin\theta\,\delta(\lambda-q)\, d\theta,$$
$$\lambda = w(\sin\theta\sin\theta'\cos\rho + \cos\theta\cos\theta'). \quad (31)$$

To carry out the integration of Eq. (31) we must go from the variables of integration ρ,θ to the variables λ,θ. Thus,

$$H(\mathbf{w}|q) = -\frac{1}{2\pi^2}\int d\theta\int\frac{\partial\rho}{\partial\lambda}\delta(\lambda-q)\, d\lambda, \quad (32)$$

where we still must specify the domain of integration in the θ–λ plane.

From the second of Eqs. (31),

$$\frac{\partial\rho}{\partial\lambda} = -\frac{1}{w\sin\rho\sin\theta\sin\theta'}. \quad (33)$$

Hence, λ is a monotonic decreasing function of ρ for $w>0$ and is a monotonic increasing function of ρ for $w<0$.

It is necessary to consider several cases.

Case 1: $w>0$, $0<\theta'<\pi/4$: First consider the domain of integration in the ρ–θ plane which is given in Fig. 1. The domain consists of a rectangle bounded by the straight lines, denoted by L_1, L_2, L_3, and L_4, which considered as curves in the plane are given by $\rho = 0$, $\rho = \pi$, $\theta = 0$, and $\theta = \pi/2$, respectively. Figure 2 shows how this domain appears in the λ–θ plane. The lines L_i map into the curves shown in Fig. 2. The line L_1 becomes the curve $\lambda = w\cos(\theta-\theta')$, the line L_2 maps into the curve $\lambda = w\cos(\theta+\theta')$, the line L_4 is the line segment $\theta = \pi/2$, while the line L_3 degenerates to the point $(0, w\cos\theta')$. The curve corresponding to L_1 has its maximum at $\theta = \theta'$. The domain of integration naturally splits up into three regions denoted by region I, region II, and region III. For any value of q, draw the line $\lambda = q$ in the θ–λ plane If $q>w$, then because of the presence of the δ function in Eq. (32), we have

$$H(\mathbf{w}|q) = 0, \quad q>w. \quad (34)$$

More generally, we see that from the second of Eqs. (31)

$$\sin\rho\sin\theta\sin\theta' = [\sin^2\theta'\sin^2\beta - (\cos\theta - \cos\beta\cos\theta')^2]^{1/2}, \quad (35)$$

where β is defined by

$$\cos\beta = \lambda/w. \quad (36)$$

In Eq. (36) and later we take

$$0\leqslant\beta\leqslant\pi. \quad (37)$$

Putting in the limits of integration in Eq. (32) gives

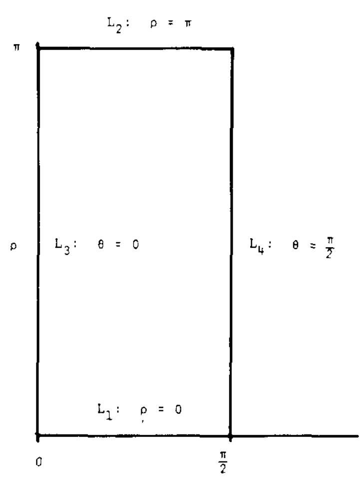


FIG. 1. Domain of integration in the $\theta = \rho$ plane.

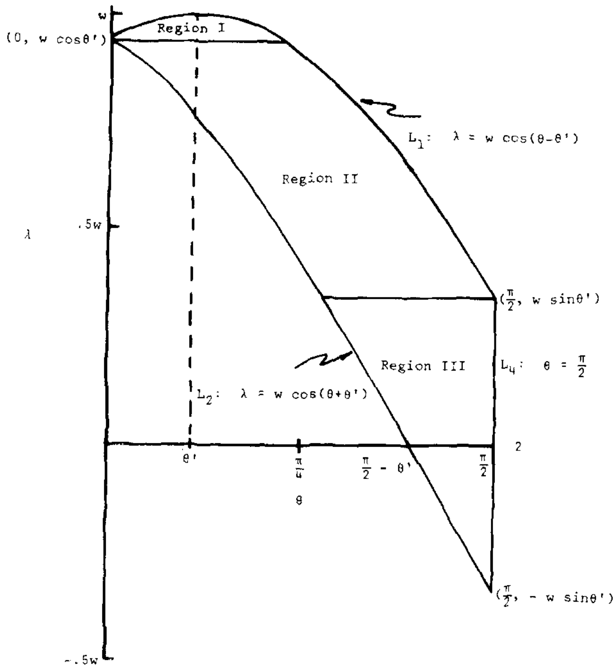


FIG. 2. Domain of integration in the θ–λ plane ($w > 0$, $\theta' < \pi/4$).

$$H(\mathbf{w}|q) = -\frac{1}{2\pi^2 w}\int_0^{\pi/2} d\theta \int_{w\cos(\theta+\theta')}^{w\cos(\theta-\theta')} \delta(\lambda - q) \times [\sin^2\theta' \sin^2\beta - (\cos\theta - \cos\beta\cos\theta')^2]^{-1/2}\, d\lambda . \tag{38}$$

We now interchange the order of integration in the double integral and thus have

$$H(\mathbf{w}|q) = -\frac{1}{2\pi^2 w}\int_{-w\sin\theta'}^{w} \delta(\lambda - q)\, d\lambda \times \int_{\theta_1(\lambda)}^{\theta_2(\lambda)} [\sin^2\theta' \sin^2\beta - (\cos\theta - \cos\beta\cos\theta')^2]^{-1/2}\, d\theta, \tag{39}$$

where $\theta_1(\lambda)$ is the curve consisting of L_2 and the left-hand branch of L_1 in Fig. 2 and $\theta_2(\lambda)$ is the curve consisting of L_4 and right-hand branch of L_1.

We can now integrate over λ and because of the δ function obtain

$$H(\mathbf{w}|q) = -\frac{1}{2\pi^2 w}\int_{\theta_1(q)}^{\theta_2(q)} [\sin^2\theta' \sin^2\alpha - (\cos\theta - \cos\alpha\cos\theta')^2]^{-1/2}\, d\theta . \tag{40}$$

In Eq. (40), α is defined here and later by

$$\cos\alpha = q/w, \quad 0 \leqslant \alpha \leqslant \pi. \tag{41}$$

The values of $\theta_1(q)$ and $\theta_2(q)$ are those values of θ which are obtained from the intersections of the straight line $\lambda = q$ with the curves $\theta_1(\lambda)$ and $\theta_2(\lambda)$, respectively.

For $w > q > w\cos\theta'$ (i.e., where $\lambda = q$ lies in region I of Fig. 2),

$$\theta_1(q) = \theta' - \alpha, \;\; \theta_2(q) = \theta' + \alpha. \tag{42}$$

We can now evaluate the integral of Eq. (40) in closed form: Let the variable of integration x be defined by

$$x = \frac{\cos\theta - \cos\alpha\cos\theta'}{\sin\theta'\sin\alpha}. \tag{43}$$

It is to be noted that x is a monotonic decreasing function of θ and thus there are no troubles with branches. Then

$$H(\mathbf{w}|q) = -\frac{1}{2\pi^2 w}\int_{-1}^{+1} (1 - x^2)^{-1/2}\, dx = -\frac{1}{\pi^2 w}\int_0^{+1} (1 - x^2)^{-1/2}\, dx,$$

or finally

$$H(\mathbf{w}|q) = -\frac{1}{2\pi w}, \quad w > q > w\cos\theta'. \tag{44}$$

For $w\sin\theta' < q < w\cos\theta'$, i.e., $\lambda = q$ is in region II of Fig. 2,

$$\theta_1(q) = \alpha - \theta', \quad \theta_2(q) = \alpha + \theta'. \tag{45}$$

On evaluating the integral of Eq. (40), one obtains the same expression for $H(\mathbf{w}|q)$ as in Eq. (44). Finally, for this case we take $w\sin\theta' > q > 0$, i.e., $\lambda = q$ is in region III.

In this case,

$$\theta_1(q) = \alpha - \theta', \quad \theta_2(q) = \pi/2. \tag{46}$$

Then

$$H(\mathbf{w}|q) = \frac{1}{2\pi^2 w}\int_1^{-\cot\alpha\cot\theta'} (1 - x^2)^{-1/2}\, dx = -\frac{1}{2\pi^2 w}\left(\int_0^1 (1 - x^2)^{-1/2}\, dx + \int_0^{\cot\alpha\cot\theta'} (1 - x^2)^{-1/2}\, dx\right),$$

or finally

$$H(\mathbf{w}|q) = -\frac{1}{2\pi^2 w}\left[\frac{\pi}{2} + \sin^{-1}(\cot\alpha\cot\theta')\right]. \tag{47}$$

In Eq. (47) the principal branch is taken of $\sin^{-1}$.

From the conditions on q we see that

$$\cos\alpha < \sin\theta' \;\text{ or }\; \cos\alpha < \cos(\pi/2 - \theta').$$

Hence,

$$\alpha > \pi/2 - \theta' \;\text{ or }\; \alpha + \theta' > \pi/2$$

and

$$\cos(\alpha + \theta') < 0 \;\text{ or }\; \cot\alpha\cot\theta' < 1. \tag{48}$$

The inequality of Eq. (48) assures us that the argument of $\sin^{-1}$ in Eq. (47) is less than unity and the principal branch is defined and real.

Thus, for $\theta' < \pi/4$, $w > 0$, we may summarize our results as follows:

$$H(\mathbf{w}|q) = -[\eta(q - w\sin\theta') - \eta(q - w)]\left(\frac{1}{2\pi w}\right) - [\eta(w\sin\theta' - q)]\left[\frac{1}{4\pi w} + \frac{1}{2\pi^2 w} \times \sin^{-1}(\cot\alpha\cot\theta')\right], \tag{46a}$$

where $\eta(x)$ is, as usual, the Heaviside function $\eta(x) = 1$ if $x > 0$, and $\eta(x) = 0$ if $x < 0$.

Case 2: $w > 0$, $\pi/4 < \theta' < \pi/2$: The domain of integration for this case is given in Fig. 3. Despite the difference in

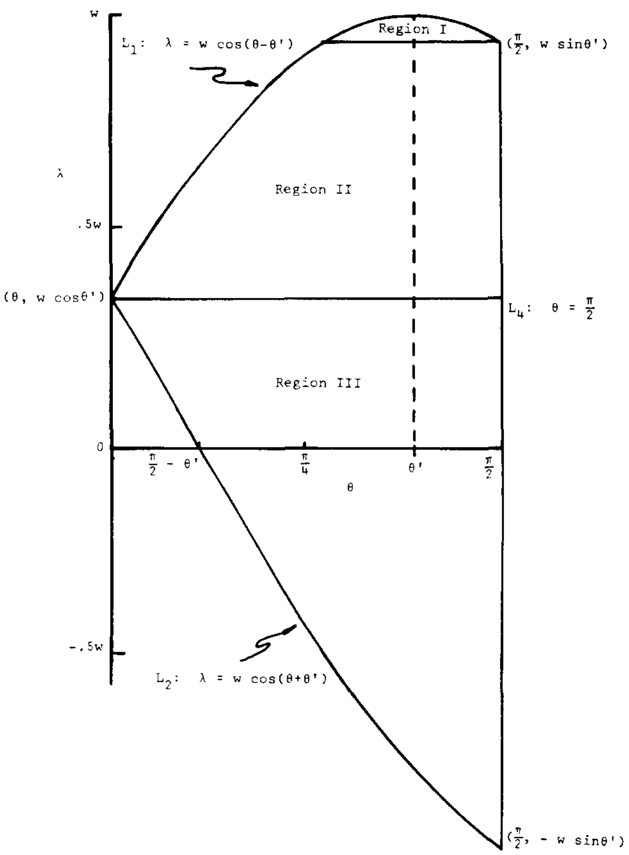


FIG. 3. Domain of integration in the θ–λ plane ($w > 0$, $\theta' > \pi/4$).

the structure of the domains of integration, the expression for $H(\mathbf{w}|q)$ for this case is identical to that for case 1, namely, Eq. (46a). The details of the calculation are omitted, since they are very similar to those for case 1.

Case 3: $w < 0$: The domain of integration is shown in Fig. 4. Again, despite the differences in the domains of integration between the present case and the previous ones, the details of the calculation are very similar. We obtain

$$H(\mathbf{w}|q) = \frac{1}{4\pi w} - \frac{1}{2\pi^2 w} \sin^{-1}(|\cot\alpha| \cot\theta'). \quad (47a)$$

We can now write the general form for $H(\mathbf{w}|q)$:

$$H(\mathbf{w}|q) = -[\eta(w-q) - \eta(w\sin\theta' - q)] \frac{1}{2\pi w} - \eta(|w|\sin\theta' - q) \times \left[\frac{1}{4\pi|w|} + \frac{1}{2\pi^2 w}\sin^{-1}(|\cot\alpha|\cot\theta')\right]. \quad (48a)$$

To summarize,

$$R(\mathbf{w}|\mathbf{z}) = \frac{1}{4\pi|z|} \nabla_{\mathbf{w}}^2 H(\mathbf{w}|\mathbf{z}), \quad (49)$$

where $H(\mathbf{w}|\mathbf{z}) \equiv H(\mathbf{w}|q)$ with $H(\mathbf{w}|\,q)$ being given by Eq. (48a) and

$$q = |z| \quad \text{and} \quad \cos\alpha = |z|/w \quad (0 < \alpha < \pi). \quad (49a)$$

Clearly, $R(\mathbf{w}|\mathbf{z})$ is a distribution, since it involves derivatives of a function.

To evaluate integrals of the form which appear in Eq. (14), we consider test functions $f(\mathbf{x})$ and first evaluate the functions

$$g_1(\mathbf{x}) = \int \frac{1}{4\pi|z|} H(\mathbf{x}|\mathbf{z}) f(\mathbf{z})\, d\mathbf{z},$$
$$g_2(\mathbf{x},\mathbf{y}) = \int \frac{1}{4\pi|\,\mathbf{y}-\mathbf{x}'|} H(\mathbf{x}-\mathbf{x}'|\,\mathbf{y}-\mathbf{x}') f(\mathbf{x}')\, d\mathbf{x}',$$
$$g_3(\mathbf{x},\mathbf{y}) \int \frac{1}{4\pi|\,\mathbf{y}-\mathbf{x}'|} H(\mathbf{x}|\,\mathbf{y}-\mathbf{x}') f(\mathbf{x}')\, d\mathbf{x}'. \quad (50)$$

For suitable test functions, these functions are well defined. We now define

$$G_1(\mathbf{x}) = \int R(\mathbf{x}|\mathbf{z}) f(\mathbf{z})\, d\mathbf{z},$$
$$G_2(\mathbf{x},\mathbf{y}) = \int R(\mathbf{x}-\mathbf{x}'|\,\mathbf{y}-\mathbf{x}') f(\mathbf{x}')\, d\mathbf{x}',$$
$$G_3(\mathbf{x},\mathbf{y}) = \int R(\mathbf{x}|\,\mathbf{y}-\mathbf{x}') f(\mathbf{x}')\, d\mathbf{x}' \quad (51)$$

by

$$G_1(\mathbf{x}) = \nabla_{\mathbf{x}}^2 g_1(\mathbf{x}),$$
$$G_2(\mathbf{x},\mathbf{y}) = \nabla_{\mathbf{x}}^2 g_2(\mathbf{x},\mathbf{y}),$$
$$G_3(\mathbf{x},\mathbf{y}) = \nabla_{\mathbf{x}}^2 g_3(\mathbf{x},\mathbf{y}), \quad (52)$$

respectively.

3. TRIANGULARITY CONDITIONS

From the explicit expression for $H(\mathbf{w}|\mathbf{q})$ given by Eq. (49), it is seen that $R(\mathbf{w}|\mathbf{z})$ vanishes in the six-dimensional w–z space when for $w > 0, |z| > w$ or for $w < 0, |z| > -w\sin\theta'$. A coarser domain in which $R(\mathbf{x}|\mathbf{z})$ vanishes is given by

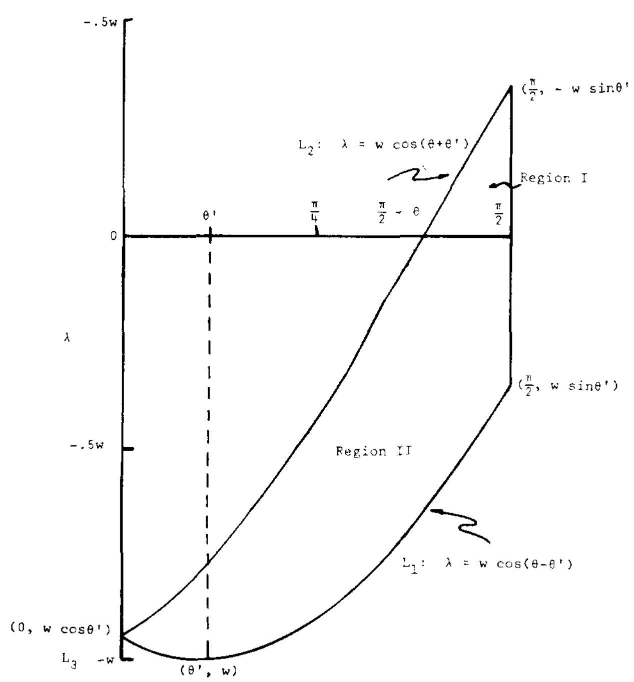


FIG. 4. Domain of integration in the θ–λ plane ($w < 0$, $\theta' < \pi/4$).

$$R(\mathbf{w}|\mathbf{z}) \equiv 0, \quad \text{for } |\mathbf{z}| > |\mathbf{w}| . \tag{53}$$

The triangularity conditions on $R(\mathbf{w}|\mathbf{z})$ should lead to triangularity conditions on $K(\mathbf{x}|\,\mathbf{y})$ through the use of Eq. (14). It seems hard to show this triangularity in general. However, for the important special case in which the potential $V(\mathbf{x})$ vanishes identically for sufficiently large $\mathbf{x}$, i.e.,

$$V(\mathbf{x}) \equiv 0, \quad \text{for } |\mathbf{x}| > R > 0, \tag{54}$$

we shall show

$$K(\mathbf{x}|\,\mathbf{y}) \equiv 0, \quad |\,\mathbf{y}| - |\mathbf{x}| > 2R \tag{55}$$

if the solution $K(\mathbf{x}|\,\mathbf{y})$ of Eq. (14) is unique.

Actually, one can see from the proof that follows that $K(\mathbf{x}|\,\mathbf{y})$ probably vanishes in a smaller domain of the six-dimensional space. The condition (55) is not inconsistent with our three-dimensional treatments of Refs. 4 and 5 in which the triangularization condition is taken as $K(\mathbf{x}|\,\mathbf{y}) \equiv 0$ for $|\,\mathbf{y}| > |\mathbf{x}|$.

We now proceed to the proof of Eq. (55) under the condition of Eq. (54), i.e., we shall take

$$|\,\mathbf{y}| - |\mathbf{x}| > 2R \tag{56}$$

and show that

$$K(\mathbf{x}|\,\mathbf{y}) = 0 \tag{57}$$

is a solution of Eq. (14). The assumption of the uniqueness of the solution then completes the theorem.

We shall first show that under the condition of Eq. (56) the first term on the right of Eq. (14) vanishes. From Eq. (56),

$$|\,\mathbf{y}|^2 - |\mathbf{x}|^2 > 2R\,(|\,\mathbf{y}| + |\mathbf{x}|). \tag{58}$$

However,

$$|\,\mathbf{y}| + |\mathbf{x}| \geqslant |\,\mathbf{y} - \mathbf{x}| \tag{59}$$

and thus

$$|\,\mathbf{y}|^2 - |\mathbf{x}|^2 > 2R\,(|\,\mathbf{y} - \mathbf{x}|) . \tag{60}$$

From Eq. (54) we only need consider $\mathbf{x}'$ on the right-hand side of Eq. (14) such that

$$|\mathbf{x}'| < R. \tag{61}$$

Thus, from Eq. (60),

$$|\,\mathbf{y}|^2 - |\mathbf{x}|^2 > 2|\mathbf{x}'|\,|\,\mathbf{y} - \mathbf{x}| > 2\mathbf{x}'\cdot(\,\mathbf{y} - \mathbf{x}). \tag{62}$$

From Eq. (62),

$$|\mathbf{y} - \mathbf{x}'|^2 - |\mathbf{x} - \mathbf{x}'|^2 > 0. \tag{63}$$

Thus, from Eq. (53) the first term on the right of Eq. (14) vanishes.

Now let us consider the second term on the right of Eq. (14). Since we are assuming Eqs. (56) and (57), the only possible contribution from the integration in the $\mathbf{x}'$ and $\mathbf{x}''$ variables comes from the six-dimensional domain

$$|\mathbf{x}'| - |\mathbf{x}''| \geqslant 2R. \tag{64}$$

However, from Eq. (61) it follows that

$$-\,|\mathbf{x}''| > R \tag{65}$$

and hence there is no domain of integration in the $\mathbf{x}',\mathbf{x}''$ variables which give a nonzero contribution to the second term. Thus, we have proved our theorem.

4. AN ALTERNATIVE FORM FOR THE DISTRIBUTION $R(\mathbf{w}|\mathbf{z})$

From Eqs. (20), (21), and (25),

$$\nabla^2_{\mathbf{w}}\, H(\mathbf{w}|\mathbf{z}) = \frac{d^2}{d\,|\mathbf{z}|^2} H(\mathbf{w}|\mathbf{z}) . \tag{66}$$

Hence, since $H(\mathbf{w}|\mathbf{z})$ in its dependence on $\mathbf{z}$ depends only on $|\mathbf{z}|$, we have on expressing the operator $\nabla^2_{\mathbf{z}}$ in terms of polar coordinates

$$4\pi R(\mathbf{w}|\mathbf{z}) = \nabla^2_{\mathbf{z}}\, [H(\mathbf{w}|\mathbf{z})/|\mathbf{z}|] . \tag{67}$$

When we apply $R(\mathbf{w}|\mathbf{z})$ to a test function and integrate by parts, we obtain the distribution $R(\mathbf{w}|\mathbf{z})$ in a form more compatible with the notion of generalized differentiation in terms of test functions

$$4\pi \int R(\mathbf{w}|\mathbf{z} - \mathbf{z}') f(\mathbf{z}')\, d\mathbf{z}' = \int [H(\mathbf{w}|\mathbf{z} - \mathbf{z}')/|\mathbf{z} - \mathbf{z}'|]\, \nabla^2_{\mathbf{z}'} f(\mathbf{z}')\, d\mathbf{z}'. \tag{68}$$

In this form the influence function $R(\mathbf{w}|\mathbf{z})$ is more easily used when solving the analog of the initial value problem for the $3+3$ ultrahyperbolic partial differential equation.

5. AN APPLICATION OF THE INFLUENCE FUNCTION TO AN EXPANSION FROM THE INVERSE PROBLEM

In a very early paper on the inverse scattering problem for the three-dimensional Schrödinger equation (Ref. 6), the author gave an expansion for the construction of the scattering potential in terms of a surprisingly small portion of the scattering amplitude. To be explicit let us consider a solution of the Schrödinger equation

$$[-\nabla^2 + V(\mathbf{x})]\, \psi(\mathbf{x}|\mathbf{k}) = \mathbf{k}^2 \psi(\mathbf{x}|\mathbf{k}) \tag{69}$$

subject to the boundary condition that it be asymptotically representable as the sum of a plane wave and an outgoing spherical wave

$$\lim_{|\mathbf{x}|\to\infty} \psi(\mathbf{x}|\mathbf{k}) = (2\pi)^{-3/2} e^{i\mathbf{k}\cdot\mathbf{x}} + b(\mathbf{k}',\mathbf{k}) \frac{e^{i|\mathbf{k}||\mathbf{x}|}}{|\mathbf{x}|}, \tag{70}$$

where

$$\mathbf{k}' = |\mathbf{k}|(\mathbf{x}/|\mathbf{x}|), \tag{70a}$$

in the usual fashion.

Let us define

$$b(\mathbf{k}) = b(-\mathbf{k},\mathbf{k}), \quad k_z \geqslant 0. \tag{71}$$

To define $b(\mathbf{k})$ for $k_z < 0$, we write

$$b(-\mathbf{k}) = b^*(\mathbf{k}). \tag{72}$$

Clearly, $b(\mathbf{k})$ defined by Eq. (71) is the amplitude of the spherical wave when observed in a direction opposite to the direction of propagation of the incident plane wave with $\mathbf{k}$ pointing in a hemisphere about the z axis. In Ref. 6 it is $b(\mathbf{k})$ which is the portion of the spherical wave amplitude which is required for the reconstruction of the scattering potential. Though a particular axis (the z axis) plays a special role in the definition of $b(\mathbf{k})$, one could use any axis with respect to which the definition could be made.

In Ref. 6 one obtains a set of equations from which one can obtain the scattering potential $V(\mathbf{x})$ in terms of $b(\mathbf{k})$. On replacing $b(\mathbf{k})$ by $\epsilon b(\mathbf{k})$ and on writing $V(\mathbf{x}) = \Sigma_{k=1}^{\infty} \epsilon^k \times V_k(\mathbf{x})$, one can obtain expressions for $V_k(\mathbf{x})$ in terms of Fourier integrals of distributions. The expression for $V_1(\mathbf{x})$ is simple and can be given immediately. The expression for $V_2(\mathbf{x})$ is more complicated and involves the use of $R(\mathbf{w}|\mathbf{z})$:

$$V_1(\mathbf{x}) = -\left(\frac{128}{\pi}\right)^{1/2} \int b(\mathbf{k})\, e^{-2i\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{k}, \tag{73}$$

as before. However, now

$$V_2(\mathbf{x}) = \frac{4}{\pi} \nabla_{\mathbf{x}}^2 \int\int \frac{H(2\mathbf{x} - \mathbf{x}' - \mathbf{x}''|\mathbf{x}' - \mathbf{x}'')}{|\mathbf{x}' - \mathbf{x}''|} \times V_1(\mathbf{x}')\, V_1(\mathbf{x}'')\, d\mathbf{x}'\, d\mathbf{x}''. \tag{74}$$

APPENDIX

In this Appendix we prove

$$\lim_{x\to-\infty} f(\mathbf{x}|\,\mathbf{p}) = \lim_{x\to-\infty} e^{i\mathbf{p}\cdot\mathbf{x}} = \frac{-2\pi i}{xp\sin\theta} e^{ipx}\,\delta(\theta-\lambda)\,\delta(\phi-\sigma), \tag{A1}$$

where we use optical coordinates for $\mathbf{x}$ and $\mathbf{p}$:

$$\mathbf{x} = x(\sin\lambda\,\cos\sigma,\ \sin\lambda\,\sin\sigma,\ \cos\lambda),$$
$$\mathbf{p} = p(\sin\theta\,\cos\phi,\ \sin\theta\,\sin\phi,\ \cos\theta). \tag{A2}$$

We start with

$$\lim_{x\to\pm\infty} e^{i\mathbf{p}\cdot\mathbf{x}} = -\frac{-2\pi i}{xp\sin\theta} e^{ipx}\,\delta(\theta-\lambda)\,\delta(\phi-\sigma). \tag{A3}$$

Equation (A3) is a well-known one which shows how plane waves can be represented asymptotically as spherical waves in the sense of distributions. It is usually proved using saddle-point methods. A heuristic method is used in Ref. 4 to prove the relation in terms of ordinary spherical coordinates.

From Eqs. (18a') and (55') and the above equations,

$$\lim_{x\to-\infty} \langle \mathbf{x}|H_0, A_0;E,a,\theta,\phi\rangle = -2\pi i[(2\pi)^{3/2} E^{1/4}(2\sin\theta)^{1/2}]^{-1} \times \frac{a}{x}\delta(\theta-\lambda)\,\delta(\phi-\sigma)\,[\exp(iaE^{1/2}x)]. \tag{A4}$$

Now the outgoing eigenfunction $\psi_-(\mathbf{x}|\mathbf{p})$ is written in terms of the energy–angle representation as

$$\langle \mathbf{x}|\,H,A;E,a,\theta,\phi\rangle_- = E^{1/4}[(\sin\theta)/2]^{1/2}\,\psi_-(\mathbf{x}|\mathbf{p}), \tag{A5}$$

where $\psi_-(\mathbf{x}|\mathbf{p})$ is given by Eq. (18').

We now want to let $x\to-\infty$ in $\langle \mathbf{x}|\,H,A;E,a,\theta,\phi\rangle_-$. In the usual elementary treatments of scattering, the asymptotic limit is expressed as the sum of a plane wave and an outgoing spherical wave whose amplitude is simply related to the scattering operator. However, in the sense of distributions, the plane wave portion of the asymptotic limit should be expressed in terms of spherical waves using Eq. (A4). Expressing the amplitude of the outgoing spherical wave in terms of the scattering operator

$$\lim_{x\to-\infty} \langle \mathbf{x}|\,H,A;E,a,\theta,\phi\rangle_- = -2\pi i[(2\pi)^{3/2} E^{1/4}(2\sin\lambda)\,x]^{-1} \times [\delta_{a,1}\, e^{iE^{1/2}x}\,\delta(\theta-\lambda)\,\delta(\phi-\sigma) - e^{-iE^{1/2}x}\langle -1,\lambda,\sigma|S(E)|a,\theta,\phi\rangle]. \tag{A6}$$

In Eq. (A6), $\langle a',\theta',\phi'|S(E)|a,\theta,\phi\rangle$ is the reduced scattering operator defined in Eq. (58').

In analogy to Eqs. (58') and (38') we define $\langle a,\theta,\phi\,|\mu_-(E)|a',\theta',\phi'\rangle$ and $\langle a,\theta,\phi\,|\mu_-^{-1}(E)|a',\theta',\phi'\rangle$ by

$$\langle H_0, A_0;E,a,\theta,\phi\,|M_-|\,H_0, A_0;E',a',\theta',\phi'\rangle = \delta(E-E')\langle a,\theta,\phi\,|\mu_-(E)|a',\theta',\phi'\rangle,$$
$$\langle H_0, A_0;E,a,\theta,\phi\,|\,M_-^{-1}|H_0, A_0;E',a',\theta',\phi'\rangle = \delta(E-E')\langle a,\theta,\phi\,|\,\mu_-^{-1}(E)|a',\theta',\phi'\rangle, \tag{A7}$$

respectively. We have from $M_-^{-1}M_- = I$ and from Eq. (46') the following two results, respectively:

$$\sum_{a''}\int_0^{2\pi} d\phi''\int_0^{\pi/2} d\theta''\,\langle a,\theta,\phi\,|\mu_-^{-1}(E)|\,a'',\theta'',\phi''\rangle \times\langle a'',\theta'',\phi''|\mu_-(E)|a',\theta',\phi'\rangle = \delta_{a,a'}\,\delta(\theta-\theta')\,\delta(\phi-\phi'), \tag{A8}$$

$$\langle a,\theta,\phi\,|\,\mu_-^{-1}(E)|a',\theta',\phi'\rangle = \delta_{a,+1}\,\delta_{a,a'}\,\delta(\theta-\theta')\,\delta(\phi-\phi') + \delta_{a,-1}\langle -1,\theta,\phi\,|\,S(E)|a',\theta',\phi'\rangle. \tag{A9}$$

Thus,

$$\lim_{x\to-\infty} \langle \mathbf{x}|\,H,A;E,a,\theta,\phi\rangle_- = \lim_{x\to-\infty}\sum_{a'}\int_0^{2\pi} d\phi'\int_0^{\pi/2} d\theta' \times\langle \mathbf{x}|\,H_0, A_0;a',\theta',\phi'\rangle \times\langle a',\theta',\phi'|\,\mu_-^{-1}(E)|a,\theta,\phi\rangle. \tag{A10}$$

From the first of Eqs. (33'),

$$\langle \mathbf{x}|\,H,A;E,a,\theta,\phi\rangle = \sum_{a'}\int_0^{2\pi} d\phi'\int_0^{\pi/2} d\theta'\,\langle \mathbf{x}|\,H,A;E,a',\theta',\phi'\rangle_- \times\langle a',\theta',\phi'|\mu_-(E)|a,\theta,\phi\rangle. \tag{A11}$$

Finally, from Eqs. (A10), (A11), and (A8),

$$\lim_{x\to-\infty} \langle \mathbf{x}|\,H, A;\,E,a,\theta,\phi\rangle = \lim_{x\to-\infty} \langle \mathbf{x}|H_0, A_0;E,a,\theta,\phi\rangle, \tag{A12}$$

which is just Eq. (A1).

[1]H.E. Moses, J. Math. Phys. **20**, 1151 (1979).
[2]H.E. Moses, J. Math. Phys. **16**, 1044 (1975).
[3]I. Kay and H.E. Moses, Nuovo Cimento **2**, 917 (1955).
[4]I. Kay and H.E. Moses, Nuovo Cimento **22**, 689 (1961).
[5]I. Kay and H.E. Moses, Commun. Pure Appl. Math. **3**, 435 (1961).
[6]H.E. Moses, Phys. Rev. **102**, 559 (1956).

Five-term WKBJ approximation

R. N. Kesarwani
Department of Mathematics, University of Ottawa, Ottawa, Canada K1N 9B4
Y.P. Varshni
Department of Physics, University of Ottawa, Ottawa, Canada K1N 9B4

(Received 18 July 1979; accepted for publication 23 August 1979)

An expression is derived for the five-term WKBJ approximation.

I. INTRODUCTION

Only a few problems in quantum mechanics can be solved exactly, and approximation methods are therefore of great practical importance. The WKBJ method[1-4] is one such important method. The single-term (first-order) WKBJ approximation is frequently used in practical applications to bound state problems.

Dunham[5] obtained the second and third nonzero terms in the WKBJ quantization condition. Kreiger *et al.*[6] have used the three-term WKBJ approximation to calculate some eigenvalues for the potentials of the form $V(x) = \lambda x^{2\nu}$. The three-term WKBJ approximation has been used by Kesarwani and Varshni,[7] and by Kirschner and Le Roy,[8] to calculate the eigenvalues for the Lennard–Jones potential; the former authors evaluated the contour integrals analytically, while the latter used a quadrature procedure. The three-term WKBJ method has been applied also to a quartic potential with a finite binding energy.[9] The fourth term in the WKBJ approximation has been derived by the authors[10] and the four-term WKBJ method has been used successfully to calculate the eigenvalues of high accuracy for the Lennard–Jones potential.[10]

In the present paper we derive an expression for the fifth nonzero term in the WKBJ approximation. The final expression for the five-term WKBJ approximation is put in a form such that the integrands occurring in the expression are free of nonintegrable singularities.

II. DERIVATION

We start with the time-independent Schrödinger equation,

$$\frac{d^2\psi}{dx^2} + \frac{2\mu}{\hbar^2}[E - V(x)]\psi = 0\,, \tag{1}$$

where the symbols have their usual meaning. The WKBJ approximation consists in seeking a solution in the form,

$$\psi = \exp\left(\frac{i}{\hbar}\int^x y(x,E,\hbar)\,dx\right), \tag{2}$$

where $y(x,E,\hbar)$ is represented by an asymptotic series expansion

$$y(x,E,\hbar) = \sum_{s=0}^{\infty}\left(\frac{\hbar}{i}\right)^s y_s(x,E)\,. \tag{3}$$

The $y_s = y_s(x,E)$ are determined in succession from

$$y_0 = \pm[2\mu(E-V)]^{1/2} \tag{4}$$

and the system of differential equations

$$y'_{n-1} = -\sum_{m=0}^{n} y_{n-m}y_m \tag{5}$$

obtained by inserting (2) in (1) and equating to zero the coefficients of successive powers of $\hbar$. Here the prime denotes the differentiation with respect to x. The recursive formulas for y_n for odd and even n take the following forms:

$$y_{2k+1} = \begin{cases} -y'_0/2y_0 & \text{if } k=0 \\ [-y_2/2y_0]' & \text{if } k=1 \\ [-y_{2k}/2y_0]' - (1/y_0)\sum\limits_{s=2}^{k} y_s y_{2k+1-s} & \text{if } k=2,3,\dots, \end{cases} \tag{6}$$

$$y_{2k} = \begin{cases} [-y_1/2y_0]' + y_1^2/2y_0 & \text{if } k=1 \\ [-y_3/2y_0]' - y_2^2/2y_0 & \text{if } k=2 \\ [-y_{2k-1}/2y_0]' - (1/y_0)\sum\limits_{s=2}^{k-1} y_s y_{2k-s} - y_k^2/2y_0 & \text{if } k=3,4,\dots\,. \end{cases}$$

The discrete energy levels are obtained from the quantization condition

$$(v+\tfrac{1}{2})h = \oint_c y_0\,dx + \sum_{s=2}^{\infty}\left(\frac{\hbar}{i}\right)^s\oint_c y_s(x)\,dx\,, \tag{7}$$

where v is the quantum number, and the domain of x is the complex plane cut along the real axis between the classical turning points and the integration is carried along a contour c enclosing the classical turning points but no other singularities of the integrands, and not crossing the cut.

We note here that the series expansion (3) is only semiconvergent and cannot yield an exact solution in all cases.[11,12] Care must be given to the degree of precision with which Eq. (7) leads to satisfactory approximations to the energy levels. It is expected, however, that the approximation will be good whenever the terms of the series on the right-hand side of (7) diminish rapidly; and that an exact result is obtained in case the series involved is convergent.[11] For the calculation of energy eigenvalues for potentials which have two classical turning points, the WKBJ approximation holds best when the two classical turning points are well separated (i.e., for high quantum numbers) and the potential is slowly varying. For most diatomic potentials for which $V(r) \propto (r-r_e)^2$ in the neighborhood of the minimum, the WKBJ approximation works very well. However, if one is dealing with a rapidly varying potential, the series in (7) may not be convergent for very small quantum numbers, and it should be terminated at an appropriate term.

Simplifying Eqs. (6) successively, we obtain $y_1 = -y'_0/2y_0$ together with

$$2y_3 = [-y_2/y_0]',$$
$$2y_5 = [\tfrac{1}{2}(y_2/y_0)^2 - y_4/y_0]',$$
$$2y_7 = [-\tfrac{1}{3}(y_2/y_0)^3 - y_6/y_0 + y_4 y_2/y_0^2]',$$
$$2y_9 = [\tfrac{1}{4}(y_2/y_0)^4 - y_8/y_0 + (y_6 y_2 + \tfrac{1}{2}y_4^2)/y_0^2 - y_4 y_2^2/y_0^3]',$$
$$2y_{11} = [-\tfrac{1}{5}(y_2/y_0)^5 - y_{10}/y_0 + (y_8 y_2 + y_6 y_4)/y_0^2 - (y_6 y_2^2 + y_4^2 y_2)/y_0^3 + y_4 y_2^3/y_0^4]',$$
$$2y_{13} = [\tfrac{1}{6}(y_2/y_0)^6 - y_{12}/y_0 + (y_{10}y_2 + y_8 y_4 + \tfrac{1}{2}y_6^2)/y_0^2 - (y_8 y_2^2 + 2y_6 y_4 y_2 + \tfrac{2}{3}y_4^3)/y_0^3 + (y_6 y_2^3 + \tfrac{3}{2}y_2^2 y_4^2)/y_0^4 - y_4 y_2^4/y_0^5]',$$

etc., and

$$2y_2 = [-y_1/y_0]' + y_1^2/y_0,$$
$$2y_4 = [-y_3/y_0]' - y_2^2/y_0,$$
$$2y_6 = [-y_5/y_0 + y_3 y_2/y_0^2]' + y_3^2/y_0 + y_2^3/y_0^2,$$
$$2y_8 = [-y_7/y_0 + y_5 y_2/y_0^2 - y_3 y_2^2/y_0^3]' - y_4^2/y_0 - 3y_3^2 y_2/y_0^2 - y_2^4/y_0^3,$$
$$2y_{10} = [-y_9/y_0 + (y_7 y_2 + y_5 y_4)/y_0^2 - (y_5 y_2^2 + \tfrac{1}{6}y_3^3)/y_0^3 + y_3 y_2^3/y_0^4]' + y_5^2/y_0 + 3y_4^2 y_2/y_0^2 + \tfrac{9}{2}y_3^2 y_2^2/y_0^3 + y_2^5/y_0^4,$$
$$2y_{12} = [-y_{11}/y_0 + (y_9 y_2 + y_7 y_4)/y_0^2 - (y_7 y_2^2 + 2y_5 y_4 y_2 - \tfrac{1}{2}y_4^2 y_3)/y_0^3 + y_5 y_2^3/y_0^4 - y_3 y_2^4/y_0^5]' - y_6^2/y_0 - (3y_5^2 y_2 - 2y_4^3)/y_0^2 - \tfrac{9}{2}y_4^2 y_2^2/y_0^3 - 7y_3^2 y_2^3/y_0^4 - y_2^6/y_0^5,$$

etc.

Since $y_3, y_5, y_7, \ldots$ reduce to become derivatives the integrals $\oint_c y_3\,dx$, $\oint_c y_5\,dx$, $\oint_c y_7\,dx, \ldots$ along a closed contour vanish. It is only y_0 and $y_2, y_4, y_6, \ldots$ that contribute nonzero values to the right-hand side of (7).

On considering contribution due to terms up to y_9, the quantization condition (7) takes the form

$$v + \tfrac{1}{2} = I_1 + I_2 + I_3 + I_4 + I_5, \tag{8}$$

where I_1, I_2, I_3, I_4, and I_5 represent the contributions due to y_0, y_2, y_4, y_6, and y_8, respectively, and are given by

$$I_1 = \frac{(2\mu)^{1/2}/\hbar}{2\pi}\oint_c (E-V)^{1/2}\,dx,$$
$$I_2 = -\frac{\hbar/(2\mu)^{1/2}}{2^6\pi}\oint_c V'^2(E-V)^{-5/2}\,dx,$$
$$I_3 = -\frac{[\hbar/(2\mu)^{1/2}]^3}{2^{12}\pi}\oint_c [49V'^4 - 16V'V'''(E-V)^2] \times (E-V)^{-11/2}\,dx,$$
$$I_4 = -\frac{[\hbar/(2\mu)^{1/2}]^5}{2^{17}\pi}\oint_c [1675V'^6(E-V)^{-17/2} + 4020V'^4V''(E-V)^{-15/2} + 48(20V'^3V''' + 49V'^2V''^2)(E-V)^{-13/2} + 64(18V'V''V''' - V''^3)(E-V)^{-11/2} + 128V'''^2(E-V)^{-9/2}]\,dx,$$

and

$$I_5 = -\frac{[\hbar/(2\mu)^{1/2}]^7}{2^{24}\pi}\oint_c [1115525V'^8(E-V)^{-23/2} + 3569680V'^6V''(E-V)^{-21/2} + 32(29140V'^5V''' + 107637V'^4V''^2)(E-V)^{-19/2} + 128(1105V'^4V^{(4)} + 11476V'^3V''V''' + 7466V'^2V''^3)(E-V)^{-17/2} + 256(884V'^2V''V^{(4)} + 754V'^2V'''^2 + 848V'V''^2V''' + 365V''^4)(E-V)^{-15/2} + 2048(28V'V'''V^{(4)} - 3V''V'''^2 + 19V''^2V^{(4)})(E-V)^{-13/2} + 4096V^{(4)^2}(E-V)^{-11/2}]\,dx. \tag{9}$$

The nonintegrable singularities at the classical turning points in the integrands of I_2, I_3, I_4, I_5 can be removed using the method of Kreiger *et al.*[6] and (8) can be written in an equivalent form with I_1, I_2, I_3, I_4, and I_5 given by

$$I_1 = \frac{(2\mu)^{1/2}/\hbar}{\pi}\int_{r_1}^{r_2}(E-V)^{1/2}\,dx,$$
$$I_2 = -\frac{\hbar/(2\mu)^{1/2}}{4!\,\pi}\frac{d}{dE}\int_{r_1}^{r_2}V''(E-V)^{-1/2}\,dx.$$
$$I_3 = \frac{[\hbar/(2\mu)^{1/2}]^3}{6!\,4\pi}\frac{d^3}{dE^3}\int_{r_1}^{r_2}(7V''^2 - 5V'V''') \times (E-V)^{-1/2}\,dx,$$
$$I_4 = -\frac{[\hbar/(2\mu)^{1/2}]^5}{9!\,2\pi}\left[\frac{d^5}{dE^5}\int_{r_1}^{r_2}(93V''^3 - 224V'V''V''' + 35V^{(4)}V'^2)(E-V)^{-1/2}\,dx + 216\frac{d^4}{dE^4}\int_{r_1}^{r_2}V'''^2(E-V)^{-1/2}\,dx\right],$$

and

$$I_5 = \frac{[\hbar/(2\mu)^{1/2}]^7}{10!\,48\pi}\left[\frac{d^7}{dE^7}\int_{r_1}^{r_2}(1143V''^4 + 2065V''^2V^{(4)} - 175V'^3V^{(5)})(E-V)^{-1/2}\,dx - \frac{d^6}{dE^6}\int_{r_1}^{r_2}(352V''V'''^2 + 6511V''^2V^{(4)}) \times (E-V)^{-1/2}\,dx - 20\frac{d^5}{dE^5}\int_{r_1}^{r_2}(29V^{(4)^2} + 173V'''V^{(5)})(E-V)^{-1/2}\,dx\right]. \tag{10}$$

Here r_1 and r_2 are, respectively, the smaller and larger positive roots of $E - V(x) = 0$, and $V^{(n)}$ represents the nth derivative of V.

Equation (8) in conjunction with (10) can be used for calculating the eigenvalues for suitable potentials to a high degree of accuracy.

ACKNOWLEDGMENT

This work was supported in part by a research grant from the Natural Sciences and Engineering Research Council of Canada to one of the authors (Y.P.V.).

[1]G. Wentzel, Z. Phys. **38**, 518 (1926).
[2]H.A. Kramers, Z. Phys. **39**, 828 (1926).
[3]L. Brillouin, C.R. Acad. Sci. **183**, 24 (1926); J. Phys. Radium **7**, 353 (1926).

[4]H. Jeffreys, Proc. London Math. Soc. Ser. 2 **23**, 428 (1925).
[5]J.L. Dunham, Phys. Rev. **41**, 713, 721 (1932).
[6]J.B. Krieger, M.L. Lewis, and C. Rosenzweig, J. Chem. Phys. **47**, 2942 (1967).
[7]R.N. Kesarwani and Y.P. Varshni, Can. J. Phys. **56**, 1488 (1978).
[8]S.M. Kirschner and R.J. LeRoy, J. Chem. Phys. **68**, 3139 (1978).
[9]R.N. Kesarwani and Y.P. Varshni, Z. Naturforsch. **33a**, 1581 (1978).
[10]R.N. Kesarwani and Y.P. Varshni, Can. J. Phys., submitted for publication.
[11]G.D. Birkhoff, Bull. Am. Math. Soc. **39**, 696 (1933).
[12]E.C. Kemble, *The Fundamental Principles of Quantum Mechanics* (Dover, New York, 1958); this is a reprint of the 1937 edition published by McGraw-Hill.

The rosette of rosettes of Hilbert spaces in the indefinite metric state space of the quantized Maxwell field

W. Gessner and V. Ernst

Sektion Physik der Universität München, Theresienstrasse 37, D 8000 München 2, Germany

(Received 28 April 1978; revised manuscript received 18 January 1979)

The indefine metric space S_m of the covariant form of the quantized Maxwell field M is analyzed in some detail. S_m contains not only the familiar pre-Hilbert space L^0 which occurs in the Gupta–Bleuler formalism of the *free M*, but a whole rosette of continuously many, isomorphic, complete, *pre-Hilbert space* L^q disjunct up to the zero element of S_m. The L^q are the maximal subspaces of S_m which allow the usual statistical interpretation. Each L^0 corresponds uniquely to one square integrable, spatial distribution $j^0(\mathbf{x})$ of the total charge $Q = 0$. If M is in any state from L^q, the bare charge $j^0(\mathbf{x})$ appears to be inseparably dressed by the quantum equivalent of its proper, classical Coulomb field $\mathbf{E}(\mathbf{x})$. The vacuum occurs only in the state space L^0 of the free Maxwell field. Each L^q contains a secondary rosette of continuously many, up to o disjunct, isomorphic *Hilbert spaces* H_g^q related to different electromagnetic gauges. The space H_0^q, which corresponds to the Coulomb gauge within the Lorentz gauge, plays a physically distinguished role in that only it leads to the usual concept fo energy. If M is in any state from H_g^q, the bare 4-current $j^0(\mathbf{x}), \mathbf{j}(\mathbf{x})$, where $\mathbf{j}(\mathbf{x})$ is any square integrable, transverse current density in space, is endowed with its proper 4-potential which depends on the chosen gauge, and with its proper, gauge independent, Coulomb–Oersted field $\mathbf{E}(\mathbf{x})$, $\mathbf{B}(\mathbf{x})$. However, these fields exist only in the sense of quantum mechanical expectation values equipped with the corresponding field fluctuations. So they are basically different from *classical* electromagnetic fields.

1. INTRODUCTION

The purpose of this work is a careful extension of the familiar Gupta–Bleuler formalism[1-9] of the free electromagnetic or Maxwell field M in the direction of interaction.

We remember that the canonical quantization of the relativistically invariant form of M can be realized only[1-9] on a state space $\mathscr{S}_M$ with an indefinite scalar product[10] $\langle\cdot|\cdot\rangle$ or metric. Gupta[1] and Bleuler[2] obtained an acceptable quantum theory of the free Maxwell field by restricting the formal theory on $\mathscr{S}_M$ to the familiar subspace $\mathscr{L}^0$ which contains only states of transverse photons with proper "admixtures" of longitudinal and scalar photons.[1-6] These "good ghosts"[8] are chosen so that the restriction $[\cdot|\cdot]$ of $\langle\cdot|\cdot\rangle$ to $\mathscr{L}^0$ is positive semidefinite. $\mathscr{L}^0$ is a *pre-Hilbert space* if we agree[11,12] that this means a complex vector space endowed with a positive semidefinite, sesquilinear[9-12] form $[\cdot|\cdot]$ so that the Cauchy–Schwarz inequality holds[9-12] on it. The theory on $\mathscr{L}^0$ therefore allows either the direct Born interpretation, or the equivalent construction[9-12] of a Hilbert space $\overline{\mathscr{H}}^0$ whose elements are equivalence classes of elements of $\mathscr{L}^0$. To avoid confusion we assume that the definition of a *Hilbert space* includes separability, completeness, and a positive definite scalar product or Hilbert metric $(\cdot|\cdot)$. These conventions apply also to the abstract.

Only a few results exist for the case of interaction.[3,7-9] The only way to a quantum theory of an interacting system seems to lead over the well known[13]

Interaction postulate: The state space $\mathscr{S}_{M+D}$ of an interacting system $M + D$ is the tensor product[13]

$$\mathscr{S}_{M+D} := \mathscr{S}_M \otimes \mathscr{S}_D \tag{1}$$

of the state spaces $\mathscr{S}_M$ and $\mathscr{S}_D$ of the respective interaction partners M and D.

This has such important consequences for the structure of the coupled system $M + D$ and the potential measurements on it that the postulate can be rightfully regarded as the chief characteristics of the notion of interaction altogether. For example, the conventional form[3-6,14] of relativistic quantum electrodynamics (QED), the theory of the interaction of M with the quantized Dirac field D, contains silently an analogous assumption.

In nonrelativistic theories[13,15] one usually assumes that both factors in (1) are Hilbert spaces. As this cannot hold for $\mathscr{S}_M$, the postulate must be reconsidered in relativistic theories like QED. A first choice for the space of *physical* states of $M + D$ is certainly $\mathscr{L}^0 \otimes \mathscr{S}_D$. However, since transverse photons and/or good ghosts[8] cannot[3] mediate the quantum equivalent of the Coulomb interaction between the quanta of D, this choice must be abandoned. So it is necessary to study states of M with admixtures of bad ghosts[8] which are not in $\mathscr{L}^0$.

In particular, we wish to learn whether $\mathscr{S}_M$ contains subspaces other than $\mathscr{L}^0$ which are also pre-Hilbert or Hilbert spaces, what these spaces mean physically, and how they are geometrically related to $\mathscr{L}^0$. The answers must be

 0022-2488/80/010093-19$01.00

sought by means of an analysis because, in close analogy to a Fock space,[14–18] the structure of $\mathcal{S}_M$ is mainly determined by the commutation relations

$$\begin{aligned} &[\Pi_\mu(\mathbf{x}), A_\nu(\mathbf{y})] = -ig_{\mu\nu}\,\delta(\mathbf{x}-\mathbf{y}), \\ &[\Pi_\mu(\mathbf{x}),\Pi_\nu(\mathbf{y})] = 0, \quad [A_\mu(\mathbf{x}),A_\nu(\mathbf{y})] = 0, \end{aligned} \tag{2}$$

of the canonical dynamical variables of M which are to be represented on $\mathcal{S}_M$. We use natural units so that $\hbar = 1 = c$, the Minkowski metric $g_{\mu\mu} = g^{\mu\mu} = g(\mu,\mu) = (-1,1,1,1,)$, and the usual summation convention so that, e.g., $\Box := \partial_\mu\,\partial^\mu = \nabla^2 - \partial^2/\partial t^2$.

The space $\mathcal{S}_M$ will be analyzed by studying the pseudointeraction of M with any prescribed, conserved, classical 4-current $j^\mu(x) = j^\mu(t,\mathbf{x})$, in close analogy to corresponding examples[15,16,19,20] on a Fock space. This is possible because in classical electrodynamics (CED) the "abstract" 4-current J plays the role of a parameter (cf. Sec. 2) whose single values $j^\mu(x)$ can be prescribed within the condition $\partial_\mu j^\mu(x) = 0$. The theory of $M + J$ is therefore defined only on $\mathcal{S}_M$, and so it is not characteristic of the actual interaction problem of QED which requires some equivalent of (1). However, it excellently serves as a physical illustration and for the mathematical parametrization of the structure of $\mathcal{S}_M$ which, being mainly determined by (2), is prior to any actual interaction!

The details on the structure of $\mathcal{S}_M$ are mainly needed for future extensions of the systematic *WW*-approach[15] to relativistic field theories like QED. In particular, they allow, and indeed necessitate, a greater flexibility in the formulation of equivalents of (1) (Sec. 12). Within the limits of the present work they suggest an elegant solution of the gauge problem (Secs. 8–12), and provide us with acceptable quantum equivalents of the classical Coulomb–Oersted fields (Sec. 10),

$$\begin{aligned} \mathbf{E}(\mathbf{x}) &:= -\nabla \int d^3x'\,\frac{j^0(\mathbf{x}')}{4\pi|\mathbf{x}-\mathbf{x}'|}, \\ \mathbf{B}(\mathbf{x}) &:= \nabla\times \int d^3x'\,\frac{\mathbf{j}(\mathbf{x}')}{4\pi|\mathbf{x}-\mathbf{x}'|}, \end{aligned} \tag{3}$$

of the subset $\mathfrak{J} = \{j^\mu(\mathbf{x})\}$ of all stationary, integrable and square integrable, classical 4-currents $j^\mu(\mathbf{x})$ which satisfy

$$Q := \int d^3x\, j^0(\mathbf{x}) = 0, \quad \nabla\mathbf{j}(\mathbf{x}) = 0. \tag{4}$$

The condition $Q = 0$ is very important (Sec. 12).

The most prospective aspect of our results is maybe that the act of the restriction of the theory to any pre-Hilbert space $\mathcal{L}^q$ in $\mathcal{S}_M$, which is necessary to allow the statistical interpretation, provides us at the same time with a natural explanation of the observed *inseparability* of any 4-current $j^\mu(\mathbf{x})\in\mathfrak{J}$ from its above *eigenfield*. This is maybe a caricature of a solution of the confinement problems of QED and QCD. Real electrons or positrons are never bare of their Coulomb eigenfields as assumed in QED for the free quanta of the Dirac field, and the bare quarks of QCD have not been observed either.

In any case, our results allow the comparison of $\mathcal{S}_M$ to a well-ordered warehouse or dressing room where the quantum mechanical eigenfields of all $j^\mu(\mathbf{x})\in\mathfrak{J}$ are kept in store for the case of interaction subject to some equivalent of (1). Actual 4-currents like $j_\mu(x)$ exist then, in general, only as *expectation values* of the Schrödinger or Heisenberg operators $J^\mu(\mathbf{x}), J^\mu(x)$, of the abstract 4-current J in appropriate states of D; compare, e.g., Ref. 15. So the present $j^\mu(\mathbf{x})$ play the role of dummies that wear the potential quantum-mechanical Coulomb–Oersted dresses of actual currents of QED. Our stock of such robes therefore retains full relevance for QED, though some more fitting will probably be necessary.

These prospects compensate by far the repellent features[9] of the indefinite metric on $\mathcal{S}_M$. The latter is mainly a consequence of the break of the symplectic symmetry of the Hamilton formalism by the Lorentz condition, and arises already in CED. This is briefly discussed in Sec. 2. $\mathcal{S}_M$ is defined in Sec. 3. In Sec. 4 we consider three eigenvalue problems which prepare the actual discussion of the structure of $\mathcal{S}_M$ in Sec. 5. Sections 6–12 contain the quoted results on the pseudointeraction of M and J and some further consequences for QED. The remainder of the work consists of the necessary proofs.

2. THE BREAK OF THE SYMPLECTIC SYMMETRY BY THE LORENTZ CONDITION IN CLASSICAL ELECTRODYNAMICS

We first cast an eye on the classical origin of all problems with the quantization of the Maxwell field M.

In any reference system with coordinates $x = (t,\mathbf{x})$ a symplectic covering theory (CCED) of CED can be obtained from the Hamilton functional

$$\begin{aligned} H_t &:= H_t\{\Pi^\mu(\mathbf{x}),\ A_\mu(\mathbf{x})\} \\ &= \tfrac{1}{2}\int d^3x\{\Pi^\mu(\mathbf{x})\Pi_\mu(\mathbf{x}) + [\nabla A^\mu(\mathbf{x})][\nabla A_\mu(\mathbf{x})] \\ &\quad + 2j^\mu(t,\mathbf{x})\,A_\mu(\mathbf{x})\}. \end{aligned} \tag{5}$$

$j^\mu(x) = (j^0(t,\mathbf{x}),\mathbf{j}(t,\mathbf{x}))$ denotes the components of any given, conserved, concrete 4-current which can indeed be prescribed like the value of a parameter J. $\Pi^\mu(\mathbf{x})$ is the canonical momentum amplitude, $A_\mu(\mathbf{x})$ the corresponding canonical position amplitude of M. The Hamilton equations of motion corresponding to H_t read

$$\begin{aligned} &\frac{\partial}{\partial t}\pi^\mu(t,\mathbf{x}) = -[\delta H_t/\delta A_\mu] = -j^\mu(t,\mathbf{x}) + \nabla^2 a^\mu(t,\mathbf{x}), \\ &\frac{\partial}{\partial t}a_\mu(t,\mathbf{x}) = [\delta H_t/\delta \Pi^\mu] = \pi_\mu(t,\mathbf{x}). \end{aligned} \tag{6}$$

The square brackets contain the usual functional derivatives and indicate further that concrete functions $\pi^\mu(t,\mathbf{x})$, $a_\mu(t,\mathbf{x})$ are to be inserted for the abstract variables $\Pi^\mu(\mathbf{x})$, $A_\mu(\mathbf{x})$. Iteration of (6) yields

$$\Box a_\mu(t,\mathbf{x}) = j_\mu(t,\mathbf{x}), \quad \Box\pi_\mu(t,\mathbf{x}) = \frac{\partial}{\partial t}j_\mu(t,\mathbf{x}). \tag{7}$$

Equations (6) and (7) show that CCED unites the symplectic symmetry of the canonical formalism with the Minkowski symmetry of relativity. The symplectic symmetry must be broken, however, if we want to arrive at Maxwell's equations. We achieve this by selecting from the symplectic set $\mathfrak{M}^C = \{a^\mu(x), j^\mu(x)\}$ of *all* solutions of (6) the subset $\mathfrak{M} \subset \mathfrak{M}^C$ of all such pairs which in addition satisfy the Lorentz condition

$$\partial^\mu a_\mu(x) = 0. \tag{8}$$

It is obviously possible to identify the sets $\mathfrak{M}^C$, $\mathfrak{M}$ with the corresponding theories CCED, CED. To any pair $\{a^\mu(x), j^\mu(x)\} \in \mathfrak{M}$ we can namely define an electromagnetic field tensor

$$f_{\mu\nu}(x) := \partial_\mu a_\nu(x) - \partial_\nu a_\mu(x) \tag{9}$$

which satisfies the covariant form of Maxwell's equations

$$\partial_\sigma f_{\mu\nu}(x) + \partial_\mu f_{\nu\sigma}(x) + \partial_\nu f_{\sigma\mu}(x) = 0, \tag{10a}$$

$$\partial^\mu f_{\mu\nu}(x) = \partial^\mu \partial_\mu a_\nu(x) - \partial^\mu \partial_\nu a_\mu(x) = \Box a_\nu(x) - \partial_\nu \partial^\mu a_\mu(x) = j_\nu(x). \tag{10b}$$

Equation (10a) is compatible with both symmetries because it is a consequence of definition (9). Eq. (10b) breaks the symplectic symmetry because it requires (8). The same can be said of current conservation

$$\partial^\mu j_\mu(x) = \partial^\mu \Box a_\mu(x) = \Box \partial^\mu a_\mu(x) = 0, \tag{11}$$

which therefore could be considered as a consequence of the break of the symplectic symmetry by the Lorentz condition. However, as usual, we regard $\partial^\mu j_\mu(x) = 0$ as an *a priori* property of the abstract 4-current J.

The break of the symplectic symmetry is symbolized by the inclusion $\emptyset \subset \mathfrak{M} \subset \mathfrak{M}^C$ which is the main cause of our trouble. The first principles of canonical field quantization[21] replace the conjugated canonical variables $P_m(\mathbf{x})$, $Q_n(\mathbf{x})$ of a Bose field by Schrödinger operators $P_m(\mathbf{x})$, $Q_n(\mathbf{x})$ which satisfy

$$[P_m(\mathbf{x}), Q_n(\mathbf{y})] = -i\,\delta_{mn}\,\delta(\mathbf{x}-\mathbf{y}), \quad \text{etc.} \tag{12}$$

Equations (6) show that $\Pi^\mu(\mathbf{x})$ and $A_\nu(\mathbf{x})$ play the role of $P_m(\mathbf{x})$ and $Q_n(\mathbf{x})$, respectively. So Eq. (12) must be replaced by $[\Pi^\mu(\mathbf{x}), A_\nu(\mathbf{y})] = -i\delta^\mu_\nu\,\delta(\mathbf{x}-\mathbf{y})$, etc., which is equivalent to (2). The respective Heisenberg operators $\Pi_\mu(x)$, $A_\nu(x)$ satisfy the corresponding equal time commutation relations. Since the classical variables $\Pi^\mu(\mathbf{x})$, $A_\nu(\mathbf{x})$ vary over the covering set $\mathfrak{M}^C$, the corresponding Schrödinger or Heisenberg operators are representatives of this $\mathfrak{M}^C$. They belong therefore to a *formal* quantum theory T^C which is more general than the desired quantum theory T corresponding to CED. This suggests that T might be some restriction of T^C.

Equations (6) show in particular that *co*- and *contra*variant variables of M are to be conjugated canonically. This is the actual origin of the factor $g_{\mu\nu}$ in (2) which necessitates the introduction of an indefinite metric $\langle\cdot|\cdot\rangle$ on $\mathscr{S}_M$. If the desired theory T is to allow the statistical interpretation, this *forces* us to look out for restrictions of T^C to some Hilbert or pre-Hilbert space in $\mathscr{S}_M$. Since the structure of $\mathscr{S}_M$ is mainly determined by (2), it was not trivial that this program could be realized altogether for the free Maxwell field,[1,2] and it is still less trivial that infinitely many other, physically prospective realizations are possible, as we shall see here.

We note finally that the quantities $\mathscr{S}_M$, $A_\mu(\mathbf{x})$, $\Pi_\mu(\mathbf{x})$ should have been ascribed to a generalization M^C of the actual Maxwell field M, but this linguistic distinction is not relevant.

3. THE STATE SPACE OF THE QUANTIZED MAXWELL FIELD

We define now the space $\mathscr{S}_M$ which carries an appropriate representation of the canonical commutation relations (2), and consider some important operators on it.

Denote by K^0 the empty set $\emptyset$, and by K^n the set $(\mathbf{k}_1,\mu_1;\dots;\mathbf{k}_n,\mu_n)$ of $n = 1,2,\cdots$ pairs of real variables $\mathbf{k}_\nu$, μ_ν, where any $\mathbf{k}_\nu$ varies continuously over the $\mathbb{R}^3$ and any μ_ν assumes the values 0, 1, 2, 3. Hereafter, define the symbol $\int dK^n$ by

$$\int dK^n \cdots := \int d^3k_1 \cdots d^3k_n \sum_{\mu_1=0}^{3} \cdots \sum_{\mu_n=0}^{3} \cdots, \tag{13}$$

where $\int d^3k$ denotes the elementary Lebesgue integral over $\mathbb{R}^3$. If necessary,[9] a weight function like $[\,|\mathbf{k}_1| \cdots |\mathbf{k}_n|\,]^{-1/2}$ can be included in the definition of $\int dK^n \cdots$. Consider also the *covering Fock space* $\mathscr{F}^C$ of all sequences

$$\alpha := \{\alpha_0(K^0),\ \alpha_1(K^1),\ \alpha_2(K^2), \cdots\} \tag{14}$$

with the following properties: The nth component of α, $\alpha_n = \alpha_n(K^n)$, is a complex number for $n = 0$, and for $n = 1,2,\cdots$ it is a complex-valued function, symmetric in the pairs $(\mathbf{k}_\nu, \mu_\nu)$, and defined in such a way that $\int dK^n\,|\alpha_n(K^n)|^2$ exists. The Hilbert scalar product on $\mathscr{F}^C$ is given by

$$(\beta\,|\,\alpha) = \beta_0^*\alpha_0 + \sum_{n=1}^{\infty} \int dK^n\,\beta_n^*(K^n)\alpha_n(K^n) \tag{15}$$

and exists for any $\alpha, \beta \in \mathscr{F}^C$ if $\alpha \in \mathscr{F}^C$ means $\|\alpha\| := (\alpha|\alpha)^{1/2} < \infty$, as usual. We ssume that $\mathscr{F}^C$ has been completed already.

The norm $\|\cdots\|$ on $\mathscr{F}^C$ defines in particular a complete Banach space $\mathscr{B}$. On this Banach space we introduce a second scalar product[9]

$$\langle\beta\,|\,\alpha\rangle := \beta_0^*\alpha_0 + \sum_{n=1}^{\infty} \int dK^n\,\beta_n^*(K^n)\,g(\mu_1,\mu_1)\cdots \times g(\mu_n,\mu_n)\,\alpha_m(K^n), \tag{16}$$

and *define* $\mathscr{S}_M$ as the pair

$$\mathscr{S}_M := (\mathscr{B}, \langle\cdot\,|\,\cdot\rangle). \tag{17}$$

This definition includes a close analogy to the Minkowski space $\mathscr{S}_M := (\mathbb{R}^4, xy)$, with $xy := x^\mu g_{\mu\nu} y^\nu$. The norm $\|\cdots\|$ on $\mathscr{B}$ plays the role of the Euclidean distance on $\mathbb{R}^4$ in that it determines a natural topology on $\mathscr{S}_M$. The latter is used *locally* to define the identity of points ($\alpha = \beta$ if $\|\alpha - \beta\| = 0$), and in the usual notions of the Cauchy convergence of a sequence, and of the density and completeness of arbitrary sets of points $\alpha \in \mathscr{S}_M$. The Banach topology is employed *globally* to define the domain $\mathbb{D}(\Omega) = \{\alpha : \|\Omega\alpha\| < \infty\}$ of an operator Ω on $\mathscr{S}_M$. The indefinite metric $\langle\cdot|\cdot\rangle$ on $\mathscr{S}_M$, the analog of xy on $\mathscr{S}_M$, is used for the definition of the adjoint Ω^+ of Ω (so that $\langle\beta|\Omega^+\alpha\rangle = \langle\Omega\beta|\alpha\rangle = \langle\alpha|\Omega\beta\rangle^*$) and for the computation of matrix elements $\langle\beta|\Omega\alpha\rangle$ and expectation values $\langle\alpha|\Omega\alpha\rangle$.

$\mathscr{S}_M$ contains the *zero-vector* $o := \{0,0,\cdots\}$ and the *nor-*

malized vacuum $\omega := \{1,0,0,\cdots\}$. $\langle\alpha|\beta\rangle$ exists for any α, $\beta\in\mathscr{S}_M$ and satisfies $|\langle\alpha|\beta\rangle|\leqslant\|\alpha\|\,\|\beta\|$ but, in general, *not* $|\langle\alpha|\beta\rangle|^2\leqslant|\langle\alpha|\alpha\rangle\langle\beta|\beta\rangle|$. $\mathscr{S}_M$ is nondegenerate[10] and decomposable,[7,10] but this will not be needed here.

In very close analogy to Fock spaces[14,16] we define destruction and creation operators $a_\mu(\mathbf{k})$, $a_\mu^+(\mathbf{k})$ by

$$(a_\mu(\mathbf{k})\alpha)_n(K^n) = (n+1)^{1/2}\alpha_{n+1}(\mathbf{k},\mu;K^n), \quad n=0,1,\cdots, \tag{18a}$$

$$(a_\mu^+(\mathbf{k})\alpha)_n(K^n) = \begin{cases} 0, & n=0, \\ \dfrac{1}{n^{1/2}}\, n \displaystyle\sum_{\nu=1}^{n} g_{\mu\mu_\nu}\delta(\mathbf{k}-\mathbf{k}_\nu)\alpha_{n-1}(K^n\setminus(\mathbf{k}_\nu,\mu_\nu)), & n=1,2,\ldots, \end{cases} \tag{18b}$$

with $(\mathbf{k},\mu;K^n) := (\mathbf{k},\mu;\mathbf{k}_1,\mu_1;\ldots;\mathbf{k}_n,\mu_n)$ and $K^n\setminus(\mathbf{k}_\nu,\mu_\nu) := (\mathbf{k}_1,\mu_1;\ldots;\mathbf{k}_{\nu-1},\mu_{\nu-1};\ \mathbf{k}_{\nu+1},\mu_{\nu+1};\ldots;\ \mathbf{k}_n,\mu_n)$. These operators satisfy

$$[a_\mu(\mathbf{k}),\, a_{\mu'}^+(\mathbf{k}')] = g_{\mu\mu'}\,\delta(\mathbf{k}-\mathbf{k}'),$$
$$[a_\mu(\mathbf{k}),\, a_{\mu'}(\mathbf{k}')] = 0 = [a_\mu^+(\mathbf{k}),\, a_{\mu'}^+(\mathbf{k}')] \tag{19}$$

and are formal adjoints of each other relative to $\langle\cdot|\cdot\rangle$. This holds in the same sense as for the corresponding creation and destruction operators on a Fock space. We also face the similar problem that $a_\mu^+(\mathbf{k})$ exists only in the sense of a distribution whereas $a_\mu(\mathbf{k})$ has for given $\mathbf{k}$ and any value of μ a dense domain $\mathbb{D}(a_\mu(\mathbf{k}))$. $\mathbb{D}(a_\mu(\mathbf{k}))$ namely encloses the dense set $\mathbb{D}_c$ of all α with a finite number of nonvanishing, everywhere continuous components $\alpha_n(K^n)$. However, $a_\mu(\mathbf{k})$ seems not to be *closable* for given μ and $\mathbf{k}$ because $a_\mu^+(\mathbf{k})$ would otherwise exist in an ordinary sense.

4. THREE EIGENVALUE PROBLEMS ON THE STATE SPACE OF THE MAXWELL FIELD

The dense core $\mathbb{D}_c$ of all $\mathbb{D}(a_\mu(\mathbf{k}))$ allows the unambiguous definition of four eigenvalue problems on $\mathscr{S}_M$ which lead us in a natural way to the main results of this work. Three of them will be considered here, the last one in Sec. 10.

It is convenient to introduce for any $\mathbf{k}\neq 0$ the usual[4–8] polarization 4-vectors $e^\mu_{(\sigma)} = e^\mu_{(\sigma)}(\mathbf{k})$, $\sigma=0,1,2,3$, as defined by $e^0_{(0)}=1$, $e^1_{(0)}=e^2_{(0)}=e^3_{(0)}=0$, $e^0_{(1)}=e^0_{(2)}=e^0_{(3)}=0$, $k_r e^r_{(1)} = k_r e^r_{(2)} = 0$, $e_{r(1)}e^r_{(2)}=0$, $e_{r(1)}e^r_{(1)} = e_{r(2)}e^r_{(2)} = 1$, and $e^r_{(3)} = k^r/|\mathbf{k}|$ for $r=1,2,3$. $e_{\mu(\sigma)}$ is given by $g_{\mu\nu}\,e^\nu_{(\sigma)}$, k^μ by $(|\mathbf{k}|,\mathbf{k})$, and repeated indices r indicate the sum over $r=1,2,3$. These vectors satisfy the orthogonality and completeness relations

$$e^\mu_{(\sigma)}\, g_{\mu\nu} e^\nu_{(\tau)} = g_{\sigma\tau}, \quad e^\mu_{(\sigma)}\, g^{\sigma\tau}\, e^\nu_{(\tau)} = g^{\mu\nu}, \tag{20}$$

$$k_\mu e^\mu_{(\sigma)} = |\mathbf{k}|(\delta^3_\sigma - \delta^0_\sigma). \tag{21}$$

With their help we introduce the annihilation operators

$$a_{(\sigma)}(\mathbf{k}) := e^\mu_{(\sigma)} a_\mu(\mathbf{k}), \quad \sigma = 0,1,2,3, \tag{22}$$

$$a_g(\mathbf{k}) := 2^{-1/2}[a_{(3)}(\mathbf{k}) - a_{(0)}(\mathbf{k})],$$
$$a_b(\mathbf{k}) := 2^{-1/2}[a_{(3)}(\mathbf{k}) + a_{(0)}(\mathbf{k})]. \tag{23}$$

Together with the corresponding creation operators they satisfy the relations $[a_{(\sigma)}(\mathbf{k}),a^+_{(\tau)}(\mathbf{k}')] = g_{\sigma\tau}\,\delta(\mathbf{k}-\mathbf{k}')$, etc., but $[a_{(g)}(\mathbf{k}),a^+_{(g)}(\mathbf{k}')] = [a_{(b)}(\mathbf{k}),a^+_{(b)}(\mathbf{k}')] = 0$, $[a_{(g)}(\mathbf{k}),a^+_{(b)}(\mathbf{k}')] = [a_{(b)}(\mathbf{k}),a^+_{(g)}(\mathbf{k}')] = \delta(\mathbf{k}-\mathbf{k}')$, etc., $a_{(1)}$ and $a_{(2)}$ destruct transverse photons, $a_{(3)}$ annihilates longitudinal, $a_{(0)}$ scalar photons. Due to the unusual commutation relations we say that a_g destructs a good, a_b a bad ghost.[8] It is namely easy to show that states α_g or α_b which respectively contain only good or only bad ghosts, satisfy $\langle\alpha_g\mid\alpha_g\rangle = 0 = \langle\alpha_b\mid\alpha_b\rangle$, but in general $\langle\alpha_g\mid\alpha_b\rangle\neq 0$. The distinction between good and bad ghosts will be motivated below. The domains of all operators (22), (23) are dense for any given $\mathbf{k}$ because they also contain the common core $\mathbb{D}_c$.

The first eigenvalue problem on $\mathscr{S}_M$ is defined either by

$$a_{(0)}(\mathbf{k})\,\alpha = 0 = a_{(3)}(\mathbf{k})\,\alpha \text{ for almost any } \mathbf{k}\in\mathbb{R}^3, \tag{24a}$$

or by

$$a_g(\mathbf{k})\,\alpha = 0 = a_b(\mathbf{k})\,\alpha \text{ for almost any } \mathbf{k}\in\mathbb{R}^3. \tag{24b}$$

We shall see in Sec. 5 that the linear space $\mathscr{F}^{\mathrm{tr}}$ of all eigenvectors α in the sense of (24) is actually a Hilbert space. For obvious reasons we call it the Fock space of transverse photons without any admixtures of ghosts or longitudinal and/or scalar photons.

The second eigenvalue problem on $\mathscr{S}_M$ is analogous to the definition of the modified vacuum states[16] or fully coherent states[22] on a Fock space. In analogy we define[23] the coherent states in $\mathscr{S}_M$ as eigenstates of the destruction operator $a_\mu(\mathbf{k})$ to the complex eigenvalue $c_\mu(\mathbf{k})$, i.e., by the equation

$$a_\mu(\mathbf{k})\,\alpha = c_\mu(\mathbf{k})\,\alpha \quad \text{for any } \mu \text{ and almost any } \mathbf{k}\in\mathbb{R}^3. \tag{25}$$

It is easily seen that α satisfies (25) if its zero component α_0 is any complex number, and its other components are of the form

$$\alpha_n(\mathbf{k}_1,\mu_1;\ldots;\mathbf{k}_n,\mu_n) = \alpha_0\, c_{\mu_1}(\mathbf{k}_1)\ldots c_{\mu_n}(\mathbf{k}_n)/(n!)^{1/2}, \quad n=1,2,\cdots. \tag{26}$$

One can show that all solutions of (25) are of this form, but this is not relevant here. α is in $\mathscr{S}_M$ if $c := c_\mu(\mathbf{k})$ is any point from the space $\mathfrak{C}$ of complex 4-vectors $c_\mu(\mathbf{k})$ whose components are square integrable in $\mathbf{k}$. For any coherent α we have namely

$$0\leqslant\langle\alpha\mid\alpha\rangle = |\alpha_0|^2\exp\int d^3k\, c^*_\mu(\mathbf{k})c^\mu(\mathbf{k}) \leqslant \|\alpha\|^2 = |\alpha_0|^2\exp\int d^3k \sum_{\mu=0}^{3}|c_\mu(\mathbf{k})|^2. \tag{27}$$

It is remarkable and important for later that $\langle\alpha|\alpha\rangle$ is positive[23] if $\alpha_0\neq 0$. The vacuum ω is coherent and corresponds to $c_\mu(\mathbf{k}) = 0$, other examples will be considered further below.

Let us finally define the *Lorentz operator* $L(\mathbf{k})$ by

$$L(\mathbf{k}) := k^\mu a_\mu(\mathbf{k}) = |\mathbf{k}|\cdot[a_{(3)}(\mathbf{k}) + a_{(0)}(\mathbf{k})] = \sqrt{2}\,|\mathbf{k}|\,a_b(\mathbf{k}). \tag{28}$$

Its domain for given $\mathbf{k}$, $\mathbb{D}(L(\mathbf{k}))$, is dense because it encloses a dense subset of $\mathbb{D}_c$. The pseudointeraction of M and J will in a natural way lead us to:

The third eigenvalue problem on $\mathscr{S}_M$, the generalized Lorentz condition

$$L(\mathbf{k})\,\alpha = q(\mathbf{k})\,\alpha \quad \text{for almost any } \mathbf{k}. \tag{29}$$

The eigenvalue $q = q(\mathbf{k})$ is any given, complex-valued function over the $\mathbb{R}^3$. We restrict the discussion to the index set $\mathfrak{Q}$ of all $q(\mathbf{k})$ given by

$$q(\mathbf{k}) = D(\mathbf{k}) \int d^3x\, e^{-i\mathbf{k}\mathbf{x}} j^0(\mathbf{x}). \tag{30}$$

$D(\mathbf{k})$ collects the usual relativistic weight factor $(2|\mathbf{k}|)^{-1/2}$ and the π's from the Fourier transformation, i.e.,

$$D(\mathbf{k}) := (16\pi^3|\mathbf{k}|)^{-1/2}. \tag{31}$$

$j^0(\mathbf{x})$ is the charge component of any $j^\mu(\mathbf{x}) \in \mathfrak{J}$ as specified in Sec. 1. The condition $Q = 0$ warrants the square integrability of $q(\mathbf{k})$ and $q(\mathbf{k})/|\mathbf{k}|$, and guarantees further on that $q(\mathbf{k})$ behaves for small $|\mathbf{k}|$ like $k^r d_r/|\mathbf{k}|^{1/2} = |\mathbf{k}|^{1/2} d_r k^r |\mathbf{k}|$, where d_r are the components of some finite vector $\mathbf{d}$. The equation

$$k^\mu c_\mu(\mathbf{k}) = q(\mathbf{k}) \tag{32}$$

therefore has a nonempty set of nontrivial solutions $c_\mu(\mathbf{k}) \in \mathfrak{C}$. The span[10] of the corresponding coherent states is obviously in the *Lorentz space* $\mathscr{L}^q$, the eigenspace of the Lorentz operator $L(\mathbf{k})$ to the eigenvalue $q = q(\mathbf{k})$. So $\mathscr{L}^q$ is not empty. These Lorentz spaces are the main objects of our analysis.

5. THE DOUBLE ROSETTE OF HILBERT SPACES IN $\mathscr{S}_M$

We are ready now to formulate the main results of this work. The proofs will be given in Secs. 13–16.

Theorem 1: *Any Lorentz space $\mathscr{L}^q$, $q \in \mathfrak{Q}$, is complete relative to the Banach norm $\|\cdots\|$ on $\mathscr{S}_M$. Any pair $\mathscr{L}^q$, $\mathscr{L}^{q'}$ of different Lorentz spaces, $q, q' \in \mathfrak{Q}$, $q \neq q'$, satisfies*

$$\mathscr{L}^q \cap \mathscr{L}^{q'} = \{o\}. \tag{33}$$

Theorem 2: *To any pair $\mathscr{L}^q$, $\mathscr{L}^{q'}$ of (different) Lorentz spaces exist bijective mappings $\mathscr{L}^q \leftrightarrow \mathscr{L}^{q'}$ which satisfy $\langle \alpha | \beta \rangle = \langle \alpha' | \beta' \rangle$ if $\alpha', \beta' \in \mathscr{L}^{q'}$ are the respective images of any $\alpha, \beta \in \mathscr{L}^q$.*

Theorem 3: *The restriction $[\cdot|\cdot]$ of the indefinite scalar product $\langle \cdot|\cdot \rangle$ on $\mathscr{S}_M$ to any Lorentz space $\mathscr{L}^q$, $q \in \mathfrak{Q}$, is positive semidefinite. Each $\mathscr{L}^q$ is therefore a pre-Hilbert space which by Theorem 1 is complete relative to the Banach norm on $\mathscr{S}_M$.*

Theorem 4: *The scalar product $\langle \cdot|\cdot \rangle$ on $\mathscr{S}_M$ does not obey the Cauchy–Schwarz inequality on any subspace of $\mathscr{S}_M$ which contains the span[10] of any pair of different Lorentz spaces $\mathscr{L}^q$, $\mathscr{L}^{q'}$.*

Equation (33) evokes the picture of a *rosette* whose leaves $\mathscr{L}^q$ are connected only in the point o. The completeness of each leaf guarantees that this picture will not be questioned by completion processes relative to the Banach norm on $\mathscr{S}_M$ that might arise in later applications. According to Theorems 1–3 the leaves of this *primary rosette* are isomorphic, complete, pre-Hilbert spaces. Since the Cauchy–Schwarz inequality holds on any Lorentz space $\mathscr{L}^q$, any $\mathscr{L}^q$ can be used as the state space of a conventional quantum theory. If the superposition principle[24] is to hold for this theory, any $\mathscr{L}^q$ is a *maximal* subspace of $\mathscr{S}_M$ which can serve this purpose, according to Theorem 4. We shall see in Sec. 11 that this has important physical consequences.

The last expression (29) for $L(\mathbf{k})$ shows that any $\mathscr{L}^q$ is also an eigenspace of the destruction operator $a_b(\mathbf{k})$ of bad ghosts. The space $\mathscr{L}^0$ which corresponds to $q = 0$, i.e., $(q(\mathbf{k}) \equiv 0)$, is the only Lorentz space without bad ghosts. The other leaves $\mathscr{L}^q$ do contain bad ghosts, but these admixtures are *controlled* by (29). According to Theorem 4, a state α from the sum of two different Lorentz spaces, i.e., an α of the form $\alpha = \alpha^q + \alpha^{q'}$, $o \neq \alpha^q \in \mathscr{L}^q$, $o \neq \alpha^{q'} \in \mathscr{L}^{q'}$, $q \neq q'$, may have a negative norm square $\langle \alpha | \alpha \rangle$. This means: If the admixture of bad ghosts is no longer controlled by (29), we get "bad behavior." This justifies the notion of *bad* ghosts, but as long as they remain bottled in any $\mathscr{L}^q$, like the ghost in Aladdin's lamp, they behave well and actually build up the quantum mechanical Coulomb robes of our dummies $j^\mu(\mathbf{x})$ as we shall see later on.

The good ghosts are not affected by (30) so that the states in any $\mathscr{L}^q$ contain arbitrary admixtures of them. They are responsible for a typical *substructure* of any $\mathscr{L}^q$, as revealed by

Theorem 5: *Each Lorentz space $\mathscr{L}^q$, $q \in \mathfrak{Q}$, contains continuously many subspaces $\mathscr{H}^q_g$ which satisfy*

$$\mathscr{H}^q_g \cap \mathscr{H}^q_{g'} = \{o\} \text{ for } g \neq g'. \tag{34}$$

The restriction $(\cdot|\cdot)$ of the semipositive scalar product $[\cdot|\cdot] = \langle \cdot|\cdot \rangle$ on $\mathscr{L}^q$ to any $\mathscr{H}^q_g$ is a positive definite Hilbert scalar product, and each $\mathscr{H}^q_g$ is indeed a Hilbert space. To each $\mathscr{H}^q_g$ exist isometric bijective mappings onto $\mathscr{H}^0_0 = \mathscr{F}^{\mathrm{tr}}$, the Fock space of transverse photons. The indices $g = g(\mathbf{k})$ vary over the set $\mathfrak{G}$ of complex gauge function $g(\mathbf{k})$ defined by the requirement $k_\mu g(\mathbf{k}) \in \mathfrak{C}$ for $\mu = 0,1,2,3$.

This shows that any primary leaf $\mathscr{L}^q$ contains a *secondary rosette* of isometric Hilbert spaces $\mathscr{H}^q_g$ which are also connected only in the point o. We shall see that different $\mathscr{H}^q_g$ correspond to different electromagnetic *gauges*.

Equations (33) and (34) evoke the picture of a rosette of rosettes, or a double rosette of Hilbert spaces in $\mathscr{S}_M$. The title of this work and this section refer to this picture. As ω is in $\mathscr{F}^{\mathrm{tr}} = \mathscr{H}^0_0$, the relations (33), (34) imply that *only one of all these Hilbert spaces is a Fock space with vacuum*. All Hilbert spaces $\mathscr{H}^q_g \neq \mathscr{H}^0_0$, as well as all pre-Hilbert spaces $\mathscr{L}^q \neq \mathscr{L}^0$, contain no vacuum relative to the commutation relations (2)!

The leaf $\mathscr{L}^0$ agrees with the familiar[1–9] Gupta-Bleuler space of transverse photons with proper admixtures of longitudinal and scalar photons. It is of course known[1–7] that different admixtures of good ghosts correspond to different gauges, but it seems to have escaped attention that proper elements of $\mathscr{L}^0$ can be grouped together to a Hilbert space $\mathscr{H}^0_g$ for any gauge. The Lorentz spaces $\mathscr{L}^q \neq \mathscr{L}^0$ and the gauge Hilbert space $\mathscr{H}^q_g$ in any $\mathscr{L}^q$ seem to be new. It appears that *controlled* admixtures of bad ghosts act like a salutary medicine so that the usual elimination[8] of all bad ghosts might be premature.

6. THE TIME EVOLUTION GENERATED ON $\mathscr{S}_M$ BY THE PSEUDOINTERACTION OF THE MAXWELL FIELD WITH ANY PRESCRIBED, CONSERVED, CLASSICAL 4-CURRENT $j_\mu(x)$

We noted already that the Schrödinger operators of a canonical quantum theory represent the dynamical *variables* of the corresponding classical theory. So they are prior to Heisenberg operators which actually are usually expressed in terms of Schrödinger operators. But we may of course *define* the latter in view of some desired Heisenberg operators. In this sense we write down the Schrödinger operators of the 4-potential, the related momentum, and of the electromagnetic field in the point $\mathbf{x}$,

$$A_\mu(\mathbf{x}) := \int d^3x\, D(\mathbf{k})[e^{i\mathbf{kx}}a_\mu(\mathbf{k}) + e^{-i\mathbf{kx}}a_\mu^+(\mathbf{k})], \tag{35a}$$

$$\Pi_\mu(\mathbf{x}) := \int d^3x\, D(\mathbf{k})[e^{i\mathbf{kx}}(-i|\mathbf{k}|a_\mu(\mathbf{k})) + e^{-i\mathbf{kx}}(i|\mathbf{k}|a_\mu^+(\mathbf{k}))], \tag{35b}$$

$$F_{\mu\nu}(\mathbf{x}) := \int d^3x\, D(\mathbf{k})[e^{i\mathbf{kx}}(ik_\mu a_\nu(\mathbf{k}) - ik_\nu a_\mu(\mathbf{k})) + \text{c.c.}], \tag{36}$$

cf. (31).

The usual Heisenberg operators of the *free* Maxwell field agree with these Schrödinger operators at $t = 0$, but the latter remain of course unchanged in the case of interaction. $A_\mu(\mathbf{x})$ and $\Pi_\mu(\mathbf{x})$ satisfy (2) in quite the same formal sense as the corresponding commutation relations of a proper quantum field theory are satisfied on a Fock space. This justifies the definition (17) of $\mathscr{S}_M$. In the following we ignore that these field operators have only a formal mathematical meaning. They become mathematically well-defined operators on $\mathscr{S}_M$ if they are smeared out, cf. Sec. 14. The pseudointeraction of M and J is defined by the Hamilton operator

$$H_t := \int d^3k\, [|\mathbf{k}|a_\mu^+(\mathbf{k})a^\mu(\mathbf{k}) + a_\mu^+(\mathbf{k})I^\mu(t,\mathbf{k}) + a^\mu(\mathbf{k})I_\mu^*(\mathbf{k})], \tag{37}$$

where $I^\mu(t,\mathbf{k})$ is given by

$$I^\mu(t,\mathbf{k}) := D(\mathbf{k}) \int d^3x\, j^\mu(t,\mathbf{x})e^{-i\mathbf{kx}}. \tag{38}$$

H_t has been obtained in the usual way by inserting (35) into (5) and omitting the so-called zero-point energy. It defines the time evolution operator $U(t)$ of the system $M + J$ by

$$i\frac{d}{dt}U(t) = H_t U(t), \quad U(0) = 1. \tag{39}$$

$U(t)$ satisfies $U^{-1}(t) = U^+(t)$, but in general $U(-t) \neq U^+(t)$, and is explicitly given in Sec. 17. The Heisenberg operators corresponding to (35), (36) are then determined by

$$A_\mu(x) := A_\mu(t,\mathbf{x}) := U^+(t)\, A_\mu(\mathbf{x})\, U(t),$$
$$\Pi_\mu(x) := U^+(t)\, \Pi_\mu(\mathbf{x})\, U(t), \tag{40}$$

$$F_{\mu\nu}(x) := U^+(t)\, F_{\mu\nu}(\mathbf{x})\, U(t), \tag{41}$$

and depend on H_t, as it must be.

We say that any operator Ω can be restricted in a natural way to some $\mathscr{L}^q$ if $\Omega\alpha$ is in $\mathscr{L}^q$ for any $\alpha \in \mathscr{L}^q \cap \mathbb{D}(\Omega)$. It is easy to show that Ω can be restricted to *any* $\mathscr{L}^q$ if $[L(\mathbf{k}), \Omega]$ vanishes for almost any $\mathbf{k}$. Ω can be restricted to some particular $\mathscr{L}^q$ if $[L(\mathbf{k}), \Omega]\alpha = 0$ for almost any $\mathbf{k}$ and any α from that $\mathscr{L}^q$. If both conditions are violated, Ω maps any $\mathscr{L}^q$ onto a set $\Omega\mathscr{L}^q$ not completely contained in $\mathscr{L}^q$, and Ω cannot be restricted to any $\mathscr{L}^q$ in this natural way. We consider some examples:

By straightforward computation we find, cf. Sec. 17,

$$[L(\mathbf{k}), A_\mu(x)] = k_\mu\, D(\mathbf{k})\, e^{-ikx}, \tag{42a}$$

$$[L(\mathbf{k}), \Pi_\mu(x)] = ik_\mu\, |\mathbf{k}|\, D(\mathbf{k})\, e^{-ikx}, \tag{42b}$$

$$[L(\mathbf{k}), F_{\mu\nu}(x)] = 0, \tag{43}$$

$$[L(\mathbf{k}), H_t] = |\mathbf{k}|\Big[L(\mathbf{k}) - \Big(I^0(t,\mathbf{k}) - \frac{i}{|\mathbf{k}|}\frac{\partial}{\partial t}I^0(t,\mathbf{k})\Big)\Big], \tag{44}$$

$$[L(\mathbf{k}), U(t)] = U(t)\{e^{-i|\mathbf{k}|t}[L(\mathbf{k}) - I^0(0,\mathbf{k})] - [L(\mathbf{k}) - I^0(t,\mathbf{k})]\}. \tag{45}$$

Equations (42)–(45) give a somewhat puzzling picture. By (42), the canonical variables $A_\mu(\mathbf{x})$, $\Pi_\mu(\mathbf{x})$ cannot be restricted to any $\mathscr{L}^q$. Equation (43) shows that $F_{\mu\nu}(\mathbf{x})$ behaves much better in that it can be restricted to *any* $\mathscr{L}^q$. This shows for the first time some merits of the definition (36). By (44), H_t can be restricted, in general, to only one $\mathscr{L}^q$, but the index q depends on t. As $j^0(t,\mathbf{x})$ is real, this dependence on t vanishes if and only if $j^0(t,\mathbf{x})$ is independent of t. In that case, both $H_t = H$ and $U(t)$ can be restricted by (44), (45) to the $\mathscr{L}^q$ whose q is related to the stationary charge density $j^0(\mathbf{x})$ by Eq. (30).

Note that $q(\mathbf{k})$ agrees with $I^0(t,\mathbf{k}) = I^0(\mathbf{k})$ in this case. In the general case we therefore also use the notation $I^0(t,\mathbf{k}) = q[t](\mathbf{k})$ which indicates that a time-dependent charge distribution $j^0(t,\mathbf{x})$ defines a curve $q[t] = q(t,\mathbf{k}) = I^0(t,\mathbf{k})$ in the index space $\mathfrak{Q}$. From (45) we easily get the relation $L(\mathbf{k}) \times (U(t)\alpha) = U(t)q[t](\mathbf{k})\alpha = q[t](\mathbf{k})(U(t)\alpha)$ for any $\alpha \in \mathscr{L}^{q[0]}$ which already proves the following:

Statement 1: *$U(t)$ maps the Lorentz space $\mathscr{L}^{q[0]}$ which by Eq. (30) corresponds to the initial charge density $j^0(0,\mathbf{x})$, onto the Lorentz space $\mathscr{L}^{q[t]}$ which by the same equation corresponds to the charge density $j^0(t,\mathbf{x})$ prescribed at time t. We write*

$$\mathscr{L}^{q[t]} = U(t)\mathscr{L}^{q[0]}. \tag{46}$$

This means that the evolution $\mathscr{L}^{q[0]} \to \mathscr{L}^{q[t]}$, as given by the prescribed evolution of $j^0(t,\mathbf{x})$, and the time evolution $\mathscr{L}^{q[0]} \to U(t)\mathscr{L}^{q[0]}$, as generated by H_t, are *compatible*. However, this result must not be overestimated, because we also have (see Sec. 17)

Statement 2: *For $j^0(t,\mathbf{x}) \neq j^0(0,\mathbf{x})$ one can find in $\mathscr{L}^{q[0]}$ a state α satisfying $\langle\alpha|\alpha\rangle = 0$, and a state β obeying $\langle\beta|\beta\rangle = 1$, such that $\langle\beta|U(t)\alpha\rangle \neq 0$.*

This means that $U(t)$ does not allow the usual statistical interpretation in this general case.

7. A REALIZATION OF THE OPERATOR MAXWELL EQUATIONS ON A LORENTZ SPACE

Against this puzzling background we consider now the same time evolution in terms of some *physical* quantities.

The Heisenberg operators $\Pi_\mu(x)$, $A_\mu(x)$ satisfy (6), and thus (7), as operator identities on $\mathscr{S}_M$. This is a consequence of the symplectic symmetry which altogether allows the canonical quantization. Since $0 = \partial^\mu j_\mu(x) = \partial^\mu \Box a_\mu(x) = \Box\partial^\mu a_\mu(x)$ holds by virtue of (7) and charge conservation, we also have on $\mathscr{S}_M$ the operator identity

$$\Box\partial^\mu A_\mu(x) = 0. \tag{47}$$

However, as $\Pi_\mu(x)$ and $A_\mu(x)$ represent the covering set $\mathfrak{M}^C$ of CED, $\partial^\mu A_\mu(x) = 0$ cannot hold as an operator identity on $\mathscr{S}_M$ because this would contradict the inclusion $\mathfrak{M} \subset \mathfrak{M}^C$. Equations (9) and (10) also require some additional attention because the *canonical* definition (41) of $F_{\mu\nu}(x)$ could be in conflict with the classical definition (9). However, we need not worry about this because we find in Sec. 18:

Statement 3: *The Heisenberg operators $A_\mu(x)$ and $F_{\mu\nu}(x)$ as defined in* (40) *and* (41), *satisfy the classical relation* (9) *as an operator identity on $\mathscr{S}_M$.*

The homogeneous Maxwell equations (10a) are therefore satisfied by $F_{\mu\nu}(x)$ as operator identities on $\mathscr{S}_M$.

So we must concern ourselves only with the Lorentz condition (8). In this connection we prove in Sec. 18 the following

Statement 4: *The matrix elements $\langle\alpha|A_\mu(x)\beta\rangle$ of $A_\mu(x)$ relative to any states α,β from the Lorentz space $\mathscr{L}^{q[0]}$, satisfy the Lorentz condition* (8).

The inhomogeneous Maxwell equation (10b) is therefore satisfied by all matrix elements $[\alpha|\,F_{\mu\nu}(x)\beta\,]$ relative to any states α, β of $\mathscr{L}^{q[0]}$. Since $F_{\mu\nu}(x)$ can in particular be restricted to $\mathscr{L}^{q[0]}$, its matrix elements can be formed by means of the positive semidefinite scalar product $[\cdot|\cdot]$ on $\mathscr{L}^{q[0]}$. Since (10a) holds even as operator identity on $\mathscr{S}_M$, it certainly holds also for all matrix elements $[\alpha|\,F_{\mu\nu}(x)\beta\,]$ relative to $\alpha,\beta\in\mathscr{L}^{q[0]}$. So we get

Statement 5: *The restriction $F^0_{\mu\nu}(x)$ of $F_{\mu\nu}(x)$ to the Lorentz space $\mathscr{L}^{q[0]}$ exists and satisfies the covariant form* (10) *of Maxwell's equations as an operator identity on $\mathscr{L}^{q[0]}$.*

We shall see below that $F_{\mu\nu}(x)$ is also *gauge invariant.* All this is the more acceptable as it has been achieved on a *maximal* subspace of $\mathscr{S}_M$ (Theorem 4) which permits the usual Born interpretation. But the result is still problematic because $U(t)$ cannot be statistically interpreted in this general case, and because $\mathscr{L}^{q[0]}$ depends sensitively on the initial time $t = 0$ in the arbitrary reference system X chosen. A change of X by a Lorentz transformation $x_\mu \to x'_\mu$, $j_\mu(x) \to j'_\mu(x')$ leads therefore to a new $j'_0(0,\mathbf{x}')$ and thus to a new, though isomorphic, Lorentz space $\mathscr{L}^{q'[0]}$. However, as $Q = 0$ holds in any X, a Lorentz transformation will at least not lead out of our rosette of Lorentz spaces. It remains to be seen how this can be reconciled with the usual[9] formulations of relativistic quantum theories in terms of representations of the Poincaré group. We shall not further investigate these problems because a part of them vanishes if J is also quantized. The role of $j^0(t,\mathbf{x})$ is then partly taken over by the Schrödinger operator $J^0(\mathbf{x})$ which is independent of t in any X, like $A_\mu(\mathbf{x})$ and $F_{\mu\nu}(\mathbf{x})$.

8. A QUANTUM MECHANICAL FOUNDATION OF ELECTROMAGNETIC GAUGES

We consider now some results which illustrate the rosette of Hilbert spaces $\mathscr{H}^q_g$ in any $\mathscr{L}^q$. Note that any $q = q[0]\in\mathfrak{Q}$ can be chosen by an appropriate choice of $j^0(0,\mathbf{x})$.

On any $\mathscr{L}^q$ we can construct in the usual way[9,11,12] the Hilbert space $\bar{\mathscr{H}}^q := \mathscr{L}^q/\mathscr{N}^q$, where $\mathscr{N}^q$ is the subspace of $\mathscr{L}^q$ which contains only α that satisfy $\langle\alpha|\alpha\rangle = 0$ (the set of all α in $\mathscr{S}_M$ which satisfy $\langle\alpha|\alpha\rangle = 0$, is *not* a linear space[10]). The elements $\bar{\alpha} = \{\alpha\}$ of $\bar{\mathscr{H}}^q$ are the corresponding equivalence classes in $\mathscr{L}^q$. Each class $\{\alpha\}$ contains precisely one element α of each Hilbert space $\mathscr{H}^q_g$ in $\mathscr{L}^q$. This means that the elements α of any particular $\mathscr{H}^q_g$ can be chosen as representatives of the classes $\bar{\alpha}$. $\bar{\mathscr{H}}^q$ is therefore isomorphic to any $\mathscr{H}^q_g$. We write $\alpha\sim\alpha'$ if $\alpha\in\mathscr{H}^q_g$ and $\alpha'\in\mathscr{H}^q_{g'}$, $g\neq g'$, are in the same class $\{\alpha\}$, and we say that α and α' are *gauge equivalent.* Gauge equivalent elements $\alpha\sim\alpha'$, $\beta\sim\beta'$ satisfy $\langle\alpha|\beta\rangle = \langle\alpha'|\beta'\rangle = (\bar{\alpha}|\bar{\beta})$ where $(\cdot|\cdot)$ is the Hilbert scalar product in $\bar{\mathscr{H}}^q$. The elements $\bar{\alpha} = \{\alpha\}$ of $\bar{\mathscr{H}}^q$ are therefore *gauge invariant and* $\bar{\mathscr{H}}^q$ can serve as the state space of a gauge invariant quantum theory. These notions are justified by

Statement 6: *Let α,β and α',β' be normalized elements from any respective Hilbert spaces $\mathscr{H}^{q[0]}_g$, $\mathscr{H}^{q[0]}_{g'}$ in $\mathscr{L}^{q[0]}$, and let them satisfy $\alpha\sim\alpha'$, $\beta\sim\beta'$, so that $\langle\alpha|\beta\rangle = \langle\alpha'|\beta'\rangle = (\bar{\alpha}|\bar{\beta})$. Then*

$$\langle\alpha\mid A_\mu(x)\beta\rangle - \langle\alpha'\mid A_\mu(x)\beta'\rangle = (\bar{\alpha}|\bar{\beta})\partial_\mu\, g(x), \tag{48}$$

where $g(x)$ is a real function satisfying $\Box g(x) = 0$, and

$$[\alpha\mid F_{\mu\nu}(x)\beta\,] - [\alpha'\mid F_{\mu\nu}(x)\beta'\,] = 0. \tag{49}$$

Equation (49) justifies once more the definition (36) of $F_{\mu\nu}(x)$ in that $F_{\mu\nu}(x)$ turns out to be *gauge invariant* whereas $A_\mu(x)$ resembles the typical behavior under gauges known from CED. The statement is therfore the core of a satisfying solution of the gauge problem which deserves some more attention:

In CED we regard two 4-potentials, say $a_\mu(x)$ and $a'_\mu(x)$:$a_\mu(x) + \partial_\mu\, g(x)$ with $\Box g(x) = 0$, as equivalent because both lead to the same $f_{\mu\nu}(x)$. But from the point of view of the canonical formalism, $a_\mu(x)$ and $a'_\mu(x)$ are only different values that have been assumed at some time t by *one and the same* dynamical variable $A_\mu(\mathbf{x})$, called the 4-potential in the point $\mathbf{x}$. As a dynamical variable, the 4-potential is gauge invariant *ab ovo.* Since the process of canonical quantization associates *one* Schrödinger operator to any classical *variable,* the Schrödinger operator $A_\mu(\mathbf{x})$ of the 4-potential in the point $\mathbf{x}$ is also gauge invariant *ab ovo.* The same holds of course for the conjugated momentum $\Pi_\mu(\mathbf{x})$ and for all given functionals of $A_\mu(\mathbf{x})$ and $\Pi_\mu(\mathbf{x})$, like any Hamiltonian $H\{\Pi_\mu(\mathbf{x}),A_\nu(\mathbf{x})\}$. Since Hamilton and Schrödinger operators determine all Heisenberg operators in a unique way, the latter are gauge invariant as well.

If this is accepted, different gauges show up only in connection with different values $a_\mu(x)$, $a'_\mu(x)$ assumed at time t by the same variable $A_\mu(\mathbf{x})$. In quantum theory these values correspond to the expectation values $\langle\alpha(t) \mid A_\mu(\mathbf{x})\alpha(t)\rangle$, $\langle\alpha'(t) \mid A_\mu(\mathbf{x})\alpha'(t)\rangle$ of the *same* operator $A_\mu(\mathbf{x})$ in *different states*, $\alpha(t) = U(t)\alpha(0)$, $\alpha'(t) = U(t)\alpha'(0)$, of the quantized system M at time t. In the Heisenberg picture these expectation values are given by $\langle\alpha(0) \mid A_\mu(x)\alpha(0)\rangle$, $\langle\alpha'(0) \mid A_\mu(x)\alpha'(0)\rangle$ if the fit to the Schrödinger picture is made at $t = 0$, as usual. In any case, different gauges correspond only to different expectation values of the same operator in different states.

How can we realize such a theory if different gauges still describe the same physical situation? The only physically indistinguishable, but mathematically different states on the unit sphere of a Hilbert space are points on one *ray* which differ by a constant phase factor of modulus 1. These states lead always to the same expectation values of the 4-potential. So the state space of a quantum theory which accounts for different gauges by different expectation values of the same variable $A_\mu(\mathbf{x})$, cannot be a Hilbert space. Otherwise we would admit that states on different rays, which according to the principles of quantum mechanical measurements[13,24,25] can always be distinguished by means of such measurements, lead only to different gauges. In other words, we would acknowledge that purely quantum mechanical measurements allow a distinction between different gauges.

It is therefore surprising that the structure of $\mathscr{S}_M$, which, we repeat, is primarily determined by (2), provides us automatically with the means to avoid this. Statement 6 describes the technical details of this solution. It suggests in particular to consider the equivalence classes $\{\alpha\}$ of elements of $\mathscr{L}^q$ as generalizations of the concept of a *ray* in a Hilbert space. In this sense our secondary rosette of Hilbert spaces in any pre-Hilbert space $\mathscr{L}^q$ allows not less than a fully quantum mechanical foundation of electromagnetic gauges.

9. COMPARISON WITH THE OPERATOR CONCEPT OF GAUGES

In view of its close relation to the canonical formalism the above gauge concept can be called *canonical*. We compare it now with the *operator concept* used frequently in field theory.[9]

Let $u(x)$ solve the Dirac equation

$$[\gamma^\mu(\partial_\mu - iea_\mu(x)) + m]u(x) = 0 \tag{50}$$

for some given potential $a_\mu(x)$. We see immediately that $u'(x) := e^{ieg(x)}u(x)$ is a solution of (50) corresponding to the potential $a'_\mu(x) = a_\mu(x) + \partial_\mu g(x)$. Though $u'(x)$ is not on the same ray as $u(x)$, the transformations $u(x) \to u'(x)$, $a_\mu(x) \to a'_\mu(x)$ are still referred to as gauge transformations. This seems to be the origin of the *operator concept* of gauges,

$$\begin{aligned} &\psi(x) \to \psi'(x) := e^{ieG(x)}\psi(x), \\ &A_\mu(x) \to A'_\mu(x) := A_\mu(x) + \partial_\mu G(x), \end{aligned} \tag{51}$$

where all quantities, $G(x)$ included,[9] are Heisenberg operators. As different Heisenberg operators $A_\mu(x)$, $A'_\mu(x)$ correspond either to different canonical variables of M, or to different Hamiltonians, or both, this suggestive and formally simple concept of gauges is so different from the canonical concept described above that we can reasonably compare only the respective advantages and disadvantages:

(i) The canonical concept is physically acceptable and, within the limits of this work, mathematically realizable. In contrast to this, the operator concept requires[9] the definition of the exponential of an operator valued distribution $G(x)$, of the product of this exponential with the operator valued distribution $\psi(x)$, and of the sum $A_\mu(x) + \partial_\mu G(x)$ of two symmetric, operator valued distributions. So far, not one of these definitions has been given satisfactorily, it seems.

(ii) In the canonical concept the commutation relations are gauge invariant *ab ovo*. In the operator concept the commutation relations for $A_\mu(x)$ and $G(x)$ must be postulated *ad hoc* if the canonical quantization principle,[21,26] so far the only successful quantization method, is to be abandoned. If it is retained, we must accept either some *ad hoc* specification of the relations between the different, but gauge-equivalent variables $A_\mu(\mathbf{x})$, $A'_\mu(\mathbf{x})$, or between the corresponding Hamiltonians, or both. Ambiguities exist in both approaches.

(iii) The operator concept shows ambiguities also with respect to the interaction postulate. In the first instance the transformations (51) should obviously be realized on $\mathscr{S}_M \otimes \mathscr{S}_D$. But we may as well think of the space $\mathscr{S}_M \otimes \mathscr{S}_G \otimes \mathscr{S}_D$ where $\mathscr{S}_G$ is the state space of the gauge field G whose field operator $G(x)$ acquires only so the same status as $A_\mu(x)$ and $\psi(x)$. In any case it is unclear what remains of (51) if the theory is later restricted to an appropriate subspace of physical states of the respective systems $M + D$ or $M + G + D$. This restriction is a necessity in any case.

(iv) We admit of course that the advantages of the canonical concept are restricted so far to the pseudointeraction with classical currents whereas the operator concept reflects an important symmetry of the actual interaction problem. However, a realization of (51) only for expectation values might also be acceptable in this case. In particular, this avoids problems similar to (ii) which occur if the Dirac field is quantized by means of a *canonical* formalism,[26] as usual.

(v) Charge conservation, which is also deeply interwoven with gauges,[9] also admits a simpler solution in the canonical formalism, cf. Sec. 12.

10. A QUANTUM THEORY OF THE ELECTROMAGNETIC FIELD OF A STATIONARY 4-CURRENT, AND A GAUGE INVARIANT APPROACH TO THE CONCEPT OF ENERGY

We assume now that a reference system X exists in which $j^0(t,\mathbf{x})$ is independent of t, $j^0(t,\mathbf{x}) = j^0(\mathbf{x})$. A time dependence of $\mathbf{j}(t,\mathbf{x})$ would not cause any troubles, but we omit it for the sake of simplicity. So we have $j^\mu(\mathbf{x}) \in \mathfrak{J}$, cf. Sec. 1. The results for this stationary case illustrate other aspects of the physical meaning of the double rosette of Hilbert spaces.

The Hamilton operator $H_t = H$ of the pseudointeraction of J and M can now assume stable *eigenvalues* and *eigenstates*, and so we hit on:

The fourth eigenvalue problem on $\mathcal{S}_M$,

$$H\beta = L\beta. \tag{52}$$

This is solved in Sec. 19 and yields

Statement 7: *H has precisely one eigenstate β in $\mathcal{S}_M$. β is fully coherent and lies in the Lorentz space $\mathcal{L}^q$ where q is related to $j^0(\mathbf{x})$ by* (30). *The corresponding eigenvalue L is given by*

$$L := -\tfrac{1}{2}\int d^3x\,[\mathbf{B}(\mathbf{x})^2 - \mathbf{E}(\mathbf{x})^2], \tag{53}$$

where $\mathbf{B}(\mathbf{x})$ *and* $\mathbf{E}(\mathbf{x})$ *are given in* (3).

We need not wonder that in a relativistic theory the eigenvalue L is not an *energy*, but a *total Lagrangian*, i.e., an integral over an invariant[27] of M. It is noteworthy, however, that the corresponding eigenstate is automatically coherent and thus in a Lorentz space, cf. Sec. 4.

As a consequence of the stationarity of the present case we get in Sec. 19 the following, certainly not unexpected

Statement 8: *The time evolution operator $U(t)$ which is generated by H, satisfies the condition*

$$U(t_1 + t_2) = U(t_1)U(t_2) \text{ for any } t_1,t_2 \tag{54}$$

of a unitary group.[28]

By (45) and Statement 1, $U(t)$ can be restricted to $\mathcal{L}^q$. We denote this restriction, which also satisfies (54), by $U^q(t)$. Statement 8 suggests to look out for subspaces of $\mathcal{L}^q$ which are Hilbert spaces so that the theorem of Stone[28] can be used for the construction of generators which play the role of restricted Hamiltonians. In this connection we obtain the following statements which are all proven in Secs. 19, 20.

Statement 9: *In the Lorentz space $\mathcal{L}^q$ that corresponds to $j^0(\mathbf{x})$, exists precisely one Hilbert space $\mathcal{H}^q_0$ which contains the eigenstate β of H according to Statement* 7, *and which is mapped onto itself by $U^q(t)$.*

This implies that $U^q(t)$ maps a cospace $\mathcal{H}^q_{g[0]} \neq \mathcal{H}^q_0$ onto *another* cospace $\mathcal{H}^q_{g[t]} \neq \mathcal{H}^q_0$ of $\mathcal{H}^q_0$. It follows therefore that $U^q(t)$ can be further restricted only to $\mathcal{H}^q_0$, and that the restriction $U^q_0(t)$ is a unitary group on a Hilbert space whose generator H^q_0 exists by the theorem of Stone.[28] The eigenstate β of H is simultaneously an eigenstate of H^q_0.

Statement 10: *The generator H^q_0 of $U^q_0(t)$ defines a self-adjoint Hamiltonian H^q_0 on $\mathcal{H}^q_0$ which is related in the usual way to the concept of energy. The eigenstate β corresponds to the lowest energy, and the Lagrangian L defines not more than a natural zero point of the energy of the quantum theory on $\mathcal{H}^q_0$. For any normalized $\alpha \in \mathcal{H}^q_0$ we get*

$$\langle \alpha \mid A_\mu(t,\mathbf{x})\alpha\rangle = -\frac{1}{4\pi}\int d^3x'\, j_\mu(\mathbf{x}')/|\mathbf{x}-\mathbf{x}'| + a_\mu(t,\mathbf{x}), \tag{55}$$

where $a_\mu(x)$ satisfies

$$a_0(x) = 0,\quad \partial^\mu a_\mu(x) = \partial^r a_r(x) = 0 \quad \Box j^0(t,\mathbf{x}) = 0. \tag{56}$$

The transverse wave $a_\mu(x)$ depends on α and vanishes if α is the eigenstate β of H^q_0.

Equations (55) and (56) show explicitly that $\mathcal{H}^q_0$ and the concept of energy are related to the *Coulomb gauge within the Lorentz gauge*. We see further on that the eigenstate indeed plays the role of a *modified vacuum state*.[16] The vacuum ω of the free theory is namely modified by controlled bad ghosts and/or transverse photons which equip the bare charge $j^0(\mathbf{x})$ with its eigen-Coulomb field $\mathbf{E}(\mathbf{x})$, and the bare current $\mathbf{j}(\mathbf{x})$ with its eigen-Oersted field $\mathbf{B}(\mathbf{x})$. However, not only ω, but also *all* other states of the natural Hilbert space $\mathcal{F}^{\text{tr}}$ of the free Maxwell field M, are modified in such a way that the *same* Coulomb–Oersted robe $\mathbf{E}(\mathbf{x})$, $\mathbf{B}(\mathbf{x})$ is obtained in *any* state $\alpha \in \mathcal{H}^q_0$.

Mathematically, this modification consists of a mapping $\mathcal{F}^{\text{tr}} \to \mathcal{H}^q_0$ of $\mathcal{F}^{\text{tr}}$ onto itself if $q = 0$ ($\mathcal{H}^0_0 = \mathcal{F}^{\text{tr}}$), and onto *another*, but isomorphic Hilbert space $\mathcal{H}^q_0$ in $\mathcal{S}_M$ if $q \neq 0$. This has the following consequences: For $q = 0$ we can achieve by a proper choice of α that the second term in (55) compensates the first one *for one instant*, say at $t = 0$, so that $\mathbf{j}(\mathbf{x})$ is *bare for one moment*. Not even this instantaneous bareness is possible for the charge $j^0(\mathbf{x})$ in the case $q \neq 0$. This is maybe an indication that the concept of charge is more fundamental than the concept of transverse current curls which indeed, as *moved charges* depend conceptually on the former. The instantaneous bareness of a transverse current from its magnetic field occurs also in CED.

For $q = 0$ our concept of dressing is equivalent to the idea of "photon binding"[15,20] on an ordinary Fock space where only the modification of *all* states has not been emphasized[15] so much. For the more fundamental case $q \neq 0$ it is superior because now the charge $j^0(\mathbf{x})$ cannot be bare of its Coulomb robe, not even for one instant $t = t_0$.

Note that the inseparability of the bare 4-current $j_\mu(\mathbf{x})$ from its quantum mechanical Coulomb–Oersted robe $\mathbf{E}(\mathbf{x})$, $\mathbf{B}(\mathbf{x})$ has been obtained as a consequence of the *restriction* of the theory from $\mathcal{S}_M$ to $\mathcal{L}^q$, and finally to $\mathcal{H}^q_0$. The former is both *necessary and possible* to allow the statistical interpretation, the latter to obtain the concept of energy.

However, the robe $\mathbf{E}(\mathbf{x})$, $\mathbf{B}(\mathbf{x})$ *exists only in the sense of quantum mechanical expectation values which are endowed with the usual quantum mechanical variances* or *quantum field fluctuations* which always *diverge* if we insist on the field strengths in one "sharp" point $\mathbf{x}$, but converge, in general, if we only ask for the forces exerted on *extended* test charges or test coils, respectively. These *quantum robes* are therefore basically different from the classical Coulomb–Oersted fields $\mathbf{E}(\mathbf{x})$, $\mathbf{B}(\mathbf{x})$. In particular, they cannot be treated like *external classical fields* which are usually represented by unit operators multiplied by the numbers $\mathbf{E}(\mathbf{x})$, $\mathbf{B}(\mathbf{x})$. This would namely lead to zero field fluctuations and deprive us of any chance of a similar dressing of realistic electrons with their respective Coulomb robes, cf. Ref. 15 and Sec. 12.

Let us finally consider the gauge invariant Hilbert space $\overline{\mathcal{H}}^q$ of equivalence classes in $\mathcal{L}^q$ as introduced in Sec. 8. We have

Statement 11: *$U^q(t)$ maps any equivalence class $\{\alpha\}$ in $\mathcal{L}^q$ onto an equivalence class $\{\alpha'\}$ again.*

Therefore, $U^q(t)$ defines on $\overline{\mathcal{H}}^q$ a unitary group $\overline{U}^q(t)$, and the theorem of Stone guarantees the existence of a self-adjoint generator $\overline{H}^q$ which plays the role of a Hamiltonian of $\overline{\mathcal{H}}^q$. As the elements α of a given class $\{\alpha\}$ correspond to different gauges, the triplet $\{\overline{\mathcal{H}}^q, \overline{U}^q(t), \overline{H}^q\}$ defines on any

$\mathscr{L}^q$ an abstract, gauge-invariant quantum theory on a Hilbert space. As any element of $\bar{\mathscr{H}}^q$ can be represented by an element of $\mathscr{H}_0^q$, this theory is isomorphic to the restricted quantum theory on $\mathscr{H}_0^q$. $\bar{H}^q$ corresponds therefore to a gauge-invariant energy! The Gupta–Bleuler origin of this theory shows up only in the lower end of the spectrum of $\bar{H}^q$ which is given by L, but plays no further role *after* the execution of the restriction to $\mathscr{H}_0^q$.

11. A NOTE ON THE DYNAMICS OF DRESSING PROCESSES

Let us finally draw some consequences for the interaction of M with any quantized partner D, where $\mathscr{S}_M$ is replaced by $\mathscr{S}_M \otimes \mathscr{S}_D$, cf. Sec. 1 For simplicity we assume that $\mathscr{S}_D$ has a Hilbert metric, as, e.g., in QED.

As long as $j^\mu(\mathbf{x})$ is a prescribed value of the parameter J we can certainly accept that the space of *physical* states of M, $\mathscr{L}^q$ or $\bar{\mathscr{H}}^q$ or $\mathscr{H}_0^q$, is prescribed as well. But if J is quantized, the operator $J^\mu(\mathbf{x})$ of the 4-current should be capable of expectation values which correspond to *different* $j^\mu(x)$. In particular, the space $\mathscr{S}_{\text{phys}}$ of physical states of $M + D$ should contain states which correspond to different charge densities. The superposition principle of quantum theory[24] requires that $\mathscr{S}_{\text{phys}}$ contains at least the span $\text{sp}\{\mathscr{L}^q\}$ of the corresponding Lorentz spaces. But since, according to Theorem 4, $\text{sp}\{\mathscr{L}^q\}$ is not a pre-Hilbert space if $\{\mathscr{L}^q\}$ contains two different $\mathscr{L}^q$, $\mathscr{S}_{\text{phys}}$ cannot be of the form $[\text{sp}\{\mathscr{L}^q\}] \otimes \mathscr{S}_D$. So we learn from Theorem 4 that *the space of physical states of the interacting system $M + D$ cannot be the product of the spaces of physical states of the partners.*

The following *might* be possible, however. Suppose (i) that there is a set $\{j^0(\mathbf{x})\}$ of "distinguished" charge distribtuions $j^0(\mathbf{x})$ which correspond to a set $\{q\}$ of distinguished indices q, that (ii) to any $q \in \{q\}$ exists a set $\{\mathbf{j}(\mathbf{x})\}$ of distinguished currents $\mathbf{j}(\mathbf{x})$, that (iii) to any $\mathbf{j}(\mathbf{x})$ exists a set of gauge functions $g(x)$ which corresponds to a set $\{g\}$ of indices $g = g^q(\mathbf{j})$, that (iv) to any distinguished triplet q, $\mathbf{j}$, $g^q(\mathbf{j})$ exists a subspace $\mathscr{S}(q,\mathbf{j},g^q(\mathbf{j}))$ of $\mathscr{S}_D$ such that $j^\mu(\mathbf{x})$ is the expectation value of $J^\mu(\mathbf{x})$ in any element of $\mathscr{S}(q,\mathbf{j},g^q(\mathbf{j}))$, and suppose finally that (v)

$$\mathscr{S}_{\text{phys}} := \text{sp}\{\mathscr{H}(q,\mathbf{j},g^q(\mathbf{j})) \otimes \mathscr{S}(q,\mathbf{j},g^q(\mathbf{j}))\} \tag{57}$$

is a Hilbert space (maybe inseparable), or at least a pre-Hilbert space, if $\mathscr{H}(q,\mathbf{j},g^q(\mathbf{j}))$ is our former $\mathscr{H}_g^q$ for $g = g^q(\mathbf{j})$. The states of D in $\mathscr{S}(q,\mathbf{j},g^q(\mathbf{j}))$ are then dressable in the sense of Sec. 10. The point is to *find* the distinguished indices q, $\mathbf{j}$, $g^q(\mathbf{j})$) such that $\mathscr{S}_{\text{phys}}$ becomes at least a pre-Hilbert space. Theorem 4 means that this is not a trivial selection problem for dressables states of any interaction partner D of M.

It is clear that *processes*, like those which possibly equip a bare electron or positron (both quanta of the Dirac field) with its permanent Coulomb robe, must be described on a state space $\mathscr{S}$ which contains dressed as well as undressed states. There is no need that $\mathscr{S}$ is already a Hilbert space; actually, this is not even desirable in order that the inseparability of a bare particle from its Coulomb field can be *automatically* achieved as a by-product of a restriction to a pre-Hilbert or a Hilbert space, which is necessary to allow the statistical interpretation.

The above selection of dressable states of any interaction partner D of M indeed shows a great similarity to the historical selection of square integrable solutions from the set of all solutions of the Schrödinger equation which first led Schrödinger[28] to the eigenvalue problem of quantum mechanics, then Born[29] to the statistical interpretation, and von Neumann[25] to the Hilbert space. In retrospection, this restriction is absolutely necessary in view of the weight we are willing to ascribe to Born's interpretation. It would be very hard to discard undressed states as unphysical if they are in a Hilbert space; we *must* discard them if they are not.

12. THE CONDITION OF VANISHING TOTAL CHARGE AND ITS IMPLICATIONS ON THE INTERACTION POSTULATE

We finally consider the condition $Q = 0$ which has been mentioned several times. It is a consequence of the convenient definition of $\mathscr{S}_M$ as a Banach (or Hilbert) space which carries the indefinite scalar product in addition to a Banach metric. Since similar spaces have been used in constructive quantum field theory,[9] it is maybe useful to first remember that in such a space we can only equip the distributions $j^0(\mathbf{x})$ of the total charge $Q = 0$ with their quantum mechanical Coulomb eigenfields. For the present theory this is certainly a defect, but it seems to be much less important in QED, as we want to now show.

If M interacts with the quantized Dirac field D, the principle of charge conservation[30] is realized by the relations

$$[Q,H_D] = 0 = [Q,H_{M+D}], \tag{58}$$

where $Q = \int d^3x\, J^0(\mathbf{x})$ is the Schrödinger operator of the total charge, and H_D, H_{M+D} are the Hamiltonians of the free D and of the coupled system $M + D$, respectively. Equations (58) mean that H_D, H_{M+D} are orthogonal sums,[31]

$$H_D = \bigoplus_{\rho=-\infty}^{\infty} H_D^{(\rho)}, \quad H_{M+D} = \bigoplus_{\rho=-\infty}^{\infty} H_{M+D}^{(\rho)}, \tag{59}$$

of *sectoral* Hamiltonians $H_D^{(\rho)}$, $H_{M+D}^{(\rho)}$ which act independently on the eigensectors $\mathscr{S}_D^{(\rho)}$, $\mathscr{S}_{M+D}^{(\rho)}$ of Q to the eigenvalues $\rho = \ldots, -1,0,1,2,\ldots,$ These eigensectors correspond to unique, Poincaré invariant decompositions

$$\mathscr{S}_D = \bigoplus_{\rho=-\infty}^{\infty} \mathscr{S}_D^{(\rho)}, \quad \mathscr{S}_{M+D} = \bigoplus_{\rho=-\infty}^{\infty} \mathscr{S}_{M+D}^{(\rho)} \tag{60}$$

of the respective state space $\mathscr{S}_D$, $\mathscr{S}_{M+D}$ of D and $M + D$. The sectors $\mathscr{S}_D^{(\rho)}$ and the corresponding Hamiltonians $H_D^{(\rho)}$ can be constructed easily, the $\mathscr{S}_{M+D}^{(\rho)}$ and $H_{M+D}^{(\rho)}$ have not yet been analyzed, it seems. Our results allow a remark to this issue.

The interaction postulate of Sec. 1 requires $\mathscr{S}_{M+D} = \mathscr{S}_M \otimes \mathscr{S}_D$ and thus

$$\begin{aligned} \mathscr{S}_{M+D} &= \mathscr{S}_M \otimes \left\{ \bigoplus_{\rho=-\infty}^{\infty} \mathscr{S}_D^{(\rho)} \right\} \\ &= \bigoplus_{\rho=-\infty}^{\infty} (\mathscr{S}_M \otimes \mathscr{S}_D^{(\rho)}) \end{aligned} \tag{61}$$

so that $\mathscr{S}^{(\rho)}_{M+D}$ must be of the form

$$\mathscr{S}^{(\rho)}_{M+D} = \mathscr{S}_M \otimes \mathscr{S}^{(\rho)}_D, \quad \rho = ..., -1,0,1,.... \tag{62}$$

But our $\mathscr{S}_M$ fits only to $\mathscr{S}^{(0)}_D$, and so we must either *extend the interaction postulate to*

$$\mathscr{S}_{M+D} = \bigoplus_{\rho=-\infty}^{\infty} (\mathscr{S}^{(\rho)}_M \otimes \mathscr{S}^{(\rho)}_D), \tag{63}$$

where $\mathscr{S}^{(0)}_M$ agrees with our $\mathscr{S}_M$ and the other $\mathscr{S}^{(\rho)}_M$ remain to be constructed (how?), or *restrict the interaction postulate* (1) to

$$\mathscr{S}_{M+D} = \mathscr{S}_M \otimes \mathscr{S}^{(0)}_D. \tag{64}$$

The second alternative seems to be more realistic. Since $\mathscr{S}^{(0)}_D$ contains all states where the electrons are "here" and the positrons "behind the moon, but not at infinity," the theory on $\mathscr{S}_M \otimes \mathscr{S}^{(0)}_D$ can account for all realistic situations, and so it is a full substitute of QED altogether. In addition, it probably avoids the infrared problem[15] and the virtual violations[30] of charge conservation implied in the usual formulation of QED of n-point functions. We learned here that it is an excellent candidate where dressing processes of quanta of D can be treated. The sectors $\mathscr{S}(q,\mathbf{j},g^q(\mathbf{j}))$ of Sec. 11 must of course be in $\mathscr{S}^{(0)}_D$ in this case. It thus appears that the restriction to $Q = 0$ has more positive than negative aspects.

13. COMPLETENESS AND DISJUNCTNESS OF THE LORENTZ SPACES

From now on we give only the missing proofs of the theorems and statements of the preceding sections. Here we verify Theorem 1.

We first remember that the missing closability of $a_\mu(\mathbf{k})$ refers to one arbitrary, but then fixed value of $\mathbf{k}$ and μ. However, in contrast to this, in Secs. 4 and 10 we needed the vectors $a_\mu(\mathbf{k})\alpha$ simultaneously *for any μ and only almost any* $\mathbf{k}\in\mathbb{R}^3$. It is therefore reasonable to define the *global domain* $\bar{\mathbb{D}}(a)$ by $\bar{\mathbb{D}}(a) := \{\alpha\in\mathscr{S}_M : |||a\alpha||| < \infty\}$ where $|||a\alpha|||$ is given by

$$|||a\alpha|||^2 := \int dK\, \|a_\mu(\mathbf{k})\alpha\|^2 : = \sum_{\mu=0}^{3} \int d^3k\, \|a_\mu(\mathbf{k})\alpha\|^2 = \sum_{n=0}^{\infty} n\|\alpha_n\|^2. \tag{65}$$

The last expression shows that $\bar{\mathbb{D}}(a)$ agrees with the *dense* domain of the operator $N^{1/2}$ given in components by $(N^{1/2}\alpha)_n = n^{1/2}\alpha_n$, $n = 0,1,2,....$ Moreover, $a_\mu(\mathbf{k})$ has the following closure property which will be needed for the proof of Theorem 1.

We consider $a_\mu(\mathbf{k})$ as a mapping of $\bar{\mathbb{D}}(a)\subseteq\mathscr{S}_M$ into the Hilbert space $\mathscr{H}'$ of the sequences $\alpha' = \{\alpha'_0(K),\alpha'_1(K|K^1), \alpha'_2(K|K^2),\cdots\}$ with the scalar product (cf. Sec. 3)

$$(\beta'|\alpha') = \sum_{n=0}^{\infty} \int dK \int dK^n\, \beta'^*_n(K|K^n)\alpha'_n(K|K^n), \tag{66}$$

where $\alpha'_n(K|K^n)$ is symmetric in the n pairs $(\mathbf{k}_\nu,\mu_\nu)$ in K^n, but not necessarily in the pair $K = (\mathbf{k},\mu)$. We have $a_\mu(\mathbf{k}) = (N+1)^{1/2}P$, where P is defined by $\alpha \to \alpha' := P\alpha$, $\alpha'_n(K|K^n) = \alpha_{n+1}(K;K^n)$, and $(N+1)^{1/2}$ by

$$\alpha' \to (N+1)^{1/2}\alpha',$$
$$((N+1)^{1/2}\alpha')_n(K|K^n) = (n+1)^{1/2}\alpha'_n(K|K^n) = (n+1)^{1/2}\alpha_{n+1}(K;K^n).$$

P is defined everywhere on $\mathscr{S}_M$ and continuous, so it is closed. $(N+1)^{1/2}$ is self-adjoint on $\mathscr{H}'$ and so it is also closed. It is clear then that $(N+1)^{1/2}P = a_\mu(\mathbf{k})$ is closed in the following sense: Let the sequence $\alpha^{(m)}$, $m = 1,2,\cdots$, $\alpha^{(m)}\in\bar{\mathbb{D}}(a)$ for any m, converge to any point α in $\mathscr{S}_M$, and let $a_\mu(\mathbf{k})\alpha^{(m)}$ converge for any given μ and almost any given $\mathbf{k}$ to some $\gamma_\mu(\mathbf{k})\in\mathscr{S}_M$ satisfying $|||\gamma|||^2 := \int dK\, \|\gamma_\mu(\mathbf{k})\|^2 < \infty$, this convergence being understood in the sense of the usual Cauchy convergence on $\mathscr{H}'$. Then α was in $\bar{\mathbb{D}}(a)$ and $a_\mu(\mathbf{k})\alpha$ is equal to $\gamma_\mu(\mathbf{k})$ for any μ and almost any $\mathbf{k}$.

It is obvious that similar sttements also hold for the operators $a_{(\sigma)}(\mathbf{k})$, $a_b(\mathbf{k})$, $a_g(\mathbf{k})$, $L(\mathbf{k})$. We need the theorem for $L(\mathbf{k})$:

Let $\bar{\mathbb{D}}(L) := \{\alpha\in\mathscr{S}_M : |||L\alpha||| < \infty\}$ be the global domain of $L(\mathbf{k})$ where $|||\cdots|||$ is now defined by $|||L\alpha|||^2 = \int d^3k\, \|L(k)\alpha\|^2$. It is clear that $\bar{\mathbb{D}}(L)$ is dense. Let $\alpha^{(m)}$, $m = 1,2,\cdots$, be any *convergent* sequence in $\bar{\mathbb{D}}(L)$. In general, its limit $\alpha\in\mathscr{S}_M$ is not in $\bar{\mathbb{D}}(L)$. However, α is in $\bar{\mathbb{D}}(L)$ if the sequence $L(\mathbf{k})\alpha^{(m)}$ converges for almost any $\mathbf{k}\in\mathbb{R}^3$ to some point $\gamma(\mathbf{k})\in\mathscr{S}_M$ satisfying $\int d^3k\, \|\gamma(\mathbf{k})\|^2 < \infty$. Then also $\gamma(\mathbf{k}) = L(\mathbf{k})\alpha$ for almost any $\mathbf{k}$.

We can show now that $\mathscr{L}^q$ is complete, i.e., that the limit of any convergent sequence in $\mathscr{L}^q$ is in $\mathscr{L}^q$ again. Let $\alpha^{(m)}\in\mathscr{L}^q$ be a sequence with the limit $\alpha\in\mathscr{S}_M$. As $L(\mathbf{k})\alpha^{(m)}$ is equal to $q(\mathbf{k})\alpha^{(m)}$ for any m, the sequence $L(\mathbf{k})\alpha^{(m)}$ converges to the point $\gamma(\mathbf{k}) := q(\mathbf{k})\alpha$ which satisfies $\int d^3k\, \|\gamma(\mathbf{k})\|^2 = \|\alpha\|^2 \int d^3k\, |q(\mathbf{k})|^2 < \infty$. It follows from the above that α was in $\bar{\mathbb{D}}(L)$ and that $L(\mathbf{k})\alpha$ is equal to $q(\mathbf{k})\alpha$, which means $\alpha\in\mathscr{L}^q$, as stated. Note that $\mathscr{L}^q\subseteq\bar{\mathbb{D}}(L)$ holds for any $q\in\mathfrak{Q}$ by the definitions of $\mathscr{L}^q$ and $\bar{\mathbb{D}}(L)$.

The proof of relation (33) is nearly trivial: Assume that α is an element contained in $\mathscr{L}^q$ as well as in $\mathscr{L}^{q'}$, $q\neq q'$. So we have $L(\mathbf{k})\alpha = q(\mathbf{k})\alpha = q'(\mathbf{k})\alpha$ and thus $[q(\mathbf{k}) - q'(\mathbf{k})]\alpha = O$ for almost any $\mathbf{k}$. Multiplying this by $q^*(\mathbf{k}) - q'^*(\mathbf{k})$, and integrating over $\mathbf{k}$ we get $\|q - q'\|^2\alpha = O$, where $\|\cdots\|$ now denotes the L_2 norm. So $\|q - q'\| > 0$ for $q\neq q'$, and all components of α must vanish nearly everywhere. This means $\alpha = o$ and thus $\mathscr{L}^q\cap\mathscr{L}^{q'} = \{o\}$, as stated.

14. SMEARED FIELD, WEYL, AND SOME OTHER OPERATORS ON $\mathscr{S}_M$

For the proof of the other theorems and statements we now define some further operators and note some useful relations between them.

For convenience we introduce contravariant creation and destruction operators by $a^{+\mu}(\mathbf{k}) := g^{\mu\nu}a^+_\nu(\mathbf{k})$, $a^\mu(\mathbf{k}) := g^{\mu\nu}a_\nu(\mathbf{k})$. Consider now the smeared field operator

$$F\{c\} := \int d^3k\, [c^\mu(\mathbf{k})a^+_\mu(\mathbf{k}) - c^*_\mu(\mathbf{k})a^\mu(\mathbf{k})], \tag{67}$$

where $c = c_\mu(\mathbf{k})$ is any element from $\mathfrak{C}$. In accordance with (18) we have

$$(F\{c\}\alpha)_n(K^n) = \frac{1}{n^{1/2}}\sum_{\nu=1}^{n} c_{\mu_\nu}(\mathbf{k}_\nu)\alpha_{n-1}(K^n\setminus(\mathbf{k}_\nu,\mu_\nu)) - (n+1)^{1/2}\int d^3k \sum_{\mu=0}^{3} g_{\mu\mu_\nu} c_\mu(\mathbf{k})\alpha_{n+1}(\mathbf{k},\mu;K^n). \tag{68}$$

We see that $F\{c\}$ maps the dense set $\mathbb{D}\supseteq\mathbb{D}_c$ of all α which have only a finite number of nonvanishing components $\alpha_n(K^n)$, into itself. Therefore, $F\{c\}$ as well as any power $(F\{c\})^m$, $m=1,2,\cdots$, is *densely defined* on $\mathscr{S}_M$ if $c\in\mathfrak{C}$. By straightforward use of (18) and (68) we get for any $c,f,g\in\mathfrak{C}$

$$[a_\mu(\mathbf{k}),(F\{c\})^m] = mc_\mu(\mathbf{k})(F\{c\})^{m-1},\quad m=1,2,\cdots, \tag{69a}$$

$$[a_\mu^+(\mathbf{k}),(F\{c\})^m] = mc_\mu^*(\mathbf{k})(F\{c\})^{m-1},\quad m=1,2,\cdots, \tag{69b}$$

$$[F\{f\},F\{g\}] = \int d^3k\,[f^\mu(\mathbf{k})g_\mu^*(\mathbf{k}) - f_\mu^*(\mathbf{k})g^\mu(\mathbf{k})], \tag{69c}$$

$$F\{c\}^+ = F\{-c\} = -F\{c\}. \tag{69d}$$

All equations hold on dense domains which enclose $\mathbb{D}$. Equation (69c) can be continued to $\mathscr{S}_M$. Equation (69b) is a consequence of (69a) and (69d); it is not relevant that the single terms in the commutator do not exist properly.

The most important operators of type (67), (68) are the cut-off field operators $\Pi_\mu^\Gamma(\mathbf{x})$, $A_\mu^\Gamma(\mathbf{x})$ obtained from (35) by restricting the integral over $\mathbf{k}$ to any *bounded* region Γ in $\mathbb{R}^3$. In analogy to (69c) they satisfy

$$[\Pi_\mu^\Gamma(\mathbf{x}),A_\nu^\Gamma(\mathbf{y})] = -i\,g_{\mu\nu}\frac{1}{(2\pi)^3}\int_{(\Gamma)} d^3k\, e^{i\mathbf{k}(\mathbf{x}-\mathbf{y})},$$
$$[\Pi_\mu^\Gamma(\mathbf{x}),\Pi_\nu^\Gamma(\mathbf{y})] = [A_\mu^\Gamma(\mathbf{x}),A_\nu^\Gamma(\mathbf{y})] = 0. \tag{70}$$

For $\Gamma\to\mathbb{R}^3$ these operators "approach" the desired field operators (35) and properly satisfy (2). In the sense of this "limit" we have represented the field commutation relations (2) on $\mathscr{S}_M$, in close analogy to the usual representation of Bose field commutation relations on a Fock space.

As any polynomial in $F\{c\}$ is densely defined on $\mathscr{S}_M$ we expect that the Weyl operator

$$W\{c\} = \exp F\{c\} = \exp\int d^3k\,[c^\mu(\mathbf{k})a_\mu^+(\mathbf{k}) - c_\mu^*(\mathbf{k})\,a^\mu(\mathbf{k})] \tag{71}$$

can be defined on $\mathscr{S}_M$ by its power series. In the Appendix we convince ourselves that this series indeed converges for any $\alpha\in\mathbb{D}$. However, in contrast to the corresponding case on a Fock space we have no proof that it can be continued to $\mathscr{S}_M$ by the usual continuation by continuity. We have namely, in general, $\|W\{c\}\alpha\|\neq\|\alpha\|$. However, we also show in the Appendix that $W\{c\}$ can be continued to any $\mathscr{L}^q$, $q\in\mathfrak{Q}$, and this is all we actually need further on.

On $\mathbb{D}$ and, if necessary, by continuation on any $\mathscr{L}^q$, we thus get for any $c,g,f\in\mathfrak{C}$ the relations

$$[a_\mu(\mathbf{k}),W\{c\}] = c_\mu(\mathbf{k})W\{c\},\ [a_\mu^+(\mathbf{k}),W\{c\}] = c_\mu^*(\mathbf{k})W\{c\}, \tag{72a}$$

$$W\{f+g\} = W\{f\}W\{g\}\exp\Big[-i\,\mathrm{Im}\int d^3k\, f^\mu(\mathbf{k})\,g_\mu^*(\mathbf{k})\Big], \tag{72b}$$

$$W\{c\} = \exp\Big[-\tfrac{1}{2}\int d^3k\, c_\mu^*(\mathbf{k})c^\mu(\mathbf{k})\Big]\times\exp\Big[\int d^3k\, c^\mu(\mathbf{k})a_\mu^+(\mathbf{k})\Big]\times\exp\Big[-\int d^3k\, c_\mu^*(\mathbf{k})a^\mu(\mathbf{k})\Big], \tag{72c}$$

$$W\{-c\} = W\{c\}^+, \tag{72d}$$

which are the main tools for our conclusions. The last formula means $\langle W\{c\}\alpha|W\{c\}\alpha\rangle = \langle\alpha|\alpha\rangle$, like on a Fock space. By (72c) we find easily that the coherent state (26) satisfying $\langle\alpha|\alpha\rangle = 1$, is given by $\alpha = W\{c\}\omega$. For any "curve" $c[t]$ in $\mathfrak{C}$ {with components $c_\mu(t,\mathbf{k})$ with the property that $\dot c[t]$ [with components $\dot c_\mu := (\partial/\partial t)c_\mu(t,\mathbf{k})$] exists in $\mathfrak{C}$ for any t in consideration} we get by means of (72d),

$$\frac{d}{dt}W\{c[t]\} = \Big[F\{\dot c[t]\} + i\,\mathrm{Im}\int d^3k\,\dot c_\mu^*(\mathbf{k})c^\mu(\mathbf{k})\Big]\times W\{c[t]\}. \tag{73}$$

We shall also need the operator (we define $|\mathbf{k}_1| + \cdots + |\mathbf{k}_m| := 0$ for $m=0$)

$$U_0(t) := \bigoplus_{m=0}^{\infty}\exp[-it(|\mathbf{k}_1| + \cdots + |\mathbf{k}_m|)], \tag{74}$$

and some relations to other operators. $U_0(t)$ is the direct sum of multiplication operators on the subspaces $\mathscr{S}_M^m$ of $\mathscr{S}_M$ which contain the elements α with only one nonvanishing component $\alpha_m(K^m)$, $m=0,1,2,\cdots$. $U_0(t)$ is defined everywhere on $\mathscr{S}_M$ and unitary, $U_0^+(t) = U_0(-t) = U_0(t)^{-1}$. By straightforward use of the definitions (18) and (74) we get

$$a_\mu(\mathbf{k})U_0(t) = U_0(t)a_\mu(\mathbf{k})e^{-it|\mathbf{k}|},\ a_\mu^+(\mathbf{k})U_0(t) = U_0(t)a_\mu^+(\mathbf{k})e^{it|\mathbf{k}|}, \tag{75a}$$

$$F\{f[t_1]\}U_0(t_2) = U_0(t_2)F\{e[t_2]\cdot f[t_1]\}, \tag{75b}$$

$$W\{f[t_1]\}U_0(t_2) = U_0(t_2)W\{e[t_2]\cdot f[t_1]\}, \tag{75c}$$

where $e[t]$ is the function $e^{it|\mathbf{k}|}$, to be considered *qua* function of $\mathbf{k}$. We note finally that $U_0(t)$ is the time evolution operator as formally generated by the direct sum[31]

$$H_0 = \bigoplus_{m=0}^{\infty}(|\mathbf{k}_1| + \cdots + |\mathbf{k}_m|) = \int d^3k\,|\mathbf{k}|a_\mu^+(\mathbf{k})a^\mu(\mathbf{k}) \tag{76}$$

of multiplication operators $|\mathbf{k}_1| + \cdots + |\mathbf{k}_m|$ on $\mathscr{S}_M^m$. It is densely defined and self-adjoint because the corresponding proof[31] can be carried over to any $\mathscr{S}_M^m$.

15. THE ISOMORPHY OF ALL LORENTZ SPACES RELATIVE TO THE INDEFINITE METRIC ON $\mathscr{S}_M$, AND THEIR PRE-HILBERT SPACE PROPERTIES

We are ready now to verify Theorems 2–4 of Sec. 5.

For Theorem 2 we consider the Weyl operator $W\{c\}$ corresponding to any $c\in\mathfrak{C}$ which satisfies $c_\mu(\mathbf{k})k^\mu = q(\mathbf{k})$, $q\in\mathfrak{Q}$. Further on, let α be any element from $\mathscr{L}^0$ so that $L(\mathbf{k})\alpha = k^\mu a_\mu(\mathbf{k})\alpha = 0$. Then $W\{c\}\alpha$ is in $\mathscr{S}_M$ and by means of (72a) we get

$$L(\mathbf{k})W\{c\}\alpha = k^\mu a_\mu(\mathbf{k})W\{c\}\alpha = k^\mu W\{c\}(c_\mu(\mathbf{k}) + a_\mu(\mathbf{k}))\alpha = q(\mathbf{k})W\{c\}\alpha \quad \text{for almost any } \mathbf{k}. \tag{77}$$

This means $W\{c\}\alpha \in \mathscr{L}^q$. The mapping $\alpha \to W\{c\}\alpha$ therefore depicts $\mathscr{L}^0$ into $\mathscr{L}^q$. Now let α' be any element from $\mathscr{L}^q$ so that $L(\mathbf{k})\alpha' = q(\mathbf{k})\alpha'$, and consider the equation

$$\begin{aligned} L(\mathbf{k})W\{-c\}\alpha' &= k^\mu a_\mu(\mathbf{k})W\{-c\}\alpha' \\ &= k^\mu W\{-c\}(-c_\mu(\mathbf{k}) + a_\mu(\mathbf{k}))\alpha' \\ &= -q(\mathbf{k})W\{-c\}\alpha' + W\{-c\}L(\mathbf{k})\alpha' = 0. \end{aligned} \tag{78}$$

The mapping $\alpha' \to W\{-c\}\alpha'$ therefore depicts $\mathscr{L}^q$ into $\mathscr{L}^0$. But because of $W\{-c\}W\{c\}\alpha = \alpha$ and $W\{c\}W\{-c\}\alpha' = \alpha'$, both mappings are onto and one-to-one, and we can write

$$\mathscr{L}^q = W\{c\}\mathscr{L}^0. \tag{79}$$

From $\langle W\{c\}\alpha | W\{c\}\beta\rangle = \langle\alpha|\beta\rangle$ we now obtain the stated isomorphy $\mathscr{L}^0 \leftrightarrow \mathscr{L}^q$, and by the iteration $\mathscr{L}^q \leftrightarrow \mathscr{L}^0 \leftrightarrow \mathscr{L}^q$ we get the isomorphy of all Lorentz spaces.

Note that the equation $k^\mu c_\mu(\mathbf{k}) = q(\mathbf{k})$ has, in general, more than one solutions $c \in \mathfrak{C}$ so that correspondingly many mappings $\mathscr{L}^0 \leftrightarrow \mathscr{L}^q$ exist. This will play a certain role later on.

We now prove Theorem 3. In view of the above it holds clearly on any $\mathscr{L}^q$ if it holds on $\mathscr{L}^0$. But on $\mathscr{L}^0$ it is equivalent to the main achievement of the conventional Gupta–Bleuler formalism. Since the statement is usually proven in a somewhat heuristical way we give a short generalization in the present terminology which will also be needed for the proof of Theorem 4.

By a straightforward application of definition (18a) of $a_\mu(\mathbf{k})$ we get on $\mathbb{D}$

$$\alpha_n(K^n) = \alpha_n(\mathbf{k}_1,\mu_1;\ldots;\mathbf{k}_n,\mu_n) = \langle\omega \mid a_{\mu_1}(\mathbf{k}_1)\ldots a_{\mu_n}(\mathbf{k}_n)\,\alpha\rangle/(n!)^{1/2}. \tag{80}$$

By continuity this holds for any α. By means of (20) and (22) we get

$$a_\mu(\mathbf{k}) = e_\mu^{(\sigma)}(\mathbf{k})a_{(\sigma)}(\mathbf{k}) \tag{81}$$

and thus

$$\alpha_n(K^n) = \frac{1}{(n!)^{1/2}} e_{\mu_1}^{(\sigma_1)}(\mathbf{k}_1)\cdots e_{\mu_n}^{(\sigma_n)}(\mathbf{k}_n) \times\langle\omega \mid a_{(\sigma_1)}(\mathbf{k}_1)\ldots a_{(\sigma_n)}(\mathbf{k}_n)\alpha\rangle. \tag{82}$$

Inserting this into (16) and using (20) we obtain for any $\alpha,\beta \in \mathscr{S}_M$,

$$\langle\beta \mid \alpha\rangle = \langle\omega \mid \beta\rangle^*\langle\omega \mid \alpha\rangle + \sum_{n=1}^{\infty}\frac{1}{n!} \times \sum_{\sigma_1,\ldots,\sigma_n=0}^{3} g(\sigma_1\sigma_1)\ldots g(\sigma_n\sigma_n) I(\sigma_1,\ldots,\sigma_n), \tag{83}$$

$$I(\sigma_1,\ldots,\sigma_n) := \int d^3k_1\cdots\int d^3k_n \langle\omega \mid a^{(\sigma_1)}(\mathbf{k}_1)\cdots a^{(\sigma_n)}(\mathbf{k}_n)\beta\rangle^* \times\langle\omega \mid a^{(\sigma_1)}(\mathbf{k}_1)\cdots a^{(\sigma_n)}(\mathbf{k}_n)\alpha\rangle. \tag{84}$$

In view of (28), definition (29) means $a_{(3)}(\mathbf{k})\alpha = -a_{(0)}(\mathbf{k})\alpha$ for any $\alpha \in \mathscr{L}^0$. Therefore, we have

$$I(\sigma_1,\ldots,\sigma_{\nu-1},0,\sigma_{\nu+1},\ldots,\sigma_n) = I(\sigma_1,\ldots,\sigma_{\nu-1},3,\sigma_{\nu+1},\ldots,\sigma_n) \tag{85}$$

for any $\alpha,\beta \in \mathscr{L}^0$, any n and $\nu \in \{1,\ldots,n\}$, and any $\sigma_1,\ldots,\sigma_{\nu-1}$, $\sigma_{\nu+1},\ldots,\sigma_n$. The contribution of the terms $\sigma_1 = 0$ and $\sigma_1 = 3$ to the n-fold sum in (83) reads

$$g(00)\sum_{\sigma_2,\ldots,\sigma_n=0}^{3} g(\sigma_2\sigma_2)\cdots g(\sigma_n\sigma_n) I(0,\sigma_2,\ldots,\sigma_n) + g(33)\sum_{\sigma_2,\ldots,\sigma_n=0}^{3} g(\sigma_2\sigma_2)\cdots g(\sigma_n\sigma_n) I(3,\sigma_2,\ldots,\sigma_n). \tag{86}$$

This vanishes because of (85) and $g(00) = -g(33)$. It follows that the factor $g(\sigma_1\sigma_1) = 1$ can be ignored further on. Repeating the same arguments for the sums over $\sigma_2,\ldots,\sigma_n$ we finally get for any $\alpha,\beta \in \mathscr{L}^0$,

$$\langle\beta \mid \alpha\rangle = \langle\omega \mid \beta\rangle^*\langle\omega \mid \alpha\rangle + \sum_{n=1}^{\infty}\frac{1}{n!}\int d^3k_1\ldots \times\int d^3k_n \sum_{\sigma_1,\ldots,\sigma_n=1}^{2} \langle\omega \mid a_{(\sigma_1)}(\mathbf{k}_1)\cdots a_{(\sigma_n)}(\mathbf{k}_n)\beta\rangle^* \times\langle\omega \mid a_{(\sigma_1)}(\mathbf{k}_1)\cdots a_{(\sigma_n)}(\mathbf{k}_n)\,\alpha\rangle. \tag{87}$$

For $\beta = \alpha$ this yields $\langle\alpha|\alpha\rangle \geqslant 0$ which proves theorem 3.

Finally consider the vector

$$\alpha^h := \int d^3k\, h(\mathbf{k})L^+(\mathbf{k})\,\omega = \{0,k_{\mu_1}h(\mathbf{k}_1),0,\ldots\} \neq o, \tag{88}$$

where $h(\mathbf{k})$ is any complex-valued, square integrable function over the $\mathbb{R}^3$. As

$$L(\mathbf{k})\alpha^h = k^\mu a_\mu(\mathbf{k})\{0,k_{\mu_1}h(\mathbf{k}),0,\ldots\} = k^\mu\{k_\mu h(\mathbf{k}),0,\cdots\} = \{0,0,\ldots\} = o,$$

α^h is in $\mathscr{L}^0$. Using $k_\mu k^\mu = 0$ once more, we get $\langle\alpha^h|\alpha^h\rangle = \int d^3k\, k^\mu k_\mu |h(\mathbf{k})|^2 = 0$. Therefore, $\alpha := W\{c\}\alpha^h$ also satisfies $\langle\alpha|\alpha\rangle = 0$ and is in $\mathscr{L}^q$ if c is any solution of $c_\mu(\mathbf{k})k^\mu = q(\mathbf{k})$.

Let c' be any solution of $c'_\mu(\mathbf{k})k^\mu = q'(\mathbf{k})$, $q' \neq q$, and consider the vector $\beta := \mathrm{W}\{c'\}\omega$ in $\mathscr{L}^{q'}$ which clearly satisfies $\langle\beta|\beta\rangle = \langle\omega|\omega\rangle = 1$. By means of (72) we get

$$\begin{aligned} \langle\alpha|\beta\rangle &= \int d^3k\, h^*(\mathbf{k})\langle\omega \mid L(\mathbf{k})W\{-c\}W\{c'\}\omega\rangle \\ &= \int d^3k\, h^*(\mathbf{k})\langle\omega \mid W\{-c\}(L(\mathbf{k}) - q(\mathbf{k}))W\{c'\}\omega\rangle \\ &= \langle\omega \mid W\{-c\}W\{c'\}\omega\rangle \int d^3k\, h^*(\mathbf{k})(q'(\mathbf{k}) - q(\mathbf{k})). \end{aligned} \tag{89}$$

The first factor is a pure exponential different from zero. As $q'(\mathbf{k}) - q(\mathbf{k})$ does not vanish by assumption, we can always find a function $h(\mathbf{k})$, e.g., $h(\mathbf{k}) = q'(\mathbf{k}) - q(\mathbf{k})$, such that the second factor is also different from zero. So we have $\langle\alpha|\beta\rangle \neq 0$.

If, in contradiction to theorem 4, the Cauchy–Schwarz inequality holds on the span of $\mathscr{L}^q$ and $\mathscr{L}^{q'}$, the above vectors α,β would always satisfy $|\langle\alpha|\beta\rangle|^2 \leqslant \langle\beta|\beta\rangle\,\langle\alpha|\alpha\rangle = 1\cdot 0 = 0$. So Theorem 4 must be true.

16. THE GAUGE HILBERT SPACES IN ANY LORENTZ SPACE

We now verify Theorem 5. In view of Theorem 2 the proof is also needed only for $\mathscr{L}^0$.

We first show that $\mathscr{F}^{\mathrm{tr}}$, as defined by (24), is indeed a Hilbert space. Its completeness could be proven in analogy to Sec. 13. It is once more contained in the following proof of the isomorphy of $\mathscr{F}^{\mathrm{tr}}$ and the usual Fock space $\mathscr{F}^{\mathrm{tr}}_{*}$ of transverse photons, which verifies all other Hilbert space properties of $\mathscr{F}^{\mathrm{tr}}$ as well.

Let φ be short for the sequence

$$\varphi = \{\varphi_0, \varphi_1(\mathbf{k}_1,\lambda_1), \varphi_2(\mathbf{k}_1,\lambda_1;\mathbf{k}_2,\lambda_2), \cdots\}, \qquad (90)$$

wehre the nth component φ_n of φ is a complex number for $n=0$, and for $n=1,2,\cdots$, a complex-valued, symmetric function $\varphi_n = \varphi_n(\mathbf{k}_1\lambda_1;\ldots;\mathbf{k}_n\lambda_n)$ of n pairs $(\mathbf{k}_\nu,\lambda_\nu)$. Each $\mathbf{k}_\nu$ varies continuously over the $\mathbb{R}^3$, and each λ_ν assumes the values 1,2. For any given $\lambda_1,\ldots,\lambda_n$, φ_n is defined almost everywhere on $\mathbb{R}^{3n}$. Let χ be defined similarly and consider the sesquilinear form

$$(\varphi \mid \chi) = \varphi_0^*\chi_0 + \sum_{n=0}^{\infty}\sum_{\lambda_1\ldots\lambda_n=1}^{2}\int d^3k_1\cdots\int d^3k_n \times\varphi_n^*(\mathbf{k}_1\lambda_1;\ldots;\mathbf{k}_n\lambda_n)\chi_n(\mathbf{k}_1\lambda_1;\ldots;\mathbf{k}_n\lambda_n). \qquad (91)$$

Denote by $\mathscr{B}_*$ the complete set of all φ for which $(\varphi|\varphi)$ exists. Equation (91) defines a Hilbert metric on $\mathscr{B}_*$, and the pair $\mathscr{F}^{\mathrm{tr}}_*:=(\mathscr{B}_*,(\cdot|\cdot))$ is indeed a Hilbert space, the usual representative of the Fock space of transverse photons. In order to get its isomorphy to $\mathscr{F}^{\mathrm{tr}}$ we see by comparison with (87) that the mapping $\alpha\to\varphi=\varphi(\alpha)$, as given by

$\varphi_0 := \langle\omega \mid \alpha\rangle$

and

$$\varphi_n(\mathbf{k}_1\lambda_1;\ldots;\mathbf{k}_n\lambda_n) := \langle\omega \mid a_{(\lambda_1)}(\mathbf{k}_1)\cdots a_{(\lambda_n)}(\mathbf{k}_n)\alpha\rangle/(n!)^{1/2} \quad (n=1,2,\ldots), \qquad (92a)$$

depicts $\mathscr{F}^{\mathrm{tr}}$ isometrically into $\mathscr{F}^{\mathrm{tr}}_*$. Moreover, the mapping $\varphi\to\alpha=\alpha(\varphi)$, as given by

$\alpha_0 = \varphi_0$

and

$$\alpha_n(\mathbf{k}_1\mu_1;\ldots;\mathbf{k}_n\mu_n) := (n!)^{1/2} \times\sum_{\lambda_1,\ldots,\lambda_n=1}^{2} e^{(\lambda_1)}_{\mu_1}(\mathbf{k}_1)\cdots e^{(\lambda_n)}_{\mu_n}(\mathbf{k}_n)\varphi(\mathbf{k}_1\lambda_1;\cdots;\mathbf{k}_n\lambda_n), \quad n=1,2,\cdots, \qquad (92b)$$

depicts $\mathscr{F}^{\mathrm{tr}}_*$ onto $\mathscr{F}^{\mathrm{tr}}$. This is easily verified by showing that $\alpha(\varphi)$ satisfies $a_{(0)}(\mathbf{k})\alpha(\varphi)=o=a_{(3)}(\mathbf{k})\alpha(\varphi)$, cf. (24). Further on, we easily find $\varphi(\alpha(\varphi))=\varphi$, $\alpha(\varphi(\alpha))=\alpha$, so that (92) defines an isometrical one-to-one mapping $\mathscr{F}^{\mathrm{tr}}\leftrightarrow\mathscr{F}^{\mathrm{tr}}_*$, which proves the isomorphy of $\mathscr{F}^{\mathrm{tr}}_*$ and $\mathscr{F}^{\mathrm{tr}}$. As $\mathscr{F}^{\mathrm{tr}}_*$ is a Hilbert space, $\mathscr{F}^{\mathrm{tr}}$ also must be a Hilbert space.

On $\mathscr{L}^0$, the mapping $\alpha\to\varphi(\alpha)$ allows the definition of the equivalence relation $\sim$: $\alpha\sim\beta$ if $\varphi(\alpha)=\varphi(\beta)$. The equivalence classes $\{\alpha\}$, $\{\beta\}$ are elements of a Hilbert space $\bar{\mathscr{H}}^0$ if the scalar product $(\{\alpha\} \mid \{\beta\})$ of two equivalence classes is defined by the scalar product of any of their elements, $(\{\alpha\} \mid \{\beta\}) = (\varphi(\alpha) \mid \varphi(\beta)) = \langle\alpha \mid \beta\rangle$, which is possible by (87). $\bar{\mathscr{H}}^0$ is obviously isomorphic to $\mathscr{F}^{\mathrm{tr}}_*$ as well as to $\mathscr{F}^{\mathrm{tr}}$ so that $\alpha\in\mathscr{F}^{\mathrm{tr}}$ can be chosen as the representative of an element of $\bar{\mathscr{H}}^0$.

Let $g=g(\mathbf{k})$ be a scalar function such that the 4-vector $kg = k^\mu g(\mathbf{k})$ is in $\mathfrak{C}$. We claim that the Hilbert space

$$\mathscr{H}^0_g := W\{kg\}\mathscr{F}^{\mathrm{tr}}, \qquad (93)$$

the image of $\mathscr{F}^{\mathrm{tr}}$ under the mapping $\alpha\to W\{kg\}\alpha$, is in $\mathscr{L}^0$, and that α and $W\{kg\}\alpha$ are equivalent, $\alpha\sim W\{kg\}\alpha$. For the proof of the relation $W\{kg\}\alpha\in\mathscr{L}^0$ if $\alpha\in\mathscr{F}^{\mathrm{tr}}$ we use $[L(\mathbf{k}),W\{kg\}]=0$, which follows immediately from $k_\mu k^\mu=0$ and (72c). So we get for any $\alpha\in\mathscr{L}^0$: $L(\mathbf{k})W\{kg\}\alpha = W\{kg\}L(\mathbf{k})\alpha = o$, which means $W\{kg\}\alpha\in\mathscr{L}^0$ in particular. To prove $\alpha\sim W\{kg\}\alpha$ we verify $\varphi(\alpha)=\varphi(W\{kg\}\alpha)$ for any $\alpha\in\mathscr{F}^{\mathrm{tr}}$. Because $k_\mu k^\mu=0$ we get from (72d)

$$W\{kg\} = \left[\exp\int d^3k\, g(\mathbf{k})L^+(\mathbf{k})\right] \times\left[\exp\left(-\int d^3k\, g^*(\mathbf{k})L(\mathbf{k})\right)\right]. \qquad (94)$$

Consider first the component

$$\begin{aligned}\varphi_0 &:= \varphi(W\{kg\}\alpha)_0 = \langle\omega \mid W\{kg\}\alpha\rangle\\ &= \left\langle\omega\left|\left[\exp\int d^3k\, g(\mathbf{k})L^+(\mathbf{k})\right]\times\left[\exp\left(-\int d^3k\, g^*(\mathbf{k})L(\mathbf{k})\right)\right]\alpha\right.\right\rangle\\ &= \left\langle\left[\exp\left(-\int d^3k\, g^*(\mathbf{k})L(\mathbf{k})\right)\right]\times\omega\left|\left[\exp\left(-\int d^3k\, g^*(\mathbf{k})L(\mathbf{k})\right)\right]\alpha\right.\right\rangle. \end{aligned} \qquad (95)$$

In the last equation we have used the fact that the factors in (94) are the adjoints of each other. Now considering the series for the exponentials in (95) we recognize that only the first term gives a contribution because in all other terms the operator $L(\mathbf{k})$ acts on an element of $\mathscr{L}^0$ which yields zero. So we get $\varphi_0=\varphi(W\{kg\}\alpha)_0=\langle\omega|\alpha\rangle=\varphi(\alpha)_0$, which is the statement for the zero-component. Now consider the nth component

$$\begin{aligned}\varphi_n &= \varphi(W\{kg\}\alpha)_n\\ &= \langle\omega \mid a_{\lambda_1}(\mathbf{k}_1)\ldots a_{\lambda_n}(\mathbf{k}_n)W\{kg\}\alpha\rangle/(n!)^{1/2}\\ &= \left\langle\omega\left|a_{\lambda_1}(\mathbf{k}_1)\ldots a_{\lambda_n}(\mathbf{k}_n)\right[\exp\int d^3k\, g(\mathbf{k})L^+(\mathbf{k})\right]\times\left.\left[\exp\int d^3k\, g^*(\mathbf{k})L(\mathbf{k})\right]\alpha\right\rangle\Big/(n!)^{1/2}. \end{aligned} \qquad (96)$$

The last exponential reproduces the state α as above. Equation(29) and the commutation relations for the $a_{(\sigma)}(\mathbf{k})$ [cf. (22)] show further that the first exponential can be exchanged with the destruction operators left of it because any λ_ν assumes only the values 1,2. Then the exponential acts on ω as above. So both exponentials can be omitted, which means $\varphi_n=\varphi(\alpha)_n$, and $\varphi(\alpha)=\varphi(W\{kg\}\alpha)$, as stated.

For the complete proof of Theorem 5 we must still verify Eq. (37): Assume that α is contained in both $\mathscr{H}^0_g$ *and*

$\mathscr{H}^0_{g'}$ so that $\alpha = W\{k g\}\alpha^1 = W\{k g'\}\alpha^2, \alpha^1, \alpha^2 \in \mathscr{F}^{tr}$. This means $\alpha^1 = W\{k(g-g')\}\alpha^2$, and as $a_{(0)}(\mathbf{k})\alpha^1 = o$, also $a_{(0)}(\mathbf{k})W\{k(g-g')\}\alpha^2$ must vanish. But by (72b) we get

$$a_{(0)}(\mathbf{k})W\{k(g-g')\}\alpha^2$$

$$= W\{k(g-g')\}[k_0(g(\mathbf{k})-g'(\mathbf{k})) + a_{(0)}(\mathbf{k})]\alpha^2$$

$$= [k_0(g(\mathbf{k})-g'(\mathbf{k}))]W\{k(g-g')\}\alpha^2. \qquad (97)$$

This vanishes if and only if $\alpha^2 = o$ which means $\alpha = o$ and thus indeed $\mathscr{H}^0_g \cap \mathscr{H}^0_{g'} = \{o\}$ if $g \neq g'$.

17. THE TIME EVOLUTION OPERATOR *U(t)*

We are ready now to verify the unproven statements in Sec. 6.

By means of (73) we can easily show that the operator

$$U(t) := e^{-i\phi(t)}U_0(t)W\{f[t]\}$$

$$= e^{-i\phi(t)}W\{e[-t]f[t]\}U_0(t) \qquad (98)$$

satisfies Eq. (39) if $U_0(t)$ is given by (74), and the functions $\phi(t), f[t]$ are defined by

$$f_\mu(t,\mathbf{k}) := -i\int_0^t dt'\, e^{it'|\mathbf{k}|} I_\mu(t,\mathbf{k}),$$

$$\Phi(t) := \mathrm{Im}\int d^3k \int_0^t dt'\, e^{it'|\mathbf{k}|} I^*_\mu(t',\mathbf{k})$$

$$\times \int_0^{t'} dt''\, e^{it''|\mathbf{k}|} I^\mu(t'',\mathbf{k}). \qquad (100)$$

$I_\mu(t,\mathbf{k})$ is given by (38) and $f_\mu(t,\mathbf{k})$ is in $\mathfrak{C}$ if $j_\mu(x)$ satisfies the conditions of Sec. 1. Because of $\partial^\mu j_\mu(x) = 0$ we get the identity

$$k^\mu f_\mu(t,\mathbf{k}) + e^{it|\mathbf{k}|} I_0(t,\mathbf{k}) = I_0(0,\mathbf{k}). \qquad (101)$$

By straightforward use of (72c) we get $U^+(t)U(t) = 1 = U(t)U^+(t)$, i.e., $U^+(t) = U^{-1}(t)$, as stated in Sec. 6. Equation (45) is obtained similarly by a straightforward use of (72) and (98)–(100). This also proves Statement 1.

For the proof of Statement 2 we use once more the state α^h given in (88). Further, let $c \in \mathfrak{C}$ be any solution of $k^\mu c_\mu(\mathbf{k}) = q[0](\mathbf{k})$, and consider the vectors $\alpha := W\{c\}\alpha^h$, $\beta := W\{c\}\omega$ which are both in $\mathscr{L}^{q[0]}$ and satisfy $\langle\alpha|\alpha\rangle = 0$, $\langle\beta|\beta\rangle = 1$, $\langle\alpha|\beta\rangle = 0$. Now consider

$$\langle\alpha \mid U(t)\beta\rangle = e^{-i\Phi(t)}\langle\alpha^h \mid W\{-c\}W\{e[-t]f[t]\}$$

$$\times U_0(t)W\{c\}\omega\rangle$$

$$= e^{-i\Phi(t)}\langle\alpha^h \mid W\{-c\}W\{e[-t]f[t]\}$$

$$\times W\{e[-t]c\}\omega\rangle, \qquad (102)$$

where use has been made of (98), (75c), and $U_0(t)\omega = \omega$. Now inserting for α^h and shifting $L(\mathbf{k})$ through the Weyl operators we get by means of (72a)

$$\langle\alpha \mid U(t)\beta\rangle$$

$$= \langle\omega | W\{-c\}W\{e[-t]f[t]\}W\{e[-t]c\}\omega\rangle$$

$$\times \int d^3k\, h^*(\mathbf{k})\, g_t(\mathbf{k}), \qquad (103)$$

where $g_t(\mathbf{k})$ is given by [recall $I^0(t,\mathbf{k}) = q[t](\mathbf{k})$!]

$$g_t(\mathbf{k}) := -q[0](\mathbf{k}) + k^\mu f_\mu(t,\mathbf{k})e^{-it|\mathbf{k}|} + e^{-it|\mathbf{k}|}q[0](\mathbf{k})$$

$$= q[t](\mathbf{k}) - q[0](\mathbf{k}). \qquad (104)$$

The first factor in (103) is always different from zero. For any time t for which $q[t](\mathbf{k}) \neq q[0](\mathbf{k})$, i.e., for $j^0(t,\mathbf{x}) \neq j^0(0,\mathbf{x})$, we have $g_t(\mathbf{k}) \neq 0$ and so we can find an $h(\mathbf{k})$ so that the second factor in (103) is also different from zero. So we obtain $\langle\alpha \mid U(t)\beta\rangle \neq 0$ despite of $\langle\alpha|\alpha\rangle = 0$, as stated.

18. THE OPERATOR MAXWELL EQUATIONS ON A LORENTZ SPACE

We prove now Statements 3–6 in Secs. 7, 8.

As $U(t)$ is explicitly given by (98), we can now easily compute the Heisenberg operators (40), (41). Using (72), (75) we so get

$$A_\mu(x) = A^f_\mu(x) + a^R_\mu(x), \quad \Pi_\mu(x) = \Pi^f_\mu(x) + \pi^R_\mu(x),$$

$$F_{\mu\nu}(x) = F^f_{\mu\nu}(x) + f^R_{\mu\nu}(x). \qquad (105)$$

$A^f_\mu(x)$, $\Pi^f_\mu(x)$, $F^f_{\mu\nu}(x)$ are the Heisenberg operators of the free Maxwell field which are obtained from the corresponding Schrödinger operators (35), (36) by replacing the exponentials $e^{\pm i\mathbf{kx}}$ by $e^{\pm ikx}$. The real functions $a^R_\mu(x)$, $\pi^R_\mu(x)$, $f^R_{\mu\nu}(x)$ are obtained from $A^f_\mu(x)$, $\Pi^f_\mu(x)$, $F^f_{\mu\nu}(x)$, by respectively replacing the operators $a_\mu(\mathbf{k})$, $a^+_\mu(\mathbf{k})$ by the functions $f_\mu(t,\mathbf{k}) f^*_\mu(t,\mathbf{k})$. By means of (105) we can now easily fill the gap in the proof of the results quoted in Sec. 6. Since $f_\mu(t,\mathbf{k})$ vanishes for $t = 0$, at $t = 0$ the operators (105) agree with Schrödinger operators (35), (36), as it must be.

We consider the functions $a^R_\mu(x)$, $\pi^R_\mu(x)$, $f^R_{\mu\nu}(x)$ in some detail. By the above prescription we obtain, e.g.,

$$a^R_\mu(t,\mathbf{x}) = \int d^3k\, D(\mathbf{k})[e^{i(\mathbf{kx}-|\mathbf{k}|t)}f_\mu(t,\mathbf{k}) + \mathrm{cc}]$$

$$= \frac{-1}{4\pi}\int \frac{d^3x'}{|\mathbf{x}-\mathbf{x}'|}$$

$$\times [j_\mu(t-|\mathbf{x}-\mathbf{x}'|,\mathbf{x}')\theta(t-|\mathbf{x}-\mathbf{x}'|)$$

$$- j_\mu(t+|\mathbf{x}-\mathbf{x}'|,\mathbf{x}')\theta(-t-|\mathbf{x}-\mathbf{x}'|)], \qquad (106)$$

where $\theta(\tau) = 1$ for $\tau \geqslant 0$ and $\theta(\tau) = 0$ for $\tau < 0$. The first (second) term vanishes therefore for $t<0$ $(t>0)$. At $t = 0$, both $a^R_\mu(t,\mathbf{x})$ and $(\partial/\partial t)a^R_\mu(t,\mathbf{x})$ vanish. At $t>0$, $a^R_\mu(t,\mathbf{x})$ is a retarded amplitude composed of "signals" which were emitted by the current $j_\mu(t',\mathbf{x}')$ in the points $\mathbf{x}'$ at the previous times $t' := t - |\mathbf{x}-\mathbf{x}'|$, $0<t'<t$. Moving with the velocity of light and being attenuated by an inverse distance law, between t' and t these signals were propagated from the points $\mathbf{x}'$ to the point $\mathbf{x}$ we look at. At time t they all arive at $\mathbf{x}$ and build up the amplitude $a^R_\mu(t,\mathbf{x})$ by interference. The amplitudes $\pi^R_\mu(t,\mathbf{x}), f^R_\mu(t,\mathbf{x})$ have similar causal structures. In addition, $\pi^R_\mu(t,\mathbf{x})$ satisfies $\pi^R_\mu(t,\mathbf{x}) = (\partial/\partial t)a^R_\mu(t,\mathbf{x})$, $\pi^R_\mu(0,\mathbf{x}) = 0$.

We find easily that the pair $\pi^R_\mu(t,\mathbf{x})$, $a^R_\mu(t,\mathbf{x})$ is a special solution of Eq. (6), and that the operators $A^f_\mu(t,\mathbf{x})$, $\Pi^f_\mu(t,\mathbf{x})$ satisfy (6) in the case $j_\mu(t,\mathbf{x}) = 0$. So the proper Heisenberg operators $\Pi_\mu(x)$, $A_\mu(x)$ satisfy (6), and thus also (7), in the sense of operator identities on $\mathscr{S}_M$, as could be expected. Consider also the field operator

$$F_{\mu\nu}(x) := \int d^3k\, D(\mathbf{k})[ik_\mu(a_\nu(\mathbf{k}) + f_\nu(t,\mathbf{k}))$$

$$- ik_\nu(a_\mu(\mathbf{k}) + f_\mu(t,\mathbf{k}))]e^{ikx} + \mathrm{cc} \qquad (107)$$

as defined by the above prescription. We can easily see that

$F_{\mu\nu}(x)$ itself as well as its operator part $F^f_{\mu\nu}(x)$ satisfy (9) as an operator identity on $\mathscr{S}_M$ whereas $a^R_\mu(\mathbf{x})$ and $f^R_{\mu\nu}(x)$ satisfy the same equation in the sense of a c-number solution. This proves Statement 3.

As $A_\mu(x)$ is now explicitly given we can *compute* the expression for $\partial^\mu A_\mu(x)$. By straightforward calculations we find

$$\partial^\mu A_\mu(x) = i\int d^3k\, D(\mathbf{k})[e^{ikx}(L(\mathbf{k}) - q[0](\mathbf{k})) + \mathrm{cc}], \tag{108}$$

where $I^0(0,\mathbf{k})$ has been replaced by $q[0](\mathbf{k})$. Since $L(\mathbf{k}) - q[0](\mathbf{k})$ is independent of t we get the identity (47), but $\partial^\mu A_\mu(x)$ is indeed not the zero operator on $\mathscr{S}_M$. However, we recognize that its restriction to $\mathscr{L}^{q[0]}$ is the zero operator on $\mathscr{L}^{q[0]}$ because the matrix elements $\langle\alpha|\partial^\mu A_\mu(x)\beta\rangle$ vanish for any x and for any $\alpha,\beta\in\mathscr{L}^{q[0]}$. This is Statement 4.

Since Statement 5 needs no further proof we look immediately at Statement 6: Let $\alpha^{\rm tr}, \beta^{\rm tr}$ be any normalized elements from $\mathscr{F}^{\rm tr}$ and let $g = g(\mathbf{k})$ and $g' = g'(\mathbf{k})$ be any gauge functions. Define by $\alpha^0 := W\{kg\}\alpha^{\rm tr}, \beta^0 := W\{kg\}\beta^{\rm tr}$ a couple of elements of $\mathscr{H}^0_g$, and by $\alpha^{0\prime} := W\{kg'\}\alpha^{\rm tr}, \beta^{0\prime} := W\{kg'\}\beta^{\rm tr}$ a couple of elements of $\mathscr{H}^0_{g'}$. These elements satisfy $\alpha^0\sim\alpha^{0\prime}, \beta^0\sim\beta^{0\prime}, \langle\alpha^0|\beta^0\rangle = \langle\alpha^{0\prime}|\beta^{0\prime}\rangle = \langle\alpha^{\rm tr}|\beta^{\rm tr}\rangle$. Let $c = c_\mu(\mathbf{k})$ be any solution of $k^\mu c_\mu(\mathbf{k}) = q[0](\mathbf{k})$ and finally define the elements $\alpha := W\{c\}\alpha^0$, $\beta = W\{c\}\beta^0, \alpha' = W\{c\}\alpha^{0\prime}, \beta' = W\{c\}\beta^{0\prime}$. These elements satisfy $\alpha\sim\alpha', \beta\sim\beta', \langle\alpha|\beta\rangle = \langle\alpha'|\beta'\rangle = \langle\alpha^{\rm tr}|\beta^{\rm tr}\rangle$, and *any* normalized elements $\alpha,\beta,\alpha',\beta'$ in $\mathscr{L}^{q[0]}$, which satisfy these relations, can be written in the form given. Consider finally the difference $\langle\alpha|A_\mu(x)\beta\rangle - \langle\alpha'|A_\mu(x)\beta'\rangle$. By (105) we may replace $A_\mu(x)$ by $A^f_\mu(x)$ without changing its value. So we get

$$\langle\alpha|A_\mu(x)\beta\rangle - \langle\alpha'|A_\mu(x)\beta'\rangle = \int d^3k\, D(\mathbf{k})[e^{ikx}z_\mu(\mathbf{k}) + \mathrm{cc}], \tag{109}$$

where $z_\mu(\mathbf{k})$ is given by

$$\begin{aligned} z_\mu(\mathbf{k}) :&= \langle\alpha|a_\mu(\mathbf{k})\beta\rangle - \langle\alpha'|a_\mu(\mathbf{k})\beta'\rangle \\ &= \langle\alpha^{\rm tr}|W\{-kg\}W\{-c\}a_\mu(\mathbf{k})\,W\{c\}W\{kg\}\beta^{\rm tr}\rangle \\ &\quad - \langle\alpha^{\rm tr}|W\{-kg'\} \\ &\qquad \times W\{-c\}a_\mu(\mathbf{k})W\{c\}W\{kg'\}\beta^{\rm tr}\rangle. \end{aligned} \tag{110}$$

Using $W\{-c\}a_\mu(\mathbf{k})W\{c\} = a_\mu(\mathbf{k}) + c_\mu(\mathbf{k})$ we see that the contributions from the term $c_\mu(\mathbf{k})$ cancel. Using the same fromula once more we get

$$z_\mu(\mathbf{k}) = k_\mu(g(\mathbf{k}) - g'(\mathbf{k}))\langle\alpha^{\rm tr}|\beta^{\rm tr}\rangle. \tag{111}$$

If this is inserted into (109), the right-hand side assumes the form of the right-hand side of (48) because $\langle\alpha^{\rm tr}|\beta^{\rm tr}\rangle = (\bar\alpha|\bar\beta)$.

By (107) we similarly get

$$\langle\alpha|F_{\mu\nu}(x)\beta\rangle - \langle\alpha'|F_{\mu\nu}(x)\beta'\rangle = \int d^3x\, D(\mathbf{k})[ik_\mu\, z_\nu(\mathbf{k}) - ik_\nu\, z_\mu(\mathbf{k})]e^{ikx} + \mathrm{cc} \tag{112}$$

with the same $z_\mu(\mathbf{k})$ as above. The right-hand side of (112) therefore vanishes, and this already completes the proof of Statement 6.

19. THE STATIONARY CASE

We now verify Statements 7–9 in Sec. 10 which are all related to the stationary case $j^\mu(t,\mathbf{x}) = j^\mu(0,\mathbf{x}), I^\mu(t,\mathbf{k}) = I^\mu(\mathbf{k}), I^0(\mathbf{k}) = q(\mathbf{k})$.

The time integrals in (99), (100) can now be computed explicitly and lead to

$$f_\mu(t,\mathbf{k}) = (1 - e^{it|\mathbf{k}|})I_\mu(\mathbf{k})/|\mathbf{k}|, \tag{113}$$

$$\Phi(t) = t\cdot L + \int d^3k\, I^*_\mu(\mathbf{k})\, I^\mu(\mathbf{k})[\sin(t\,|\mathbf{k}|)]/|\mathbf{k}|^2, \tag{114}$$

where L is given in (53). Inserting this into (98) we get a simpler form for $U(t)$. Statement 8 can then be easily proved by a straightforward verification.

By completing the square in (37) we obtain for the Hamiltonian $H = H_t$ the expression

$$H = \int d^3k\,|\mathbf{k}|\,[a^+_\mu(\mathbf{k}) + I^*_\mu(\mathbf{k})/|\mathbf{k}|\,] \times [a^\mu(\mathbf{k}) + I^\mu(\mathbf{k})/|\mathbf{k}|\,] + L. \tag{115}$$

By means of (72b) we can write this in the form

$$H = W\{\bar c\}(H_0 + L)W\{-\bar c\}, \tag{116}$$

where H_0 is the Hamilton operator if M is free, $H_0\omega = o$, and $\bar c$ is the special 4-vector $\bar c_\mu(\mathbf{k}) := -I_\mu(\mathbf{k})/|\mathbf{k}|$. Because $\mathbf{kI(k)} = 0$, it satisfies $k^\mu\bar c_\mu(\mathbf{k}) = -k^\mu I_\mu(\mathbf{k})/|\mathbf{k}| = I^0(\mathbf{k}) = q(\mathbf{k})$ so that the coherent state $\beta := W\{\bar c\}\omega$ is in $\mathscr{L}^q$. Moreover, we have

$$H\beta = W\{\bar c\}(H_0 + L)W\{-\bar c\}W\{\bar c\}\omega = LW\{\bar c\}\omega = L\beta,$$

so that β is indeed an eigenstate of H to the eigenvalue L.

For the completion of the proof of Statement 7 we must only show that it is the only eigenstate of H in $\mathscr{L}^q$. Let α be any element of $\mathscr{L}^q$. We show in the Appendix that $W\{\bar c\}\alpha$ exists. By means of (98) we further find

$$\begin{aligned} U(t)W\{\bar c\}\alpha &= e^{-i\Phi(t)}W\{e[-t\,]f[t\,]\}U_0(t)W\{\bar c\}\alpha \\ &= e^{-i\Phi(t)}W\{e[-t\,]f[t\,]\} \\ &\quad \times W\{e[-t\,]\bar c\}U_0(t)\,\alpha \\ &= e^{-itL}W\{\bar c\}U_0(t)\alpha. \end{aligned} \tag{117}$$

For $\alpha = \omega$ this reproduces the eigenstate property of β as $U_0(t)\omega = \omega$. As ω is the only state in $\mathscr{S}_M$ with this property, $\beta = W\{\bar c\}U_0(t)\omega$ is the only state in $\mathscr{L}^q$ which satisfies $U(t)\beta = e^{-itE}\beta$, where E is a number, in our case $E = L$. So β is indeed the only eigenstate of H.

We come now to Statement 9: In Secs. 15 and 16 we defined $\mathscr{H}^q_g$ by $\mathscr{H}^q_g := W\{c\}W\{kg\}\mathscr{F}^{\rm tr} = W\{c + kg\}\mathscr{F}^{\rm tr}$, where $c\in\mathfrak{C}$ was *any* solution $c_\mu(\mathbf{k})$ of $k^\mu c_\mu(\mathbf{k}) = q(\mathbf{k})$, and $g = g(\mathbf{k})$ was any gauge function. But as $c' = c + kg'$ with any $g'(\mathbf{k})$ is also a solution of $k^\mu c'_\mu(\mathbf{k}) = q(\mathbf{k})$, we must either specify c or g. A natural choice is $c = \bar c$ with the above $\bar c$. So we finally *define* the gauge Hilbert spaces in $\mathscr{L}^q$ by

$$\mathscr{H}^q_g = W\{\bar c + kg\}\mathscr{F}^{\rm tr}. \tag{118}$$

For the proof of Statement 9 we must show that g can be chosen so that $U(t)\mathscr{H}^q_g = \mathscr{H}^q_g$, and that β is in this $\mathscr{H}^q_g$. We claim that this is achieved for $g(\mathbf{k}) = 0$ if we adopt the definition (118). As $\beta = W\{\bar{c}\}\omega$ is in $\mathscr{H}^q_0$, we must only verify that $U(t)\mathscr{H}^q_g = \mathscr{H}^q_g$ holds only for $g = 0$. The relation $\mathscr{H}^q_g = U(t)\mathscr{H}^q_g$ is obviously equivalent to the equation $\mathscr{F}_t = \mathscr{F}^{\text{tr}}$ where $\mathscr{F}_t$ is given by

$$\mathscr{F}_t := W\{-\bar{c} - kg\}U(t)W\{\bar{c} + kg\}\mathscr{F}^{\text{tr}}. \tag{119}$$

So $\mathscr{F}_t$ must satisfy $a^{(0)}(\mathbf{k})\mathscr{F}_t = o = a^{(3)}(\mathbf{k})\mathscr{F}_t$. By means of (98), (75c), $U_0(t)\mathscr{F}^{\text{tr}} = \mathscr{F}^{\text{tr}}$, (72b), and finally by (101) and the definition $\bar{c}_\mu := -I_\mu(\mathbf{k})/|\mathbf{k}|$ we find

$$\begin{aligned}\mathscr{F}_t &= W\{-(\bar{c} + kg\}W\{e[-t]f[t]\}U_0(t)W\{\bar{c} + kg\} \\ &\quad\times\mathscr{F}^{\text{tr}} \\ &= W\{-(\bar{c} + kg\}W\{e[-t]f[t]\} \\ &\quad\times W\{e[-t](\bar{c} + kg)\}\mathscr{F}^{\text{tr}} \\ &= W\{(e[-t] - 1)(\bar{c} - \bar{c} + kg)\}\mathscr{F}^{\text{tr}} \\ &= W\{(e[-t] - 1)kg\}\mathscr{F}^{\text{tr}}.\end{aligned} \tag{120}$$

By (72a), (94a), the conditions to be satisfied are

$$\begin{aligned}o &= -a_{(0)}(\mathbf{k})\mathscr{F}_t = (e^{-it|\mathbf{k}|} - 1)k_0\, g(\mathbf{k})\mathscr{F}_t, \\ o &= a_{(3)}(\mathbf{k})\mathscr{F}_t = (e^{-it|\mathbf{k}|} - 1)(k^r k_r/|\mathbf{k}|)g(\mathbf{k})\mathscr{F}_t.\end{aligned} \tag{121}$$

As $e^{-it|\mathbf{k}|} - 1$ does not vanish identically in $\mathbf{k}$ and the same holds for k_0 and $k^r k_r = |\mathbf{k}|^2$, the conditions (121) are met if and only if $g(\mathbf{k}) = 0$.

the choice $c = \bar{c}$ is arbitrary relative to the transverse part of c which however only depicts $\mathscr{H}^q_0$, as defined by (118), onto itself. The *space* $\mathscr{H}^q_0$ is therefore *uniquely* defined by the requirements $U(t)\mathscr{H}^q_0 = \mathscr{H}^q_0$ and $\beta\in\mathscr{H}^q_0$. This completes the proof of Statement 9.

20. THE GAUGE INVARIANT ENERGY

In this last section we verify Statements 10 and 11 related to the concept of energy.

For statement 10 we note that the restriction H^q_0 of H to $\mathscr{H}^q_0$ which exists by the general arguments quoted in Sec. 10, is of course determined by the equation $H^q_0\,\alpha = H\alpha$ for any $\alpha\in\mathscr{H}^q_0\cap\mathbb{D}(H)$. This α can be written in the form $\alpha = W\{\bar{c}\}\alpha^{\text{tr}}$, where $\alpha^{\text{tr}}\in\mathscr{F}^{\text{tr}}$ is conversely given by $\alpha^{\text{tr}} := W\{-\bar{c}\}\alpha$. As $\gamma := H^q_0\alpha$ is in $\mathscr{H}^q_0$ for any $\alpha\in\mathbb{D}(H^q_0) = \mathscr{H}^q_0\cap\mathbb{D}(H)$, it can be similarly written in the form $\gamma = W\{\bar{c}\}\gamma^{\text{tr}}$, and so we get by means of (116)

$$\begin{aligned}\gamma^{\text{tr}} &= W\{-\bar{c}\}HW\{\bar{c}\}\alpha^{\text{tr}} = (H_0 + L)\alpha^{\text{tr}} \\ &= \left[L + \int d^3k\,|\mathbf{k}|a^+_\mu(\mathbf{k})a^\mu(\mathbf{k})\right]\alpha^{\text{tr}} \\ &= \Big[L + \int d^3k\,|\mathbf{k}|\{a_{(1)+}(\mathbf{k})a_{(1)}(\mathbf{k}) + a_{(2)+}(\mathbf{k})a_{(2)}(\mathbf{k}) \\ &\quad + a_{(3)+}(\mathbf{k})a_{(3)}(\mathbf{k}) - a_{(0)+}(\mathbf{k})a_{(0)}(\mathbf{k})\}\Big]\alpha^{\text{tr}}.\end{aligned} \tag{122}$$

The contributions from the last couple of terms vanish because of (24a). What remains is the usual "energy"-Hamiltonian H^{tr}_0 of free, transverse photons. So we obtain H^q_0 in the form

$$H^q_0 = W\{\bar{c}\}(H^{\text{tr}}_0 + L)W\{-\bar{c}\} \tag{123}$$

which verifies Statement (10) up to Eqs. (55) and (56) which, however, are straightforward consequences of the definition (118) of $\mathscr{H}^q_0$.

Let us finally return to $\mathscr{L}^q$ and verify the last Statement 11 which can also be written in the form: $U^q(t)\alpha \sim U^q(t)\beta$ if $\alpha\sim\beta$, $\alpha,\beta\in\mathscr{L}^q$. This is equivalent to: $U^q(t)\times(\alpha - \beta)\sim o$ if $(\alpha - \beta)\sim o$. Now let $\gamma\sim o$, $\gamma\in\mathscr{L}^q$ be given, and consider $\langle\delta\,|\,U^q(t)\gamma\rangle$ for *any* $\delta\in\mathscr{L}^q$. By the Cauchy–Schwarz inequality which holds here because $U^q(t)\gamma$ is in $\mathscr{L}^q$, we get

$$\begin{aligned}|\langle\delta|U^q(t)\gamma\rangle|^2 &= |\langle U^q(-t)\delta|\gamma\rangle|^2 \\ &\leqslant\langle U^q(-t)\delta|U^q(-t)\delta\rangle\,\langle\gamma|\gamma\rangle = 0,\end{aligned}$$

i.e., $\langle\delta|U^q(t)\gamma\rangle = 0$ for *any* $\delta\in\mathscr{L}^q$. In particular, for $\delta = U^q(t)\gamma$ we get $\langle U^q(t)\gamma|U^q(t)\gamma\rangle = 0$ so that $U^q(t)\gamma = U^q(t)(\alpha - \beta)\sim o$, as stated.

APPENDIX: ON THE DOMAIN OF A WEYL OPERATOR ON $\mathscr{S}_M$

We show first and in the usual way that $\mathbb{D}(W\{c\})$ is dense, cf. (71).

Let $\alpha = \alpha_n$ be any vector in $\mathscr{S}_M$ with only one nonvanishing component $\alpha_n(K^n)$, $n = 0,1,\cdots$. Denote by $F_1\{c\}$ and $F_2\{c\}$ the first and second term in (67) so that $F\{c\} = F_1\{c\} + F_2\{c\}$. We obviously have the relations

$$\|F_1\{c\}\alpha_n\|\leqslant(n + 1)^{1/2}\|c\|\|\alpha_n\|$$

and

$$\|F_2\{c\}\alpha_n\|\leqslant n^{1/2}\|c\|\|\alpha_n\| < (n + 1)^{1/2}\|c\|\|\alpha_n\|$$

so that

$$\|F\{c\}\alpha_n\|\leqslant(n + 1)^{1/2}2\|c\|\|\alpha_n\|, \tag{A1}$$

where $\|c\|$ is given by

$$\|c\|^2 = \sum_{\mu=0}^{3}\int d^3k\,|c_\mu(\mathbf{k})|^2. \tag{A2}$$

By induction we can easily obtain the relation

$$\|F\{c\}^\nu\alpha_n\|\leqslant[(n + \nu)\cdots(n + 1)]^{1/2}(2\|c\|)^\nu\|\alpha_n\| \tag{A3}$$

and thus

$$\begin{aligned}\|W\{c\}\alpha_n\| &= \left\|\sum_{\nu=0}^{\infty}\frac{1}{\nu!}(F\{c\})^\nu\alpha_n\right\| \\ &\leqslant\|\alpha_n\|\sum_{\nu=0}^{\infty}\frac{1}{\nu!}[(n + \nu)\ldots(n + 1)]^{1/2}(2\|c\|)^\nu \\ &\leqslant\|\alpha_n\|\sum_{\nu=0}^{\infty}\frac{1}{(\nu!)^{1/2}}[(n + 1)^{1/2}2\|c\|]^\nu.\end{aligned} \tag{A4}$$

The last sum converges for any n. So $W\{c\}$ is defined at least on the dense set $\mathbb{D}$ of all α with only a finite number of nonvanishing components $\alpha_n(K^n)$.

Now we convince ourselves that $W\{c\}\alpha$ is in $\mathscr{S}_M$ if $c\in\mathfrak{C}$ and if α is any element from any Lorentz space $\mathscr{L}^q$.

For the proof of this we use the Hilbert metric $(\cdot|\cdot)$ defined by (15). Then we get

$$\left(\int d^3k\,h^*_\mu(\mathbf{k})a^\mu(\mathbf{k})\alpha\,|\beta\right) = \left(\alpha|\int d^3k\,h^\mu(\mathbf{k})\,g_{\mu\mu}a^+_\mu(\mathbf{k})\beta\right) \tag{A5}$$

if either $\alpha\in\mathbb{D}$ or $\beta\in\mathbb{D}$. Under the same condition we obviously get:

$$\left(\exp\int d^3k\, h_\mu^*(\mathbf{k})a^\mu(\mathbf{k})\alpha\,|\,\beta\right) = \left(\alpha\,|\,\exp\int d^3k\, h^\mu(\mathbf{k})\, g_{\mu\mu}a_\mu^+(\mathbf{k})\beta\right), \tag{A6}$$

$$(F\{c\}\alpha\,|\,\beta) = (\alpha\,|\,F\{c'\}\beta), \tag{A7}$$

$$(W\{c\}\alpha\,|\,\beta) = (\alpha\,|\,W\{c'\}\beta), \tag{A8}$$

where c' is given by

$$c'_\mu(\mathbf{k}) = -g_{\mu\mu}c_\mu(\mathbf{k}). \tag{A9}$$

As $W\{c\}\alpha$ exists for any $\alpha\in\mathbb{D}$, we get by means of (A6)

$$\begin{aligned}\|W\{c\}\alpha\|^2 &= (W\{c\}\alpha\,|\,W\{c\}\alpha) = (\alpha\,|\,W\{c'\}W\{c\}\alpha)\\ &= (\alpha\,|\,W\{c+c'\}\alpha),\end{aligned} \tag{A10}$$

where $c+c'$ has the components $(c+c')^0(\mathbf{k}) = 2c^0(\mathbf{k})$, $(c+c')^r(\mathbf{k}) = 0$ for $r = 1,2,3$. By means of (72c) we thus get

$$\begin{aligned}W\{c+c'\} &= \left(\exp 2\int d^3k\, |c^0(\mathbf{k})|^2\right)\\ &\quad\times\left(\exp 2\int d^3k\, c^0(\mathbf{k})a_0^+(\mathbf{k})\right)\\ &\quad\times\left(\exp\left[-2\int d^3k\, c_0^*(\mathbf{k})a^0(\mathbf{k})\right]\right).\end{aligned} \tag{A11}$$

By the use of (A6) we therefore arrive at

$$\begin{aligned}\|W\{c\}\alpha\| &= \left(\exp\int d^3k\, |c^0(\mathbf{k})|^2\right)\\ &\quad\times\left\|\exp\left[-2\int d^3k\, c_0^*(\mathbf{k})a^0(\mathbf{k})\right]\alpha\right\|.\end{aligned} \tag{A12}$$

Now consider any $\alpha\in\mathscr{L}^q$, $q\in\mathfrak{Q}$. Then

$$a^0(\mathbf{k})\alpha = \frac{1}{|\mathbf{k}|}\,[k^r a_r(\mathbf{k}) - q(\mathbf{k})]\alpha \tag{A13}$$

holds for almost any $\mathbf{k}\in\mathbb{R}^3$. As $a^0(\mathbf{k})$ commutes with any $a_r(\mathbf{k})$, in any term of the Taylor series for the last term in (A12) we can replace $a^0(\mathbf{k})\alpha$ by the right-hand side of (A13), i.e.,

$$\begin{aligned}&\exp\left[-2\int d^3k\, c_0^*(\mathbf{k})a^0(\mathbf{k})\right]\alpha\\ &= \exp\left\{-2\int d^3k\, [c_0^*(\mathbf{k})/|\mathbf{k}|][k^r a_r(\mathbf{k}) - q(\mathbf{k})]\right\}\alpha\\ &= \left(\exp 2\int d^3k\, c_0^*(\mathbf{k})q(\mathbf{k})/|\mathbf{k}|\right)\\ &\quad\times\left(\exp\left[-2\int d^3k\, c_0^*(\mathbf{k})a_r(\mathbf{k})/|\mathbf{k}|\right]\right)\alpha.\end{aligned} \tag{A14}$$

So we get

$$\begin{aligned}&\|W\{c\}\alpha\|\\ &= \left(\exp\int d^3k\, |c^0(\mathbf{k})|^2\right)\\ &\quad\times\exp\left(2\int d^3k\, c_0^*(\mathbf{k})q(\mathbf{k})/|\mathbf{k}|\right)\\ &\quad\times\left\|\exp\left[-2\int d^3k\, c_0^*(\mathbf{k})a_r(\mathbf{k})k_r/|\mathbf{k}|\right]\alpha\right\|.\end{aligned} \tag{A15}$$

By (A6) and (45) we obtain

$$\begin{aligned}&\left\|\exp\left[-2\int d^3k\, c_0^*(\mathbf{k})a^r(\mathbf{k})k_r/|\mathbf{k}|\right]\alpha\right\|^2\\ &= \left(\alpha\left|\exp\left[-2\int d^3k\, c_0(\mathbf{k})a_r^+(\mathbf{k})k^r/|\mathbf{k}|\right]\right.\right.\\ &\quad\left.\times\exp\left[-2\int d^3k\, c_0^*(\mathbf{k})a_r(\mathbf{k})/|\mathbf{k}|\right]\alpha\right)\\ &= (\alpha\,|\,W\{g\}\alpha)\exp\left[-2\int d^3k\, |c_0(\mathbf{k})|^2\right],\end{aligned} \tag{A16}$$

where g is the 4-vector with the components $g^0(\mathbf{k}) = 0$, $g^r(\mathbf{k}) = 2c^0(\mathbf{k})k^r/|\mathbf{k}|$, $r = 1,2,3$. By the Cauchy–Schwarz inequality, which of course holds for the metric $(\cdot|\cdot)$, we get

$$(\alpha\,|\,W\{g\}\alpha)\,|\leqslant\|\alpha\|\,\|W\{g\}\alpha\| = \|\alpha\|^2, \tag{A17}$$

where use has been made of $\|W\{g\}\alpha\| = \|\alpha\|$ which holds by (A10) because $g^0(\mathbf{k}) = 0$. Inserting (A17) into (A16) and this into (A15), we finally get

$$\begin{aligned}\|W\{c\}\alpha\| &\leqslant \|\alpha\|\,\left|\exp 2\int d^3k\, c_0^*(\mathbf{k})q(\mathbf{k})/|\mathbf{k}|\right|\\ &= \|\alpha\|\exp\left[2\,\mathrm{Re}\int d^3k\, c_0^*(\mathbf{k})q(\mathbf{k})/|\mathbf{k}|\right].\end{aligned}$$

This proves the statement because the second factor is independent of α.

[1] S. Gupta, Proc. Phys. Soc. London A **63**, 681 (1950); **64**, 850 (1951).
[2] K. Bleuler, Helv. Phys. Acta **23**, 567 (1950).
[3] W. Heitler, *The Quantum Theory of Radiation*, 3rd ed. (Clarendon, Oxford, 1954).
[4] J.M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley, Cambridge, Mass., 1955).
[5] N.N. Bogoljubow and D.V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959).
[6] S.S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, New York, 1964).
[7] K.L. Nagy, *State vector spaces with indefinite metric in quantum field theory* (Adadémiai Kiadó, Budapest, 1966).
[8] H.P. Dürr and E. Rudolph, Nuovo Cimento A **62**, 411 (1969).
[9] F. Strocchi and A.S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[10] J. Bognár, *Indefinite Inner Product Spaces* (Springer, Berlin 1974).
[11] N. Bourbaki, *Espaces vectorielles topologiques* (Hermann, Paris, 1955).
[12] U. Mutze, Fortschr. Phys. **23**, 113 (1975).
[13] J.M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).
[14] K. Hepp, *Théorie de la renormalisation* (Springer, Berlin, 1969).
[15] V. Ernst, J. Phys. A: Math. Gen. **11**, 311 (1978).
[16] K.O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience, New York, 1953).
[17] J.M. Cook, Trans. Am. Math. Soc. **74**, 222 (1953).
[18] D. Kastler, *Introduction a l'électrodynamique quantique* (Dunod, Paris, 1961).
[19] R.J. Glauber, Phys. Rev. **84**, 395 (1951).
[20] J.M. Cook, J. Math. Phys. **2**, 33 (1961).
[21] W. Heisenberg and W. Pauli, Z. Phys. **56**, 1 (1929); **59**, 169 (1930).
[22] R.J. Glauber, Phys. Rev. **130**, 2529 (1963); **131**, 2766 (1963).
[23] J. Gomatam, Phys. Rev. D **6**, 1292 (1971).
[24] P.A.M. Dirac, *The Principles of Quantum Mechanics*, 4th ed. (Oxford U. P., Oxford, 1958).

[25] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).
[26] G. Wentzel, *Quantum Theory of Fields* (Interscience, New York, 1949).
[27] A. Sommerfeld, *Elektrodynamik* (Akademische Verlagsgesellschaft, Leipzig, 1967).
[28] E. Hille and R.S. Phillips, *Functional Analysis and Semi-Groups* (American Mathematical Society, Providence, R.I., 1957).
[29] M. Born, Z. Phys. **58**, 803 (1926).
[30] I. Bialynicki-Birula, in *Mathematical Physics and Physical Mathematics*, edited by K. Maurin and R. Raczka, (Reidel, Dordrecht, 1976), p. 39.
[31] N.J. Achieser and I.M. Glasman, *Theorie der linearen Operatoren im Hilbertraum*, 5th ed. (Akademie, Berlin, 1968).

Summation of partial wave expansions in the scattering by short-range potentials [a)]

Fernando F. Grinstein
Instituto de Física, Universidade Federal Fluminense, Outeiro de São João Batista s/n, Caixa Postal 296, 24000 Niterói-RJ, Brazil

(Received 20 March 1979; accepted for publication 23 August 1979)

Previous theorems on the convergence of the punctual Padé approximant to the scattering amplitude are extended. The new proofs correspond to the case of potentials having a short-range tail of the type $V(r)_{r\to\infty} \sim V_0 r^{-\rho-1}\exp[-\mu r]$, where V_0 is a constant, ρ an integer and $\mu > 0$, and are restricted to within the Lehmann ellipse, in the complex $\cos\theta$ plane, where the partial wave expansion converges. Asymptotic estimates are obtained for the error of the approximants.

1. INTRODUCTION

A usual method for the calculation of differential cross sections is that based on the partial wave expansion of the scattering amplitude (PWESA). For two-body potential scattering, if we restrict ourselves to spherically symmetric interactions, this series is given by

$$f(\cos\theta) = \sum_{l=0}^{\infty} a_l P_l(\cos\theta) = \sum_{l=0}^{\infty} (2l+1)\frac{[\exp(2i\delta_l)-1]}{2ik} P_l(\cos\theta), \quad (1.1)$$

where k is the magnitude of the wave vector, the $\{\delta_l\}$ are the phase shifts, and the $\{P_l\}$ the Legendre polynomials. The number of terms in (1.1) which contribute significantly to the series can be estimated by the semiclassical relation

$$N_{\max} \sim kr_0 , \quad (1.2)$$

where r_0 is the effective range of the potential. Thus, the method is of great value when relatively low values of r_0, of the energy, and of the reduced mass of the system are involved. Otherwise, the convergence of the PWESA is slow, and a considerable number of phase shifts must be calculated in order to obtain accurate values for the scattering amplitude. Typical of these situations are atomic and molecular collision processes, where long range interactions are present, and nuclear systems described by short range potentials in the intermediate energy region.

Several alternative methods of calculation can be used for some of these cases, such as the semiclassical and distorted wave approximations. An interesting global approach, however, is to keep the original PWESA, and to find adequate mathematical methods to resume efficiently the information contained in the terms of the series. Such methods are available within the framework of Padé-type rational approximations, and some of them have been proposed for that purpose.[1-4] Of particular interest is the punctual Padé approximant (PPA) approach,[2] which has the attractive feature, from the numerical point of view, of counting on algorithms which allow for a recurrent calculation of the successive approximations.

The convergence of the PPA to $f(\cos\theta)$ was proven[2,3] in the case of long range potentials having the asymptotic behavior

$$V(r) \underset{r\to\infty}{\sim} V_0/r^{\alpha+2}, \quad \alpha \geqslant -1 , \quad (1.3)$$

where α is an integer and V_0 a constant. The proofs included all of the cases for which $f(\cos\theta)$ has finite meaningful values in the physical interval $-1 \leqslant \cos\theta \leqslant 1$, where the convergence of the PWESA is, in principle, restricted. The efficiency of the approach was shown, as a convergence acceleration procedure, when the PWESA is slowly convergent, and, as a regularization method, when it is divergent or oscillating. Furthermore, its importance from the practical point of view was verified in a set of typical examples of situations involving this type of potentials.[4]

In this work, we deal with potentials having the short range tail

$$V(r) \underset{r\to\infty}{\sim} V_0 r^{-\rho-1}\exp(-\mu r) , \quad (1.4)$$

where V_0 is a constant, ρ an integer, and $\mu > 0$. We prove the convergence of the PPA to $f(\cos\theta)$, for $\cos\theta$ within the Lehmann ellipse in the complex $\cos\theta$ plane, where the convergence of the PWESA is restricted, and obtain asymptotic estimates for the error of the approximants. It is shown that the rate of convergence of the nontrivial PPA's is asymptotically greater than that of the sequence of partial wave sums of the PWESA.

In Sec. 2 we briefly outline the PPA method, and prove two theorems regarding the convergence of the PPA when applied to certain sequences. The asymptotic behavior of the sequence of partial wave sums of the PWESA within the Lehmann ellipse is investigated in Sec. 3, and it is shown that its study essentially reduces to that of the sequences considered in Sec. 2, and, thus, we are able to establish the main results of this work.

2. THE PUNCTUAL PADÉ APPROXIMANT (PPA)

A. The method

Given a formal power series

$$g(z) = \sum_{n=0}^{\infty} a_n z^n , \quad (2.1)$$

[a)]Partially supported by CNPq (Brazil).

the $[N, M]_{g(z)}$, Padé approximant (PA) to $g(z)$, is defined as a quotient of polynomials,[5]

$$[N, M]_{g(z)} = R_M(z)/Q_N(z) = (r_0 + r_1 z + r_2 z^2 + \cdots + r_M z^M)/(1 + q_1 z + q_2 z^2 + \cdots + q_N z^N), \quad (2.2)$$

where the coefficients $\{r_i\}$ and $\{q_i\}$ are uniquely determined by the requirement

$$[N, M]_{g(z)} - g(z) = \mathcal{O}[z^{M+N+1}]. \quad (2.3)$$

The PPA $[N, M]_{f(\cos\theta)}$ is defined[2] as the PA $[N, M]_{F(\cos\theta, z)}$ to the series

$$F(\cos\theta, z) = \sum_{l=0}^{\infty} a_l P_l(\cos\theta) z^l,$$

calculated at $z = 1$. In this way, a doubly infinite array of rational approximants to $f(\cos\theta)$ are introduced, the PPA table, from which many different sequences may be chosen. Of particular interest, are the row sequences, consisting of the PPA $[n, n+m]_{f(\cos\theta)}$ for fixed $n \geqslant 0$. By defining

$$S_m(\cos\theta) = \sum_{l=0}^{m} a_l P_l(\cos\theta), \quad (2.4)$$

they can be shown to coincide with the $E_n(S_m)$ nonlinear transforms introduced by Shanks[5] as formal generalizations of Aitken's extrapolation formula, which corresponds to $n = 1$. Moreover, for $n = 0$, i.e., for the first row of the table, one has

$$E_0(S_m) = S_m .$$

Thus, the sequence of partial wave sums is one of the particular sequences of approximations to $f(\cos\theta)$ which may be chosen within the PPA table.

The $[n, n+m]_f$ may be expressed in a compact form:

$$[n, n+m]_f = H_{n+1}^{(m)}\{S_r\}/H_n^{(m)}\{\Delta^2 S_r\}, \quad (2.5)$$

where $\Delta^0 S_r = S_r$, $\Delta^p S_r = \Delta^{p-1} S_{r+1} - \Delta^{p-1} S_r$ for $p > 0$, and the $H_k^{(s)}$ are the Hankel determinants, defined for a given sequence $\{f_r\}$ as

$$H_k^{(s)}\{f_r\} = \begin{bmatrix} f_s & f_{s+1} & \cdots & f_{s+k-1} \\ f_{s+1} & f_{s+2} & \cdots & f_{s+k} \\ \vdots & \vdots & & \vdots \\ f_{s+k-1} & f_{s+k} & \cdots & f_{s+2k-2} \end{bmatrix},$$

for $k > 0$ and $H_0^{(s)}\{f_r\} \equiv 1$.

By inspection of Eq. (2.5) it is clear that the PPA $[n, n+m]_f$ may be seen, equivalently, as a nonlinear transformation of the sequence $\{S_m\}$ (given $S_m, S_{m+1}, \ldots, S_{m+2n}$), or of the series $F(\cos\theta, 1)$ (given $a_0, a_1, \ldots, a_{m+2n}$). Because of this equivalence, reference will be made, alternatively, to the series or to the sequence, and we shall define

$$[n, n+m]_{\{S_r\}} \equiv [n, n+m]_{f(\cos\theta)} .$$

Furthermore, let us state a useful property of the PPA which may be easily deduced from Eq. (2.5). If

$$S_m = A + BS'_m ,$$

where A and B are constants, then

$$[n, n+m]_{\{S_r\}} = A + B\,[n, n+m]_{\{S'_r\}} . \quad (2.6)$$

A final remark can be made with regards to the calculation of the PPA, which becomes very involved, from the numerical point of view, when n is large and Eq. (2.5) is used. The PPA table may be generated recurrently by means of certain algorithms,[4] which, while doing so efficiently, allow us as well to check the convergence of the successive approximations being calculated.

B. Convergence of the PPA for certain sequences

Definition: Given the sequences $\{A_m\}$ and $\{B_m\}$, we shall say that A_m asymptotically approaches B_m, that is,

$$A_m \underset{m\to\infty}{\sim} B_m ,$$

if, given any small positive quantity ϵ, an integer $m_0 > 0$ can be found such that, for $m \geqslant m_0$

$$|A_m - B_m|/|A_m| < \epsilon .$$

Lemma 2.1: Given the sequence $\{D_r\}$ with asymptotic representation

$$D_r \underset{r\to\infty}{\sim} (r + \tfrac{1}{2})^{-\nu}, \quad (2.7)$$

where ν is an integer, then

$$H_k^{(p)}\{D_r\} \underset{p\to\infty}{\sim} (p + \tfrac{1}{2})^{-k(\nu+k-1)} \prod_{t=0}^{k-1} [-\nu]_t (-1)^t t!, \quad (2.8)$$

where $[-\nu]_t = (-\nu)(-\nu-1)\cdots(-\nu-t+1)$ if $t > 0$ and $[-\nu]_0 = 1$.

Proof: We have

$$H_k^{(p)}\{D_r\} \underset{p\to\infty}{\sim} H_k^{(p)}\{(r+\tfrac{1}{2})^{-\nu}\} = H_k^{(0)}\{\Delta^r (p+\tfrac{1}{2})^{-\nu}\} . \quad (2.9)$$

Recalling now that

$$\Delta^r (p + \tfrac{1}{2})^{-\nu} \underset{p\to\infty}{\sim} [-\nu]_r (p + \tfrac{1}{2})^{-\nu-r} \quad (2.10)$$

and the equality[3]

$$H_n^{(m)}\{[\rho]_r\} = \prod_{t=0}^{n-1} [\rho]_{m+t} (-1)^t t!, \quad (2.11)$$

valid for any ρ, we obtain from (2.9), using (2.10),

$$H_k^{(p)}\{D_r\} \underset{p\to\infty}{\sim} H_k^{(0)}\{\Delta^r (p+\tfrac{1}{2})^{-\nu}\} \underset{p\to\infty}{\sim} H_k^{(0)}\{[-\nu]_r (p+\tfrac{1}{2})^{-\nu-r}\} = (p+\tfrac{1}{2})^{-k(\nu+k-1)} H_k^{(0)}\{[-\nu]_r\},$$

from which (2.8) follows, by using (2.11).

Theorem 2.1: Given the sequence $\{C_r\}$ whose members have the asymptotic representation

$$C_r \underset{r\to\infty}{\sim} R^{-\nu} q^R, \quad (2.12)$$

where $R = r + \frac{1}{2}$, $|q| < 1$, and ν is an integer, the PPA $[n, n+m]_{\{C_r\}}$ has, for fixed $n \geqslant 0$ and large m, the following asymptotic behavior:

$$[n, n+m]_{\{C_r\}} \underset{m\to\infty}{\sim} q^{(m+\frac{1}{2}+2n)} [-\nu]_n (-1)^n n!/(q-1)^{2n}$$

$$\times (m+\tfrac{1}{2})^{\nu+2n}. \quad (2.13)$$

Proof: Starting from Eq. (2.5), we have

$$[n, n+m]_{\{C_r\}} = H^{(m)}_{n+1}\{C_r\}/H^{(m)}_{n}\{\Delta^2 C_r\}. \quad (2.14)$$

Let us evaluate $\Delta^2 C_r$ asymptotically for large r:

$$\Delta C_r = C_{r+1} - C_r = q^R R^{-\nu}[q(1+1/R)^{-\nu} - 1]$$

$$\underset{r\to\infty}{\sim} q^R R^{-\nu}(q-1) \underset{r\to\infty}{\sim} (q-1)C_r;$$

consequently,

$$\Delta^2 C_r = \Delta C_{r+1} - \Delta C_r \underset{r\to\infty}{\sim} (q-1)^2 C_r. \quad (2.15)$$

By replacing (2.15) in (2.14), we then have

$$\begin{aligned}[n, n+m]_{\{C_r\}} &\underset{m\to\infty}{\sim} H^{(m)}_{n+1}\{C_r\}/H^{(m)}_{n}\{(q-1)^2 C_r\} \\ &= H^{(m)}_{n+1}\{C_r\}/(q-1)^{2n}H^{(m)}_{n}\{C_r\} \\ &\underset{m\to\infty}{\sim} q^{(n+1)(m+\frac{1}{2}+n)}H^{(m)}_{n+1}\{D_r\}/ \\ &\qquad (q-1)^{2n}q^{(n)(m+\frac{1}{2}+n-1)}H^{(m)}_{n}\{D_r\},\end{aligned} \quad (2.16)$$

where $\{D_r\}$ is the sequence of Lemma 2.1. Using now Eq. (2.8) in (2.16), Eq. (2.13) follows.

Lemma 2.2: Given the sequence $\{g_r\}$ defined by

$$g_r = g_r^{0,\epsilon} = (r+\tfrac{1}{2})^{\epsilon}\sin\eta_r, \quad (2.17)$$

where ϵ is an *arbitrary* complex number, $\eta_r = r\theta + b$, with b and θ real constants, the sequences $\{g_r^{j+1,\epsilon}\}$ defined for fixed $j\geqslant 0$ by the recurrent formulas

$$g_r^{j+1,\epsilon} = g_{r+1}^{j,\epsilon} + g_{r-1}^{j,\epsilon} - 2\cos\theta\, g_r^{j,\epsilon}, \quad (2.18)$$

have, for large r, the following asymptotic representations:

$$g_r^{j,\epsilon} \underset{r\to\infty}{\sim} (-2\sin\theta)^j[\epsilon]_j(r+\tfrac{1}{2})^{\epsilon-j}\sin(\eta_r - j\pi/2), \quad (2.19)$$

with $[\epsilon]_n$ defined as in Lemma 2.1.

Proof: It follows from an immediate extension of that of Lemma 2.1 of Ref. 3, by noting that, in the latter, only the linear dependence of η_r on r is essential, and that its results are actually valid for *any* ϵ.

Lemma 2.3: Given the sequence $\{g_r\}$ of Lemma 2.2, we have for fixed $n\geqslant 0$ and large m:

$$\begin{aligned}H^{(m)}_{n+1}\{g_r\} &\underset{m\to\infty}{\sim} (-1)^{(N+1)(n-2N)}(\sin\theta)^{2(n-N)} \\ &\times(\sin\eta_{n+m})^{2N-n+1}H^{(0)}_{N+1}\{h_r\}H^{(0)}_{n-N}\{h_r\},\end{aligned} \quad (2.20)$$

where $N = n/2$ for even n, $N = (n-1)/2$ for odd n, and the sequence $\{h_r\}$ is defined by

$$h_r = (-2\sin\theta)^r[\epsilon]_r(m+\tfrac{1}{2})^{\epsilon-r}. \quad (2.21)$$

Proof: It follows as an extension of that of Eq. (4.4) of Ref. 2, by noting that in its algebraic derivation it is only essential that a family of sequences be associated to $\{g_r\}$ as in Eq. (2.18) and having the asymptotic behaviors given by (2.19).

Lemma 2.4: Given the sequence $\{T_r\}$ whose members have the asymptotic representation

$$T_r \underset{r\to\infty}{\sim} R^{\epsilon}\exp[-(r+1)\alpha]\sin\Lambda_r^{(+)}, \quad (2.22)$$

where $R = r+\frac{1}{2}$, $\alpha>0$, $\Lambda_r^{(+)} = (r+1)\theta + \gamma + \pi/4$, $0<\theta<\pi$, and ϵ is an arbitrary complex number, with γ such that $\tan\gamma = \tan(\theta/2)$. $\coth(\alpha/2)$ $(0<\gamma<\pi/2)$, then

$$\Delta^2 T_r \underset{r\to\infty}{\sim} Q(R+1)^{\epsilon}\exp[-(r+2)\alpha]\sin\Lambda_{r+1}^{(-)}, \quad (2.23)$$

where $Q = 2(\cosh\alpha - \cos\theta)$ and $\Lambda_{r+1}^{(-)} = \Lambda_{r+1}^{(+)} - 2\gamma$.

Proof: We have

$$\begin{aligned}\Delta^2 T_r &= T_{r+2} + T_r - 2T_{r+1} \\ &\underset{r\to\infty}{\sim} (R+2)^{\epsilon}\exp[-(r+3)\alpha]\sin\Lambda_{r+2}^{(+)} \\ &\quad + R^{\epsilon}\exp[-(r+1)\alpha]\sin\Lambda_r^{(+)} \\ &\quad - 2(R+1)^{\epsilon}\exp[-(r+2)\alpha]\sin\Lambda_{r+1}^{(+)} \\ &\underset{r\to\infty}{\sim} 2(R+1)^{\epsilon}\exp[-(r+2)\alpha] \\ &\quad \times\{(\cos\theta\cosh\alpha - 1)\sin\Lambda_{r+1}^{(+)} \\ &\quad - \sin\theta\sinh\alpha\cos\Lambda_{r+1}^{(+)}\} \\ &= 2(R+1)^{\epsilon}\exp[-(r+2)\alpha] \\ &\quad \times(\cosh\alpha - \cos\theta)\sin\Lambda_{r+1}^{(-)},\end{aligned}$$

where we have used the identities

$$\cos 2\gamma = \frac{\cos\theta\cosh\alpha - 1}{\cosh\alpha - \cos\theta}, \quad \sin 2\gamma = \frac{\sin\theta\sinh\alpha}{\cosh\alpha - \cos\theta},$$

valid for γ defined as above.

Theorem 2.2: Given the sequence $\{T_r\}$ of Lemma 2.4, the PPA $[n, n+m]_{\{T_r\}}$ has, for fixed $n\geqslant 0$ and large m, the following asymptotic behavior:

$$\begin{aligned}[n, n+m]_{\{T_r\}} &\underset{m\to\infty}{\sim} (-Q)^{-n}\exp[-(m+n+1)\alpha]2^{2N} \\ &\times(\sin\theta)^{2(n-N)}[\epsilon]_N N! \\ &\times\frac{(\sin\Lambda_{m+n}^{(+)})^{2N-n+1}}{(\sin\Lambda_{m+n}^{(-)})^{n-2N}}(m+\tfrac{1}{2})^{\epsilon-2N},\end{aligned} \quad (2.24)$$

where N is defined as in Lemma 2.3, and $\Lambda_{m+n}^{(+)}$, $\Lambda_{m+n}^{(-)}$, and Q are those of Lemma 2.4.

Proof: Starting from Eq. (2.5) and using Lemma 2.4, we have, with $R = r+\frac{1}{2}$,

$$\begin{aligned}[n, n+m]_{\{T_r\}} &= \frac{H^{(m)}_{n+1}\{T_r\}}{H^{(m)}_{n}\{\Delta^2 T_r\}} \\ &\underset{m\to\infty}{\sim} \frac{H^{(m)}_{n+1}\{R^{\epsilon}\exp[-(r+1)\alpha]\sin\Lambda_r^{(+)}\}}{H^{(m)}_{n}\{\Delta^2 R^{\epsilon}\exp[-(r+1)\alpha]\sin\Lambda_r^{(+)}\}} \\ &\underset{m\to\infty}{\sim} H^{(m)}_{n+1}\{R^{\epsilon}\exp[-(r+1)\alpha]\sin\Lambda_r^{(+)}\}/ \\ &\qquad H^{(m)}_{n}\{Q(R+1)^{\epsilon}\exp[-(r+2)\alpha]\sin\Lambda_{r+1}^{(-)}\} \\ &= \frac{\exp[-(n+1)(m+n+1)\alpha]H^{(m)}_{n+1}\{R^{\epsilon}\sin\Lambda_r^{(+)}\}}{Q^n\exp[-n(m+n+1)\alpha]H^{(m+1)}_{n}\{R^{\epsilon}\sin\Lambda_r^{(-)}\}},\end{aligned} \quad (2.25)$$

from which Eq. (2.24) follows, by using Lemma 2.3 to evaluate asymptotically the Hankel determinants in (2.25), noting that for $k\geqslant 1$, fixed m, and $\{h_r\}$ defined by (2.21),

$$H^{(0)}_{k}\{h_r\} = H^{(0)}_{k}\{[-2\sin\theta/(m+\tfrac{1}{2})]^r(m+\tfrac{1}{2})^{\epsilon}[\epsilon]_r\}$$

$$= [2\sin\theta/(m+\tfrac{1}{2})]^{k(k-1)}(m+\tfrac{1}{2})^{k\epsilon}H_k^{(0)}\{[\epsilon]_r\}, \tag{2.26}$$

and taking account of Eq. (2.11) in (2.26).

Let us remark, to be more precise, that Eq. (2.24) does not actually hold, when $\sin\Lambda_{m+n}^{(+)} = 0$ and n is even, or when $\sin\Lambda_{m+n}^{(-)} = 0$ and n is odd. For these isolated singular cases the predictions are $[n, n+m] \sim 0$, and $[n,n+m] \sim \infty$, respectively, these results being related to the fact that only the leading asymptotic term was considered for $g_r^{j,\epsilon}$ in Eq. (2.19) and further on in the calculations which lead to Eq. (2.20).

3. THE PARTIAL WAVE EXPANSION OF THE SCATTERING AMPLITUDE (PWESA) FOR SHORT RANGE POTENTIALS

A. Asymptotic representation of the sequence of partial wave sums within its region of convergence in the complex cosθ plane

Let us consider short range potentials with the behavior

$$V(r) \underset{r\to\infty}{\sim} V_0 r^{-\rho-1}\exp(-\mu r), \tag{3.1}$$

where V_0 is a constant, ρ is an integer, and $\mu > 0$. For these cases, it has been shown[7] that the corresponding phase shifts have, for large values of angular momentum, the following asymptotic representation:

$$\delta_l \underset{l\to\infty}{\sim} \frac{A\exp(-L\alpha)}{L^{\rho+1/2}}[1+o(1/L)], \tag{3.2}$$

where α is a positive number defined by

$$\cosh\alpha = 1 + \mu^2/2k^2,$$

$$L = l + \tfrac{1}{2},$$

and

$$A = (-V_0/2k)(\pi/2)^{\frac{1}{2}}(\sinh\alpha)^{\rho-1/2}(k^2/\mu)^\rho.$$

Owing to this fact, the PWESA given by Eq. (1.1) converges within an elliptical region Z in the complex $\cos\theta$ plane (the Lehmann ellipse), characterized by having focii at ± 1 and major axis equal to $2\cosh\alpha$. Let us note that *within* Z the sequence of partial wave sums of the PWESA defined by Eq. (2.4) can be expressed in the form

$$S_m(\cos\theta) = f(\cos\theta) - \sum_{l=m+1}^{\infty} s_l, \tag{3.3}$$

with $s_l = a_l P_l(\cos\theta)$, and investigate its large m behavior, which in turn, is responsible for the convergence of expansion (1.1). In order to do this, it will prove useful to study the asymptotic representation of the partial wave amplitudes a_l. According to (1.1), and noting that $\delta_l \to 0$ when $l\to\infty$, we have

$$\begin{aligned} a_l &= (2l+1)[\exp(2i\delta_l)-1]/(2ik) \\ &= 2L(2i\delta_l - 4\delta_l^2 + \cdots)/(2ik) \\ &\underset{l\to\infty}{\sim} BL^{-\rho+\frac{1}{2}}\exp(-\alpha L)[1+o(1/L)], \end{aligned} \tag{3.4}$$

where $B = 2A/k$, and Eq. (3.2) has been used. Starting now from (3.4), let us obtain the large-l behavior of s_l within Z.

(i) Consider $Z_1 = \{Z - [-1, 1]\}$: Here, $\cos\theta = \cosh\xi$ with

$$\mathrm{Re}\alpha > \mathrm{Re}\xi > 0, \tag{3.5}$$

and[8]

$$P_l(\cosh\xi) \underset{l\to\infty}{\sim} \frac{\exp(L\xi)}{(2\pi\sinh\xi)^{\frac{1}{2}}L^{\frac{1}{2}}}[1+o(1/L)], \tag{3.6}$$

and, consequently,

$$s_l \underset{l\to\infty}{\sim} C_1 L^{-\rho}\exp(-\beta L)[1+o(1/L)], \tag{3.7}$$

with $C_1 = B(2\pi\sinh\xi)^{-1/2}$ and $\beta = \alpha - \xi$. In particular, we note that, because of (3.5), $\mathrm{Re}\beta > 0$ within this region.

(ii) Consider $Z_2 = \{\pm 1\}$. Here we use

$$P_l(\pm 1) = (\pm 1)^l = \exp(i\pi l)$$

to obtain

$$s_l \underset{l\to\infty}{\sim} BL^{-\rho+1/2}\exp(-\alpha L)[1+o(1/L)] \quad (\cos\theta = 1), \tag{3.8a}$$

$$s_l \underset{l\to\infty}{\sim} -iBL^{-\rho+1/2}\exp[-(\alpha - i\pi)L][1+o(1/L)] \quad (\cos\theta = -1). \tag{3.8b}$$

(iii) Finally, consider $Z_3 = \{(-1, 1)\}$. Here[9]

$$\begin{aligned} P_l(\cos\theta) &\underset{l\to\infty}{\sim} [(2\pi\sin\theta)^{1/2}L^{1/2}]^{-1} \\ &\times[\exp(i\Omega_L) + \exp(-i\Omega_L)][1+o(1/L)], \end{aligned} \tag{3.9}$$

with $\Omega_L = L\theta - \pi/4$, and, then,

$$\begin{aligned} s_l \underset{l\to\infty}{\sim} C_2 &\left\{ \frac{\exp[-L\alpha^{(-)} - i\pi/4]}{L^\rho} + \frac{\exp[-L\alpha^{(+)} + i\pi/4]}{L^\rho} \right\} \\ &\times[1+o(1/L)], \end{aligned} \tag{3.10}$$

where $\alpha^{(\pm)} = \alpha \pm i\theta$ and $C_2 = B(2\pi\sin\theta)^{-1/2}$.

The explicit dependence of $S_m(\cos\theta)$ on m can be formally obtained from Eq. (3.3) by using the Euler–McLaurin summation formula,[10] in the form adequate for our case:

$$S_m(\cos\theta) = f(\cos\theta) - \left\{ \int_m^\infty s_l\,dl - \frac{s_m}{2} - \sum_{n=1}^{\infty} \frac{B_{2n}}{(2n)!}\frac{d^{(2n-1)}s_l}{dl^{2n-1}}\bigg|_{l=m} \right\} \tag{3.11}$$

where the $\{B_{2n}\}$ are the Bernoulli numbers. In what follows, we shall obtain an asymptotic representation for $S_m(\cos\theta)$, by replacing the estimates for s_l given by Eqs. (3.7)–(3.10) into (3.11). In order to do so, it suffices to note, that for complex u such that $\mathrm{Re}u > 0$, we have, with $M = m + \frac{1}{2}$,

$$\int_m^\infty \frac{\exp(-uL)}{L^\rho}dl \underset{m\to\infty}{\sim} \frac{1}{u}\frac{\exp(-uM)}{M^\rho}\left\{1 + o\left(\frac{1}{M}\right)\right\}$$

and

$$\sum_{n=1}^{\infty}\frac{B_{2n}}{(2n)!}\frac{d^{2n-1}}{dl^{2n-1}}\frac{\exp(-uL)}{L^\rho}\bigg|_{l=m} \underset{m\to\infty}{\sim} -\frac{\exp(-uM)}{M^\rho}\frac{1}{u}\sum_{n=1}^{\infty}\frac{B_{2n}}{(2n)!}u^{2n}\left[1 + o\left(\frac{1}{M}\right)\right]$$

$$= -\frac{\exp(-uM)}{M^{\rho}}\left\{\frac{1}{2}\coth\frac{u}{2}-\frac{1}{u}\right\}\left[1+o\left(\frac{1}{M}\right)\right].$$

Making use of these results, we obtain, after some algebra:

$$S_m(\cosh\xi) \underset{m\to\infty}{\sim} f(\cosh\xi)+E_1\frac{\exp(-\beta M)}{M^{\rho}}\left[1+o\left(\frac{1}{M}\right)\right]$$

$$(\mathrm{Re}\beta>0,\ \mathrm{Re}\xi>0), \tag{3.12}$$

with $E_1=-2A\{[\exp(\beta)-1]^2(2\pi\sinh\xi)k^2\}^{-1/2}$;

$$S_m(1) \underset{m\to\infty}{\sim} f(1)+E_2\frac{\exp(-\alpha M)}{M^{\rho-1/2}}\left[1+o\left(\frac{1}{M}\right)\right], \tag{3.13a}$$

with $E_2=-2A\{[\exp(\alpha)-1]k\}^{-1}$;

$$S_m(-1) \underset{m\to\infty}{\sim} f(-1)+E_3(-1)^m\frac{\exp(-\alpha M)}{M^{\rho-1/2}}\times\left[1+o\left(\frac{1}{M}\right)\right], \tag{3.13b}$$

with $E_3=2A\{[\exp(\alpha)+1]k\}^{-1}$;

$$S_m(\cos\theta) \underset{m\to\infty}{\sim} f(\cos\theta)+E_4\frac{\exp[-\alpha(m+1)]}{M^{\rho}}\times\sin\Lambda_m^{(+)}\left[1+o\left(\frac{1}{M}\right)\right], \tag{3.14}$$

with

$E_4=-2A\{(2\pi k^2\sin\theta)[\sinh^2(\alpha/2)+\sin^2(\theta/2)]\}^{-1/2}$;

$\Lambda_m^{(+)}=(m+1)\theta+\gamma+\pi/4$,

$\tan\gamma=\tan(\theta/2)\coth(\alpha/2)\quad(0<\gamma<\pi/2)$.

B. Asymptotic behavior of the PPA to the PWESA

By comparison of Eqs. (3.12)–(3.13) with (2.12), and of Eq. (3.14) with (2.22), it is readily seen that, with a proper choice of the parameters which define the large-r behaviors of sequences $\{C_r\}$ and $\{T_d\}$ in Sec. 2, we have, *in a first order asymptotic approximation*:

$$S_m(\cosh\xi) \underset{m\to\infty}{\sim} f(\cosh\xi)+E_1C_m\quad(\mathrm{Re}\alpha>\mathrm{Re}\xi>0), \tag{3.15}$$

with $\nu=\rho$ and $q=\exp(-\beta)=\exp(\xi-\alpha)$ in (2.12):

$$S_m(1) \underset{m\to\infty}{\sim} f(1)+E_2C_m, \tag{3.16a}$$

with $\nu=\rho-\frac{1}{2}$ and $q=\exp(-\alpha)$ in (2.12);

$$S_m(-1) \underset{m\to\infty}{\sim} f(-1)+iE_3C_m, \tag{3.16b}$$

with $\nu=\rho-\frac{1}{2}$ and $q=\exp(-\alpha+i\pi)$ in (2.12);

$$S_m(\cos\theta) \underset{m\to\infty}{\sim} f(\cos\theta)+E_4T_m\quad(-1<\cos\theta<1), \tag{3.17}$$

with $\epsilon=-\rho$ in (2.22).

Taking account of property (2.6) of the PPA and of Theorems 2.1 and 2.2, we may now state the following:

Theorem 3.1: The PPA $[n,n+m]_{f(\cos\theta)}$ corresponding to a central potential having a short range tail such as that given by Eq. (3.1), has, for fixed $n\geqslant 0$, $m\to\infty$, and $\cos\theta$ within the Lehmann ellipse, the following asymptotic behavior:

$$[n,n+m]_{f(\cosh\xi)} \underset{m\to\infty}{\sim} f(\cosh\xi)-\frac{B}{(2\pi\sinh\xi)^{1/2}}\frac{(-1)^n[-\rho]_n n!}{[\exp(\beta)-1]^{2n+1}}\times\frac{\exp\left[-\beta\left(m+\frac{1}{2}\right)\right]}{(m+\frac{1}{2})^{\rho+2n}},$$

$$(\cos\theta=\cosh\xi,\text{ with }\mathrm{Re}\xi>0\text{ and }\mathrm{Re}\beta=\mathrm{Re}(\alpha-\xi)>0); \tag{3.18}$$

$$[n,n+m]_{f(1)} \underset{m\to\infty}{\sim} f(1)-B\frac{(-1)^n[-\rho+\frac{1}{2}]_n n!}{[\exp(\alpha)-1]^{2n+1}}\times\frac{\exp[-\alpha(m+\frac{1}{2})]}{(m+\frac{1}{2})^{\rho+1/2+2n}}; \tag{3.19a}$$

$$[n,n+m]_{f(-1)} \underset{m\to\infty}{\sim} f(-1)+B\frac{(-1)^n[-\rho+\frac{1}{2}]_n n!}{[\exp(\alpha)+1]^{2n+1}}\times\frac{(-1)^m\exp[-\alpha(m+\frac{1}{2})]}{(m+\frac{1}{2})^{\rho+1/2+2n}}; \tag{3.19b}$$

$$[n,n+m]_{f(\cos\theta)} \underset{m\to\infty}{\sim} f(\cos\theta)-\frac{B}{\{(2\pi\sin\theta)[\sinh^2(\alpha/2)+\sin^2(\theta/2)]\}^{1/2}}\times\frac{(-2)^{2N-n}(\sin\theta)^{2(n-N)}[-\rho]_N N!}{[\exp(\alpha)(\cosh\alpha-\cos\theta)]^n}\times\frac{\exp[-\alpha(m+1)](\sin\Lambda_{m+n}^{(+)})^{2N-n+1}}{(m+\frac{1}{2})^{\rho+2N}(\sin\Lambda_{m+n}^{(-)})^{n-2N}}$$

$$(-1<\cos\theta<1), \tag{3.20}$$

where

$B=2A/k=-(V_0/k^2)(\pi/2)^{1/2}(\sinh\alpha)^{\rho-1/2}(k^2/\mu)^{\rho}$,

$\Lambda_r^{(\pm)}=(r+1)\theta\pm\gamma+\pi/4$,

$\tan\gamma=\tan(\theta/2)\coth(\alpha/2)\quad(0<\gamma<\pi/2)$,

$N=n/2$ for even n, and $N=(n-1)/2$ for odd n.

As in the discussion which followed the proof of Theorem 2.2, for a given θ we must exclude from Eq. (3.20) the isolated cases for which $\sin\Lambda_{m+n}^{(+)}=0$ for even n or $\sin\Lambda_{m+n}^{(-)}=0$ for odd n. Furthermore, by inspection of Eqs. (3.18) and (3.20), it is seen that, for potentials for which $\rho\leqslant 0$, we have $[-\rho]_j=0$ if $j\geqslant|\rho|+1$, and the first predicts

$$[n,n+m]_{f(\cosh\xi)} \underset{m\to\infty}{\sim} f(\cosh\xi)+0$$

for $n\geqslant|\rho|+1$, while the second gives

$$[n,n+m]_{f(\cos\theta)} \underset{m\to\infty}{\sim} f(\cos\theta)+0$$

for $N\geqslant|\rho|+1$, i.e., for $n\geqslant 2|\rho|+2$ (even n) or $n\geqslant 2|\rho|+3$ (odd n). In order to obtain, for finite large m, a nonvanishing asymptotic estimate for the error of the PPA, higher order terms in (3.12) and (3.14) should be considered in the calculations, and, hence, a more detailed knowledge of those expansions is required. Since important potentials from the physical point of view, such as the Yukawa ($\rho=0$) and exponential ($\rho = -1$), fall into this category, it is worthwhile to investigate these singular cases with some more detail. We shall do so only for $\cos\theta$ within the Lehmann ellipse and out of the real segment $[-1, 1]$, i.e., for $\cos\theta \in Z_1$, where the algebra is somewhat less involved.

By using the Born approximation for the phase shifts, we show in Appendix A that for $\rho \leqslant 0$ and $\cos\theta \in Z_1$, Eq. (3.12) can be written in a more detailed fashion as

$$S_m(\cosh\xi) \underset{m\to\infty}{\sim} f(\cosh\xi) - \frac{B}{(2\pi\sinh\xi)^{1/2}[\exp(\beta) - 1]} \times \frac{\exp[-\beta(m+\frac{1}{2})]}{(m+\frac{1}{2})^{\rho}} \sum_{j=0}^{\infty} \frac{x_j}{(m+\frac{1}{2})^j}, \tag{3.21}$$

where the x_j are independent of m and, in particular, $x_0 = 1$. Then, using again property (2.6) and Eq. (B4), obtained in Appendix B as a proper extension of (2.13) of Theorem 2.1, we expect for the singular cases, i.e., for $\rho \leqslant 0$, $n \geqslant |\rho| + 1$, and $\cosh\xi = \cos\theta \in Z_1$,

$$[n, n+m]_{f(\cosh\xi)} \underset{m\to\infty}{\sim} f(\cosh\xi) - x_{|\rho|+1} \frac{B}{(2\pi\sinh\xi)^{1/2}} \times \frac{(n+\rho-1)!(n-\rho+1)!}{[\exp(\beta)-1]^{2n+1}} \frac{\exp[-\beta(m+\frac{1}{2})]}{(m+\frac{1}{2})^{2n+1}}; \tag{3.22}$$

and it is seen that the first $|\rho| + 1$ terms of expansion (3.21) are "filtered" by the PPA tranformation.

For $-1 < \cos\theta < 1$ we can expect, at least qualitatively, analogous asymptotic predictions when $\rho \leqslant 0$ and $N \geqslant |\rho| + 1$. The algebra involved, however, is much more complicated, particularly so with regard to the required extension of Theorem 2.2.

4. CONCLUSIONS

The results of Theorem 3.1 of the preceeding section prove the convergence of the PPA $[n, n+m]_f$, for fixed n and $m \to \infty$, to the scattering amplitude $f(\cos\theta)$ corresponding to short range potentials with behaviors given by Eq. (3.1) and for $\cos\theta$ within the Lehmann ellipse, where the PWESA converges. Moreover, by inspection of Eqs. (3.18) and (3.19), it is seen that for $\cos\theta \in \{Z_1 \cup Z_2\}$, the asymptotic rate of convergence of the sequence $[n, n+m]_f$ with fixed $n > 0$, i.e., the nth row of the PPA table, is greater than that of the first now ($n = 0$) consisting of the sequence of partial wave sums $S_m(\cos\theta)$ of the PWESA, by a factor of order *at least* $(1/m)^{2n}$. In the singular cases, i.e., for $\rho \leqslant 0$, $n \geqslant |\rho| + 1$, and $\cos\theta \in Z_1$, Eq. (3.22) predicts a corresponding factor of order $(1/m)^{2n+|\rho|+1}$, while the first $|\rho| + 1$ terms of the expansion (3.21) of $S_m(\cos\theta)$ are "filtered" by the transformation procedure. Furthermore, for $\cos\theta \in Z_3$ and according to Eq. (3.20), the factors involved are of the order of *at least* $(1/m)^n$ and $(1/m)^{n-1}$, for even and odd n, respectively. They are expected to have larger exponents in the cases in which $\rho \leqslant 0$ and $n \geqslant 2|\rho| + 2$ (even n) or $n \geqslant 2|\rho| + 3$ (odd n).

Thus, the traditional method of summation of the PWESA by considering the sequence of its partial wave sums, i.e., the first row of the PPA table, turns out to be one of the poorest with regards to its rate of convergence, as was also verified in the case of long range potentials in previous papers. This fact shows that the importance of the PPA approach is independent of the range of the potential involved.

Our proofs have been restricted to $\cos\theta$ within the elliptical domain where the PWESA converges, and, hence, where Eq. (3.3) is valid. Thus, we are not able, at present, to obtain the behavior of the PPA out of that domain. However, owing to the rational nature of the approximations and their good behavior within the Lehmann ellipse, we can expect them to be an important method for the approximate analytical continuation of the scattering amplitude in the complex $\cos\theta$, starting from its partial wave expansion.

APPENDIX A: Asymptotic expansion for S_m ($\cosh\xi$) ($\mathrm{Re}\,\alpha > \mathrm{Re}\,\xi > 0$)

The plan of this appendix is to first obtain an asymptotic expansion for the partial wave amplitudes and then, by using the corresponding ones for the Legendre polynomials $P_l(\cosh\xi)$, and repeating the procedures of Sec. 3A, obtain the form of an asymptotic expansion for $S_m(\cosh\xi)$ valid for large m.

For a short range potential, and, in particular, for the one considered in this work, with behavior

$$V(r) \underset{r\to\infty}{\sim} V_0 r^{-\rho-1}\exp(-\mu r), \tag{A1}$$

with constant V_0, integer ρ, and $\mu > 0$, it is well known that the phase shifts δ_l tend asymptotically to their Born approximation[11]:

$$\delta_l \underset{l\to\infty}{\sim} -(\pi/2)\int_0^{\infty} r[J_L(kr)]^2 V(r)dr, \tag{A2}$$

where $L = l + \frac{1}{2}$ and $J_L(kr)$ is the Bessel function of order L and argument kr. Furthermore, noting that for $l \to \infty$ only the large r tail of $V(r)$ contributes significantly to the integral in (A2), we can write

$$\delta_l \underset{l\to\infty}{\sim} (-\pi V_0/2)\int_0^{\infty} r^{-\rho}\exp(-\mu r)[J_L(kr)]^2 dr. \tag{A3}$$

Let ρ be such that $\rho \leqslant 0$, and define $\sigma = |\rho|$. Then,

$$\delta_l \underset{l\to\infty}{\sim} \frac{-\pi V_0}{2}(-1)^{\sigma}\frac{d^{\sigma}}{d\mu^{\sigma}}\int_0^{\infty}\exp(-\mu r)[J_L(kr)]^2 dr = \frac{-V_0}{2k}(-1)^{\sigma}\frac{d^{\sigma}}{d\mu^{\sigma}}Q_l\left(1 + \frac{\mu^2}{2k^2}\right), \tag{A4}$$

where Q_l is the Legendre function of the second kind of order l. Let us recall a property for the derivatives of this function[12]:

$$\frac{d^{\sigma}Q_l(z)}{dz^{\sigma}} = (z^2-1)^{-\sigma/2}Q_l^{\sigma}(z) \tag{A5}$$

and the fact that for $z > 1$, fixed σ, and $l \to \infty$, we have,[13] with $z = \cosh\alpha$,

$$Q_l^{\sigma}(\cosh\alpha) = (-1)^{\sigma}\left(\frac{\pi}{2\sinh\alpha}\right)^{1/2}\frac{\Gamma(l+\sigma+1)}{\Gamma(l+\frac{3}{2})}\exp[-\alpha(l+\tfrac{1}{2})] \times F\left(\sigma+\tfrac{1}{2}, \tfrac{1}{2}-\alpha; l+\tfrac{3}{2}; \frac{1}{1-\exp(2\alpha)}\right) \tag{A6}$$

$$\underset{l\to\infty}{\sim} (-1)^{\sigma}\left(\frac{\pi}{2\sinh\alpha}\right)^{1/2}(l+\tfrac{1}{2})^{\sigma-1/2}\exp[-\alpha(l+\tfrac{1}{2})]$$

$$\times \sum_{n=0}^{\infty} \frac{\gamma_n}{(l+\frac{1}{2})^n},$$

where the $\{\gamma_n\}$ are *independent* of l, and in particular, $\gamma_0 = 1$. Then, using (A5) and (A6) in (A4) and taking account of Eq. (3.4), it follows that the partial wave amplitudes a_l can be asymptotically represented in the form:

$$a_l \underset{l\to\infty}{\sim} \frac{2}{k}(l+\tfrac{1}{2})\delta_l$$

$$\underset{l\to\infty}{\sim} B(l+\tfrac{1}{2})^{\sigma+1/2}\exp[-\alpha(l+\tfrac{1}{2})] \sum_{n=0}^{\infty} \frac{\beta_n}{(l+\frac{1}{2})^n}, \quad \text{(A7)}$$

where $\beta_0 = 1$, the $\{\beta_n\}$ are *independent* of l, and B is defined as in Eq. (3.4).

Furthermore, for $\mathrm{Re}\xi > 0$, we have[8]

$$P_l(\cosh\xi) \underset{l\to\infty}{\sim} \frac{\Gamma(l+\frac{1}{2})\exp[\xi(l+\frac{1}{2})]}{l!(2\pi\sinh\xi)^{1/2}} \times F\left(\tfrac{1}{2},\tfrac{1}{2};\,-l+\tfrac{1}{2};\,-\frac{\exp(-\xi)}{2\sinh\xi}\right) \quad \text{(A8)}$$

$$\underset{l\to\infty}{\sim} \frac{\exp[\xi(l+\frac{1}{2})]}{(2\pi\sinh\xi)^{1/2}(l+\frac{1}{2})^{1/2}} \sum_{n=0}^{\infty} \frac{\varphi_n}{(l+\frac{1}{2})^n},$$

where $\varphi_0 = 1$ and the $\{\varphi_n\}$ are *independent* of l. By using (A7) and (A8) we can now obtain an asymptotic expansion for $s_l = a_l P_l(\cosh\xi)$, and, making use of the Euler–McLaurin summation procedure as in Sec. 3A, we are able to obtain *the form* of such an expansion for $S_m(\cosh\xi)$, when $\mathrm{Re}\alpha > \mathrm{Re}\xi > 0$:

$$S_m(\cosh\xi) = f(\cosh\xi) - \sum_{l=m+1}^{\infty} s_l$$

$$\underset{m\to\infty}{\sim} f(\cosh\xi) + E_1 \frac{\exp[-\beta(m+\frac{1}{2})]}{(m+\frac{1}{2})^{\rho}} \times \sum_{n=0}^{\infty} \frac{x_n}{(m+\frac{1}{2})^n}, \quad \text{(A9)}$$

where $x_0 = 1$, the $\{x_n\}$ are *independent* of m, $\beta = \alpha - \xi$, and E_1 is defined as in Eq. (3.12).

APPENDIX B: Extension of Theorem 2.1

Let the sequence $\{D_r\}$ have the asymptotic representation

$$D_r \underset{r\to\infty}{\sim} R^{-\nu} \sum_{t=0}^{\infty} x_t R^{-t} \quad \text{(B1)}$$

where $x_0 = 1$, the $\{x_t\}$ are independent of r, $R = r + \frac{1}{2}$, and ν is an integer.

If only the first order term in the above expansion is considered, we have, according to Lemma 2.1,

$$H_k^{(p)}\{D_r\} \underset{p\to\infty}{\sim} (p+\tfrac{1}{2})^{-k(\nu+k-1)} \prod_{t=0}^{k-1} [-\nu]_t (-1)^t t!. \quad \text{(B2)}$$

Moreover, for $\nu \leqslant 0$ and $t \geqslant |\nu| + 1$, we have $[-\nu]_t = 0$, and hence, for $k \geqslant |\nu| + 2$, Eq. (B2) predicts

$$H_k^{(p)}\{D_r\} \underset{p\to\infty}{\sim} 0,$$

and, in order to have a nonidentically zero asymptotic estimate for $H_k^{(p)}\{D_r\}$, higher order terms must be considered in (B1).

Let ν be such that $\nu \leqslant 0$ and $k \geqslant \sigma + 2$, with $\sigma = |\nu|$. Then,

$$H_k^{(p)}\{D_r\} \underset{p\to\infty}{\sim} H_k^{(p)}\left\{\sum_{t=0}^{\infty} x_t (r+\tfrac{1}{2})^{-\nu-t}\right\} = H_k^{(0)}\left[\sum_{t=0}^{\infty} x_t \Delta^r (p+\tfrac{1}{2})^{-\nu-t}\right],$$

and, after some algebra, we can show that, in the lowest asymptotic approximation which allows for a nonzero estimate, we have

$$H_k^{(0)}\left\{\sum_{t=0}^{\infty} x_t \Delta^r (p+\tfrac{1}{2})^{-\nu-t}\right\}$$

$$\underset{p\to\infty}{\sim} (-1)^{\sigma(\sigma+1)/2}(\sigma!)^{\sigma+1} H_{k-\sigma-1}^{(2\sigma+2)}\{x_{\sigma+1}\Delta^r(p+\tfrac{1}{2})^{-1}\}$$

$$\underset{p\to\infty}{\sim} (-1)^{\sigma(\sigma+1)/2}(\sigma!)^{\sigma+1} \times H_{k-\sigma-1}^{(2\sigma+2)}\{x_{\sigma+1}[-1]_r (p+\tfrac{1}{2})^{-1-r}\}$$

$$= (-1)^{\sigma(\sigma+1)/2}(\sigma!)^{\sigma+1}(x_{\sigma+1})^{k-\sigma-1}(p+\tfrac{1}{2})^{-[k^2-(\sigma+1)^2]} \times H_{k-\sigma-1}^{(2\sigma+2)}\{[-1]_r\}.$$

By using Eq. (2.11) and rearranging the final expression, we obtain

$$H_k^{(p)}\{D_r\} \underset{p\to\infty}{\sim} (x_{\sigma+1})^{k-\sigma-1}(p+\tfrac{1}{2})^{-[k^2-(\sigma+1)^2]} \times (-1)^{\sigma(\sigma+1)/2}(\sigma!)^{\sigma+1} \prod_{t=0}^{k-\sigma-2} (2\sigma+t+2)!t!. \quad \text{(B3)}$$

Let $\{C_r\}$ be such that $C_r = q^R D_r$, with $|q| < 1$, and let us investigate the asymptotic behavior of $[n, n+m]_{\{C_r\}}$ for fixed n and $m \to \infty$.

As in the proof of Theorem 2.1, we can show that

$$\Delta C_r = C_{r+1} - C_r \underset{r\to\infty}{\sim} q^R \sum_{t=0}^{\infty} R^{-\nu-t}\left[q\left(1+\frac{1}{R}\right)^{-\nu-t} - 1\right] x_t \underset{r\to\infty}{\sim} q^R \sum_{t=0}^{\infty} R^{-\nu-t}(q-1)x_t \underset{r\to\infty}{\sim} (q-1)C_r,$$

hence,

$$\Delta^2 C_r \underset{r\to\infty}{\sim} (q-1)^2 C_r;$$

consequently, as in that case,

$$[n, n+m]_{\{C_r\}} \underset{m\to\infty}{\sim} \frac{q^{(m+1/2+2n)}}{(q-1)^{2n}} \frac{H_{n+1}^{(m)}\{D_r\}}{H_n^{(m)}\{D_r\}} \underset{m\to\infty}{\sim} \frac{x_{|\nu|+1} q^{(m+1/2+2n)}(n+\nu-1)!(n-\nu+1)!}{(q-1)^{2n}(m+\frac{1}{2})^{2n+1}} \quad \text{(B4)}$$

where Eq. (B3) has been used in the last step. Equation (B4) extends Eq. (2.13) of Theorem 2.1 in order to have a *nonidentically zero* asymptotic estimate for the PPA $[n, n+m]_{\{C_r\}}$ in the singular case in which $\nu \leqslant 0$ and $n \geqslant |\nu| + 1$.

[1]J. Fleischer, Nucl. Phys. B **37**, 59 (1972); Erratum, Nucl. Phys. B **44**, 640 (1972); K. Alder, D. Trautmann, and R.D. Viollier, Z. Naturwiss. **28a**, 321 (1973); J. Fleischer, J. Math. Phys. **14**, 276 (1973); O.D. Corbella, C.R. Garibotti, and F.F. Grinstein, Z. Phys. A **277**, 1 (1976); A.K. Common and T. Stacey, J. Phys. A **11**, 259, 275 (1978).
[2]C.R. Garibotti and F.F. Grinstein, J. Math. Phys. **19**, 821 (1978).
[3]C.R. Garibotti and F.F. Grinstein, J. Math. Phys. **19**, 2405 (1978).
[4]C.R. Garibotti and F.F. Grinstein, J. Math. Phys. **20**, 141 (1979).
[5]G.A. Baker, Jr., *Essentials of Padé Approximants* (Academic, New York, 1975).
[6]D. Shanks, J. Math. & Phys. **34**, 1 (1955).
[7]B. Jaksić and N. Limić, J. Math. Phys. **7**, 88 (1966).
[8]E.T. Copson, *Theory of Functions of a Complex Variable* (Oxford, London, 1957), p. 284.
[9]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), p. 331.
[10]Reference 9, p. 804.
[11]V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965), p. 83.
[12]Reference 9, p. 334.
[13]Reference 9, p. 336.

Einstein spaces and homothetic motions. I

W. D. Halford

Department of Mathematics, Massey University, Palmerston North, New Zealand

R. P. Kerr

Department of Mathematics, University of Canterbury, Christchurch, New Zealand

(Received 24 July 1978)

Algebraically special, nonflat vacuum Einstein spaces with an expanding and/or twisting geodesic principal null congruence are considered. These spaces are assumed to possess locally a homothetic symmetry as well as two or more Killing vectors. All metrics of such spaces are determined along with the form of the homothetic Killing vector admitted. All but one of the metrics are twist free. It is proved that two of the NUT metrics do not admit a homothetic motion.

1. INTRODUCTION

Two Riemannian spaces $M, \tilde{M}$ endowed with metrics g, $\tilde{g}$ are *conformally related* if

$$\tilde{g} = e^{2\phi} g \tag{1.1}$$

for some function ϕ. In general, ϕ is not constant. The relationship is said to be *homothetic* if ϕ is constant, and *isometric* if $\phi = 0$.

If X is a vector field on M which generates locally a one-parameter group of infinitesimal *conformal motions* on M, then in local coordinates x^{α},

$$\mathscr{L}_X g_{\mu\nu} = \psi g_{\mu\nu} \tag{1.2}$$

for some positive function $\psi = \psi(x^{\alpha})$ on M, where $\mathscr{L}_X$ denotes the Lie derivative with respect to X, and $g_{\mu\nu}$ are the components of the metric g. The motion is *homothetic* when $\psi = \text{constant} \neq 0$, and *isometric* when $\psi = 0$. If (1.2) holds in M, we say that M admits a *conformal Killing vector* (CKV), a *homothetic Killing vector* (HKV), or a *Killing vector* (KV) according as whether ψ is a nonzero function, a nonzero constant, or zero.

The amount of interest in the use of the conformal group in physics has increased a great deal in the last decade.[1] In microphysics much attention is being given to such matters as the breaking of conformal invariance.[2] The study of conformal motions as an external symmetry group in the theory of gravitation and cosmology is also developing.[3] A survey of the use of the conformal group from its beginnings to the present time will be presented elsewhere. However, the following brief remarks will serve to set the background for the rest of the paper.

Brinkmann[4] long ago determined all Einstein spaces which can be conformally mapped nontrivially (i.e., nonhomothetically) on Einstein spaces.

Collinson[5] proved that the only curvature collineations[6] admitted by a vacuum space–time not of Petrov type N are conformal motions. However, type N vacuum spaces do admit more general types of symmetry.

Collinson and French[7] showed for nonflat vacuum space–times that (i) a conformal motion must be homothetic unless the space–time is type N with a twist-free principal null congruence, and (ii) for each Petrov type the maximum order of the group of conformal motions is at most one greater than the maximum order of the group of isometries.

The only type N vacuum fields which admit nontrivial conformal motions are the *pp* waves.[5,8] That *pp* waves admit homothetic motions has been demonstrated by McIntosh,[9] who also proved that nonflat vacuum space-times can admit a nontrivial[10] HKV only if the HKV is nonnull and not hypersurface orthogonal, is shear-free and has constant expansion; such HKV H has either (a) a nonnull homothetic bivector, in which case H is not tangent at a geodesic, or (b) a null homothetic bivector, in which case the space–time is necessarily Petrov type III or N.

Godfrey[11] has classified Weyl metrics according to the homothetic motions which they admit.

A result attributable to Yano[12] is the following: If H_1, H_2 are two HKV's in a Riemannian space, then their commutator $[H_1, H_2]$ is a Killing vector. This result has been mentioned by Eardley[13] and McIntosh,[9,14] and specializes to the cases where either or both of H_1, H_2 are improper, i.e., are Killing vectors.

Several authors[9,13–19] have discussed the physical importance of homothetic motions. Mathematically, the HKV should be treated with the same respect as a Killing vector, for it plays the same role as the KV in the solving of Einstein's field equations viz. reducing the order of the partial differential equations; if enough symmetry is present (KV's plus HKV's), the field equations reduce to ordinary differential equations. This paper supports the case for the seeking of solutions of the Einstein equations by requiring symmetry of a higher order than an isometry.

The above results imply that, when looking for conformally symmetric nonflat solutions to the vacuum field equations[20]

$$R_{\mu\nu} = 0, \tag{1.3}$$

we can restrict our attention to those spaces which admit at most one nonnull HKV, unless the space is Petrov type N with a twist-free geodesic ray congruence. In this paper we

 0022-2488/80/010120-09$01.00

further confine our attention to nonflat *vacuum spaces which are algebraically special, possessing an expanding and/or twisting principal null geodesic congruence, and which admit one HKV plus some KV's.* Here we treat the case of two, three, or four KV's. The case of one or no KV's plus the HKV will be dealt with in a separate paper.

This work has been inspired by the papers of Debney, Kerr, and Schild,[21] and Kerr and Debney[22] who constructed all possible isometry groups for nonflat vacuum Einstein spaces which possess an expanding and/or twisting, shear-free null geodesic congruence. It is a natural extension of that work. It is also believed to be the first *systematic* search for all vacuum spaces of the above type which admit homothetic motions.

2. FORMALISM

The basic tetrad formalism is that developed by Kerr, Debney, and Schild.[21,22] Only the main features of it are given here as a prelude to the development in the next section.

A complex, null orthonormal tetrad $\{e_a\}$ is defined over a four-dimensional Riemannian manifold M with metric g and signature $(+\ +\ +\ -)$. The vectors e_3 and e_4 are real, while e_1 and e_2 are complex conjugates, $e_2 = \bar{e}_1$. The dual basis $\{\epsilon^a\}$ is defined by the inner product

$$(\epsilon^a, e_b) = \delta^a{}_b.$$

The tetrad components of the metric tensor are given by

$$g_{ab} = (e_a, e_b) = g_{\mu\nu} e_a{}^\mu e_b{}^\nu$$

with

$$(g_{ab}) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (g^{ab}).$$

The rotation coefficients $\Gamma^m{}_{ab}$ are defined by

$$g_{am}\, \Gamma^m{}_{bc} = \Gamma_{abc} = -\epsilon_{a\mu;\nu} e_b{}^\mu e_c{}^\nu$$

and are related to the connection forms $\omega^a{}_b$ by

$$\omega^a{}_b = \Gamma^a{}_{bc} \epsilon^c.$$

The requirement that the tetrad be "rigid" implies

$$\omega_{ab} = -\omega_{ba}, \quad \Gamma_{abc} = -\Gamma_{bac}.$$

The tetrad vector e_4 is chosen to be a Debever principal null vector in an algebraically special space–time. The necessary and sufficient condition for this to be so are the vanishing of two of the five complex conformal quantities of Sachs[23]: $C^{(5)} = C^{(4)} = 0$. The Goldberg–Sachs theorem requires the vanishing of geodesy and shear,

$$\Gamma_{424} = 0 = \Gamma_{422}$$

and their complex conjugates $\Gamma_{414} = \Gamma_{411} = 0$. The remaining optical information is contained in the statement

$$\omega_{42} = -\rho \epsilon_2 + \Gamma_{423} \epsilon_4,$$

where $\rho = \theta + i\omega = -\Gamma_{421}$ is the complex divergence (assumed nonzero), θ and ω being the rate of expansion and twist respectively. One can choose to set $\Gamma_{423} = 0$ by means of a proper, orthochronous Lorentz transformation of the null tetrad, and then the following tetrad freedom remains:

$$e_1' = e^{-iB} e_1, \quad e_3' = e^{-A} e_3, \quad e_4' = e^A e_4, \tag{2.1}$$

where A and B are real functions. The (4,2) curvature equation [see Kerr and Debney,[22] Eq. (1.5)] now enables one to write

$$\omega_{42} = -d\zeta, \quad \epsilon^1 = \epsilon_2 = \rho^{-1} d\zeta, \tag{2.2}$$

where ζ is a local smooth complex function on M. Under a tetrad transformation (2.1), ω_{42} transforms as

$$\omega_{42}' = e^{A+iB} \omega_{42}. \tag{2.3}$$

A coordinate system $(\zeta, \bar{\zeta}, u, v)$ is chosen, where u, v are real functions satisfying

$$u_{,3} = 1, \quad u_{,4} = 0, \quad v = \mathrm{Re}(\rho^{-1}), \tag{2.4}$$

where ",a" denotes the directional derivative in the e_a direction. In these coordinates the metric takes the form

$$d\tau^2 = 2\epsilon_1\epsilon_2 + 2\epsilon_3\epsilon_4, \tag{2.5}$$

where

$$\begin{aligned} \epsilon_2 &= (v + \Delta)d\zeta, \quad \epsilon_1 = \bar{\epsilon}_2, \\ \epsilon_3 &= dv - 2\,\mathrm{Re}\{[(v - \Delta)\dot{\Omega} + D\Delta]d\zeta\} \\ &\quad + \mathrm{Re}(D\dot{\bar{\Omega}} + \mu\rho)\epsilon_4, \\ \epsilon_4 &= du + \Omega\, d\zeta + \bar{\Omega}\, d\bar{\zeta}. \end{aligned} \tag{2.6}$$

The functions Ω, μ, Δ are independent of the coordinate v and μ is the "complex mass." In (2.6) we use the notation $\dot{\Omega} \equiv \partial\Omega/\partial u \equiv \partial_u \Omega$, and the complex D operator is defined by

$$D \equiv \partial_\zeta - \Omega\partial_u.$$

The pure imaginary function Δ is defined by

$$\Delta \equiv i\,\mathrm{Im}(\bar{D}\Omega) = i\,\mathrm{Im}(\rho^{-1}),$$

whence the Debever vector e_4 is hypersurface orthogonal if and only if $\Delta = 0$.

The functions Ω and μ must satisfy the field equations

$$\begin{aligned} &\bar{D}\mu = 3\dot{\bar{\Omega}}\mu, \\ &\mathrm{Im}(\mu - \bar{D}\bar{D}D\Omega) = 0, \\ &\partial_u(\mu - \bar{D}\bar{D}D\Omega) = |\partial_u D\Omega|^2. \end{aligned} \tag{2.7}$$

The remaining conformal quantities of Sachs are given by Kerr and Debney,[22] Eq. (1.16), and we note

$$C^{(3)} = \mu\rho^3. \tag{2.8}$$

The space M is flat if and only if

$$\mu = \bar{D}\partial_u\, D\Omega = \partial_u\, \partial_u D\Omega = 0. \tag{2.9}$$

It was also shown by Kerr and Debney that if the space admits a KV of the type $K = e^{-p}\partial_u$, where $p = p(\zeta, \bar{\zeta})$, then a coordinate system $(\zeta, \bar{\zeta}, s, r)$ could be chosen such that

$$s = e^p u, \quad r = e^{-p} v, \quad K = \partial_s,$$

and the metric takes the form

$$\tfrac{1}{2} d\tau^2 = (r^2 + d^2)e^{2p}\, d\zeta\, d\bar{\zeta} + [dr + i(d_{\bar{\zeta}}\, d\bar{\zeta} - d_\zeta\, d\zeta)]\kappa$$

$$+ \{R^{(2)} + \mathrm{Re}[m/(r + id)]\}\kappa^2, \tag{2.10}$$

where

$$\kappa = e^{p}\epsilon_4 = ds + \Lambda\, d\zeta + \bar{\Lambda}\, d\bar{\zeta}, \quad \Lambda = e^{p}(\Omega - p_{\zeta} u), \tag{2.11}$$

$$d = -i\Delta e^{-p} = e^{\ 2p}\,\mathrm{Im}(\Lambda_{\bar{\xi}}), \quad m = \mu e^{-3p},$$

and $R^{(2)}$ is the 2-curvature of the 2-metric $e^{2p}\, d\zeta\, d\bar{\zeta}$, given by the "generalized Liouville equation"

$$R^{(2)} = e^{-2p} p_{\zeta\bar{\zeta}}, \tag{2.12}$$

where coordinate subscripts denote partial derivatives. Also, $\Lambda = \Lambda(\zeta,\bar{\zeta})$ and $d = d(\zeta,\bar{\zeta}) = \bar{d}$. The field equations in this coordinate system are

$$\partial_{\bar{\zeta}} m = 0, \quad R^{(2)}{}_{\zeta\bar{\zeta}} = 0, \quad \mathrm{Im}(m) = e^{-2p} d_{\zeta\bar{\zeta}} - 2R^{(2)} d. \tag{2.13}$$

3. COORDINATE AND TETRAD FREEDOM

We now determine the coordinate and tetrad freedom available under a homothetic change of metric (1.1).

Consider a diffeomorphism $\varphi: M \to M^*$ from one connected manifold M to another M^*. Write $q = \varphi(p) \in M^*$ for each point $p \in M$. Let $\{e^*{}_a\}$ be a basis in the tangent space $T_q(M^*)$ at q, corresponding to $\{e_a\}$ in $T_p(M)$ at p. φ defines a linear map φ_*,

$$\varphi_*: T_p(M) \to T_q(M^*).$$

Let $\{\epsilon^{*a}\}$ be a basis in the dual space $T_q{}^*(M^*)$ at q, corresponding to $\{\epsilon^a\}$ in $T_p{}^*(M)$ at p. φ induces the linear map φ^*,

$$\varphi^*: T_q{}^*(M^*) \to T_p{}^*(M),$$

by the requirement that the inner product (,) is preserved, thus

$$(\epsilon^a, e_b)(p) = (\epsilon^{*a}, e^*{}_b)(q) = \delta^a{}_b.$$

Let $\{\hat{e}_a\} = \{e^{-\phi} e^*{}_a\}$ be a new basis in $T_q(M^*)$, and let $\{e'_a\}$ be a basis in $T_q(M^*)$ obtained from $\{\hat{e}_a\}$ by a Lorentz transformation which leaves the direction of $\hat{e}_4$ unchanged. Then, by demanding $\Gamma_{423} = 0$, we may use (2.1) to write

$$\begin{aligned} e'_1 &= e^{-iB}\hat{e}_1 = e^{-\phi} e^{-iB} e^*{}_1, \\ e'_3 &= e^{-A}\hat{e}_3 = e^{-\phi} e^{-A} e^*{}_3, \\ e'_4 &= e^{A}\hat{e}_4 = e^{-\phi} e^{A} e^{*4}. \end{aligned} \tag{3.1}$$

Also let $\{\hat{\epsilon}^a\} = \{e^{\phi}\epsilon^{*a}\}$ be a new basis in $T_q{}^*(M^*)$ and introduce the basis $\{\epsilon'^a\}$ in $T_q{}^*(M^*)$ by

$$(e'^a, e'_b)(q) = \delta^a{}_b.$$

The structure constants $C^m{}_{ab}$ and the rotation coefficients $\Gamma^m{}_{ab}$ are related (in M) by

$$[e_a, e_b] = C^m{}_{ab} e_m, \quad 2\Gamma_{abm} = C_{abm} + C_{bma} - C_{mab}. \tag{3.2}$$

Under the mapping φ we require

$$\varphi_*[e_a, e_b] = [\varphi_* e_a, \varphi_* e_b].$$

Define, at $q \in M^*$,

$$\begin{aligned} \hat{C}^m{}_{ab}\hat{e}_m &= [\hat{e}_a, \hat{e}_b] = [e^{-\phi} e^*{}_a, e^{-\phi} e^*{}_b] \\ &= [e^{-\phi}\varphi_* e_a, e^{-\phi}\varphi_* e_b] \\ &= e^{-2\phi}\varphi_*[e_a, e_b] \quad (\text{since } \phi \text{ is constant}) \\ &= e^{-2\phi} C^{*m}{}_{ab} e^*{}_m \\ &= e^{-\phi} C^{*m}{}_{ab}\hat{e}_m, \end{aligned} \tag{3.3}$$

where $\varphi_* C^m{}_{ab} = C^{*m}{}_{ab}$. If $\Gamma'^m{}_{ab}, \hat{\Gamma}^m{}_{ab}, \Gamma^{*m}{}_{ab}$ are the rotation coefficients with respect to bases $\{e'_a\}, \{\hat{e}_a\}, \{e^*{}_a\}$ on M^*, then (3.2) and (3.3) give $\hat{\Gamma}_{abc} = e^{-\phi}\Gamma^*{}_{abc}$. Hence the respective connection coefficients are related by

$$\begin{aligned} \hat{\omega}_{ab}(q) &= (\hat{\Gamma}_{abc}\hat{\epsilon}^c)(q) = (\Gamma^*{}_{abc}\epsilon^{*c})(q) \\ &= \omega^*{}_{ab}(q) = \varphi^{*-1}(\omega_{ab}(p)). \end{aligned}$$

Using (2.2) we have

$$\begin{aligned} \hat{\omega}_{42}(q) &= \omega_{42}{}^*(q) = -\varphi^{*-1}(d\zeta(p)) = -d(\varphi^{*-1}\zeta(p)) \\ &= -d\zeta^*(q), \end{aligned}$$

where $\zeta^* = \zeta\circ\varphi^{-1}$ is a differentiable function on M^*.

Also, by the reasoning leading to (2.2) and (2.3) we have for a differentiable function ζ' on M^*,

$$\omega'_{42}(q) = -d\zeta'(q) = e^{A+iB}\hat{\omega}_{42}(q) = -e^{A+iB} d\zeta^*$$

which implies

$$\zeta' = \Phi(\zeta^*), \tag{3.4}$$

where

$$e^{A+iB} = \partial_{\zeta^*}\Phi \equiv \Phi_{\zeta^*}.$$

The function ζ^* is thus coupled to the tetrad $\{e^*{}_a\}$ through $\Phi_{\zeta^*} = e^{A+iB}$.

Define functions u^*, v^*, Ω^* on M^* by

$$u^* = u\circ\varphi^{-1}, \quad v^* = v\circ\varphi^{-1}, \quad \Omega^* = \Omega\circ\varphi^{-1}$$

and define

$$\epsilon^*{}_4 = du^* + \Omega^*\, d\zeta^* + \bar{\Omega}^*\, d\bar{\zeta}^*. \tag{3.5}$$

Let u', Ω' be smooth functions on M^* such that

$$\epsilon'_4 = du' + \Omega'\, d\zeta' + \bar{\Omega}'\, d\bar{\zeta}' \tag{3.6}$$

and require u' to satisfy [cf. (2.4)]

$$u'{}_{,3} = 1, \quad u'{}_{,4} = 0. \tag{3.7}$$

Since $u' = u'(\zeta^*, \bar{\zeta}^*, u^*, v^*)$ and $\epsilon'_4 = e^{\phi}e^{A}\epsilon^*{}_4$, we find from (3.5), (3.6), and (3.7) that

$$u' = e^{\phi}|\Phi_{\zeta^*}|(u^* + S^*), \tag{3.8}$$

where $S^*(\zeta^*, \bar{\zeta}^*)$ is a real function on M^*, and also

$$\Omega' = e^{\phi}|\Phi_{\zeta^*}|\Phi_{\bar{\zeta}^*}^{-1}[\Omega^* - S^*{}_{\zeta^*} - \tfrac{1}{2}\Phi_{\zeta^*\zeta^*}\Phi_{\zeta^*}^{-1}$$

$$(u^* + S^*)]. \tag{3.9}$$

The complex divergence relative to bases $\{e'_a\}$ and $\{e^*{}_a\}$ is given by

$$\rho' = -\Gamma'_{421} = -e^{-\phi}e^{A}\Gamma^*{}_{421} = e^{-\phi}|\Phi_{\zeta^*}|\rho^*, \tag{3.10}$$

where $\rho^* = \rho\circ\varphi^{-1}$. Let v', Δ' be smooth functions on M^* defined by

$$v' = \mathrm{Re}(\rho')^{-1}, \quad \rho' = (v' + \Delta')^{-1}.$$

Then, using (3.10), we have

$$v' = e^{\phi}\,|\Phi_{\zeta^*}|^{-1}v^* \tag{3.11}$$

and

$$\Delta' = e^{\phi}\,|\Phi_{\zeta^*}|^{-1}\Delta^*, \tag{3.12}$$

where

$$\Delta^* = \Delta\circ\varphi^{-1}.$$

Let $\mu^* = \mu\circ\varphi^{-1}$, where $\mu(\zeta,\bar\zeta,u)$ is the complex mass function of Sec. 2. Under the mapping φ the quantity $C^{(3)}$ of Sachs maps as $\varphi^*C^{(3)*} = C^{(3)}$, where $C^{(3)*} = \mu^*\rho^{*3}$. If $\hat C^{(3)}$ is the corresponding quantity relative to the basis $\{\hat e_a\}$ at $q\in M^*$, then[21]

$$\hat C^{(3)} = 2\hat\rho(\hat\Gamma_{123} + \hat\Gamma_{343}) = e^{-2\phi}C^{(3)*}.$$

Also, by the theory of Sec. 2, we have

$$C^{(3)\prime} = \mu'\rho'^3.$$

Under a tetrad transformation (3.1) we find

$$C^{(3)\prime} = \hat C^{(3)}$$

and so, with the aid of (3.10), we obtain

$$\mu' = e^{\phi}\,|\Phi_{\zeta^*}|^{-3}\mu^*. \tag{3.13}$$

Now let us view φ as a mapping of M into itself,

$$\varphi: M\to M.$$

Then we can interpret the transformation equations (3.4), (3.8), (3.9), (3.11), (3.12), and (3.13) in the following ways:

(i) $\varphi =$ identity map ($q\equiv p$), $\phi = 0$. This corresponds to a change of coordinates at the point $p\in M$.

(ii) $\varphi\neq$identity map, $\phi\neq 0$. This corresponds to a proper homothetic motion, where the coordinate system is "dragged along" by φ. The symbols with an asterisk are to be identified with the symbols without.

Summarizing, the remaining coordinate and tetrad freedoms under a homothetic change of metric (1.1) on M are:

$$\zeta' = \Phi(\zeta),\quad u' = e^{\phi}\,|\Phi_\zeta|(u+S),\quad v' = e^{\phi}\,|\Phi_\zeta|^{-1}v, \tag{3.14}$$

where $S(\zeta,\bar\zeta)$ is a real function,

$$\begin{aligned}\Omega' &= e^{\phi}\,|\Phi_\zeta|\Phi_\zeta^{-1}[\Omega - S_\zeta - \tfrac12\Phi_{\zeta\zeta}\Phi_\zeta^{-1}(u+S)],\\ \mu' &= e^{\phi}\,|\Phi_\zeta|^{-3}\mu,\\ \Delta' &= e^{\phi}\,|\Phi_\zeta|^{-1}\Delta,\end{aligned} \tag{3.15}$$

and

$$\begin{aligned}e_1' &= e^{-\phi}\,|\Phi_\zeta|\Phi_\zeta^{-1}e_1,\\ e_3' &= e^{-\phi}\,|\Phi_\zeta|^{-1}e_3,\\ e_4' &= e^{-\phi}\,|\Phi_\zeta|e_4.\end{aligned} \tag{3.16}$$

When $\phi = 0$ the homothety becomes an isometry (when φ is not the identity map) and Eqs. (3.14)–(3.16) reduce to those of Kerr and Debney.[22]

In the $(\zeta,\bar\zeta,s,r)$ coordinate system the transformation equations corresponding to (3.14) and (3.15) are

$$\zeta' = \Phi(\zeta),\quad s' = e^{\phi}C_0(s+A),\quad r' = e^{\phi}C_0^{-1}r, \tag{3.17}$$

where $A(\zeta,\bar\zeta)$ is a real function and C_0 (and ϕ) are real constants, and

$$\begin{aligned}A' &= e^{\phi}C_0\Phi_\zeta^{-1}(A - A_\zeta),\quad m' = e^{\phi}C_0^{-3}m,\\ d &= e^{\phi}C_0^{-1}d,\quad e^{P'} = C_0|\Phi_\zeta|^{-1}e^{P}.\end{aligned} \tag{3.18}$$

4. HOMOTHETIC KILLING VECTORS

It is well known[24] that every vector field X on M generates a local one-parameter group of local (infinitesimal, identity-connected) transformations φ_t, and conversely. Geometrically, if $U\subset M$, then φ_t takes each point $p\in U$ a parameter distance t along the integral curves of X. Suppose $X\equiv H$, a homothetic Killing vector, and the local one-parameter group φ_t of homothetic motions generated by H is defined by $x^\mu\to x'^\mu = f^\mu(x^\nu,t)$, where $(x^\nu) = (\zeta,\bar\zeta,u,v)$ and the x'^μ are given by

$$\begin{aligned}\zeta' &= \Phi(\zeta;t),\\ u' &= e^{\phi(t)}|\Phi_\zeta(\zeta;t)|\,[u + S(\zeta,\bar\zeta;t)],\\ v' &= e^{\phi(t)}|\Phi_\zeta(\zeta;t)|^{-1}v.\end{aligned} \tag{4.1}$$

Then $(x'^\mu)_{t=0} = x^\mu$ implies

$$\Phi(\zeta;0) = \zeta,\quad \Phi_\zeta(\zeta;0) = 1,\quad S(\zeta,\bar\zeta;0) = 0 = \phi(0).$$

The HKV has components H^μ relative to a local coordinate basis $\{\partial_\mu = \partial/\partial x^\mu\}$ at $p(t=0)$, where

$$H^\mu = \left[\frac{\partial x'^\mu}{\partial t}\right]_{t=0}. \tag{4.2}$$

Defining

$$\alpha(\zeta) = \left[\frac{\partial\Phi}{\partial t}\right]_{t=0},\quad R(\zeta,\bar\zeta) = \left[\frac{\partial S}{\partial t}\right]_{t=0} = \bar R,$$

and

$$a = \left[\frac{\partial\phi}{\partial t}\right]_{t=0} = \text{real constant},$$

we obtain the following form which a HKV must take if it is present in the type of space which we are investigating,

$$H = \alpha\partial_\zeta + \bar\alpha\partial_{\bar\zeta} + \mathrm{Re}(\alpha_\zeta)(u\partial_u - v\partial_v) + R\partial_u + a(u\partial_u + v\partial_v). \tag{4.3}$$

For H to exist, the constant a must be nonzero. If we set $a = 0$, the form (4.3) becomes that of a KV admitted by such a space [Kerr and Debney,[22] Eq. (3.5)].

The functions $\alpha(\zeta)$ and $R(\zeta,\bar\zeta)$ in (4.3) will transform under a coordinate transformation (3.14) (with $\phi = 0$) according to

$$\alpha' = \Phi_\zeta\alpha,\quad R' = |\Phi_\zeta|\{R - [\mathrm{Re}(\alpha_\zeta) + a]S + HS\}. \tag{4.4}$$

By solving the equations $R' = 0$, $\alpha' = 1$ when $\alpha\neq 0$, or by choosing a new u coordinate when $\alpha = 0$, we can put the HKV into either of the two mutually exclusive *canonical forms*

$$H = \partial_\zeta + \partial_{\bar\zeta} + u\partial_u + v\partial_v, \tag{4.5a}$$

$$H = u\partial_u + v\partial_v. \tag{4.5b}$$

In the $(\zeta,\bar{\zeta},s,r)$ coordinates the expressions corresponding to (4.3) and (4.4) are

$$H = \alpha\partial_\zeta + \bar{\alpha}\partial_{\bar{\zeta}} + a_0(s\partial_s - r\partial_r) + T\partial_s + a(s\partial_s + r\partial_r), \tag{4.6}$$

where α is a function of ζ only, $T(\zeta,\bar{\zeta})$ is a real function, a_0 and a are real constants (a nonzero for homothetic motions); and

$$\alpha' = \Phi_\zeta\, \alpha, \quad T' = C_0[T - (a_0 + a)\Lambda + H\Lambda\,]. \tag{4.7}$$

The two canonical forms of H in these coordinates are

$$H = \partial_\zeta + \partial_{\bar{\zeta}} + (a + a_0)s\partial_s + (a - a_0)r\partial_r, \tag{4.8a}$$

$$H = (a + a_0)s\partial_s + (a - a_0)r\partial_r. \tag{4.8b}$$

5. HOMOTHETIC KILLING EQUATIONS

The HKV (4.3) must satisfy the homothetic Killing equations (1.2) for $\psi =$ nonzero constant. We shall express these equations, and their first order integrability conditions, in a form which involves only the functions Ω and μ (and their derivatives) since these are the unknown functions in the field equations.

Let $y(x)$ denote the components of a geometrical object at a point of M with local coordinates (x^μ). Under a mapping φ of M let the transformed object be denoted by y'. A symmetry of the local object is defined by

$$y'(x') = y(x'), \tag{5.1}$$

where $\varphi : x \rightarrow x'$. If φ is the local one-parameter transformation φ_t generated by the HKV (4.3), then $\varphi_t : x \rightarrow x'(x,t)$. Identifying y with Ω and μ in turn in the derivative of both sides of (5.1) with respect to t, also using (3.15), we obtain (at $t = 0$) the homothetic Killing equations

$$(H - a)\Omega + \tfrac{1}{2}(\alpha_\zeta - \bar{\alpha}_{\bar{\zeta}})\Omega + \tfrac{1}{2}\alpha_{\zeta\zeta}u + R_\zeta = 0 \tag{5.2}$$

and (5.5) listed below, where H is given by (4.3). The integrability conditions are obtained by successive differentiation of (5.2) and (5.5) with respect to ζ, $\bar{\zeta}$, and u. Writing (5.2) in different form, we now list the set of homothetic Killing equations and their first order integrability conditions:

$$\text{(I)}\ H(\Omega - u\dot{\Omega}) + (\tfrac{1}{2}(\alpha_\zeta - \bar{\alpha}_{\bar{\zeta}}) - a)(\Omega - u\dot{\Omega}) + R\dot{\Omega} + R_\zeta = 0, \tag{5.3}$$

$$\text{(II)}\ H\dot{\Omega} + \alpha_\zeta\,\dot{\Omega} + \tfrac{1}{2}\alpha_{\zeta\zeta} = 0, \tag{5.4}$$

$$\text{(III)}\ H\mu + (3\,\mathrm{Re}(\alpha_\zeta) - a)\mu = 0, \tag{5.5}$$

$$\text{(IVa)}\ H\dot{\mu} + 4\,\mathrm{Re}(\alpha_\zeta)\dot{u} = 0, \tag{5.6}$$

$$\text{(IVb)}\ H(\bar{D}\dot{\Omega}) + 2\,\mathrm{Re}(\alpha_\zeta)\bar{D}\dot{\Omega} = 0, \tag{5.7}$$

$$\text{(IVc)}\ H\Delta + (\mathrm{Re}(\alpha_\zeta) - a)\Delta = 0, \tag{5.8}$$

$$\text{(IVd)}\ H\ddot{\Omega} + (\alpha_\zeta + \mathrm{Re}(\alpha_\zeta) + a)\ddot{\Omega} = 0. \tag{5.9}$$

The references in Roman numerals are used to match the corresponding equations for isometries in the Kerr and Debney paper[22] to which the above set reduces on putting $a = 0$.

Since there are essentially six unknowns $(\alpha,\bar{\alpha},\alpha_\zeta,\bar{\alpha}_{\bar{\zeta}},R,a)$ in the above set of equations, we have the following extension of the Kerr–Debney Lemma, based on the same reasoning with the standard form of these equations:

Theorem 1: *The dimension of the group of homothetic motions of an algebraically special vacuum space with nonzero complex divergence is at most six. When the dimension is six, all integrability conditions are zero identically and the space is flat.*

This result confirms the Collinson–French theorem[7] that the maximum order of the group H_m of homothetic motions admitted by a nonflat empty space–time is at most one greater than the maximum order of the group G_n of isometries.

In the $(\zeta,\bar{\zeta},s,r)$ coordinate system the equations corresponding to (5.3)–(5.9) are

$$\text{(I)}\ H\Lambda + (\alpha_\zeta - a_0 - a)\Lambda + T_\zeta = 0, \tag{5.10}$$

$$\text{(II)}\ Hp_\zeta + \alpha_\zeta\, p_\zeta + \tfrac{1}{2}\alpha_{\zeta\zeta} = 0, \tag{5.11}$$

$$\text{(II')}\ Hp + \mathrm{Re}(\alpha_\zeta) = a_0, \tag{5.12}$$

$$\text{(III)}\ Hm + (3a_0 - a)m = 0, \tag{5.13}$$

$$\text{(IVb)}\ HR^{(2)} + 2a_0R^{(2)} = 0, \tag{5.14}$$

$$\text{(IVc)}\ Hd + (a_0 - a)d = 0, \tag{5.15}$$

where $R^{(2)}$ is the 2-curvature given by (2.12), and H is given by (4.6).

6. SPACES WITH HIGH SYMMETRY

Since (see Sec. 1) there is at most one independent HKV in the space, we concern ourselves here with spaces which admit one HKV plus two, three, or four KV's, four KV's being the maximal number in a nonflat space. In order to find the metrics of those spaces possessing this high symmetry, we use the results of Kerr and Debney.[22] The $(\zeta,\bar{\zeta},s,r)$ coordinate system with the field equations and the homothetic Killing equations in the form (2.13) and (5.10)–(5.15) will be used except for the Cases VIII and IX below.

Case I (4 KV's + 1 HKV):

$$\Omega = \Lambda = id_0\bar{\zeta}, \quad \mu = m = m_0 = \text{nonzero real constant},$$

$$\Delta = id_0, \quad d = d_0 = \text{constant}, \quad p = 0 = R^{(2)}.$$

(i) $m_0 \neq 0$, $d_0 \neq 0$. Equations (5.13) and (5.15) give $a = 0$, so there is no HKV in this case.

(ii) $m_0 \neq 0$, $d_0 = 0$. Equations (5.11) and (5.12) yield $\alpha = \alpha_0\zeta + \beta$, $\mathrm{Re}(\alpha_0) = a_0$. There is enough coordinate freedom available to transform α to $\alpha = a_0\zeta$ and m_0 to 1. Equation (5.10) gives $T = T_0$, real constant, while (5.13) gives $3a_0 = a$. Equations (5.14) and (5.15) are trivially satisfied. Thus we have found an HKV of the form

$$H = a_0(\zeta\partial_\zeta + \bar{\zeta}\partial_{\bar{\zeta}} + 4s\partial_s + 2r\partial_r) + T_0\partial_s.$$

But a HKV is determined only up to an additive KV, and because $K = \partial_s$ is already present in the space, we may take the HKV to be

$$H = \zeta\partial_\zeta + \bar{\zeta}\partial_{\bar{\zeta}} + 4s\partial_s + 2r\partial_r. \tag{6.1}$$

The metric which admits this is

$$\tfrac{1}{2}d\tau^2 = r^2\, d\zeta\, d\bar{\zeta} + dr\, ds + r^{-1}\, ds^2, \tag{6.2}$$

which is Petrov type D. The 4 KV's are

$$K_1 = \partial_s, \quad K_2 = \partial_\zeta + \partial_{\bar\zeta}, \quad K_3 = i(\partial_\zeta - \partial_{\bar\zeta}),$$
$$K_4 = i(\zeta\partial_\zeta - \bar\zeta\partial_{\bar\zeta}).$$

The metric (6.2) is of Kerr–Schild type[25] since $D\Omega = 0$, and can be made manifestly of this type by choosing new coordinates (x,y,z,t), where

$$\zeta(z-t) = x + iy, \quad \sqrt{2}r = z - t, \quad \sqrt{2}s = z + t + \zeta\bar\zeta r.$$

There is a plane of singularities $z - t = 0$, so this is the metric for a nullicle.[26] The metric is known from the Collinson and French paper,[7] where it is in their Class C, Case (i), $\psi_2^0 = 1$ with $U^0 = 0$; it is, in fact, the NUT metric[27] with $\mu^0 = 0 = \rho^0$.

(iii) $m_0 = 0 = d_0$. The space is flat.

Case II (4 KV's + 1 HKV):

$$A = id_0\bar\zeta/R_0(\zeta\bar\zeta - R_0), \quad m = m_0 = \text{constant},$$
$$d = d_0 = -\operatorname{Im}(m_0)/2R_0, \quad R^{(2)} = R_0 = \text{real constant},$$
$$e^{-p} = \zeta\bar\zeta - R_0.$$

(i) $m_0 \neq 0, d_0 \neq 0$. Equations (5.13) and (5.15) give $a = 0$, so there is no HKV.

(ii) $m \neq 0, d_0 = 0$. Equation (5.14) gives $a_0 = 0$, and then (5.13) gives $a = 0$, so there is no HKV. This result confirms the well-known fact that *the Schwarzschild metric (m_0 real) does not admit a HKV*, and also proves that *the same is true of two of the NUT metrics (m_0 complex)*.

(iii) $m_0 = 0 = d_0$. The space is flat.

Case III (3 KV's + 1 HKV):

$$A = m = d = 0, \quad R^{(2)} = \zeta + \bar\zeta, \quad e^{-2p} = \tfrac{2}{3}(\zeta + \bar\zeta)^3.$$

Equation (5.14) gives $\alpha = -2a_0(\zeta + ib_0)$, where b_0 is a real constant. There is still a linear transformation in ζ left to transform α to the form $\alpha = -2a_0\zeta$. Then equations (5.11) and (5.12) are satisfied identically, while (5.13) and (5.15) are trivial. The remaining equation (5.10) gives $T = T_0$, real constant. Thus we arrive at the metric

$$\tfrac{1}{2}d\tau^2 = \tfrac{3}{2}r^2(\zeta + \bar\zeta)^{-3}\,d\zeta\,d\bar\zeta + dr\,ds + (\zeta + \bar\zeta)\,ds^2 \quad (6.3)$$

which is Petrov type III and admits the HKV

$$H = s\partial_s + r\partial_r, \quad (6.4)$$

where we have accounted for the KV's present when writing down the form of H. Putting $8\zeta = 3(x + iy)$ we can write the metric in the form

$$d\tau^2 = r^2x^{-3}(dx^2 + dy^2) + 2dr\,ds + \tfrac{3}{2}x\,ds^2, \quad (6.5)$$

which admits the HKV (6.4) plus the KV's

$$K_1 = \partial_s, \quad K_2 = \partial_y, \quad K_3 = 2(x\partial_x + y\partial_y) - s\partial_s + r\partial_r.$$

This metric is that given by Kerr and Debney,[22] Eq. (5.19); it has been noted by McIntosh[9]; it is the so-called "singular" metric in the Collinson and French paper,[7] Case (i) with $\psi_2^0 = 0$.

Case IV (2 KV's + 1 HKV):

$$A = i\bar\zeta e^{2p}[-\tfrac{1}{2}\operatorname{Im}(m_0)R^2 + C_1 + C_2(2\log R + R_0R^{-2})$$
$$+ C_3(R^2 + R_0^2R^{-2})],$$

$m = m_0$ = constant such that *either* $\operatorname{Re}(m_0) = 1$ or $\operatorname{Im}(m_0) = 1$,

$$d = e^{-2p}\operatorname{Im}(A_{\bar\zeta}), \quad R^{(2)} = R_0 = \text{constant},$$
$$e^{-p} = \zeta\bar\zeta - R_0,$$

where $R = |\zeta|$ and C_1, C_2, C_3 are real constants.

The space is flat when $m_0 = 0$.

(1) $m_0 \neq 0, R_0 \neq 0$. Equation (5.14) gives $a_0 = 0$, which together with (5.13) gives $a = 0$, so there is no HKV.

(2) $m_0 \neq 0, R_0 = 0$. Equation (5.14) is the zero identity, while (5.13) gives $3a_0 = a$. Equation (5.12) yields

$$\alpha = \alpha_0\zeta^2 - a_0\zeta,$$

where α_0 is a complex constant, while (5.11) is satisfied identically. Equation (5.15) is satisfied when either (i) $C_1 = C_2 = 0$; or (ii) $\alpha_0 = C_2 = 0$, $5a_0 = a$; or (iii) $\alpha_0 = C_1 = C_2 = 0$, $5a_0 = a$.

Using (ii) and (iii) with $3a_0 = a$ implies $a = 0$, and no HKV.

When (i) obtains we find

$$A = i\beta_0\zeta^{-1}, \quad d = 0, \quad \beta_0 = C_3 - \tfrac{1}{2}\operatorname{Im}(m_0).$$

Equation (5.10) and $3a_0 = a$ now give

$$T = i\beta_0[\tfrac{4}{3}a\log(\zeta/\bar\zeta) - \alpha_0\zeta + \bar\alpha_0\bar\zeta].$$

Using this form of T in (4.6) and taking the commutator of H with each of the two KV's

$$K_1 = \partial_s, \quad K_2 = i(\zeta\partial_\zeta - \bar\zeta\partial_{\bar\zeta})$$

and applying the theorem $[H,K_i] = mK_1 + nK_2$ leads to $\alpha_0 = 0$. Choosing $\operatorname{Re}(m_0) = 1$ as we may, we obtain the Petrov type D metric

$$\tfrac{1}{2}d\tau^2 = r^2(\zeta\bar\zeta)^{-2}\,d\zeta\,d\bar\zeta + dr\,ds + i\beta_0(\zeta^{-1}\,d\zeta - \bar\zeta^{-1}\,d\bar\zeta)dr$$
$$+ r^{-1}[ds + i\beta_0(\zeta^{-1}\,d\zeta - \bar\zeta^{-1}\,d\bar\zeta)]^2, \quad (6.6)$$

where β_0 is an arbitrary real constant, which admits the KHV

$$H = \zeta\partial_\zeta + \bar\zeta\partial_{\bar\zeta} - 4s\partial_s - 2r\partial_r + 4i\beta_0\log(\bar\zeta/\zeta)\partial_s. \quad (6.7)$$

This metric is not of the Kerr–Schild type since $D\Omega \neq 0$, but it does collapse to the Kerr–Schild metric (6.2) when $\beta_0 = 0$. It must be the special form of the "C" metric[8] of Ehlers and Kundt: in their coordinates

$$d\tau^2 = (x + y)^{-2}(x^{-3}\,dx^2 + x^3\,d\varphi^2 + y^{-3}\,dy^2 - y^3\,dt^2)$$

which admits the HKV

$$H = x\partial_x + y\partial_y - 3\varphi\partial_\varphi + 3t\partial_t.$$

Case V (2 KV's + 1 HKV):

$$A = i[c_0x^{-5/2}\sinh\tfrac{1}{2}\sqrt{13}(x - x_0) + \tfrac{3}{4}\operatorname{Im}(m_0)x^{-3}],$$

$m = m_0$ = complex constant, $\quad c_0$ = arbitrary real constant,

$$d = e^{-2p}\operatorname{Im}(A_{\bar\zeta}), \quad R^{(2)} = \zeta + \bar\zeta = x, \quad e^{-2p} = 2x^3/3.$$

The space cannot be flat since $\bar D\partial_u D\Omega = e^{2p}$.

As in Case III Eq. (5.14) leads to $\alpha = -2a_0\zeta$, while

(5.11) and (5.12) are satisfied identically. However, Eq. (5.15) gives rise to three possibilities:

(1) $c_0 = 0 = 3a_0 - a$, $\operatorname{Im}(m_0)\neq 0$ $(a_0\neq 0)$;

(2) $c_0 = 0 = \operatorname{Im}(m_0)$, $3a_0 - a\neq 0$, a_0 arbitrary;

(3) $c_0 = 0 = 3a_0 - a = \operatorname{Im}(m_0)$ $(a_0\neq 0)$.

Taking these in turn we have

(1) If $a_0 = 0$, then $a = 0$ and there is no HKV. If $a_0\neq 0$, equation (5.13) is satisfied identically and (5.10) gives $T = T_0$, real constant. Thus we arrive at the Petrov type II metric

$$d\tau^2 = 3x^{-3}(t^2 + A^2x^{-2})(dx^2 + dy^2) + 4(dt + Ax^{-2}\,dy)\kappa + \left\{2x + \operatorname{Re}\left(\frac{m_0x}{tx - iA}\right)\right\}\kappa^2, \tag{6.8}$$

where $\kappa = ds - Ax^{-3}\,dy$, $2\zeta = x + iy$, $t = r/2$ and A is an arbitrary, nonzero, real constant given by $4A = 3\operatorname{Im}(m_0)$. The HKV admitted by this metric is

$$H = x\partial_x + y\partial_y - 2s\partial_s - t\partial_t. \tag{6.9}$$

Since $\Delta = i\,de^{P}\neq 0$ in this case, the metric (6.8) is that of a type II vacuum space with twist; it has not yet been identified.

(2) In this case Eq. (5.13) gives $m_0 = 0$, while (5.10) yields $T = T_0$, real constant. Thus we recover the solution (6.3), (6.4) of Case III, which admits 3 KV's.

(3) Equation (5.13) allows the (real) constant m_0 to be arbitrary, but if zero the space would be flat. We may use the transformation (3.18) to set $m_0 = 1$. Equation (5.10) gives $T = T_0$, real constant. Thus, on putting $8\zeta = 3(X + iY)$, we arrive at the Petrov type II metric

$$d\tau^2 = r^2X^{-3}(dX^2 + dY^2) + 2dr\,ds + \left(\frac{3X}{2} + \frac{2}{r}\right)ds^2, \tag{6.10}$$

which admits the HKV

$$H = X\partial_X + Y\partial_Y - 2s\partial_s - r\partial_r. \tag{6.11}$$

The metric (6.10) is given in the Collinson-French paper[7] as their class C, Case (i) "singular" solution with $\psi_2^0 = 1$, $U^0 = \zeta + \bar{\zeta}$.

The two KV's admitted by (6.8) and (6.10) are

$$K_1 = \partial_s,\quad K_2 = i(\partial_\zeta - \partial_{\bar\zeta}) = \partial_y.$$

Case VI (2 KV's + 1 HKV):

$$\Lambda = \lambda_0\bar{\zeta}\zeta^{-1/\alpha_0} + m_0\bar{\zeta}^2\zeta^{(\alpha_0-3)/\alpha_0},$$
$$m = 2m_0\alpha_0^{-1}(\alpha_0 - 3)\zeta^{-3/\alpha_0},$$
$$d = \operatorname{Im}(\Lambda_{\bar\zeta}),\quad p = 0 = R^{(2)},$$

where $\operatorname{Re}(\alpha_0) = 1$ (α_0 invariant), λ_0 is a complex constant, and m_0 is a complex constant which can be made real if the remaining coordinate freedom is used. However, we prefer to use this freedom to transform α obtained from Eqs. (5.11) and (5.12) to

$$\alpha = \beta_0\zeta,\quad \operatorname{Re}(\beta_0) = a_0.$$

Now condition (2.9) requires $m_0\neq 0$ for nonflat space. Then Eq. (5.13) gives $(3a_0 - a)\alpha_0 - 3\beta_0 = 0$. Adding this to its complex conjugate yields $a = 0$, so there is no HKV in this case.

Case VII 2 KV's + 1 HKV):

$$\Lambda = i(\zeta + \bar{\zeta})^{-1}L(\theta),\quad \zeta = \rho e^{i\theta},$$
$$m = \mu_0\zeta^{3/2},\quad \mu_0 \text{ complex constant},$$
$$d = (\zeta + \bar{\zeta})^{1/2}F(\theta),\quad F = \bar{F},$$
$$R^{(2)} = \zeta + \bar{\zeta},\quad e^{-2p} = \tfrac{2}{3}(\zeta + \bar{\zeta})^3.$$

Kerr and Debney did not obtain a solution for L and F, where

$$\tfrac{1}{2}\sin 2\theta\,(dL/d\theta) - L = e^{2p}F,$$

and

$$\operatorname{Im}[\mu_0(1 + 2e^{-2i\theta})^{-3/2}] = e^{2P}\cos^2\theta\,(d^2F/d\theta^2) - (2 + \tfrac{1}{4}e^{-2P})F,$$

$$e^{-2P}\left(\cos^2\theta\,\frac{d^2P}{d\theta^2} + \frac{3}{2}\right) = 1,\quad P = P(\theta).$$

The only known solution of the last equation is $e^{-2P} = 2/3$. Just as in Case III, Eq. (5.14) leads to $\alpha = -2a_0\zeta$. Then (5.11) and (5.12) are satisfied identically.

If $m\neq 0$, Eq. (5.13) now gives $a = 0$ and there is no HKV.

If $m = 0$, Eq. (5.13) gives no information. But (5.15) gives either $a = 0$ and no HKV; or $F(\theta) = 0$ when we have $L = C\tan\theta$, where C is a real constant, so that

$$\Lambda = C(\zeta - \bar{\zeta})(\zeta + \bar{\zeta})^{-2}.$$

Equation (5.10) now is an ordinary differential equation in $T = T(\zeta - \bar{\zeta})$ if either (1) $a_0 + a = 0$, or (2) $C = 0$, or (3) $a_0 + a = 0 = C$, in which cases T can only be a real constant.

(1) Putting $8\zeta = 3(x + iy)$ we obtain the Petrov type III metric

$$d\tau^2 = r^2x^{-3}(dx^2 + dy^2) + 2dr\,ds - 2Cx^{-2}y\,dy\,dr + \tfrac{3}{2}x(ds - Cx^{-2}y\,dy)^2, \tag{6.12}$$

where C is an arbitrary, real, nonzero constant, which admits the HKV

$$H = x\partial_x + y\partial_y + r\partial_r. \tag{6.13}$$

This metric appears to be the "singular" solution in the Collinson–French paper,[7] Class C, Case (ii) with $\psi_2^0 = 0$, $\nabla U^0\neq 0$.

(2) This is a degenerate case in that the metric is (6.3) with 3 KV's. However, there remains the possibility that $a_0 = a$, when the metric (6.3) admits the HKV

$$H = x\partial_x + y\partial_y - s\partial_s. \tag{6.14}$$

(3) Again degenerate to (6.3), which is now found to admit also the HKV (6.13).

Case VIII (2 KV's + 1 HKV):

In this case, and in Case IX below, the metrics do not admit a KV of the type $K = e^{-P}\partial_u$, so we revert to the $(\zeta,\bar{\zeta},u,v)$ coordinate system. The appropriate sets of equations to use are (3.14), (4.3), and (5.3)–(5.9). The functions Ω and μ are known to the extent that they are dependent upon μ only:

$$\Omega = \Omega(u), \quad \mu = \mu_0\bar{\Omega}^{-3}, \tag{6.15}$$

where μ_0 is complex constant. The two remaining field equations are

$$\dot{E} = |(\Omega^2)^{\cdot\cdot}|^2 \tag{6.16}$$

and

$$\mathrm{Im}(E) = 0, \tag{6.17}$$

and

$$E(u) = 4\mu_0\bar{\Omega}^{-3} + 2\bar{\Omega}[\bar{\Omega}(\Omega^2)^{\cdot\cdot}]^{\cdot} \tag{6.18}$$

and the dot denotes differentiation with respect to u. The two KV's present are $K_1 = \partial_\zeta + \partial_{\bar{\zeta}}$, $K_2 = i(\partial_\zeta - \partial_{\bar{\zeta}})$. Now the commutator of a HKV H with each of these is

$$[K_i,H] = a_iK_1 + b_iK_2 \quad (i = 1,2), \tag{6.19}$$

where the a_i, b_i are real constants. Applying this constraint and using the transformation equations (3.14) and (4.4) we find seven possible forms for the HKV. However, upon using the field equations and homothetic Killing equations with each of these forms, we find either no solution or flat space except when H takes the form

$$H = \zeta\partial_\zeta + \bar{\zeta}\partial_{\bar{\zeta}} + (a+1)u\partial_u + (a-1)v\partial_v \quad (a \neq -1,0). \tag{6.20}$$

Using this form, Eqs. (5.3) and (5.4) give

$$\Omega = Cu^{a/(a+1)},$$

where C is a complex constant which we take nonzero, for $C = 0$ implies flat space. But since

$$H\mu = H(\mu_0\bar{\Omega}^{-3}) = -3\mu\bar{\Omega}^{-1}H\bar{\Omega} = -3a\mu,$$

Eq. (5.5) gives $(3-4a)\mu = 0$, so that *either* (1) $\mu = 0$, *or* (2) $a = \frac{3}{4}$, *or* (3) $3 - 4a = 0 = \mu$.

(1) The field equations (6.16) and (6.17) are satisfied if $a = 1$ (flat space) or $a = \frac{3}{2}$. The remaining equations (5.6)–(5.9) are trivial. Thus we arrive at the Petrov type III metric

$$d\tau^2 = 2v^2\,d\zeta\,d\bar{\zeta} + 2[dv - \tfrac{3}{5}u^{-2/5}v(C\,d\zeta + \bar{C}\,d\bar{\zeta}) + C\bar{C}u^{-4/5}\epsilon_4]\epsilon_4 \tag{6.21}$$

where

$$\epsilon_4 = du + u^{3/5}(C\,d\zeta + \bar{C}\,d\bar{\zeta}),$$

and C is an arbitrary, nonzero, complex constant. This metric admits the HKV

$$H = 2\zeta\partial_\zeta + 2\bar{\zeta}\partial_{\bar{\zeta}} + 5u\partial_u + v\partial_v. \tag{6.22}$$

The form (6.21) of the metric appears to be explicitly new.

(2) The field equations are satisfied if

$$343\mu = -16(C\bar{C})^2u^{-9/7}.$$

The remaining equations (5.6)–(5.9) are trivial. We have obtained the Petrov type II metric

$$d\tau^2 = 2v^2\,d\zeta\,d\bar{\zeta} + 2[dv - \tfrac{3}{7}u^{-4/7}v(C\,d\zeta + \bar{C}\,d\bar{\zeta}) + \tfrac{4}{343}C\bar{C}u^{-9/7}(21u^{1/7} - 4C\bar{C}v^{-1})\epsilon_4]\epsilon_4, \tag{6.23}$$

where

$$\epsilon_4 = du + u^{3/7}(C\,d\zeta + \bar{C}\,d\bar{\zeta}),$$

and C is an arbitrary, nonzero complex constant. This metric admits the HKV

$$H = 4\zeta\partial_\zeta + 4\bar{\zeta}\partial_{\bar{\zeta}} + 7u\partial_u - v\partial_v. \tag{6.24}$$

The metric does not appear to have been written down explicitly before.

(3) The field equation (6.16) is not satisfied, so this case does not yield a solution.

Case IX (2 KV's + 1 HKV):

The functions Ω and μ were not fully determined by Kerr and Debney, but it is known that

$$\Omega = \Omega(t), \quad \mu = u^{-3}\nu(t), \quad t = u/\mathrm{Im}(\zeta).$$

The two KV's are

$$K_1 = \partial_\zeta + \partial_{\bar{\zeta}}, \quad K_2 = \zeta\partial_\zeta + \bar{\zeta}\partial_{\bar{\zeta}} + u\partial_u - v\partial_v.$$

By using the constraint (6.19) with each of these KV's and by making a transformation $u \to u + S$ we find two allowable forms for H, only one of which leads to nonflat solutions, namely,

$$H = u\partial_u + v\partial_v. \tag{6.25}$$

Equations (5.3), (5.4), and (5.5) now yield

$$\Omega = Ct = Cuy^{-1}, \quad \mu = Au^{-3}t^4 = Auy^{-4},$$

where A, C are complex constants and $\zeta = x + iy$. The field equations (2.7) are satisfied if and only if $A = 0$ and either (1) $C = 0$, or (2) $C = i/2$, or (3) $C = 3i/4$. Condition (2.9) shows that nonflat space results only in case (3), when we have the Petrov type III metric

$$d\tau^2 = 2v^2(dx^2 + dy^2) + 2[dv + \tfrac{3}{2}vy^{-1}\,dy - \tfrac{3}{8}y^{-2}(du - \tfrac{3}{2}uy^{-1}dy)](du - \tfrac{3}{2}uy^{-1}\,dy). \tag{6.26}$$

This metric is of the Robinson, Robinson, and Zund form[28] if one puts $P = v$, $Z = 3v/2(\zeta - \bar{\zeta})$, $S = -3/2(\zeta - \bar{\zeta})^2$, $a = 1$, $b = -3u/2(\zeta - \bar{\zeta})$, $\sigma = u$, $\rho = v$ in their Eqs. (2.12) and (2.13). It does not seem to have been written down explicitly before.

CONCLUSION

The class of nonflat, vacuum, algebraically special metrics which are expanding and/or twisting and which admit an H_m symmetry group ($m = 3,4,5$) is small. It consists of the metrics (6.2), (6.5), (6.6), (6.8), (6.10), (6.12), (6.21), (6.23), and (6.26). All are hypersurface orthogonal except (6.8), which has not yet been identified. Three others in the Robinson–Trautman class are believed to be explicitly new, namely, (6.21), (6.23), and (6.26).

ACKNOWLEDGMENT

Thanks are due to Dr. Colin McIntosh and Dr. Tony

Lun for useful discussions in the course of preparing this paper.

[1]A full discussion is given in the first author's Ph.D. thesis, Massey University (1977).
[2]See, e.g., S. Ferrara, R. Gatto, and A.F. Grillo, "Conformal Algebra in Space–Time," in *Tracts in Modern Physics* (Springer-Verlag, Berlin, 1973), Vol. 67.
[3]See, e.g., L. Infeld and A. Schild, Phys Rev. **68**, 250 (1945); **70**, 410 (1946) for an early discussion. A recent discussion is to be found in Refs. 13 and 14.
[4]H.W. Brinkmann, Math. Ann. **94**, 119 (1925).
[5]C.D. Collinson, J. Math. Phys. **11**, 818 (1970).
[6]G.H. Katzin, J. Levine, and W.R. Davis, J. Math. Phys. **10**, 617 (1969); **11**, 1578 (1970).
[7]C.D. Collinson and D.C. French, J. Math. Phys. **8**, 701 (1967).
[8]J. Ehlers and W. Kundt, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), p. 80; A. Thompson, Phys. Rev. Lett. **34**, 507 (1975).
[9]C.B.G. McIntosh, Gen. Rel. Grav. 7, 199, 215 (1976).
[10]The HKV's considered in this paper are all nontrivial (i.e., are not Killing vectors) unless otherwise specified.
[11]B.B. Godfrey, Gen. Rel. Grav. **3**, 3 (1972).
[12]K. Yano, J. Indian Math. Soc. (N.S.) **15**, 105 (1951).
[13]D.M. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[14]C.B.G. McIntosh, Phys. Lett. A **50**, 429 (1975).
[15]A. Einstein, Univ. Nacional de Tucuman Rev. Ser. A **2**, 11 (1941).
[16]M.E. Cahill and A. Taub, Commun. Math. Phys. **21**, 1 (1971).
[17]A.H. Taub, in *General Relativity*, edited by L. O'Raifeartaigh (Oxford U.P., London, 1972), pp. 135–50.
[18]F. Hoyle and J.V. Narlikar, Mon. Not. R. Astron. Soc. **155**, 305 (1972).
[19]J. Winicour, in *Lecture Notes in Physics* (Springer-Verlag, Berlin, 1972), Vol. 14, p. 145.
[20]Throughout this paper Greek indices $\lambda, \mu, \nu, \cdots$ will be used to denote components relative to a coordinate frame. Latin indices $i, j, k\cdots$ will be used to denote tetrad components. All indices run from 1 to 4.
[21]G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[22]R.P. Kerr and G.C. Debney, J. Math. Phys. **11**, 2807 (1970).
[23]R.K. Sachs, Proc. R. Soc. London Ser. A **264**, 309 (1961); **270**, 103 (1962).
[24]See, e.g., S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963).
[25]R.P. Kerr and A. Schild, *Proceedings of Symposia in Applied Mathematics* (Amer. Math. Soc., Providence, R.I., 1965), Vol. 17, p. 199.
[26]P.A. Hogan, Int. J. Theor. Phys. **11**, 419 (1974).
[27]E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[28]I. Robinson, J. Robinson, and J.D. Zund, J. Math. Mech. **18**, 881 (1969).

Einstein spaces and homothetic motions. II

W. D. Halford

Department of Mathematics, Massey University, Palmerston North, New Zealand
(Received 6 December 1978)

Algebraically special, nonflat vacuum Einstein spaces with an expanding and/or twisting geodesic principal null congruence are considered. These spaces are assumed to possess locally a homothetic symmetry as well as an isometry (one Killing vector). Nine metrics are obtained, six of which are twisting Petrov type II.

1. INTRODUCTION

In a previous paper,[1] referred to as Paper I, we determined all algebraically special nonflat vacuum Einstein spaces with an expanding and/or twisting geodesic principal null congruence, and which admitted a homothetic Killing vector (HKV) plus two, three, or four Killing vectors (KVs). It was evident that the addition of a homothetic symmetry enabled one to obtain solutions of the vacuum field equations more readily. In this paper the same claim is made; an HKV is placed in the space along with just one KV, and the field equations are reduced in complexity. Even so, they are still of a sufficiently difficult nature that in some cases a complete solution has not been obtained.

The formalism used in this paper is that of Paper I. It is an extension of the formalism developed by Debney, Kerr, and Schild[2] and by Kerr and Debney.[3] In order to make this paper self-contained and easier to read, the main equations and results of Paper I will be summarized next.

The general form of an HKV admitted by a vacuum space–time with nonzero complex divergence is, in local coordinates $(\zeta,\bar{\zeta},u,v)$,

$$H = \alpha\partial_\zeta + \bar{\alpha}\partial_{\bar{\zeta}} + \mathrm{Re}(\alpha_\zeta)(u\partial_u - v\partial_v) + R\partial_u + a(u\partial_u + v\partial_v), \tag{1.1}$$

where $\alpha = \alpha(\zeta)$, $R = R(\zeta,\bar{\zeta}) = \bar{R}$, and a is a nonzero real constant. In any space–time there is only one independent HKV.[4] Of importance is the following result: A Killing vector K and a homothetic Killing vector H obey the commutation relation

$$[K,H] = KH - HK = \lambda K, \tag{1.2}$$

where λ is a constant. This restriction on the geometry is used to determine more precisely the form of H for a given K. Kerr and Debney[3] showed that K may take one of the canonical forms

$$\text{(i)}\quad K = e^{-p}\partial_u, \quad p = p(\zeta,\bar{\zeta}), \tag{1.3}$$

$$\text{(ii)}\quad K = \partial_\zeta + \partial_{\bar{\zeta}} \tag{1.4}$$

in the $(\xi,\bar{\xi},u,v)$ coordinates, in which the field equations take the form

$$\bar{D}\mu = 3\dot{\bar{\Omega}}\mu, \tag{1.5}$$

$$\mathrm{Im}(\mu - \bar{D}\bar{D}D\Omega) = 0, \tag{1.6}$$

$$\partial_u(\mu - \bar{D}\bar{D}D\Omega) = |\partial_u D\Omega|^2. \tag{1.7}$$

Here Ω and μ are functions of only three coordinates $\zeta,\bar{\zeta},u$ and the operator D is defined by

$$D = \partial_\zeta - \Omega\partial_u.$$

The dot appearing in (1.5) and in the sequel denotes differentiation with respect to u. The bar over a symbol denotes complex conjugation.

The metric which admits the KV and the HKV is of the form

$$d\tau^2 = 2\epsilon_1\epsilon_2 + 2\epsilon_3\epsilon_4, \tag{1.8}$$

where

$$\begin{aligned} \epsilon_2 &= (v + \Delta)\, d\zeta, \quad \epsilon_1 = \bar{\epsilon}_2, \\ \epsilon_3 &= dv - 2\,\mathrm{Re}\{[(v - \Delta)\dot{\Omega} + D\Delta]\, d\zeta\} + \mathrm{Re}(D\dot{\bar{\Omega}} + \mu\rho)\epsilon_4, \\ \epsilon_4 &= du + \Omega\, d\zeta + \bar{\Omega}\, d\bar{\zeta}, \end{aligned} \tag{1.9}$$

and Δ is defined by

$$\Delta = i\,\mathrm{Im}(\bar{D}\Omega). \tag{1.10}$$

The complex divergence $\rho(=\theta + iw)$ is related to the coordinate v by $v = \mathrm{Re}(\rho^{-1})$.

In order to obtain the explicit forms of the metric (1.8), our procedure is to take one of the forms (1.3), (1.4) for K, solve the field equations (1.5)–(1.7) for Ω and μ subject to the following homothetic Killing equations and their integrability conditions:

$$(H - a)(\Omega - \mu\dot{\Omega}) + \tfrac{1}{2}(\alpha_\zeta - \bar{\alpha}_{\bar{\zeta}})(\Omega - \mu\dot{\Omega}) + R\dot{\Omega} + R_\zeta = 0, \tag{1.11}$$

$$H\dot{\Omega} + \alpha_\zeta\dot{\Omega} + \tfrac{1}{2}\alpha_{\zeta\zeta} = 0, \tag{1.12}$$

$$H\mu + [3\,\mathrm{Re}(\alpha_\zeta) - a]\mu = 0, \tag{1.13}$$

$$H\dot{\mu} + 4\,\mathrm{Re}(\alpha_\zeta)\dot{\mu} = 0, \tag{1.14}$$

$$H(\bar{D}\dot{\Omega}) + 2\,\mathrm{Re}(\alpha_\zeta)(\bar{D}\dot{\Omega}) = 0, \tag{1.15}$$

$$H\Delta + [Re(\alpha_\zeta) - a]\Delta = 0, \tag{1.16}$$

$$H\ddot{\Omega} + [\alpha_\zeta + \mathrm{Re}(\alpha_\zeta) + a]\ddot{\Omega} = 0, \tag{1.17}$$

where $\alpha(\zeta)$ and $R(\zeta,\bar{\zeta})$ are the functions occurring in H, which are found initially from the geometric condition (1.2). This leaves a certain amount of coordinate freedom which may be used to simplify the form of the metric. The function Δ is computed from (1.10) and checked for consistency with (1.16). The Killing equations and their integrability conditions to be satisfied for a selected K are obtained from (1.11)–(1.17) by putting $a = 0$ and by using α and R as in (1.3) or (1.4) as the case may be.

When the KV is present in the form (1.3), Kerr and Debney[3] showed that it was convenient to use $(\zeta,\bar{\zeta},s,r)$ as local coordinates, where $s = e^{p}u$ and $r = e^{-p}v$. The form of

 0022-2488/80/010129-06$01.00

the KV is then simply

$$K = \partial_s, \tag{1.18}$$

and the metric takes the form

$$\tfrac{1}{2}d\tau^2 = (r^2 + d^2)e^{2p}\, d\zeta\, d\bar{\zeta} + [dr + i(d_{\bar{\zeta}}\, d\bar{\zeta} - d_\zeta\, d\zeta)]\kappa + \{R^{(2)} + \mathrm{Re}[m/(r + id)]\}\kappa^2, \tag{1.19}$$

where

$$\kappa = ds + \Lambda\, d\zeta + \bar{\Lambda}\, d\bar{\zeta}, \quad \Lambda = e^p(\Omega - p_\zeta u), \tag{1.20}$$
$$d = -i\Delta e^{-p} = e^{-2p}\,\mathrm{Im}(\Lambda_{\bar{\zeta}}), \quad m = \mu e^{-3p}.$$

$R^{(2)}$ is the curvature of the two-dimensional metric $e^{2p}\, d\zeta\, d\bar{\zeta}$ and is given by the generalized Liouville equation

$$R^{(2)} = e^{-2p} p_{\zeta\bar{\zeta}}, \tag{1.21}$$

where, as usual, coordinate subscripts denote partial derivatives. Also $\Lambda = \Lambda(\zeta,\bar{\zeta})$ and $d = d(\zeta,\bar{\zeta}) = \bar{d}$. The field equations (1.5)–(1.7) become

$$m_{\bar{\zeta}} = 0, \tag{1.22}$$
$$R^{(2)}{}_{\zeta\bar{\zeta}} = 0, \tag{1.23}$$
$$\mathrm{Im}(m) = e^{-2p} d_{\zeta\bar{\zeta}} - 2R^{(2)} d \tag{1.24}$$

The homothetic Killing equations (1.11)–(1.17) become

$$H\Lambda + (\alpha_\zeta - a_0 - a)\Lambda + T_\zeta = 0, \tag{1.25}$$
$$Hp_\zeta + \alpha_\zeta\, p_\zeta + \tfrac{1}{2}\alpha_{\zeta\zeta} = 0, \tag{1.26}$$
$$Hp + \mathrm{Re}(\alpha_\zeta) = a_0, \tag{1.27}$$
$$Hm + (3a_0 - a)m = 0, \tag{1.28}$$
$$HR^{(2)} + 2a_0 R^{(2)} = 0, \tag{1.29}$$
$$Hd + (a_0 - a)d = 0, \tag{1.30}$$

where the HKV has the general form

$$H = \alpha\partial_\zeta + \bar{\alpha}\partial_{\bar{\zeta}} + (a + a_0)s\partial_s + (a - a_0)r\partial_r + T\partial_s. \tag{1.31}$$

Here $\alpha = \alpha(\zeta)$, $T = T(\zeta,\bar{\zeta}) = \bar{T}$, and a_0, a are real constant with $a \neq 0$. In Paper I it was shown that H could be put into one or other of the canonical forms

(i) $$H = (a + a_0)s\partial_s + (a - a_0)r\partial_r, \tag{1.32}$$

(ii) $$H = \partial_\zeta + \partial_{\bar{\zeta}} + (a + a_0)s\partial_s + (a - a_0)r\partial_r. \tag{1.33}$$

This is always possible in the presence of $K = \partial_s$, but the coordinate freedom is reduced and depends upon which of the two canonical forms of H we take.

We now proceed to determine those nonflat vacuum space–times with expanding and/or twisting rays which admit an HKV and one KV.

2. SPACES ADMITTING $K = \partial_s$ AND H AS IN (1.32)

Equation (1.27) gives $a_0 = 0$. The form of the HKV reduces to

$$H = s\partial_s + r\partial_r. \tag{2.1}$$

Equations (1.25), (1.28), (1.30) and field equation (1.22) give

$$\Lambda = m = d = 0,$$

and the Eqs. (1.26), (1.29), and (1.24) are satisfied identically. It remains to determine $p(\zeta,\bar{\zeta})$ from field equation (1.23). To do this, we consider separately the possibilities $R^{(2)} = 0$, $R^{(2)} =$ nonzero const, $R^{(2)} \neq$ const. It turns out that the only case leading to nonflat space is the last. Then $R^{(2)} = 2\,\mathrm{Re}[F(\zeta)]$ for some analytic function F of ζ. Using available coordinate freedom (see Paper I), the 2-curvature transforms as

$$R^{(2)} \to R^{(2)\prime} = C_0^{-2} R^{(2)}, \tag{2.2}$$

where C_0 is a real constant. Setting $\zeta' = C_0^{-2}F(\zeta)$ in (2.2) and dropping the primes (since we shall be working with the new functions hereafter), we have

$$R^{(2)} = \zeta + \bar{\zeta} = e^{-2p} p_{\zeta\bar{\zeta}}. \tag{2.3}$$

This generalized Liouville equation possesses only one known solution, namely,

$$3e^{-2p} = 2(\zeta + \bar{\zeta})^3. \tag{2.4}$$

Substituting (2.4) into (1.19), we have the metric

$$d\tau^2 = 3r^2(\zeta + \bar{\zeta})^{-3}\, d\zeta\, d\bar{\zeta} + 2dr ds + 2(\zeta + \bar{\zeta})\, ds^2$$

or, putting $8\zeta = 3(x + iy)$,

$$d\tau^2 = r^2x^{-3}(dx^2 + dy^2) + 2dr\, ds + \tfrac{3}{2}x\, ds^2. \tag{2.5}$$

This is the Petrov type III hypersurface-orthogonal metric of Kerr and Debney[3] which, in fact, admits not just one KV but three of them:

$$K_1 = \partial_s, \quad K_2 = \partial_y, \quad K_3 = 2(x\partial_x + y\partial_y) - s\partial_s + r\partial_r.$$

Because it also admits the HKV (2.1) it has appeared already in Paper I.

3. SPACES ADMITTING $K = \partial_s$ AND H NOT IN FORM (1.32)

H may be taken in form (1.33), but we shall not do so because we prefer to have the full coordinate freedom available for future use. We take H in the general form (1.31) with $\alpha \neq 0$. The constraint (1.2) requires $\lambda = a_0 + a$. We shall split the analysis according to whether the 2-curvature $R^{(2)}$ is constant or not.

A. $R^{(2)} = \mathrm{const} = R_0$

Equation (1.29) gives $a_0 R_0 = 0$. Consider in turn the possibilities $R_0 = 0$, $R_0 \neq 0$.

Case I: $R_0 = 0$. We have $p_{\zeta\bar{\zeta}} = 0$, implying $p = \mathrm{Re}[G(\zeta)]$ for some analytic function G of ζ. The allowed coordinate transformation $\zeta \to \zeta' = \Phi(\zeta)$, under which $e^p \to e^{p'} = C_0|\Phi_\zeta|^{-1}e^p$ (see Paper I), enables us to set $p = 0$ by choosing $\Phi_\zeta = C_0 e^G$. The coordinate freedom left is a linear transformation in ζ, with complete freedom on s and r. Now (1.26) and (1.27) give $\alpha = \gamma_0\zeta + \alpha_0$, where γ_0 and α_0 are constants such that $\mathrm{Re}(\gamma_0) = a_0$. There are two possibilities:

(i) $\gamma_0 \neq 0$, (ii) $\gamma_0 = 0$ ($\Rightarrow a_0 = 0$).

In (i), if $\alpha_0 \neq 0$ we can use the coordinate freedom in ζ to transform α to $\alpha = \gamma_0\zeta$. Also, a_0 may or may not be zero. If $a_0 = 0$, then $\gamma_0 = ib_0$, where b_0 is a nonzero real constant which we may take equal to 1, since a multiplicative constant can be absorbed into the HKV. Thus, if $a_0 = 0$, we have $\alpha = i\zeta$.

In (ii), $\alpha = \alpha_0 \neq 0$ and a_0 is necessarily zero.

In both (i) and (ii), field equations (1.22) and (1.24) give, quite generally,

$$2id = 2\bar{\zeta}\beta - 2\zeta\bar{\beta} + \varphi - \bar{\varphi}, \tag{3.1}$$

where $m = 2\beta_\zeta$, $\beta = \beta(\zeta)$ and $\varphi(\zeta)$ is a function yet to be determined. The definition of d in (1.20) then implies

$$\Lambda = \bar{\zeta}^2\beta + \bar{\zeta}\varphi + B_\zeta,$$

where $B(\zeta,\bar{\zeta})$ is a real function of integration which can be eliminated by letting $s \to s + B$, as is allowed. Then

$$\Lambda = \bar{\zeta}^2\beta + \bar{\zeta}\varphi. \tag{3.2}$$

We must now discuss (i) and (ii) separately.

Case I (i): $R_0 = 0$, $\alpha = \gamma_0\zeta$, $\mathrm{Re}(\gamma_0) = a_0$.

Equations (1.22) and (1.28) give

$$m(\zeta) = N\zeta^c, \quad c = (a - 3a_0)/\gamma_0,$$

where N and c are complex constants. Integration yields β, and then (3.1) gives d. Substituting d into (1.30), we get

$$\mathrm{Im}\{\bar{\zeta}[2\gamma_0\zeta\beta_\zeta + 2\bar{\gamma}_0\beta + 2(a - a_0)\beta] + [\gamma_0\zeta\varphi_\zeta + (a - a_0)\varphi]\} = 0,$$

which gives rise to the following solutions:

(A) $\varphi = 2M_0\zeta + D_0\zeta^{(a - a_0)/\gamma_0} \quad (a \neq a_0)$,

(B) $\varphi = B_0\gamma_0^{-1}\log\zeta + 2M_0\zeta + E_0 \quad (a = a_0)$,

where B_0, D_0, E_0, and M_0 are constants, B_0 real. In (B) one can redefine the function φ to absorb E_0, and in both (A) and (B) the constant M_0 can be eliminated by an allowed coordinate transformation $s \to s + 2\,\mathrm{Re}(M_0\zeta\bar{\zeta}^2)$. Having done this, we determine $T(\zeta,\bar{\zeta})$ from (1.25). We obtain respectively

(A) $T = T_0$, (B) $T = T_0 - B_0\zeta\bar{\zeta}$,

where T_0 is a real constant which can be eliminated because of the presence of $K = \partial_s$ in the space.

Thus there are two vacuum metrics (1.19) which admit the Killing vector $K = \partial_s$ and a homothetic Killing vector H, specified by

Case I (i)(A):

$a \neq a_0$, $(a + a_0)$ may or may not equal zero,

$$p = 0 = R^{(2)},$$
$$\Lambda = N_0\zeta^{(a - 3a_0 + \gamma_0)/\gamma_0}\bar{\zeta}^2 + D_0\zeta^{(a - a_0)/\gamma_0}\bar{\zeta}, \tag{3.3}$$
$$m = 2(a - 3a_0 + \gamma_0)\gamma_0^{-1}N_0\zeta^{(a - 3a_0)/\gamma_0},$$
$$d = \tfrac{1}{2}i(\bar{\Lambda}_\zeta - \Lambda_{\bar{\zeta}}),$$

where $a_0, a, \gamma_0, N_0, D_0$ are all constants (γ_0, D_0 complex) such that $a \neq 0$, $\mathrm{Re}(\gamma_0) = a_0$ and the HKV admitted is

$$H = \gamma_0\zeta\partial_\zeta + \bar{\gamma}_0\bar{\zeta}\partial_{\bar{\zeta}} + (a + a_0)s\partial_s + (a - a_0)r\partial_r. \tag{3.4}$$

A simplification occurs if $a_0 = 0$. For, as pointed out above, we may then take $\gamma_0 = i$

Case I (i)(B):

$a = a_0$, $(a_0 + a)$ cannot equal zero,

$$p = 0 = R^{(2)},$$
$$\Lambda = N_0\zeta^{(\gamma_0 - 2a)/\gamma_0}\bar{\zeta}^2 + B_0\gamma_0^{-1}\bar{\zeta}\log\zeta, \tag{3.5}$$
$$m = 2(\gamma_0 - 2a)\gamma_0^{-1}N_0\zeta^{-2a/\gamma_0},$$
$$d = \tfrac{1}{2}i(\bar{\Lambda}_\zeta - \Lambda_{\bar{\zeta}}),$$

where a, γ_0, N_0, B_0 are constants, with γ_0 complex such that $\mathrm{Re}(\gamma_0) = a \neq 0$, and the HKV admitted is

$$H = \gamma_0\zeta\partial_\zeta + \bar{\gamma}_0\bar{\zeta}\partial_{\bar{\zeta}} + (2as - B_0\zeta\bar{\zeta})\partial_s. \tag{3.6}$$

In both Case I(i)(A) and Case I(i)(B) the remaining coordinate freedom may be used to make N_0 a real constant. Both metrics (3.3) and (3.5) are Petrov type II with twist, unless $N_0 = 0$ when the space is flat.

Note: (a) The metric (3.3) with $a = 3a_0(\neq 0)$ and $D_0 \neq 0$ is Zund's metric[5] [his Eqs. (1), (6), and (9)].

(b) Formally putting $a = 0$ in (3.3) and (3.5) produces vacuum metrics which admit Killing vectors.

In Case I(i)(A), we regain the Kerr–Debney metric [Ref. 3, Eq. (6.16)] which admits *two* KVs, when $a_0 = 1$, $a = 0$. In case I(i)(B), $a_0 = 0 = a$ gives, for $B_0 \neq 0$, one of the Kerr–Debney metrics [Ref. 3, Eq. (6.10)] after suitable coordinate transformations; it admits two KVs. If, further, we put $B_0 = 0$ in this last case, we obtain one of the NUT metrics [see Paper I, Eq. (6.2)] which is Petrov type D and admits four KVs and an HKV.

Case I (ii): $R_0 = 0$, $\alpha = \alpha_0(\neq 0)$, $a_0 = 0$.

Equations (1.22) and (1.28) give $m(\zeta) = Ne^{\zeta/\alpha_0}$, where N is an arbitrary complex constant. Integrating $m = 2\beta_\zeta$, we get β, and then (3.1) gives d. Substituting d into (1.30), we get

$$\mathrm{Im}\{2[\bar{\zeta}(\alpha_0\beta_\zeta - \beta) + \alpha_0\beta] + (\alpha_0\varphi_\zeta - \varphi)\} = 0,$$

with the solution, for β as found,

$$\varphi(\zeta) = A_0 - N(k + \bar{\alpha}_0\zeta)e^{\zeta/\alpha_0} + 2\bar{\alpha}_0M_0 + 2\bar{M}_0(\zeta + \alpha_0),$$

where A_0, k, M_0 are constants, A_0 real. However, M_0 and A_0 can be eliminated by the allowed coordinate transformations $s \to s + 2\,\mathrm{Re}(M_0\zeta\bar{\zeta}^2)$, $s \to s + A_0\zeta\bar{\zeta}$ respectively. Having done this, we determine $T(\zeta,\bar{\zeta})$ from (1.25), obtaining

$$T = |\alpha_0|^2[N(k - |\alpha_0|^2 + \bar{\alpha}_0\zeta)e^{\zeta/\alpha_0} + \bar{N}(\bar{k} - |\alpha_0|^2 + \alpha_0\bar{\zeta})e^{\bar{\zeta}/\bar{\alpha}_0}]. \tag{3.7}$$

Thus we arrive at the metric (1.19) with

$$p = 0 = R^{(2)},$$
$$\Lambda = N(\tfrac{1}{2}\alpha_0\bar{\zeta}^2 - k\bar{\zeta} - \bar{\alpha}_0\zeta\bar{\zeta})e^{\zeta/\alpha_0}, \tag{3.8}$$
$$m = Ne^{\zeta/\alpha_0},$$
$$d = \mathrm{Im}[N(\alpha_0\bar{\zeta} - \bar{\alpha}_0\zeta - k)e^{\zeta/\alpha_0}],$$

where α_0, N, and k are complex constants, $\alpha_0 \neq 0$. This metric admits the Killing vector $K = \partial_s$ and the HKV

$$H = \alpha_0\partial_\zeta + \bar{\alpha}_0\partial_{\bar{\zeta}} + s\partial_s + r\partial_r + T\partial_s, \tag{3.9}$$

where T is given in (3.7). The metric is Petrov type II with twist, unless $N = 0$ when the space–time is flat.

Case II: $R_0 \neq 0$ (const).

Equation (1.29) requires $a_0 = 0$, so $(a_0 + a)$ cannot equal zero for proper homothetic motions.

By means of the transformation (2.2) with $C_0^2 = |R_0|$ we can make $R_0 = \pm 1$. The metric $e^{2p}\,d\zeta\,d\bar{\zeta}$ is then that of a sphere or pseudosphere and so the coordinates can be chosen so that $e^{-p} = \zeta\bar{\zeta} - R_0$. Then $\bar{D}\partial_u D\Omega = 0 = \partial_u\partial_u D\Omega$ and $\mu = m(\zeta\bar{\zeta} - R_0)^{-3}$, so unless $m = 0$ (corresponding to flat space) Petrov type II solutions are possible. The coordinate freedom left is a bilinear transformation in ζ and complete freedom in s and r.

Equations (1.26) and (1.27) give

$$\alpha = \alpha_0\zeta^2 + ib_0\zeta - \bar{\alpha}_0R_0, \tag{3.10}$$

where α_0 and b_0 are constants, b_0 real. There are two cases to consider:

(i) $R_0 = -1$ (sphere), (ii) $R_0 = +1$ (pseudosphere)

Case II (i): $R_0 = -1$.

It is always possible to reduce the form (3.10) to

$$\alpha = k(\zeta^2 + 1), \tag{3.11}$$

where k is a constant. The coordinate freedom left on ζ is $\zeta' = (K\zeta + L)/(K - L\zeta)$, $K^2 + L^2 = 1$. There is still complete freedom in s and r. We can use part of this freedon in s, after making the substitution

$$\zeta = \tan z,$$

to put the HKV in the form

$$H = \partial_z + \partial_{\bar z} + a(s\partial_s + r\partial_r), \tag{3.12}$$

where a is a real constant. The freedom remaining on the coordinate s is $s \to C_0(s + A)$, $A = \bar A = e^{az}\psi(z - \bar z)$.

The remaining equations (1.22), (1.24), (1.25), (1.28), and (1.30) lead to m as given below in (3.14), and

$$\Lambda = e^{az}\cos^2 z \cdot f(z - \bar z),$$

where f is a function of $(z - \bar z)$ satisfying

$$(1 + x^2)f''' + (1 + 2x)f'' - 2f' = m_0.$$

Here m_0 is a complex constant, $x = \tan(z - \bar z)$, and the prime denotes differentiation with respect to x. On putting $f'(x) = \eta(q)$, $q = \frac{1}{2}(1 + ix)$, this yields the solution

$$\eta = \eta_H - \tfrac{1}{2} m_0,$$

where

$$\begin{aligned} \eta_H &= C_1 F(\alpha,\beta,\gamma,q) \\ &\quad + C_2 q^{1-\gamma} F(\alpha - \gamma + 1, \beta - \gamma + 1, 2 - \gamma, q), \\ \alpha &= 2, \quad \beta = -1, \quad \gamma = 1 - \tfrac{1}{2} i, \end{aligned}$$

C_1, C_2 are arbitrary constants, and F is the hypergeometric function. There is no solution expressible in finite terms other than the trivial one $\eta_H = 0$. Thus

$$f(x) = \int \eta_H(x)\, dx + (N - \tfrac{1}{2} m_0 x),$$

where N is a complex constant.

The particular choice $\eta_H = 0$ gives the metric

$$\begin{aligned} \tfrac{1}{2} d\tau^2 &= (r^2 + d^2)\sec^2(z - \bar z)\, dz\, d\bar z \\ &\quad + [dr + i(d_{\bar z}\, d\bar z - d_z\, dz)]\kappa \\ &\quad + \{\mathrm{Re}[m/(r + id)] - 1\}\kappa^2, \end{aligned} \tag{3.13}$$

where $\kappa = ds + \Lambda \sec^2 z\, dz + \bar\Lambda \sec^2 \bar z\, d\bar z$, and

$$\begin{aligned} \Lambda &= e^{az}\cos^2 z[N - \tfrac{1}{2} m_0 \tan(z - \bar z)], \\ m &= m_0 e^{az}, \\ d &= (-i/4)(m_0 e^{az} - \bar m_0 e^{a\bar z}), \end{aligned} \tag{3.14}$$

where m_0, N are arbitrary complex constants. Since $\Delta = ide^{P}$, this metric is Petrov type II with twist, unless m_0 is zero when the space–time is flat. The HKV admitted is (3.12). In $(\zeta,\bar\zeta,s,r)$ coordinates, this metric takes the form (1.19) with

$$\begin{aligned} e^{-P} &= \zeta\bar\zeta + 1, \quad R^{(2)} = -1, \\ \Lambda &= (\zeta^2 + 1)^{-1} e^{-iaQ_0}[N - \tfrac{1}{2} m_0(\zeta - \bar\zeta)(\zeta\bar\zeta + 1)^{-1}], \\ m &= m_0 e^{-iaQ_0}, \quad Q_0 = Q_0(i\zeta) \\ d &= -(i/4)(m_0 e^{iaQ_0} - \bar m_0 e^{ia\bar Q_0}), \end{aligned} \tag{3.15}$$

where $Q_n(y)$ is the Legendre function of the second kind, and $Q_0(i\zeta) = \frac{1}{2}\log[(1 + i\zeta)/(1 - i\zeta)]$. The HKV now admitted is

$$H = (\zeta^2 + 1)\partial_\zeta + (\bar\zeta^2 + 1)\partial_{\bar\zeta} + a(s\partial_s + r\partial_r). \tag{3.16}$$

Case II (ii): $R_0 = 1$.

It is not always possible to reduce (3.10) to a single form. We consider separately

(A) $\alpha_0 \neq 0$, $b_0 = 0$, (B) $\alpha_0 = 0$, $b_0 \neq 0$,

(C) $\alpha_0 \neq 0$, $b_0 \neq 0$.

Case II (ii)(A): By means of a transformation $\zeta \to \zeta' = (\alpha_0/\bar\alpha_0)^{1/2}\zeta$ we can reduce α to the form

$$\alpha_0 = |\alpha_0|(\zeta^2 - 1).$$

It turns out that there is no nonflat solution to Eqs. (1.22), (1.23), (1.24), (1.25), (1.28), and (1.30) in this case.

Case II (ii)(B): Now $\alpha = ib_0\zeta$ and we can absorb the constant b_0 into the form of the HKV, and also eliminate $T(\zeta,\bar\zeta)$ by using some of the s-coordinate freedom, to obtain

$$H = i\zeta\partial_\zeta - i\bar\zeta\partial_{\bar\zeta} + a(s\partial_s + r\partial_r). \tag{3.17}$$

Equations (1.22)–(1.25), (1.28), and (1.30) lead to m as given below in (3.20), and

$$\Lambda = \zeta^{-1-ia} f(x), \quad x = \zeta\bar\zeta - 1, \tag{3.18}$$

where

$$f(x) = B - \tfrac{1}{2} m_0(1 + ia)^{-1}x^{-2} + \int x^{-3}G(x)\, dx. \tag{3.19}$$

Here B is a constant which can be made real by the transformation $s \to s + \mathscr{A}$, where $\mathscr{A} = a^{-1}(\zeta^{-ia} + \bar\zeta^{ia})\,\mathrm{Im}(B)$. The function $G(x)$ is given by

$$\begin{aligned} G(x) &= kF(-1, -1 - ia, -2, -x) \\ &\quad + lx^3 F(2, 2 - ia, 4, -x), \end{aligned}$$

valid for $|x| < 1$, where k and l are arbitrary real constants and F is the hypergeometric function.

There is no solution expressible in finite terms except the trivial one $G = 0$. In this particular case we have the metric (1.19) with

$$\begin{aligned} e^{-P} &= \zeta\bar\zeta - 1, \quad R^{(2)} = 1, \\ \Lambda &\equiv \Lambda_p \\ &= \zeta^{-1-ia}[B - \tfrac{1}{2} m_0(1 - ia)(1 + a^2)^{-1}(\zeta\bar\zeta - 1)^{-2}], \\ m &= m_0\zeta^{-ia}, \\ d &= -\tfrac{1}{2} i(1 + a^2)^{-1}(\zeta\bar\zeta - 1)^{-1} \\ &\quad \times [(1 - ia)m_0\zeta^{-ia} - (1 + ia)\bar m_0\bar\zeta^{ia}], \end{aligned} \tag{3.20}$$

where a, B, m_0 are constants, with m_0 complex and $a \neq 0$. The space–time will be flat iff $m_0 = 0$; otherwise, Petrov type II with twist. The HKV admitted is (3.17).

More generally, we should have Λ as in (3.18) and (3.19).

Case II (ii)(C): In this case we can reduce the form (3.10) to $\alpha = |\alpha_0|(\zeta^2 + ib_0\zeta - 1)$ by means of the transformation $\zeta \to \zeta' = (\alpha_0/\bar\alpha_0)^{1/2}\zeta$, where α_0 and b_0 are nonzero constants, b_0 real. If $\mathrm{Im}(\alpha_0) \neq 0$, we can reduce α further by transforming b_0 to zero, and then we have Case II(ii)(A) with no nonflat solution. Otherwise, α_0 is a real nonzero constant, and we have

$$\alpha = \alpha_0[(\zeta + \tfrac{1}{2} ib_0)^2 + (\tfrac{1}{4} b_0^2 - 1)].$$

(*a*) *If* $b_0 = 2$, we put $\zeta + i = -(\alpha_0 z)^{-1}$ giving $\alpha = (\alpha_0 z^2)^{-1}$ and we may use a coordinate transformation to put the HKV in the form

$$H = \partial_z + \partial_{\bar z} + a(s\partial_s + r\partial_r). \tag{3.21}$$

(*b*) *If* $b_0 > 2$, putting $z = \arctan[(\zeta + \tfrac{1}{2}ib_0)/k]$, where $k^2 = \tfrac{1}{4} b_0^2 - 1$, gives $\alpha = \alpha_0 k^2 \sec^2 z$, and we may take the HKV in the form (3.21) after absorbing the factor $\alpha_0 k$.

(*c*) *If* $b_0 < 2$, putting $z = \arctan[(\zeta + ib_0)/(iM)]$, where $M^2 = 1 - \tfrac{1}{4} b_0^2$, gives $\alpha = -\alpha_0 M^2 \sec^2 z$, and we may take the HKV in the form

$$H = i(\partial_z - \partial_{\bar z}) + a(s\partial_s + r\partial_r), \tag{3.22}$$

after absorbing the factor $\alpha_0 M$.

We treat each of these cases in turn.

Case II (ii)(C)(a): Equations (1.22)–(1.25), (1.28), and (1.30) lead to the metric

$$\tfrac{1}{2}d\tau^2 = \alpha_0^2 x^{-2}(r^2 + d^2)dz\,d\bar z + [dr + i(d_{\bar z}\,d\bar z - d_z\,dz)]\kappa + \{1 + \mathrm{Re}[m_0 e^{az}(r + id)^{-1}]\}\kappa^2, \tag{3.23}$$

where

$$x = 1 + i\alpha_0(z - \bar z) = \bar x,$$

$$\kappa = ds + \alpha_0^{-1}[e^{az} f(x)dz + e^{a\bar z}\bar f(x)d\bar z],$$

$$d = -\tfrac{1}{2}\alpha_0^{-2} x^2[e^{az} f'(x) + e^{a\bar z}\bar f'(x)].$$

The function f is given by

$$f(x) = (A\alpha_0 a^{-2} - 2iB\alpha_0^2 a^{-3} - \tfrac{1}{2} m_0\alpha_0^3 a^{-1})x^{-2} + (iAa^{-1} + 2B\alpha_0 a^{-2})x^{-1} + iBa^{-1} + Cx^{-2}e^{iax/\alpha_0}, \tag{3.24}$$

where $a,\alpha_0 A,B,C,m_0$ are arbitrary constants, with a and α_0 nonzero real. The metric (3.23) is Petrov type II with twist, unless $m_0 = 0$ when the space–time is flat. The HKV admitted is (3.21).

Case II (ii)(C)(b): Equations (1.22)–(1.25), (1.28), and (1.30) give

$$m = m_0 e^{az}, \quad m_0 \text{ arbitrary complex constant},$$

$$\Lambda = e^{az}\cos^2 z \cdot f(\theta), \quad \theta = z - \bar z,$$

where

$$z = \arctan[(\zeta + \tfrac{1}{2} ib_0)/k], \quad 4k^2 = b_0^2 - 4, \ b_0 > 2,$$

and the function $f(\theta)$ is to be determined from

$$\begin{aligned} &k(k + ib_0 \tan\theta)^2 \\ &\times\{2[(2k^2 - \tfrac{1}{2} ib_0 ka)\tan^2\theta - (2ak^2 + a + 2ib_0 k)\tan\theta \\ &- (2k^2 + 2 - \tfrac{1}{2} ib_0 ka)]f' + [(2ib_0 k - ak^2 - a)\tan^2\theta \\ &+ (8k^2 + 4 + ib_0 ka)\tan\theta - (2ib_0 k - ak^2)]f'' \\ &+ (k + \tfrac{1}{2} ib_0 \tan\theta)^2 f'''\} = m_0(1 + \tan^2\theta)^2, \end{aligned} \tag{3.25}$$

for arbitrary $b_0\ (>2)$, $a\ (\neq 0)$, and m_0. Solutions $f(\theta)$ to (3.25), other than the flat-space one $m_0 = 0, f = \text{const}$, have not been found.

Case II (ii)(C)(c): The analysis is essentially the same as in Case II(ii)(C)(b), with a similar conclusion.

B. $R^{(2)} \neq$ const

We may proceed as in Sec. 2 to obtain $R^{(2)}$ and $p(\zeta,\bar\zeta)$ as in (2.3) and (2.4) respectively. Equation (1.29) gives $\alpha = -2a_0(\zeta + ie_0)$, where e_0 is a real constant. There is still enough coordinate freedom on ζ to bring α into the form $\alpha = -2a_0\zeta$, and we can use some s-coordinate freedom to transform the function $T(\zeta,\bar\zeta)$ in (1.31) to zero. Equations (1.26) and (1.27) are satisfied identically. The remaining equations (1.22)–(1.25), (1.28), and (1.30) must be solved for the two possibilities $a_0 = 0$, $a_0 \neq 0$.

Case III (i): $a_0 = 0$.

We regain the metric (2.5), which admits not just one Killing vector, but three, so that this case is degenerate.

Case III (ii): $a_0 \neq 0$.

After suitable allowed transformations, we obtain

$$\Lambda = \lambda\zeta^c, \quad -2a_0 c = 3a_0 + a, \quad \lambda \text{ real const},$$

$$m = m_0 \text{ real const}, \quad d = 0.$$

This gives the metric

$$d\tau^2 = 3r^2(\zeta + \bar\zeta)^{-3}d\zeta\,d\bar\zeta + 2dr\,ds + 2\lambda dr(\zeta^c d\zeta + \bar\zeta^c d\bar\zeta) + 2(\zeta + \bar\zeta + m_0 r^{-1})[ds + \lambda(\zeta^c d\zeta + \bar\zeta^c d\bar\zeta)]^2, \tag{3.26}$$

which admits $K = \partial_s$ and the HKV

$$H = 2(\zeta\partial_\zeta + \bar\zeta\partial_{\bar\zeta}) - a_0^{-1}[(a + a_0)s\partial_s + (a - a_0)r\partial_r]. \tag{3.27}$$

The constants $a_0\ (\neq 0)$, $a\ (\neq 0)$ and λ are all real and arbitrary, and either (A) $c \neq -3$ (i.e., $a \neq 3a_0$) when $m_0 = 0$ in order to satisfy the field equations, or (B) $c = -3$ (i.e., $a = 3a_0$) when m_0 is real and arbitrary. In case (A), the metric is Petrov type III and twist-free. In case (B), it is twist-free and type II ($m_0 \neq 0$) or type III ($m_0 = 0$).

4. SPACES ADMITTING $K = \partial_\zeta + \partial_{\bar\zeta}$ AND AN HKV

Elsewhere[6] we showed that the HKV in the presence of this K must be of the form, in $(\zeta,\bar\zeta,u,v)$ coordinates,

$$H = \zeta\partial_\zeta + \bar\zeta\partial_{\bar\zeta} + (a + 1)u\partial_u + (a - 1)v\partial_v. \tag{4.1}$$

From Eqs. (1.11)–(1.13) and their Killing counterparts, we obtain two possibilities:

$$\text{(i)} \quad a \neq -1, \quad \Omega = u^{a/(a+1)}g, \quad \mu = y^{a-3}h, \tag{4.2}$$

$$\text{(ii)} \quad a = -1, \quad \Omega = y^{-1}f, \quad \mu = y^{-4}\Gamma, \tag{4.3}$$

where $\zeta = x + iy$, f and Γ are both functions of u, and g and h are both functions of $(y^{a+1}u^{-1})$.

Case IV (i): Equations (1.15) and (1.17) require $g = 0$. Field equation (1.5) is satisfied if *either* $h = 0$ (flat space solution) *or* $a = 3$ and h is a constant which field equation (1.6) requires to be real. Then $\Omega = 0$, μ is a real constant which can be transformed to 1, and Eq. (1.16) gives $\Delta = 0$. All other conditions are satisfied, and so we obtain the Petrov type D, twist-free metric

$$\tfrac{1}{2}\,d\tau^2 = v^2\,d\zeta\,d\bar\zeta + du\,dv + v^{-1}du^2, \tag{4.4}$$

which we recognize as the NUT metric [(6.2) of Paper I]. This case is therefore degenerate in the sense that the metric admits

$$H = \zeta\partial_\zeta + \bar\zeta\partial_{\bar\zeta} + 4u\partial_u + 2v\partial_v \tag{4.5}$$

and *four* Killing vectors.

Case IV (ii): The remaining conditions (1.14), (1.15), and (1.17) on Ω and μ are satisfied. The field equation (1.5) gives

$$(2i + 3\dot{\bar{f}})\Gamma + \bar{f}\dot{\Gamma} = 0, \tag{4.6}$$

which may be expressed in terms of a new complex function $G(u)$ as follows:

$$f = 2i/\dot{\bar{G}}, \quad \Gamma = -ie^{G}\dot{G}^{3}/8. \tag{4.7}$$

Introducing

$$E(u) \equiv if - 2f\dot{f}, \quad F(u) \equiv iE + \bar{f}\dot{E}, \tag{4.8}$$

the field equations (1.6) and (1.7) become the ordinary differential equations

$$4\dot{\Gamma} - (3iF + 2\bar{f}\dot{F})^{\cdot} = |\dot{E}|^{2}, \tag{4.9}$$

$$4(\Gamma - \bar{\Gamma}) = 3i(F + \bar{F}) + 2(\bar{f}\dot{F} - f\dot{\bar{F}}). \tag{4.10}$$

By substituting (4.7) into (4.8) and (4.9), we get a fifth order equation in the complex function $G(u)$, subject to the constraint (4.10). Apart from some obvious special solutions which represent flat space–time, I do not have any solutions in this case.

5. CONCLUSION

The number of nonflat, expanding and/or twisting vacuum solutions of Einstein's field equations which admit one Killing vector and a homothetic Killing vector is small. In the foregoing sections we have obtained nine such metrics, but two of these, namely, (2.5) and (4.4), are degenerate in the sense that they admit more than one Killing vector. The other seven are given at Eqs. (3.3), (3.5), (3.8), (3.14), (3.20), (3.23), and (3.26), all of which are Petrov type II with twist except (3.26), which is a Petrov type III hypersurface-orthogonal member of the Robinson–Trautman family[7] when $m_0 = 0$ and type II without twist otherwise. Except for a special case of (3.3), all these seven metrics appear to be explicitly new.

The list is not quite complete since only particular solutions to the field equations are given in Cases II(i) and II(ii)(B), and no(nonflat) solutions have been obtained in Cases II(ii)(C)(b), (c) and IV(ii).

Apart from the intrinsic interest in metrics admitting homothetic motions, the technique of introducing this higher-order symmetry into the space–time has again proved useful in finding vacuum solutions which admit Killing vectors. If one formally puts $a = 0$ in the solutions obtained in this paper, one may pull out algebraically special vacuum solutions which admit only isometries. The process works for (3.3), (3.5), (3.14), (3.20), and (3.26), but not for (3.8) and (3.23) on account of apparent singularities. Each solution thus obtained admits at least two Killing vectors.

ACKNOWLEDGMENTS

I am indebted to Professor Roy Kerr for encouragement in this work, and to Dr. Colin McIntosh for useful discussions.

[1]W.D. Halford and R.P. Kerr, J. Math. Phys. (1979), (to be published); See also W.D. Halford, Ph.D thesis Massey University (1977), unpublished.
[2]G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[3]R.P. Kerr and G.C. Debney, J. Math. Phys. **11**, 2807 (1970).
[4]K. Yano, J. Indian Math. Soc. **15**, 105 (1951); See also T. Suguri and S. Ueno, Tensor (N.S) **24**, 253 (1972); D.M. Eardley, Commun. Math. Phys. **37**, 287 (1974); C.B.G. McIntosh, Gen. Rel. Grav. **7**, 199 (1976).
[5]J.D. Zund, Lett. Nuovo Cimento **4**, 879 (1972).
[6]W.D. Halford, J. Math. Phys. (1979) (to be published).
[7]I. Robinson and A. Trautman, Proc. Roy. Soc. A **265**, 463 (1961).

Structure of the gravitational field at spatial infinity. I. Asymptotically Euclidean spaces

S. Persides

Institute of Astronomy, University of Cambridge, Cambridge, England and Department of Astronomy, University of Thessaloniki, Thessaloniki, Greece

(Received 17 October 1978)

A new formulation for the study of the asymptotic structure of a gravitational field at spatial infinity is presented. First, the disadvantages of the existing formulations are identified in order to recognize the underlying causes and exclude them from the new formulation. It is concluded that neither conformal nor projective completion should be used. From a study of the Euclidean space we obtain a method of completion of a three-dimensional space $(\mathscr{H},\mathbf{g})$ with positive-definite metric $\mathbf{g}$ so that a two-dimensional boundary $\mathscr{L}$ is attached to the space at infinity and a three-dimensional positive-definite C^∞ metric $\hat{\mathbf{g}}$ exists near and on $\mathscr{L}$. The whole method is based on replacing the conformal transformation of the conformal completion by the relation $\Omega^{-2}g^{ij} - \Omega^{-4}g^{im}g^{jn}\Omega_{;m}\Omega_{;n} = \hat{g}^{ij} - \hat{g}^{im}\hat{g}^{jn}\Omega_{|m}\Omega_{|n}$. Thus the concept of asymptotic simplicity is defined. Then the additional conditions are determined for the space to be asymptotically Euclidean. The asymptotic symmetries and the uniqueness of the boundary are examined briefly.

I. INTRODUCTION

Let us consider a bounded source of gravity, e.g., our solar system, a binary neutron star, a globular cluster, a star undergoing gravitational collapse, etc., alone in the universe. In the framework of general relativity it seems reasonable to assume that inside or near the source the space–time is curved while as we go away from the source the curvature decreases so that at infinity we recover somehow our familiar Minkowskian geometry. To describe such a system (or more generally a system whose energy density decreases appropriately as we go to large affine distances from a central region) we introduce the concept of an asymptotically flat space–time. Naively perhaps, we picture as asymptotically flat a space–time whose metric approaches in some sense the Minkowskian metric as the affine distance from the region of large energy density becomes infinite.

To make this concept precise we have to go to infinity. Hence, because of the Lorenzian signature of the metric there are three choices: We can go to infinity along null, spacelike, or timelike directions. Thus we can study (future or past) null infinity, spatial (or spacelike) infinity, and (future or past) timelike infinity.

Modern studies of null infinity have been started essentially by Bondi, van der Burg, and Metzner,[1] Sachs,[2] and Penrose,[3] while of spatial infinity by Arnowit, Deser, and Misner[4] and Geroch.[5,6] Studies of timelike infinity, as well as attempts to unify the approaches to null and spatial infinity, have started relatively recently.[7] Thus for null and spatial infinity there is a rich collection of results, particularly remarkable given the inherent difficulties of general relativity. However, comparing the existing formulations for null and spatial infinity we can definitely say that the formulation of null infinity is conceptually clearer, aesthetically nicer, and practically better than that of spatial infinity. Thus for null infinity a precise definition (a kind of "fine tuning") in tensor as well as coordinate language has been made possible,[8] while for spatial infinity such a procedure seems impossible at this stage. This situation persists even after considerable improvements for spatial infinity by Ashtekar and Hansen[9] and Sommers.[10] Can we do anything about this situation? It seems we can. In this and the subsequent paper we shall propose a "better" formulation of the asymptotic structure at spatial infinity. However, we have first to highlight the disadvantages of the existing formulations and state clearly the final objectives. This is done in the next section. The conclusion can be summarized as follows: A better formulation (i) should not be based on the idea of conformal completion (for a three- or four-dimensional space) and (ii) should result in the definition of an unphysical metric which is C^∞ everywhere including the boundary of the space–time. In Sec. 3 we establish the new method for completing a three-dimensional manifold and introduce a generalized definition of asymptotic simplicity. In Sec. 4 we present some theorems and define rigorously the concept of asymptotically Euclidean space, that is a space with a positive-definite metric which approaches in an appropriate sense the Euclidean metric at infinity. In Sec. 5 we examine briefly the asymptotic symmetries of asymptotically Euclidean spaces and the question of uniquesness of the boundary. Finally, in Sec. 6 we present some remarks and conclusions and we comment on future developments.

In our notation Latin indices run from 1 to 3 while Greek indices from 0 to 3. A semicolon denotes covariant differentiation with respect to $\tilde{g}_{ij}$, while a vertical rule with respect to $\hat{g}_{ij}$. The special symbol $\hat{=}$ has been used instead of the equality symbol in equations which hold only on the boundary of the space. The symbol O_n denotes a quantity $O(\Omega^n)$. In a coordinate system (ω,θ,φ) in which $\Omega = \omega$ such a quantity is $O(\omega^n)$ near $\omega = 0$, its derivatives with respect to

 0022-2488/80/010135-07$01.00

θ and φ are also $O(\omega^n)$, while its derivative with respect to ω is $O(\omega^{n-1})$.

2. FORMULATION OF THE PROBLEM

To specify our objectives for spatial infinity we outline briefly the formulation of null infinity emphasizing those features which we would like to have in a formulation of spatial infinity. Then we present the existing formulations of spatial infinity in a way that underlines their deficiencies. This unfair presentation has the advantage of isolating the features which should be changed, if possible. It is expected that most of the physically meaningful results of the existing framework which are not mentioned here can be easily recast into the new formulation.

The formulation of null infinity starts with the concept of asymptotic simplicity.[3,8] This concept guarantees the existence of a boundary and defines from the physical space–time[11] ($\mathscr{M}$, g) the unphysical space–time ($\widehat{\mathscr{M}}$, $\hat{\mathbf{g}}$) with $\hat{g}_{\mu\nu} = \Omega_0{}^2 g_{\mu\nu}$, where Ω_0 is a scalar field and $\hat{g}_{\mu\nu}$ and Ω_0 are C^∞ on a neighborhood of the boundary $\mathscr{I}$ of $\widehat{\mathscr{M}}$. Further conditions specify the intrinsic structure of the boundary and how it is attached to the space–time. Thus we end up with an unphysical space–time having the following properties:

(A) The boundary has dimension $n-1$, where n is the dimension of the space (in the particular case of the physical space–time $n=4$).

(B) The unphysical metric is C^∞ on an open neighborhood of the boundary.

(C) On an open neighborhood of the boundary the unphysical metric is determined completely and uniquely from the physical metric and the scalar field Ω and, vice versa, the physical metric can be determined completely and uniquely from the unphysical metric and Ω.

In general, if we are able to compactify a manifold in a way satisfying the above conditions we will say that we have attached a *natural boundary* to the manifold or that the manifold admits a natural boundary. As naive as they seem to be these three properties are the essential reasons for the superiority of the formulation of null infinity. And they are both obtained following an idea originated by Penrose[3]: Bring infinity "metrically" close by a conformal transformation of the metric. The advantage resulting from the existence of a natural boundary is essentially the possibility to study the geometry at infinity by applying local differential geometry on the boundary, since tensor fields on the manifold can be extended smoothly to the boundary. Thus the boundary serves simultaneously as a boundary of the space–time manifold, a space where asymptotic symmetries can be studied and a manifold in its own right where asymptotic fields register. The whole formulation is simple, elegant, and practical.

For spatial infinity Geroch's formulation[5,6] of the Arnowitt–Deser–Misner approach[4] is based on the concept of asymptotically flat initial data sets and (again) Penrose's idea of conformal completion. We give here the basic formal definition in order to pinpoint the undesired features. In Sec. 4 we will show that the new formulation satisfies the conditions of this definition.

Let $\mathscr{T}$ be a three-dimensional spacelike hypersurface of the physical space–time with a smooth positive–definite metric $q_{\mu\nu}$ (the induced metric) and extrinsic curvature $p_{\mu\nu}$. The set ($\mathscr{T}$, $q_{\mu\nu}$, $p_{\mu\nu}$) is an initial data set. It is asymptotically flat at spatial infinity if there exist a three-dimensional manifold $\widetilde{\mathscr{T}}$ with a preferred point Λ and positive-definite metric $\tilde{q}_{\mu\nu}$, a scalar field Ω_0 and a diffeomorphism ψ from $\mathscr{T}$ to $\widetilde{\mathscr{T}} - \Lambda$ (which identifies $\mathscr{T}$ with $\widetilde{\mathscr{T}} - \Lambda$) such that:

(i) At Λ, $\widetilde{\mathscr{T}}$ is C^1, $\tilde{q}_{\mu\nu}$ is C^0, and Ω_0 is C^2, while all are C^∞ on $\widetilde{\mathscr{T}} - \Lambda$.

(ii) On $\mathscr{T}$, $\tilde{q}_{\mu\nu} = \Omega_0^2 q_{\mu\nu}$.

(iii) At Λ, $\Omega_0 = 0$, $\widetilde{D}_\mu \Omega_0 = 0$, $\widetilde{D}_\mu \widetilde{D}_\nu \Omega_0 = 2\tilde{q}_{\mu\nu}$ ($\widetilde{D}_\mu$ is the derivative operator with respect to $\tilde{q}_{\mu\nu}$).

(iv) The relations

$$\mathbf{p}_{\mu\nu} = \lim \Omega_0 \, p_{\mu\nu},$$

$$\Omega_{\mu\nu} = \lim \Omega_0^{-1/2} (\widetilde{D}_\mu \widetilde{D}_\nu \Omega_0 - 2\tilde{q}_{\mu\nu}),$$

$$\mathscr{R}_{\mu\nu} = \lim \Omega_0^{1/2} \cdot \widetilde{\mathscr{R}}_{\mu\nu} \quad (\widetilde{\mathscr{R}}_{\mu\nu} \text{ is the Ricci tensor of } \tilde{q}_{\mu\nu})$$

define *direction-dependent* tensors at Λ (i.e., tensors which depend on the unit tangent vector of the smooth curve we follow in going to Λ).

The first undesired features of this formulation to be noticed are its three-dimensional character, its dependence on the evolution of initial data sets, and all the inherent problems which arise from them. To eliminate this problem Ashtekar and Hansen[9] have proposed an improved formulation along the same lines. Their key idea is to attach to the space–time (which is to be regarded as asymptotically flat) a single point i^0 at spatial infinity (as a replacement of all possible Λ) and describe the asymptotic structure of the gravitational field in terms of the behavior of the various tensor fields at i^0. In particular they recast Geroch's definition in a four-dimensional language replacing essentially ($\mathscr{T}$, $q_{\mu\nu}$, $p_{\mu\nu}$) by the physical space–time, ($\widetilde{\mathscr{T}}$, $\tilde{q}_{\mu\nu}$, $\tilde{p}_{\mu\nu}$) by the unphysical space–time ($\widehat{\mathscr{M}}$, $\hat{\mathbf{g}}$) and condition (i) by the following:

(i′) At i^0, $\widehat{\mathscr{M}}$ is $C^{>1}$, $\hat{g}_{\mu\nu}$ is $C^{>0}$ and Ω is C^2, while all are C^∞ on $\widehat{\mathscr{M}} - i^0$.

($C^{>n}$ states essentially that the differential structure of the object is such that its derivatives of order $n+1$ have direction-dependent limits at i^0.) The condition (iv) is not needed.

Although this formulation is a definite improvement (since it is four-dimensional in character and global problems of evolution do not arise), it makes more acute the remaining deficiencies of the original formulation: the awkward differential structure of the manifold and the metric at i^0, the dependence of tensor fields on the direction of approach to i^0, the single-point spatial boundary (although the space is four-dimensional), the inability to set up at i^0 a coordinate system in which everything behaves nicely, the fact that all asymptotic properties of spatial infinity register at a single point, etc. To study further spatial infinity Geroch introduces a three-dimensional manifold $\mathscr{S}_G$ (consisting in his formulation of all "points" at spatial infinity) with an appropriate metric. Ashtekar and Hansen propose a suitable "blowing up" of i^0. Their result is a four-dimensional manifold called Spi constructed from various inextendible spacelike curves "regular" at i^0. But both these manifolds are somehow artificial and neither of them is a boundary surface of the space–time itself. Thus, e.g., it is meaningless to ask whether or not a curve of the physical space–time $\mathscr{M}$ has an end point on $\mathscr{S}_G$ or Spi, or whether a tensor field defined on $\mathscr{M}$ has a limit on $\mathscr{S}_G$ or Spi.

The above raise many puzzling questions: Why do we have to worry about differentiability class, something which almost has no physical significance[6]? Is it reasonable to model the space–time in such a way that going to infinity along different directions implies ending up at the same point? Is spatial infinity represented faithfully by a single point on the Penrose diagram[3]? Is it reasonable to "shrink" spatial infinity to a point and then "blow" it up? Why are we restricted to consider only the structure of first and second order (metric and its first derivatives) at i^0? The conclusion is inescapable: *All these are inevitable results of the one-point compactification. The conformal completion shrinks spatial infinity too much. If we want to avoid the previous undesired features, we have to abandon the method of conformal completion.*

The basic idea in Sommers' approach[10] is to attach to an asymptotically flat space–time at spatial infinity a three-dimensional boundary $\mathscr{P}$ using projective rather than conformal completion. The boundary $\mathscr{P}$ is assumed to be the unit timelike hyperboloid of the Minkowski metric. Instead of Ω_0 a scalar field Σ on $\mathscr{M}$ is used (Σ is essentially the inverse of a spacelike distance) to define the Σ-foliation, that is the family of timelike hypersurfaces $\Sigma = \text{const}\neq 0$, with known intrinsic properties. This family suggests an intrinisic metric for the limiting case $\Sigma = 0$ which represents $\mathscr{P}$. Attaching this surface $\mathscr{P}$ to the manifold we have a new manifold $\bar{\mathscr{M}} = \mathscr{M}\cup\mathscr{P}$. Finally, it is asked that the metric and the negative of the extrinsic curvature of a $\Sigma = \text{const}$ hypersurface tend in a continuous way to the metric of $\mathscr{P}$ as $\Sigma \to 0$.

This second approach (i.e., Sommers' formulation) does not have any of the deficiencies of the first approach (i.e., the Arnowitt–Deser–Misner–Geroch–Ashtekar–Hansen formulation). No awkward differentiability requirements are needed, no direction-dependent limits are to be taken. It has, however, another perhaps more serious deficiency: There is no four-dimensional metric on the boundary $\mathscr{P}$. Thus to study the physical fields near $\mathscr{P}$ we have to decompose them into their tangential and normal components with respect to the Σ-foliation and treat each component differently. A spacelike curve which defines a point p on $\mathscr{P}$ is not spacelike or anything else at p simply because there is no 4-metric there. Thus we reach the following conclusion: *The projective completion results in a 4-metric which diverges on the spatial boundary. If we want a metric regular everywhere, we have to abandon the projective completion.*

In summary, the first approach seems to be more practical (more fruitful) while the second more simple and elegant. Of the requirements for a natural boundary the first approach satisfies completely (C), partially (B) (it gives a metric $C^{>0}$ at i^0), and violates (A). The second approach satisfies completely (A) and (C) (with $\bar{g}_{\mu\nu} = \Sigma^2 g_{\mu\nu}$), but violates flagrantly (B). Is there any approach that will have the advantages of both and none of their disadvantages? Is there any way of attaching a natural boundary to the space–time? The objective of this and the subsequent paper[12] is to provide such a framework. As an extra (and nonanticipated) bonus we get a "unified" formulation which applies to null as well as to spatial infinity. Before, however, we start to formulate the new approach let us give two of the final results of this paper: The *new* unphysical metric for the three-dimensional Euclidean space is

$$\hat{h}_{ij} = \operatorname{diag}[(1-\omega^2)^{-1}, 1, \sin^2\theta] \tag{1}$$

while for any three-dimensional $t = \text{const}$ submanifold of the Schwarzschild space–time is ($r_s = 2Gmc^{-2}$)

$$\hat{g}_{ij} = \operatorname{diag}[(1-\omega^2 + r_s\,\omega^3)^{-1}, 1, \sin^2\theta] \tag{2}$$

with $\omega = 0$ on the boundary.

3. THE UNPHYSICAL METRIC

In the previous section we have formulated the problem in a rather negative way: We specified what deficiencies the new formulation should not have. In this section we give a positive contribution: We specify how the new unphysical metric should be determined. To discover that "how" we have to be faithful to the old principle that in general relativity "metric is the foundation of all."

The fact that the awkward conditions in Geroch's formulation of the Arnowitt–Deser–Misner study appear from the beginning (that is, when we are dealing with asymptotically flat initial data sets) shows that the bad behavior, is not only a property of four-dimensional space–times with Lorentzian signature but also of the innocent three-dimensional Euclidean space. Thus we set ourselves an apparently simpler problem: Define the concept of an asymptotically Euclidean space. But even that seems difficult. So we consider as a first step an even simpler problem: Complete the three-dimensional Euclidean space with a natural boundary.

In coordinates (r,θ,φ) the metric of E^3 gives the line element

$$dS^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2). \tag{3}$$

We consider as "infinity" whatever we reach by leaving $r\to\infty$ with θ and φ constant. Let us bring this "infinity" close at least in terms of coordinates. This is simple and can be done by a transformation of coordinates, e.g., $\rho = r(r+1)^{-1}$ and θ and φ unchanged. The whole space is now mapped on the interior of the sphere $0\leqslant\rho<1$. We consider the surface $\rho = 1$ as representing infinity and denoted by $\mathscr{L}$. A further transformation $\rho' = 1-\rho = (r+1)^{-1}$ sends $\rho = 1$ (that is $r = \infty$) to $\rho' = 0$. Since ρ' behaves as r^{-1} near $\rho' = 0$, the simplest transformation which will bring infinity to $\omega = 0$ is $\omega = r^{-1}$. From (3) we have the components of the physical metric tensor in coordinates (ω,θ,φ)

$$h_{ij} = \operatorname{diag}[\omega^{-4}, \omega^{-2}, \omega^{-2}\sin^2\theta]. \tag{4}$$

It is clear that if we want to complete conformally this space we have to multiply h_{ij} by Ω_0^2, where Ω_0 behaves as ω^2 (or r^{-2}) near $\omega = 0$. Such an action will result in "excessive shrinking": While the area of a surface $r = \text{const}$ tends originally to ∞ as $r\to\infty$, after a conformal completion will tend to zero. We can, however, set $\Omega = \omega = r^{-1}$ and define the *conformal metric*

$$\tilde{h}_{ij} = \Omega^2 h_{ij} = \mathrm{diag}[\omega^{-2}, 1, \sin^2\theta]. \tag{5}$$

Thus we have "shrinked" the surface $r = \text{const}$ but not too much: Its area is constant and remains constant as $r \to \infty$.[13] However, $\tilde{h}_{ij}$ still diverges on $\mathscr{L}$.

Equation (5) suggests two things. First, since every surface $\Omega = \text{const} \neq 0$ has intrinsic metric diag $[1, \sin^2\theta]$, we should give somehow the same intrinsic metric to the boundary $\mathscr{L}$.[14] Second, since $\tilde{h}_{11}$ diverges on $\mathscr{L}$ while $\tilde{h}_{22}$ and $\tilde{h}_{33}$ behave nicely there, we should find a way to treat $\tilde{h}_{11}$ "differently." Here we face the essential and perhaps the only difficult part of the problem, whether we deal with Euclidean space, Minkowskian space–time, or a general asymptotically flat space–time. How can we treat $\tilde{h}_{11}$ differently using only tensor relations? To answer this question let us consider the contravariant form of the conformal metric

$$\tilde{h}^{ij} = \Omega^{-2} h^{ij} = \mathrm{diag}[\omega^2, 1, \sin^2\theta]. \tag{6}$$

If $\hat{h}^{ij}$ is the contravariant form of the *unphysical metric* we must have $\hat{h}^{11} \neq 0$ on $\mathscr{L}$. A change in the scale of ω will give $\hat{h}^{11} = 1$. We observe now that the desired behavior of $\hat{h}^{ij}$ on $\mathscr{L}$ can be obtained *if we add* 1 *to* $\tilde{h}^{11}$ *only*, that is if we write

$$\hat{h}^{ij} \mathrel{\hat{=}} \tilde{h}^{ij} + \hat{h}^{im}\hat{h}^{jn}\delta^1_m\delta^1_n. \tag{7}$$

But this relation can be easily written in tensor form as

$$\tilde{h}^{ij} \mathrel{\hat{=}} \hat{h}^{ij} - \hat{h}^{im}\hat{h}^{jn}\Omega_{|m}\Omega_{|n}. \tag{8}$$

We have determined $\hat{h}^{ij}$ on $\mathscr{L}$ but not on E^3. Obviously (8) can be written in the form

$$\tilde{h}^{ij} + X^{ij} = \hat{h}^{ij} - \hat{h}^{im}\hat{h}^{jn}\Omega_{|m}\Omega_{|n}, \tag{9}$$

where $X^{ij} \mathrel{\hat{=}} 0$. What is the best choice for X^{ij}? It appears that the answer to this question is $X^{ij} = -\tilde{h}^{im}\tilde{h}^{jn}\Omega_{;m}\Omega_{;n}$. Thus (6) becomes

$$\hat{h}^{ij} - \hat{h}^{im}\hat{h}^{jn}\Omega_{|m}\Omega_{|n} = \tilde{h}^{ij} - \tilde{h}^{im}\tilde{h}^{jn}\Omega_{;m}\Omega_{;n}. \tag{10}$$

This is the relation which determines $\hat{h}^{ij}$ from $\tilde{h}^{ij}$ or h^{ij}. Essentially it replaces the conformal transformation of the metric in the conformal completion. Of course many other relations have been tried. None of them exhibited the nice and fruitful properties of (10) which will become apparent in this and the next paper.

Before we proceed any further it is appropriate to give here the definition of asymptotic simplicity at spatial infinity.

Definition: A pair $(\mathscr{H}, \mathbf{g})$ of a three-dimensional manifold $\mathscr{H}$ (without boundary) with a positive-definite metric $\mathbf{g}$ is *asymptotically simple* (at spatial infinity) iff there exist:

(a) A (three-dimensional) space $\hat{\mathscr{H}}$ with a two-dimensional boundary $\mathscr{L}$ ($\mathscr{L} \subset \hat{\mathscr{H}}$) and a metric $\hat{\mathbf{g}}$ positive definite and C^∞ on an open neighborhood $\hat{U}$ of $\mathscr{L}$ ($\mathscr{L} \subset \hat{U} \subset \hat{\mathscr{H}}$).

(b) A C^∞ scalar field Ω on $\hat{U}$, positive on $\hat{U} - \mathscr{L}$ and zero on $\mathscr{L}$, such that[15]

$$\hat{g}^{ij} - \hat{g}^{im}\hat{g}^{jn}\Omega_{|m}\Omega_{|n} = \Omega^{-2}g^{ij} - \Omega^{-4}g^{im}g^{jn}\Omega_{;m}\Omega_{;n}. \tag{11}$$

Thus asymptotic simplicity is essentially a guarantee that the space $\mathscr{H}$ can be imbedded in $\hat{\mathscr{H}}$ with a boundary $\mathscr{L}$ in such a way that the first two conditions for $\mathscr{L}$ being a natural boundary are satisfied. As it will turn out in the next paper this definition serves its purpose for a locally Minkowskian space–time too, if we drop the term "positive-definite" and consider $i, j, \cdots$ as taking values from 0 to 3. Furthermore, we will see that we can easily generalize this definition to cover null infinity by dropping from it the term "spatial infinity" while keeping (11) as it is. At present, however, we have to find a condition such that condition (C) for a natural boundary is satisfied. For this the following theorem will help us.

Theorem 1: If an asymptotically simple space satisfies the conditions $\Omega^{-2}\tilde{g}^{ij}\Omega_{;i}\Omega_{;j} \mathrel{\hat{=}} \hat{g}^{ij}\Omega_{|i}\Omega_{|j} \mathrel{\hat{=}} 1$, then in a coordinate system (ω,θ,φ) in which $\Omega = \omega$ we have $(A,B = 2,3)$

$$\hat{g}^{11} = 1 - \tilde{g}^{11}, \tag{12}$$

$$\hat{g}^{1A} = \tilde{g}^{1A}(1/\tilde{g}^{11} - 1), \tag{13}$$

$$\hat{g}^{AB} = \tilde{g}^{AB} + \tilde{g}^{1A}\tilde{g}^{1B}[(1/\tilde{g}^{11} - 1)^2 - 1]. \tag{14}$$

Proof: In coordinates (ω,θ,φ) with $\Omega = \omega$ we have $\Omega_{;m} = \Omega_{|m} = \delta^1_m$. Hence (11) becomes

$$\hat{g}^{11} - \hat{g}^{i1}\hat{g}^{j1} = \tilde{g}^{ij} - \tilde{g}^{i1}\tilde{g}^{j1}. \tag{15}$$

For $i = j = 1$ this gives a second degree algebraic equation with solutions $\hat{g}^{11} = \tilde{g}^{11}$ and $\hat{g}^{11} = 1 - \tilde{g}^{11}$. The first is rejected because it does not satisfy the conditions $\Omega^{-2}\tilde{g}^{ij}\Omega_{;i}\Omega_{;j} \mathrel{\hat{=}} \hat{g}^{ij}\Omega_{|i}\Omega_{|j} \mathrel{\hat{=}} 1$. Thus we have only the second solution, that is (12). For $i = 1, j = A$ and $i = A, j = B$ we have from (12) the other two relations (13) and (14). It should be noted that in this coordinate system we obtain $\tilde{g}^{ij}$ from $\hat{g}^{ij}$ by simply interchanging $\tilde{g}^{ij}$ and $\hat{g}^{ij}$.

An immediate consequence of this theorem is that if we want condition (C) for an asymptotically simple space to be satisfied, it is enough to assume that $\Omega^{-2}\tilde{g}^{ij}\Omega_{;i}\Omega_{;j} \mathrel{\hat{=}} \hat{g}^{ij}\Omega_{|i}\Omega_{|j} \mathrel{\hat{=}} 1$. Thus we have the following theorem.

Theorem 2. An asymptotically simple space which satisfies the conditions $\Omega^{-2}\tilde{g}^{ij}\Omega_{;i}\Omega_{;j} \mathrel{\hat{=}} \hat{g}^{ij}\Omega_{|i}\Omega_{|j} \mathrel{\hat{=}} 1$ admits a natural boundary.

Before we define in the next section the concept of an asymptotically Euclidean space it is useful to summarize the essence of the proposed changes. The most important point is the replacement of the relation $\hat{g}^{ij} = \Omega_0{}^2 g^{ij}$ (conformal completion) by the relation (11). They are both arbitrary and artificial relations in the sense that they have been invented by us to register at infinity information about the structure of the space–time. They are not suggested by the space–time itself. The best is that which is more convenient, more practical, more useful. At spatial infinity this is relation (11), although $\hat{g}_{ij} = \Omega_0{}^2 g_{ij}$ seems at first glance simpler. Finally, it should be noted that the behavior of Ω as r^{-1} (instead of Ω_0 as r^{-2}) could have been anticipated from Geroch's formulation of asymptotic flatness at spatial infinity. In that formulation $\Omega_0^{1/2}$ appears too often to be pure coincidence.

4. ASYMPTOTICALLY EUCLIDEAN SPACES

The original, naive but basic idea for an asymptotically flat space is that somehow the metric approaches the flat metric as we go further and further away from the source. Thus we expect that the asymptotic behavior of the metric is described somehow by (3)–(6). Perhaps the weaker condition covering this case is that an asymptotically Euclidean space should admit a suitable completion with a natural boundary similar to the boundary $\mathscr{L}$ of the Euclidean space. The model suggested for the latter in the previous section is a two-dimensional manifold isometric to the unit two sphere S^2, a manifold on its own right. Thus in addition to asymptotic simplicity we should impose this boundary to the space.[16]

A definition of a particular kind of space, e.g., of flat or asymptotically flat space, should be always given, if possible, in terms of tensor conditions as well as in terms of the existence of a coordinate system in which the metric has a particularly simple and useful form. The following theorem will help us to do so in the definition of an asymptotically Euclidean space.

Theorem 3: For an asymptotically simple space the following conditions are equivalent:

(a) The boundary $\mathscr{L}$ is isometric to S^2 and

$$\Omega^{-2}\tilde{g}^{ij}\Omega_{;i}\,\Omega_{;j} \mathrel{\widehat{=}} \hat{g}^{ij}\Omega_{|i}\,\Omega_{|j} \mathrel{\widehat{=}} 1. \tag{16}$$

(b) There exists a coordinate system (ω,θ,φ) in which on $\hat{U}$ (an open neighborhood of $\mathscr{L}$) we have $\Omega=\omega$, $\tilde{g}^{11}=\omega^2+O_3$ and (λ, μ arbitrary functions of θ, φ)

$$\hat{g}_{ij}=\begin{bmatrix} 1+\lambda^2+\mu^2\sin^{-2}\theta+O_1 & \lambda+O_1 & \mu+O_1 \\ \lambda+O_1 & 1+O_1 & O_1 \\ \mu+O_1 & O_1 & \sin^2\theta+O_1 \end{bmatrix}. \tag{17}$$

(c) There exists a coordinate system (ω,θ,φ) in which on $\hat{U}$ we have $\Omega=\omega$, $\tilde{g}^{11}=\omega^2+O_3$, and

$$\hat{g}^{ij}=\begin{bmatrix} 1+O_1 & -\lambda+O_1 & -\mu\sin^{-2}\theta+O_1 \\ -\lambda+O_1 & 1+\lambda^2+O_1 & \lambda\mu\sin^{-2}\theta+O_1 \\ -\mu\sin^{-2}\theta+O_1 & \lambda\mu\sin^{-2}\theta+O_1 & \sin^{-2}\theta+\mu^2\sin^{-2}\theta+O_1 \end{bmatrix}. \tag{18}$$

(d) There exists a coordinate system (ω,θ,φ) in which on $\hat{U}$ we have $\Omega=\omega$, $\hat{g}^{11}=1+O_1$, and

$$\tilde{g}_{ij}=\begin{bmatrix} \omega^{-2}+O_{-1} & \lambda+O_1 & \mu+O_1 \\ \lambda+O_1 & 1+O_1 & O_1 \\ \mu+O_1 & O_1 & \sin^2\theta+O_1 \end{bmatrix}. \tag{19}$$

(e) There exists a coordinate system (ω,θ,φ) in which on $\hat{U}$ we have $\Omega=\omega$, $\hat{g}^{11}=1+O_1$, and

$$\tilde{g}^{ij}=\begin{bmatrix} \omega^2+O_3 & -\lambda\omega^2+O_3 & -\mu\sin^{-2}\theta\cdot\omega^2+O_3 \\ -\lambda\omega^2+O_3 & 1+O_1 & O_1 \\ -\mu\sin^{-2}\theta\cdot\omega^2+O_3 & O_1 & \sin^{-2}\theta+O_1 \end{bmatrix}. \tag{20}$$

Proof: Let (a) be true. Then since the space is asymptotically simple and $\mathscr{L}$ is isometric to S^2, there is a coordinate system (ω,θ,φ) in which $\Omega=\omega$ and on an open neighborhood of $\mathscr{L}$

$$\hat{g}^{ij}=\begin{bmatrix} \kappa+O_1 & \lambda+O_1 & \mu+O_1 \\ \lambda+O_1 & 1+O_1 & O_1 \\ \mu+O_1 & O_1 & \sin^2\theta+O_1 \end{bmatrix}. \tag{21}$$

From (16) we have $\hat{g}^{11}\mathrel{\widehat{=}}1$, $\omega^{-2}\tilde{g}^{11}\mathrel{\widehat{=}}1$ which give $\kappa=1+\lambda^2+\mu^2\sin^{-2}\theta$ and $\tilde{g}^{11}=\omega^2+O_3$. Hence (a) implies (b). Inverting (17) we have (18). Hence (a) implies (c). From (12)–(14) we find (20). Hence (a) imples (e). Inverting (20) we find (19). Hence (a) implies (d). Thus (a) implies (b), (c), (d), (e). Let now (b) be true. Then $\mathscr{L}$ is isometric to S^2. From (17) we find $\hat{g}^{11}=1+O_1$, which means that (16) is satisfied. Hence (b) implies (a) and consequently (c), (d), and (e). If (c) is true, then inverting (18) we obtain (17), etc. If (d) is true we obtain (20) and then use the relations (12)–(14). Thus we find (18), etc.

We are now ready to define an asymptotically Euclidean space.

Definition: A pair $(\mathscr{H},\mathbf{g})$ of a three-dimensional manifold $\mathscr{H}$ (without boundary) with a positive-definite metric $\mathbf{g}$ is *asymptotically Euclidean* iff it is asymptotically simple and satisfies one of the conditions (a)–(e) of the previous theorem.

A word of caution should be added to the above definition. As it is , the definition contains only the minimum requirements for a space to resemble asymptotically the Euclidean space. There is a rather small probability that usefullness will suggest at a later stage the addition of other conditions which hold in Euclidean space, such as $R_{ij}=0$ near $\mathscr{L}$. However, in the case of a space–time with Lorentzian signature it is quite possible that some additional conditions will be needed in order that certain physical quantities can be defined on the boundary.

To familiarize ourselves with the new formulation and increase our confidence let us determine $\hat{g}_{ij}$ for a few well-known spaces which must be asymptotically Euclidean whatever the definition is. From (12)–(14) it is obvious that (in coordinates $\Omega = \omega,\theta,\varphi$) if $\tilde{g}^{ij}$ is diagonal, then $\hat{g}^{ij}$ is also diagonal with $\hat{g}^{ij} = \tilde{g}^{ij}$ except $\hat{g}^{11} = 1 - \tilde{g}^{11}$. Thus for the Euclidean metric (3)–(6) we find (1). For the spatial metric ($\omega = r^{-1}$, $r_s = 2Gmc^{-2}$)

$$g_{ij} = \mathrm{diag}[(1 - r_s\,\omega)^{-1}, \omega^{-2}, \omega^{-2}\sin^2\theta] \tag{22}$$

of the Schwarzschild space–time we find (2). For the spatial metric

$$g_{ij} = \mathrm{diag}[\rho^2 \Delta^{-1}\omega^{-4},\ \rho^2,\ \omega^{-2}\sin^2\theta + (1 + r_s\omega^{-1}\rho^{-2}\sin^2\theta)\alpha^2\sin^2\theta] \tag{23}$$

of the Kerr space–time with $\rho^2 = \omega^{-2} + a^2\cos^2\theta$, $\Delta = \omega^{-2} - r_s\omega^{-1} + \alpha^2$ we find

$$\hat{g}_{ij} = \mathrm{diag}[(1 - \rho^{-2}\Delta\omega^2)^{-1}, \rho^2\omega^2, \sin^2\theta + (\omega^2 + r_s\omega\rho^{-2}\sin^2\theta)\alpha^2\sin^2\theta]. \tag{24}$$

For the spatial metric

$$g_{ij} = \mathrm{diag}[\omega^{-4}e^{2(\beta-\alpha)}, \omega^{-2}e^{2(\beta-\alpha)}, \omega^{-2}\sin^2\theta\cdot e^{-2\alpha}] \tag{25}$$

of the Weyl solution we find

$$\hat{g}_{ij} = \mathrm{diag}[(1 - \omega^2 e^{2(\alpha-\beta)})^{-1}, e^{2(\beta-\alpha)}, \sin^2\theta\cdot e^{-2\alpha}]. \tag{26}$$

It is obvious that all $\hat{g}_{ij}$ of (1), (2), (24), (26) are C^∞ at $\omega = 0$. Finally, we observe that if from (19) we obtain $g_{ij} = \omega^{-2}\tilde{g}_{ij}$, transform it to coordinates (r,θ,φ) and then to coordinates (x,y,z) (with the usual transformation from spherical to Cartesian coordinates) we find a metric of the form $\mathrm{diag}(1,1,1) + O_1$.[4,17] So, after all, the original naive assumption was not wrong.

Before we close this section it is useful to see if there is any relation between the new definition and Geroch's definition of an asymptotically flat initial data set. In fact we can prove the following:

Theorem 4: An asymptotically Euclidean space satisfies all Geroch's conditions which refer only to the intrinsic geometry of the three-dimensional hypersurface $\mathscr{T}$.

Proof: The conformal factor for the conformal completion is $\Omega_0 = \Omega^2$. Hence in coordiantes (ω,θ,φ) with $\Omega = \omega$ we have the "conformal" metric of the conformal completion" $\tilde{q}_{ij} = \Omega_0^2 g_{ij} = \Omega^2\tilde{g}_{ij} = \omega^2\tilde{g}_{ij}$ and from (19)

$$\tilde{q}_{ij} = \begin{bmatrix} 1 + O_1 & O_2 & O_2 \\ O_2 & \omega^2 + O_3 & O_3 \\ O_2 & O_3 & \omega^2\sin^2\theta + O_3 \end{bmatrix}. \tag{27}$$

Obviously $\tilde{q}_{ij}$ is C^0 at Λ (that is for $\omega = 0$) and Ω_0 is C^2. Hence condition (i) (Sec. 2) is satisfied. Note, however, that $\tilde{q}_{11,1} = O_0$ which depends on θ and φ at $\omega = 0$. Hence $\tilde{q}_{ij}$ is not C^1 at Λ. Also $\Omega_0 = \omega^2 = 0$ and $\widetilde{D}_i\,\Omega_0 = 2\omega\delta_i^1 = 0$ at Λ. Calculating the Christoffel symbols we find $\widetilde{\Gamma}^1{}_{ij} = O_0$ and

$$\widetilde{D}_i\,\widetilde{D}_j\Omega_0 - 2\tilde{q}_{ij} = -2\omega\widetilde{\Gamma}^1{}_{ij} = 0$$

at Λ. Hence condition (iii) is satisfied. Further calculations show that $\mathbf{\Omega}_{ij}$ and $\mathbf{R}_{ij}$ of condition (iv) are direction-dependent tensors at Λ. The condition on the extrinsic curvature cannot be checked since it does not refer only to the intrinsic geometry of $\mathscr{T}$. Note that we have also proved condition (i') of Ashtekar and Hansen that $\tilde{q}_{ij}$ is $C^{>0}$ at Λ.

5. ASYMPTOTIC SYMMETRIES AND UNIQUENESS

In this section we consider two somehow related problems which arise naturally for any asymptotically flat space. The first refers to the group of transformations which leave the asymptotic behavior of the space unchanged. This group is intimately related to the physical quantities which register on the boundary and are conserved in some sense in the four-dimensional case. The second refers to the uniqueness of the boundary. It is important to know to what extent the boundary is determined by the space $(\mathscr{H},\mathrm{g})$, since otherwise we cannot know whether our statements express properties of $(\mathscr{H},\mathrm{g})$ or of the procedure in constructing the boundary.

The group of asymptotic symmetries can be defined as the group of coordinate transformations which preserve the asymptotic form of the physical metric. From (19) the asymptotic form of the physical metric $g_{ij} = \Omega^{-2}\tilde{g}_{ij}$ is

$$g_{ij} = \begin{bmatrix} \omega^{-4} + O_{-3} & O_{-2} & O_{-2} \\ O_{-2} & \omega^{-2} + O_{-1} & O_{-1} \\ O_{-2} & O_{-1} & \omega^{-2}\sin^2\theta + O_{-1} \end{bmatrix} \tag{28}$$

Assuming a transformation $(\omega,\theta,\varphi)\to(\omega',\theta',\varphi')$ of the form

$$\omega = \omega_1\omega' + O_2, \tag{29}$$

$$\theta = \theta_0 + \theta_1\omega' + O_2, \tag{30}$$

$$\varphi = \varphi_0 + \varphi_1\omega' + O_2, \tag{31}$$

where $\theta_0, \varphi_0, \omega_1, \theta_1, \varphi_1$, etc., are functions of θ and φ, we calculate the physical metric g'_{ij} in coordinates $(\omega',\theta',\varphi')$. Demanding that each component of g'_{ij} starts with the same term or power of ω' as the corresponding component of g_{ij} we find that the transformation (29)–(31) should satisfy the conditions[18,19]

$$\left(\frac{\partial\theta_0}{\partial\theta'}\right)^2 + \left(\frac{\partial\varphi_0}{\partial\theta'}\right)^2\sin^2\theta_0 = 1, \tag{32}$$

$$\left(\frac{\partial\theta_0}{\partial\varphi'}\right)^2 + \left(\frac{\partial\varphi_0}{\partial\varphi'}\right)^2\sin^2\theta_0 = \sin^2\theta', \tag{33}$$

$$\frac{\partial\theta_0}{\partial\theta'}\frac{\partial\theta_0}{\partial\varphi'} + \frac{\partial\varphi_0}{\partial\theta'}\frac{\partial\varphi_0}{\partial\varphi'}\sin^2\theta_0 = 0, \tag{34}$$

and $\omega_1 = 1$ (we have assumed that on $\mathscr{H}$ we have $\omega > 0$ and $\omega' > 0$). Hence the group of asymptotic symmetries is the group which leaves the intrinsic metric $d\theta^2 + \sin^2\theta\,d\varphi^2$ of $\mathscr{L}$ unchanged.

The meaning of the term "unique boundary"[6] is not yet settled in the literature, since a definition of uniqueness alone is not enough: It should be accompanied by a theorem which will establish its usefulness. To compare two boundaries $\mathscr{L}_1$ and $\mathscr{L}_2$ corresponding to conformal factors Ω_1 and Ω_2 we require first that they refer to the same asymptotic region[20] by asking that $\hat{U}_1\cap\hat{U}_2\cap\mathscr{H}\neq 0$ for arbitrary neighborhoods $\hat{U}_1$ and $\hat{U}_2$ of $\mathscr{L}_1$ and $\mathscr{L}_2$, respectively. Next we ask whether or not the conformal metric on the boundary depends on the conformal factor we have chosen. In Geroch's formulation the values of $\tilde{q}_{ij}$ at Λ do not depend on the choice of Ω_0, as it is obvious from (27). This is not true, however, for $\hat{g}_{ij}$ of (17), because of λ and μ which carry information hidden in the higher orders of $\tilde{g}_{ij}$. Nevertheless, there are two encouraging points. First, for the transformation (29)–(31) we found $\omega_1 = 1$. Thus ω and ω' or Ω and Ω' are equal to first order. Second, the formulation contains the relations (16) which limit the scale of Ω (they act as gauge conditions). In any case it would be very useful to prove that Ω_1 and Ω_2 give compatible asymptotes in the sense of Geroch[6] and that compatible asymptotes are equivalent.

6. CONCLUSIONS

The contribution of this paper can be summarized in one phrase as follows: We have determined from g_{ij} and Ω a unique metric $\hat{g}_{ij}$ which carries all the information of g_{ij} and is C^∞ on a neighborhood of a two-dimensional boundary attached to the space itself. It is obvious that this formulation does not have the disadvantages of the other formulations, although it appears to have the right strength (i.e., it covers as many spaces as it should be covering) as the specific examples [Eqs. (1), (2), (24), (26)] and Theorem 4 show. We are free to exploit the C^∞ structure of $\hat{\mathscr{H}}$ and $\hat{g}_{ij}$ at infinity. Explicit calculations can be carried out on $\mathscr{L}$. To study the gravitational field itself we can use $\hat{g}_{ij}$ or $\tilde{g}_{ij}$ or g_{ij}. Every direction-dependent field at Λ appears to become a smooth tensor field on $\mathscr{L}$. Thus a (static) electric field E_i after multiplication by the appropriate power of Ω will register as a tensor field on $\mathscr{L}$ and its surface integral will give the total charge of the space. Furthermore, the smoothness on $\mathscr{L}$ will probably facilitate a definition of multipole moments for a gravitational field with no Killing vectors.[21] Higher dimensional asymptotically Euclidean spaces can be studied along the same lines.

Other advantages of this formulation will become clearer in the next paper. These include an automatic unification of the formulations at null and spatial infinity, a possibility of extending this procedure to timelike infinity, etc. In the case of a four-dimensional space–time the real problem will be to determine the additional conditions needed on the boundary. Quite probably conditions of the form $R_\mu{}^\nu = 0$ or $R_\mu{}^\nu = O_n$ will not be appropriate and some delicate requirements will be needed on the Weyl tensor. In that respect the results and the experience gained from the other formulations will be extremely useful.

Does the formulation presented here have any new disadvantages? It appears it has only one. Because of the nonlinear (although algebraic) relation (11) between g_{ij} and $\hat{g}_{ij}$ differential equations on $(\mathscr{H},g)$ cannot be transformed easily to differential equations on $(\hat{\mathscr{H}},\hat{g})$. Even the zero order terms of $\hat{g}_{ij}$ on $\mathscr{L}$ [Eqs. (17) and (18)] are complicated and this is expected to make the calculations difficult. Is there any chance that a physical condition will simplify the situation implying, e.g., $\lambda = \mu = 0$? In any case the new difficulties are only calculational and require some extra effort from us. It seems that at spatial infinity we have to put more effort than at null infinity to discover physically less important quantities.

ACKNOWLEDGMENTS

Most of this work has been carried out at the Institute of Astronomy of the University of Cambridge, where the author was a Senior Visiting Fellow, and it has been partially supported by the U.K. Science Research Council. The hospitality and support of the members and the staff of the Institute of Astronomy is gratefully acknowledged.

[1] H. Bondi, M.G.J. van der Burg, and A.W.K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962).
[2] R.K. Sachs, Proc. R. Soc. London Ser. A **270**, 103 (1962).
[3] R. Penrose, Phys. Rev. Lett. **10**, 66 (1963); Proc. R. Soc. London Ser. A **284**, 159 (1965).
[4] R. Arnowitt, S. Deser, and C.W. Misner, Phys. Rev. **117**, 1595 (1960); **118**, 1100 (1960); **121**, 1556 (1961); **122**, 997 (1961); the article in *Gravitation, an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[5] R. Geroch, J. Math. Phys. **13**, 956 (1972).
[6] A recent review for both null and spatial infinity is given by R. Geroch, in *Asymptotic Structure of Space-Time*, edited by F.P. Esposito and L. Witten (Plenum, New York, 1977).
[7] D. Eardley and R.K. Sachs, J. Math. Phys. **14**, 209 (1973); B.G. Schmidt, Commun. Math. Phys. **36**, 73 (1974).
[8] S. Persides, J. Math. Phys. **20**, 1731 (1979).
[9] A. Ashtekar and R.O. Hansen, J. Math. Phys. **19**, 1542 (1978).
[10] P. Sommers, J. Math. Phys. **19**, 549 (1978).
[11] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U. P., Cambridge, 1973).
[12] After completing this paper I discovered a remarkable, almost prophetic, question raised by Geroch in p. 100 of his 1977 review (Ref. 6): "Is there any formulation which combines the best of both?" This and the subsequent paper are essentially an answer to that question.
[13] The conformal completion can be thought of as a two step process: multiplication by r^{-2} and then again multiplication by r^{-2}. We take only the first step. The second is replaced by the determination of $\hat{g}_{ij}$ from $\tilde{g}_{ij}$.
[14] Since $\mathscr{L}$ is not contained in the original space E^3, there is no meaning of "induced metric" on it.
[15] It is enough that $\hat{\mathscr{H}}$ is C^4 and g_{ij} C^3. For simplicity we assume both C^∞.
[16] The idea of imposing *first* the intrinsic metric on the boundary has been used by Sommers and Persides (Refs. 10 and 8).
[17] R. Geroch, J. Math. Phys. **11**, 2580 (1970).
[18] S. Persides, Prog. Theor. Phys. **58**, 829 (1977).
[19] S. Persides, J. Gen. Rel. Grav. **10**, 609 (1979).
[20] A. Ashtekar, "Asymptotic Structure of the Gravitational Field at Spatial Infinity" (preprint).
[21] R.O. Hansen, J. Math. Phys. **15**, 46 (1974).

Structure of the gravitational field at spatial infinity. II. Asymptotically Minknowskian space–times

S. Persides
Institute of Astronomy, University of Cambridge, Cambridge, England and Department of Astronomy, University of Thessaloniki, Thessaloniki, Greece

(Received 13 November 1978)

A new formulation is established for the study of the asymptotic structure at spatial infinity of asymptotically Minkowskian space–times. First, the concept of an asymptotically simple space–time at spatial infinity is defined. This is a (physical) space–time $(\mathscr{M},\mathbf{g})$ which can be imbedded in an unphysical space–time $(\widehat{\mathscr{M}},\hat{\mathbf{g}})$ with a boundary $\mathscr{S}$, a C^∞ metric $\hat{\mathbf{g}}$ and a C^∞ scalar field Ω such that $\Omega = 0$ on $\mathscr{S}$, $\Omega > 0$ on $\widehat{\mathscr{M}} - \mathscr{S}$, and $\hat{g}^{\mu\nu} + \hat{g}^{\mu\lambda}\hat{g}^{\nu\rho}\Omega_{|\lambda}\Omega_{|\rho} = \Omega^{-2}g^{\mu\nu} + \Omega^{-4}g^{\mu\lambda}g^{\nu\rho}\Omega_{;\lambda}\Omega_{;\rho}$ on $\mathscr{M}$. Then an almost asymptotically flat space–time (AAFS) is defined as an asymptotically simple space–time for which $\mathscr{S}$ is isometric to the unit timelike hyperboloid and $\hat{g}^{\mu\nu}\Omega_{|\lambda}\Omega_{|\nu} = \Omega^{-4}g^{\mu\nu}\Omega_{;\mu}\Omega_{;\nu} = -1$ on $\mathscr{S}$. Equivalent definitions are given in terms of the existence of coordinate systems in which $g_{\mu\nu}$ or $\hat{g}_{\mu\nu}$ have simple explicitly given forms. The group of asymptotic symmetries of $(\mathscr{M},\mathbf{g})$ is studied and is found to be isomorphic to the Lorentz group. The asymptotic behavior of an AAFS is studied. It is proven that the conformal metric $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ gives $\tilde{C}_{\lambda\mu\rho\nu} = 0$, $\Omega^{-1}\tilde{C}^{\lambda\mu\rho\nu}\Omega_{;\mu} = 0$, $\Omega^{-2}C_{\lambda\mu\rho\nu}\Omega_{;\mu}\Omega_{;\nu} = 0$ on $\mathscr{S}$

1. INTRODUCTION

In a previous paper[1] (referred hereafter as Paper I) we presented the reasons supporting the need for a new formulation to study the asymptotic structure of the gravitational field at spatial infinity. We also introduced a method for solving the problem, and we applied that method in the case of a three-dimensional asymptotically Euclidean space. In this paper we set up the formalism for studying the structure at spatial infinity for a four-dimensional space-time.

The whole problem can be summarized as follows: An "isolated" material system, e.g., a binary neutron star, is modeled in general relativity, or any other metric theory of gravity, by a space–time[2] $(\mathscr{M},\mathbf{g})$, whose metric approaches in an "appropriate sense" the Minkowskian metric or whose Riemann tensor tends to zero as the affine distance from the source becomes infinite. In particular, at infinite spacelike distance from the source we have "spatial infinity." To study spatial infinity, we apply a prescription that has been very successful for null infinity: We attempt to bring spatial infinity somehow close and then create a structure (differential or otherwise) that will enable us to apply well-known techniques of local geometry.

Ashtekar and Hansen[3] have used the method of conformal completion to bring infinity metrically close and then added additional structure. Their method is a four-dimensional formulation of Geroch's[4,5] version of the Arnowitt–Deser–Misner[6] approach. According to this method spatial infinity is brought metrically close by a conformal transformation of the physical metric and is represented by a single point i^0. This point is attached to the space–time and serves as its spatial boundary. The formal definition of Ashtekar and Hansen which will be used for comparison in Sec. 5 is as follows[3,7]: A space–time $(\mathscr{M},\mathbf{g})$ is asymptotically flat at spatial infinity if there exists a space–time $(\tilde{\tilde{\mathscr{M}}},\tilde{\tilde{\mathbf{g}}})$ with $\tilde{\tilde{\mathscr{M}}} = \mathscr{M}\cup i^0$, where i^0 is a single point (spacelike related to all points of $\mathscr{M}$), an imbedding of $\mathscr{M}$ into $\tilde{\tilde{\mathscr{M}}}$ (by which $\mathscr{M}$ is identified with its image in $\tilde{\tilde{\mathscr{M}}}$) and a scalar field Ω_0 on $\tilde{\tilde{\mathscr{M}}}$ satisfying the following conditions:

(i) On $\mathscr{M}$, $\Omega_0 > 0$ and $\tilde{\tilde{g}}_{\mu\nu} = \Omega_0^2\, g_{\mu\nu}$

(ii) At i^0, $\tilde{\tilde{\mathscr{M}}}$ is $C^{>1}$, $\tilde{\tilde{g}}_{\mu\nu}$ is $C^{>0}$ and Ω_0 is C^2, while all are C^∞ on $\mathscr{M}$.

(iii) At i^0, $\Omega_0 = 0$, $\tilde{\tilde{\nabla}}_\mu\Omega_0 = 0$, and $\tilde{\tilde{\nabla}}_\mu\tilde{\tilde{\nabla}}_\nu\Omega_0 = 2\tilde{\tilde{g}}_{\mu\nu}$.

The advantage of this formulation over Geroch's version is that it is four-dimensional in spirit (it does not refer to initial-value and evolution problems). Furthermore, it does not need additional conditions on the metric and the Weyl tensor to ensure finiteness of the 4-momentum. It preserves, however, some highly undesired features, all unescapable consequences of the fact that spatial infinity is represented by a single point: These features appear in the definition as the awkward differentiability conditions (ii), lead inevitably to the use of direction-dependent tensors, and impose limitations on the order of asymptotic structure we can study .

The key idea in Sommers' formulation[8] is to attach to the space–time an appropriate boundary $\mathscr{P}$ using projective rather than conformal completion. The boundary is constructed from Minkowski's space-time as the set of all "end points" of (equivalence classes of) spacelike curves. His formal definition can be stated as follows:

A space-time $(\mathscr{M},\mathbf{g})$ is asymptotically flat iff there exist a manifold $\bar{\mathscr{M}}$ with a boundary $\mathscr{P}$ and $\bar{\mathscr{M}} = \mathscr{M}\cup\mathscr{P}$ and a scalar field Σ on an open neighborhood $\bar{U}$ of $\mathscr{P}$ satisfying the following conditions:

(i) $\mathscr{P}$ is equipped with a metric $\mathring{\mathbf{h}}$ such that $(\mathscr{P}, \mathring{\mathbf{h}})$ is isometric with the unit timelike hyperboloid in Minkowski's space-time.

 0022-2488/80/010142-10$01.00

(ii) $\Sigma = 0$ on $\mathscr{P}$ and $\Sigma > 0$ on $\bar{U} - \mathscr{P}$.

(iii) If $\mathring{h}_{\mu\nu}$ is the induced metric and $p_{\mu\nu}$ the extrinsic curvature[9] of a hypersurface $\Sigma = \text{const} \neq 0$ (a leaf of the Σ-foliation) then $\Sigma^2 \mathring{h}_{\mu\nu}$ and $\Sigma p_{\mu\nu}$ have continuous extensions to $\mathscr{P}$ equal both to $\mathring{h}_{\mu\nu}$.

(iv) The curves orthogonal to the hypersurfaces $\Sigma = \text{const}$ (the Σ-foliation) are the restrictions of a congruence on $\bar{U}$ which meets $\mathscr{P}$ transversely.

(v) If $E_{\mu\nu}$ and $B_{\mu\nu}$ are the "electric" and "magnetic" parts of the Weyl tensor with respect to the hypersurfaces $\Sigma = \text{const}$, $\Sigma^{-1} E_{\mu\nu}$, and $\Sigma^{-1} B_{\mu\nu}$ admit smooth limits on $\mathscr{P}$.

The advantages of this formulation with respect to that of Ashtekar and Hansen are mainly two. First, there are no awkward differentiability conditions, no direction-dependent tensors. Second, the boundary of the space–time is three-dimensional and the limits of fields on the space–time become, if they exist, ordinary tensor fields on $\mathscr{P}$. However, a new essential disadvantage appears: There is no four metric on $\mathscr{P}$. Thus, there is no sense of "distance" near $\mathscr{P}$, tensor fields have to be decomposed to tangential and normal components with respect to the Σ-foliation, etc. Another important disadvantage is that there is no way to determine the order of tensor fields from the metric near $\mathscr{P}$. This is manifested by the necessity to include condition (v) in the above definition in order that 4-momentum can be defined.

The Ashtekar–Hansen formulation appears to lead to more practical results while the Sommers formulation seems to be more simple and elegant. Both need an additional condition (that $\Sigma^{-2} B_{\mu\nu}$ admits a limit on the boundary) in order that angular momentum can be defined. None of them leads to a natural definition of asymptotic symmetries. In the Ashtekar–Hansen formulation a "blowing up" of i^0 is necessary to obtain a four-dimensional manifold called Spi on which the group of asymptotic symmetries is defined. In Sommers' formulation the whole concept of asymptotic symmetries appears nonapplicable at least in its present form. None of these formulations seems to lead to a satisfactory unified formulation for spatial, null, and timelike infinity (although Ashtekar and Hansen[3] have presented a "unified" version for null and spatial infinity). All the above remarks suggest that we ask for a new formulation which will incorporate the advantages of the previous formulations, will be free from their disadvantages and will have, if possible, other desirable features. The objective of this paper is to introduce such a formulation for the study of spatial infinity in the four-dimensional case.

In Paper I we introduced the new method for a three-dimensional asymptotically Euclidean space. Conformal completion or projective completion should not be employed in the new formulation. The basic requirement is that *the space admits a natural boundary*. For a space–time $(\mathscr{M}, \mathbf{g})$ this requirement can be described roughly as follows: There is an "unphysical" space–time $(\hat{\mathscr{M}}, \hat{\mathbf{g}})$ with boundary $\mathscr{S}$ such that $\hat{\mathscr{M}} = \mathscr{M} \cup \mathscr{S}$ and the following conditions are satisfied:

A. $\mathscr{S}$ is three dimensional.

B. The unphysical metric $\hat{\mathbf{g}}$ is C^∞ on an open neighborhood $\hat{U}$ of $\mathscr{S}$.

C. On $\hat{U}$ the unphysical metric $\hat{\mathbf{g}}$ is determined uniquely from the physical metric $\mathbf{g}$ and a scalar field Ω and vice versa.

The four-dimensional case can be apparently formulated in two different approaches: A first approach can be based on the concept of a three-dimensional asymptotically Euclidean manifold (defined in Paper I) and consider the space–time as a family of such spacelike 3-surfaces as Geroch's approach considers the space–time as a family of asymptotically flat initial data sets. A second approach can use only the underlying features and the methodology employed in Paper I and consider the space–time as a four-dimensional manifold right from the beginning. Obviously the second approach has many advantages and will be followed in this paper. The strategy that will be followed in order to arrive at a definition of asymptotic flatness at spatial infinity is similar to the one followed for null infinity.[10] It consists essentially of four successive steps which impose on the space–time additional structure with a reasonable order. In the first step we demand the existence of an appropriate space–time with a boundary, a metric regular on the boundary, and its interior representing the original space–time. In the second step we specify the intrinsic structure of the boundary. In the third step we determine the conditions needed for a "smooth fastening" of the boundary to the space–time. Finally, in the fourth step we determine other "physical" conditions needed for the space–time to have rich enough and physically interesting structure. In the first two steps only the "geometrical" fields are involved. These are the fields that provide a "background" or "asymptotic" geometry for the manifold and have no direct physical significance. In the third step the physical fields enter and dominate the fourth step. These fields register at infinity information about the interior of the space–time. In this paper we examine the geometrical fields and some immediate consequences of the formulation on the physical fields. The later as well as the physical quantities they define will be examined in a subsequent paper.

Before we proceed to the presentation of the details of the new formulation is seems useful to present briefly its main results and advantages. The new formulation is based on Penrose's idea of asymptotic simplicity[10,11] appropriately modified (Sec. 2). Adding some boundary conditions we have in Sec. 3 the concept of an almost asymptotically flat (at spatial infinity) space–time. This is a space–time $(\mathscr{M}, \mathbf{g})$ whose corresponding unphysical space–time $(\hat{\mathscr{M}}, \hat{\mathbf{g}})$ has a boundary $\mathscr{S}$ isometric to the unit timelike hyperboloid and a C^∞ 4-metric $\hat{\mathbf{g}}$ on $\mathscr{S}$. This structure suffices for the study of the asymptotic symmetries in Sec. 4. In Sec. 5 we study some properties of almost asymptotically flat space–times and sketch a definition of asymptotically flat space–times (at spatial infinity). Some concluding remarks and future problems are presented in Sec. 6.

The advantages of the present formulation are obvious and numerous. It contains no awkward differentiability conditions, no direction-dependent tensors, no intricate limiting procedures. The boundary $\mathscr{S}$ is suggested by the space–time (not by any arbitrary construction) and is a real boundary of the space–time itself as $\mathscr{I}$ is at null infinity. Tensor fields on $\mathscr{M}$ induce tensor fields on $\mathscr{S}$. On an open neighborhood $\hat{U}$ of $\mathscr{S}$ there is a C^∞ 4-metric $\hat{\mathbf{g}}$, which induces an invertible metric on $\mathscr{S}$, that of the unit timelike hyperboloid. Local differential geometry can be used on $\mathscr{S}$ as on any other regular subspace of $\hat{\mathscr{M}}$. The group of asymptotic symmetries emerges naturally and easily from the asymptotic structure with no additional construction. As it turns out this is the Lorentz group. The order of tensor fields near $\mathscr{S}$ is easily obtained. No regularity or other additional condition on the physical fields (e.g., on $E_{\mu\nu}$, $B_{\mu\nu}$, $C_{\lambda\mu\rho\nu}$, etc.) is needed to imply existence of the ADM 4-momentum (this is not so for the angular momentum). Most of the asymptotic behavior of these fields is a consequence of the smoothness of the unphysical metric. Further and unexpected advantages emerge: A unified framework for null and spatial infinity is already established in this paper. The whole method appears to be extendable to timelike infinity with minor modifications. The possibility of a unified treatment of spatial, null, and timelike infinity is now in sight. The smoothness and universality of the whole formulation open the possibility of finding stronger evidence for the interpretation of the physical quantities. Perhaps the most important of all it appears that we can solve now problems which could not have been attacked before.

In our notation Greek indices λ, μ, ν, etc., take values 0,1,2,3 while Latin indices i, j, k, etc, take values 0,2,3. Covariant derivatives with respect to the physical metric are denoted by ∇_μ, with respect to the conformal metric by $\widetilde{\nabla}_\mu$ or a semicolon and with respect to the unphysical metric by $\hat{\nabla}_\mu$ or a vertical rule. A quantity Ψ is said[5] to vanish asymptotically to order n or is of order n near $\mathscr{S}$ iff $\Omega^{-n}\Psi$ admits a smooth extension to $\mathscr{S}$. We assume also that in a coordinate system $\chi, \omega, \theta, \varphi$ in which $\Omega = \omega$ the order of $\partial\Psi/\partial\chi$, $\partial\Psi/\partial\theta, \partial\Psi/\partial\varphi$ is n (the same as that of Ψ) while the order of $\partial\Psi/\partial\omega$ is $n-1$. This will be denoted by $\Psi = O_n$. Finally the symbol $\triangleq$ will be used instead of the equality symbol to denote tensor relations which hold only on the boundary of the space–time.

2. ASYMPTOTICALLY SIMPLE SPACE–TIMES

It is obvious that the key element in the new formulation is the relation which will determine the unphysical metric $\hat{g}_{\mu\nu}$ from the physical metric $g_{\mu\nu}$. To discover this relation we work as in the three-dimensional case of Paper I.

In coordinates t, r, θ, φ the physical metric of the Minkowski space–time is $(c = 1)$

$$\mathrm{diag}[1, -1, -r^2, -r^2\sin^2\theta]. \tag{1}$$

Since we want to study spatial infinity, it is reasonable to use as a coordinate the spacelike distance $\gamma = (r^2 - t^2)^{1/2}$ from the origin. To reach spatial infinity, we let γ increase with $|t| < r$. For that region of the space–time we can set

$$t = \gamma\sinh\chi, \quad r = \gamma\cosh\chi. \tag{2}$$

Thus we have a transformation $(t, r, \theta, \varphi) \to (\chi, \gamma, \theta, \varphi)$. With $\gamma \to \infty$, while χ, θ, φ remain constant, we reach spatial infinity. If we set $\omega = \gamma^{-1}$, we have from (1) the *physical metric* in coordinates $\chi, \omega, \theta, \varphi$

$$h_{\mu\nu} = \mathrm{diag}[\omega^{-2}, -\omega^{-4}, -\cosh^2\chi\cdot\omega^{-2}, -\cosh^2\chi\cdot\sin^2\theta\cdot\omega^{-2}]. \tag{3}$$

Obviously a simple multiplication by a scalar field (conformal completion) will not produce a regular metric at $\omega = 0$. Nevertheless, let Ω be a scalar field such that in coordinates $\chi, \omega, \theta, \varphi$ we have $\Omega = \omega$. Then the *conformal metric* defined by the relation $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ is in coordinates $\chi, \omega, \theta, \varphi$

$$\tilde{h}_{\mu\nu} = \mathrm{diag}[1, -\omega^{-2}, -\cosh^2\chi, -\cosh^2\chi\sin^2\theta]. \tag{4}$$

Its contravariant form is

$$\tilde{h}^{\mu\nu} = \mathrm{diag}[1, -\omega^2, -\cosh^{-2}\chi, -\cosh^{-2}\chi\sin^{-2}\theta]. \tag{5}$$

The situation is similar to that in Paper I. Equation (5) suggests the following: To find a "nonsingular at $\omega = 0$" metric, we have to change somehow $\tilde{h}^{11}$ (and only that) to zero order in ω. This can be done if we add -1 to $\tilde{h}^{11}$ only, that is if we write

$$\hat{h}^{\mu\nu} = \tilde{h}^{\mu\nu} - \hat{h}^{\mu\lambda}\hat{h}^{\nu\rho}\delta_\lambda{}^1\delta_\rho{}^1, \tag{6}$$

where $\hat{h}^{\mu\nu}$ is the (contravariant form of the) unphysical metric we seek to determine. The previous relation can be written in tensor form as

$$\hat{h}^{\mu\nu} + \hat{h}^{\mu\lambda}\hat{h}^{\nu\rho}\Omega_{|\lambda}\Omega_{|\rho} = \tilde{h}^{\mu\nu}, \tag{7}$$

Although this relation is designed to hold only for $\omega = 0$, we can arbitrarily generalize it to hold near $\omega = 0$ adding at the same time any term which vanishes for $\omega = 0$. After several attempts the best choice appears to be the relation

$$\hat{h}^{\mu\nu} + \hat{h}^{\mu\lambda}\hat{h}^{\nu\rho}\Omega_{|\lambda}\Omega_{|\rho} = \tilde{h}^{\mu\nu} + \tilde{h}^{\mu\lambda}\tilde{h}^{\nu\rho}\Omega_{;\lambda}\Omega_{;\rho}. \tag{8}$$

This is the key relation of the whole formulation. It determines $\hat{g}_{\mu\nu}$ from $g_{\mu\nu}$ and Ω. It is reasonable but essentially arbitrary as the conformal transformation is in the method of conformal completion. Using (8), we can define asymptotic simplicity[12] as follows:

Definition: A space–time $(\mathscr{M}, \mathrm{g})$ is *asymptotically simple* iff there exist:

(a) a space $(\hat{\mathscr{M}}, \hat{\mathrm{g}})$ with a nonempty boundary $\mathscr{B}$ $(\mathscr{B} \subset \hat{\mathscr{M}})$ and a metric $\hat{\mathrm{g}}$ which is C^∞ on some open neighborhood $\hat{U}$ of $\mathscr{B}$ $(\mathscr{B} \subset \hat{U})$,

(b) a diffeomorphism $f: U \to \hat{U} - \mathscr{B}$ from an open subset U of $\mathscr{M}$ to $\hat{U} - B$

(c) a C^∞ scalar field[13] Ω on $\hat{U}$, positive on $\hat{U} - \mathscr{B}$ and zero on $\mathscr{B}$, such that

$$\hat{g}^{\mu\nu} + \hat{g}^{\mu\lambda}\hat{g}^{\nu\rho}\Omega_{|\lambda}\Omega_{|\rho} = \Omega^{-2}g^{\mu\nu} + \Omega^{-4}g^{\mu\lambda}g^{\nu\rho}\Omega_{;\lambda}\Omega_{;\rho}. \tag{9}$$

The first and more important consequence of the above definition is that an asymptotically simple space–time can be imbedded in a space–time $\hat{\mathscr{M}}$ with a boundary $\mathscr{B}$ which satisfies condition B of Sec. 1 for being a natural boundary. Another far reaching consequence of the above definition is given by the following theorem:

Theorem 1: For an asymptotically simple space–time the following hold:

(a) If on some part $\mathscr{N}$ of $\mathscr{B}$

$$\Omega^{-2}\tilde{g}^{\mu\nu}\Omega_{;\mu}\Omega_{;\nu} \triangleq -1, \quad \hat{g}^{\mu\nu}\Omega_{|\mu}\Omega_{|\nu} = 0, \tag{10}$$

then on some open neighborhood of $\mathscr{N}$ we have

$$\hat{g}^{\mu\nu} = \tilde{g}^{\mu\nu}. \tag{11}$$

(b) If on some part $\mathscr{S}$ of $\mathscr{B}$

$$\Omega^{-2}\tilde{g}^{\mu\nu}\Omega_{;\mu}\Omega_{;\nu} \triangleq -1, \quad \hat{g}^{\mu\nu}\Omega_{|\mu}\Omega_{|\nu} \triangleq -1, \tag{12}$$

then on some open neighborhood of $\mathscr{S}$ we have in a coordinate system x^μ with $\Omega = x^1$ $(i,j = 0,2,3)$

$$\hat{g}^{11} = -1 - \tilde{g}^{11}, \tag{13}$$

$$\hat{g}^{i1} = \tilde{g}^{i1}(-1/\tilde{g}^{11} - 1), \tag{14}$$

$$\hat{g}^{ij} = \tilde{g}^{ij} + \tilde{g}^{i1}\tilde{g}^{j1}[1 - (1/\tilde{g}^{11} + 1)^2]. \tag{15}$$

Proof: It is enough to show (11) and (13)–(15) in a coordinate system x^μ with $\Omega = x^1$. In such a system $\Omega_{;\mu} = \Omega_{|\mu} = \delta_\mu{}^1$ and (8) or (9) gives

$$\hat{g}^{\mu\nu} + \hat{g}^{\mu 1}\hat{g}^{\nu 1} = \tilde{g}^{\mu\nu} + \tilde{g}^{\mu 1}\tilde{g}^{\nu 1}. \tag{16}$$

For $\mu = \nu = 1$ we have a second degree algebraic equation which gives two solutions $\hat{g}^{11} = \tilde{g}^{11}$ and $\hat{g}^{11} = -1 - \tilde{g}^{11}$. If we assume (10), then $\tilde{g}^{11} \mathrel{\hat{=}} 0$ and $\hat{g}^{11} = 0$. Hence we *have* to take the solution $\hat{g}^{11} = \tilde{g}^{11}$. Then (16) gives $\hat{g}^{i1} = \tilde{g}^{i1}$ and $\hat{g}^{ij} = \tilde{g}^{ij}$. Hence in this case $\hat{g}^{\mu\nu} = \tilde{g}^{\mu\nu}$ that is (11). If we assume (12), then $\tilde{g}^{11} \mathrel{\hat{=}} 0$, $\hat{g}^{11} \mathrel{\hat{=}} -1$ and we *have* to take the solution $\hat{g}^{11} = -1 - \tilde{g}^{11}$. Then for $\mu = i$, $\nu = 1$, and $\mu = i$, $\nu = j$, (16) gives Eqs. (14) and (15) respectively. It should be noted that in this coordinate system $\tilde{g}_{\mu\nu}$ is expressed in terms of $\hat{g}_{\mu\nu}$ by the relations obtained from (11) and (13)–(15) after interchanging $\hat{g}_{\mu\nu}$ and $\tilde{g}_{\mu\nu}$.

Since the relations (10) specify the part of $\mathscr{B}$ that is null infinity[10] and the relations (12) specify the part of $\mathscr{B}$ which will be called spatial infinity, we have already a *unified formulation* for null and spatial infinity: The general definition of asymptotic simplicity is that given above. In the case of null infinity the relations (10) and (11) hold and the definition of asymptotic simplicity is equivalent to that given for null infinity.[10] To obtain the formulation for spatial infinity, we will adopt in Sec. 3 the relations (12) and (13)–(15).

Another consequence of the previous theorem is that an asymptotically simple space–time which satisfies the conditions (12) admits a boundary which fulfills the requirements B and C (Sec. 1) for being a natural boundary. Thus only requirement A remains to be satisfied.

The unphysical metric for the Minkowski space–time (at spatial infinity) can be obtained easily from (5) and (13)–(15). We find

$$\hat{h}^{\mu\nu} = \operatorname{diag}[1, -1 + \omega^2, -\cosh^2\chi, -\cosh^2\chi\sin^2\theta] \tag{17}$$

and

$$\hat{h}_{\mu\nu} = \operatorname{diag}[1, (-1 + \omega^2)^{-1}, -\cosh^{-2}\chi, -\cosh^{-2}\chi\sin^{-2}\theta]. \tag{18}$$

Obviously $\hat{h}_{\mu\nu}$ and $\hat{h}^{\mu\nu}$ are C^∞ on the boundary $\mathscr{S}$ (represented by $\omega = 0$) and induce on $\mathscr{S}$ the metric of the unit timelike hyperboloid. This fact will be used in the next section. Note that the coordinates χ, θ, φ can take any values in the intervals $-\infty < \chi < +\infty$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \varphi \leqslant 2\pi$ ($\varphi = 0$ is identified with $\varphi = 2\pi$). This will be implicitly assumed throughout this paper.

3. ALMOST ASYMPTOTICALLY FLAT SPACE-TIMES

The concept of asymptotic flatness is introduced in order to capture the idea that the space–time of an isolated and bounded source behaves as the Minkowskian space–time as we go further and further away from the source, that is, as we go to infinity. Since the Minkowski space–time is asymptotically simple, it is reasonable to postulate as the first requirement for asymptotic flatness that of asymptotic simplicity. This condition guarantees the existence of a boundary at infinity and a 4-metric on the boundary. The next step is to specify the intrinsic geometry of the boundary. Again Minkowski's spacetime serves as a guide. It has been found in the previous section that its boundary is isometric to the unit timelike hyperboloid. Thus it is reasonable again to ask that an asymptotically flat space–time has a boundary isometric to the unit timelike hyperboloid (or to the boundary of Minkowski's space–time).

For practical as well as aesthetic reasons it is always preferable to give a definition of a particular kind of space in terms of tensor relations as well as in terms of the existence of a coordinate system in which the metric obtains a simple and useful form. Thus we seek a theorem which will establish the equivalence of tensor relations (i.e., relations independent of the coordinate system) to the existence of one or more (generally a class of) coordinate systems. Such a theorem, which is very similar to the corresponding theorem for the three-dimensional case of Paper I, is the following:

Theorem 2: For an asymptotically simple space–time the following conditions are equivalent:

(a) A part $\mathscr{S}$ of the boundary $\mathscr{B}$ is isometric to the unit timelike hyperboloid and on $\mathscr{S}$ the conditions (12) hold.

(b) There exists a coordinate system[14] $(\chi, \omega, \theta, \varphi)$ in which on an open neighborhood $\hat{U}$ of a part $\mathscr{S}$ of the boundary $\mathscr{B}$ we have $\Omega = \omega$, $\tilde{g}^{11} = -\omega^2 + O_3$ and

$$\hat{g}_{\mu\nu} = \begin{bmatrix} 1 + O_1 & \alpha + O_1 & O_1 & O_1 \\ \alpha + O_1 & \beta + O_1 & \gamma + O_1 & \delta + O_1 \\ O_1 & \gamma + O_1 & -\cosh^2\chi + O_1 & O_1 \\ O_1 & \delta + O_1 & O_1 & -\cosh^2\chi\sin^2\theta + O_1 \end{bmatrix} \tag{19}$$

with α, γ, δ arbitrary functions of χ, θ, φ and

$$\beta = \alpha^2 - \gamma^2\cosh^{-2}\chi - \delta^2\cosh^{-2}\chi\sin^{-2}\theta - 1. \tag{20}$$

(c) There exists a coordinate system $(\chi, \omega, \theta, \varphi)$ in which on $\hat{U}$ we have $\Omega = \omega$, $\tilde{g}^{11} = -\omega^2 + O_3$ and ($\alpha, \beta, \gamma, \delta$ as before)

$$\tilde{g}^{\mu\nu} = \begin{bmatrix} 1 - \alpha^2 + O_1 & \alpha + O_1 & \alpha\gamma\cosh^{-2}\chi + O_1 & \alpha\delta\cosh^{-2}\chi\sin^{-2}\theta + O_1 \\ \alpha + O_1 & -1 + O_1 & -\gamma\cosh^{-2}\chi + O_1 & -\delta\cosh^{-2}\chi\sin^{-2}\theta + O_1 \\ \alpha\gamma\cosh^{-2}\chi + O_1 & -\gamma\cosh^{-2}\chi + O_1 & (-1 - \gamma^2\cosh^{-2}\chi)\cosh^{-2}\chi + O_1 & -\gamma\delta\cosh^{-4}\chi\sin^{-2}\theta + O_1 \\ \alpha\delta\cosh^{-2}\chi\sin^{-2}\theta + O_1 & -\delta\cosh^{-2}\chi\sin^{-2}\theta + O_1 & -\gamma\delta\cosh^{-4}\chi\sin^{-2}\theta + O_1 & (-1 - \delta^2\cosh^{-2}\chi\sin^{-2}\theta)\cosh^{-2}\chi\sin^{-2}\theta + O_1 \end{bmatrix} \tag{21}$$

(d) There exists a coordinate system $(\chi, \omega, \theta, \varphi)$ in which $\Omega = \omega$, $\hat{g}^{11} = -1 + O_1$ and

$$\tilde{g}_{\mu\nu} = \begin{bmatrix} 1+O_1 & \alpha+O_1 & O_1 & O_1 \\ \alpha+O_1 & -\omega^{-2}+O_{-1} & \gamma+O_1 & \delta+O_1 \\ O_1 & \gamma+O_1 & -\cosh^2\chi+O_1 & O_1 \\ O_1 & \delta+O_1 & O_1 & -\cosh^2\chi\sin^2\theta+O_1 \end{bmatrix} \tag{22}$$

(e) There exists a coordinate system $(\chi, \omega, \theta, \varphi)$ in which $\Omega = \omega$, $\hat{g}^{11} = -1 + O_1$, and

$$\tilde{g}^{\mu\nu} = \begin{bmatrix} 1+O_1 & \alpha\omega^2+O_3 & O_1 & O_1 \\ \alpha\omega^2+O_3 & -\omega^2+O_3 & -\gamma\cosh^{-2}\chi\cdot\omega^2+O_3 & -\delta\cosh^{-2}\chi\sin^{-2}\theta\cdot\omega^2+O_3 \\ O_1 & -\gamma\cosh^{-2}\chi\cdot\omega^2+O_3 & -\cosh^{-2}\chi+O_1 & O_1 \\ O_1 & -\delta\cosh^{-2}\chi\cdot\omega^2+O_3 & O_1 & -\cosh^{-2}\chi\sin^{-2}\theta+O_1 \end{bmatrix} \tag{23}$$

Proof: Let us assume first that (a) is true. Since the space–time is asymptotically simple and $\mathscr{S}$ is isometric to the unit timelike hyperboloid, there is a coordinate system $(\chi, \omega, \theta, \varphi)$ in which $\Omega = \omega$, and on an open neighborhood $\hat{U}$ of $\mathscr{S}$ the unphysical metric is given by (19) with α, γ, δ and β functions of χ, θ, φ. From (19) we find

$$\hat{g}^{11} = (\beta - \alpha^2 + \gamma^2\cosh^{-2}\chi + \delta^2\cosh^{-2}\chi\sin^{-2}\theta)^{-1} + O_1. \tag{24}$$

But from the second of conditions (12) we have $\hat{g}^{11} = -1 + O_1$. Thus we obtain the relation (20). Also the first of conditions (12) gives $\tilde{g}^{11} = -\omega^2 + O_3$. Hence (a) implies (b). Inverting (19), we obtain (21). Hence (a) implies (c). Since (12) hold, so the relations (13)–(15) do. From them we obtain (23), and inverting we find (22). Hence (a) implies (d) and (e). Let now (b) be true. Obviously $\mathscr{S}$ is isometric to the unit timelike hyperboloid and the first of conditions (12) is satisfied. From (19) and (20) we have $\hat{g}^{11} = -1 + O_1$, which means that the second of conditions (12) is satisfied. Hence (b) implies (a) and consequently (c), (d), and (e). If (c) is true, the proof is similar. If (d) is true, then inverting we find (23). Thus conditions (12) hold. Using (23) and (13)–(15), we find (21). Hence (d) implies (c) and consequently (a), (b), and (e). When (c) is true, the proof is similar.

This theorem is useful in two ways. First, it suggests the minimum conditions to be included in a definition of asymptotic flatness. Second, it provides a practical framework for doing calculations near $\mathscr{S}$. Should its conditions alone be included in a definition of asymptotic flatness or more conditions are needed? Since this question cannot be answered easily and for other reasons to be presented in Sec. 5, it is preferable to define at this stage an "intermediate" concept, that of an almost asymptotically flat space–time (AAFS).

Definition: A space–time is *almost asymptotically Minkowskian or flat*[15] iff it is asymptotically simple and satisfies one of the conditions (a)–(e) of Theorem 2.

It is obvious that an AAFS admits a natural boundary in the sense of Sec. 1. Furthermore, it is possible to give a general expression of its physical metric near the spatial boundary. Using the relation $g_{\mu\nu} = \Omega^{-2}\tilde{g}_{\mu\nu}$ we have from (22)

$$g_{\mu\nu} = \begin{bmatrix} \omega^{-2}+O_{-1} & O_{-2} & O_{-1} & O_{-1} \\ O_{-2} & -\omega^{-4}+O_{-3} & O_{-2} & O_{-2} \\ O_{-1} & O_{-2} & -\cosh^2\chi\omega^{-2}+O_{-1} & O_{-1} \\ O_{-1} & O_{-2} & O_{-1} & -\cosh^2\chi\sin^2\theta\omega^{-2}+O_{-1} \end{bmatrix}. \tag{25}$$

This is the most general AAFS at spatial infinity. Expression (25) contains explicitly the "geometrical" part of $g_{\mu\nu}$. The "physical" part is hidden in the terms O_n. Further calculations will quite probably need some of the "physical" part to be written out explicitly. However, this structure is enough for a study of the group of asymptotic symmetries in the next section. Note that from (25) a general expression of the physical metric in coordinates t, r, θ, φ or even t, x, y, z can be found.

4. ASYMPTOTIC SYMMETRIES

On the subject of asymptotic symmetries at spatial infinity there is considerable disagreement in the literature. This situation is in sharp constrast with that at null infinity. In the later case there is a clear cut definition (in fact more than one and equivalent definitions) of the concept of asymptotic symmetries and a single group of transformations, the BMS group, describes these symmetries.

The main reasons for this situation at spatial infinity is the lack (up to now) of a general expression for the physical metric near the boundary. On the contrary at null infinity a general expression for the physical metric has been available from the beginning.[16] Geroch,[4,5] in his formulation of asymptotic structure, based the definition of asymptotic symmetries on a three-dimensional manifold $\mathscr{S}_G$. To construct this manifold, Geroch considers a four-dimensional vector space V with a metric of Lorentz signature. Then $\mathscr{S}_G$ is the submanifold of V consisting of all unit spacelike vectors. Thus $\mathscr{S}_G$ turns out to have the metric of a unit timelike hyperboloid. The group of asymptotic symmetries is then defined as the group of isometries of $\mathscr{S}_G$ and this group turns out to be isomorphic to the Lorentz group. Although the whole construction is related to the physical space–time (each initial data set gives a two-dimensional vector space U_S of unit vectors tangent at Λ and all U_S are embedded in $\mathscr{S}_G$), it appears to be artificial. Consequently, it is not completely justified why the resulting group of isometries represents the asymptotic symmetries of the physical space–time.

Ashtekar and Hansen[3] have followed a similar prescription: They construct a manifold call Spi and take the isometries of Spi to be the asymptotic symmetries of the physical space–time. To construct Spi, a "blowing up" of i^0 (the single-point spatial boundary of the physical space–time) is needed. For this they consider the collection of equivalence classes of all "regular" inextendible spacelike curves which reach i^0. This is Spi and has the structure of a fiber bundle with base space the unit timelike hyperboloid and the additive group of reals as the structure group. However, any "blowing up" process introduces arbitrariness in the construction and the interpretation of the results. The mere fact that Spi is four-dimensional makes it unattractive. Furthermore, the resulting group of symmetries at spatial infinity is infinite-dimensional while at the same time there is an invertible metric. Nevertheless, if we restrict ourselves to the symmetries of the base space, we end again with a group isomorphic to the Lorentz group.

In Sommers' approach[8] we have at spatial infinity only projective structure, and this introduces difficulties for an unambiguous definition of asymptotic symmetries. Since, however, the $\Sigma = 0$ hypersurface has the metric of the unit timelike hyperboloid (and this metric is equal to $\Sigma p_{\mu\nu}$), it appears that the asymptotic symmetry group should be that of preserving the metric on $\Sigma = 0$, that is (as in the Geroch approach) a group isomorphic to the Lorentz group.

In the present formulation of asymptotic flatness we can determine the group of asymptotic symmetries in a rather unambiguous and unique way. This is possible because the spatial boundary $\mathscr{S}$ serves as a real boundary of the space–time manifold with a C^∞ 4-metric on it, as a manifold where asymptotic fields register and as a space on which asymptotic symmetries can be described. Nevertheless, it will be useful to define the group of asymptotic symmetries as much as naturally is possible and with more than one equivalent definitions.

The first and apparently more natural definition can be based on the physical space–time. The most important and with physical significance tensor field on $\mathscr{M}$ which also contains all the information about $\mathscr{M}$ is the physical metric tensor. Hence a first definition can be given exactly similar to the original definition which gave the BMS group at null infinity[16,17]: The group of asymptotic symmetries at spatial infinity is the group of coordinate transformations $(\chi, \omega, \theta, \varphi) \to (\chi', \omega', \theta', \varphi')$ which preserve the asymptotic form of the physical metric. It is obvious that this group must depend only on the geometrical fields which emerge from $g_{\mu\nu}$ at infinity. To find the group, we use expression (25) of the physical metric near $\mathscr{S}$. The most general (differentiable) coordinate transformation $(\chi, \omega, \theta, \varphi) \to (\chi', \omega', \theta', \varphi')$ which maps $\omega = 0$ to $\omega' = 0$ and the region $\omega > 0$ to the region $\omega' > 0$ is

$$\chi = \chi_0 + O_1, \tag{26}$$

$$\omega = \omega_1 \omega' + O_1, \quad \text{with } \omega_1 > 0, \tag{27}$$

$$\theta = \theta_0 + O_1, \tag{28}$$

$$\varphi = \varphi_0 + O_1. \tag{29}$$

According to the definition, we have to find $g'_{\mu\nu}$ and demand that $g'_{\mu\nu}$ has diagonal elements with leading terms $\omega'^{-2}, \omega'^{-4}, -\cosh^2\chi' \cdot \omega'^{-2}, -\cosh^2\chi' \cdot \sin^2\theta' \omega'^{-2}$ and nondiagonal elements of the same order as in (25). The condition $g'_{11} = \omega'^{-4} + O_3$ gives $\omega_1 = 1$. Then for the requirements on g'_{00}, g'_{02}, and g'_{03} we have

$$\left(\frac{\partial\chi_0}{\partial\chi'}\right)^2 - \left(\frac{\partial\theta_0}{\partial\chi'}\right)^2 \cosh^2\chi_0 - \left(\frac{\partial\varphi_0}{\partial\chi'}\right)^2 \cosh^2\chi_0 \sin^2\theta = 1, \tag{30}$$

$$\frac{\partial\chi_0}{\partial\chi'}\frac{\partial\chi_0}{\partial\theta'} - \frac{\partial\theta_0}{\partial\chi'}\frac{\partial\theta_0}{\partial\theta'}\cosh^2\chi_0 - \frac{\partial\varphi_0}{\partial\chi'}\frac{\partial\varphi_0}{\partial\theta'}\cosh^2\chi_0\sin^2\theta_0 = 0, \tag{31}$$

$$\frac{\partial\chi_0}{\partial\chi'}\frac{\partial\chi_0}{\partial\varphi'} - \frac{\partial\theta_0}{\partial\chi'}\frac{\partial\theta_0}{\partial\varphi'}\cosh^2\chi_0 - \frac{\partial\varphi_0}{\partial\chi'}\frac{\partial\varphi_0}{\partial\varphi'}\cosh^2\chi_0\sin^2\theta_0 = 0. \tag{32}$$

Similarly the requirements on g'_{22}, g'_{23}, g'_{33} give

$$\left(\frac{\partial\chi_0}{\partial\theta'}\right)^2 - \left(\frac{\partial\theta_0}{\partial\theta'}\right)^2 \cosh^2\chi_0 - \left(\frac{\partial\varphi_0}{\partial\theta'}\right)^2 \cosh^2\chi_0 \sin^2\theta_0 = -\cosh^2\chi', \tag{33}$$

$$\left(\frac{\partial\chi_0}{\partial\varphi'}\right)^2 - \left(\frac{\partial\theta_0}{\partial\varphi'}\right)^2 \cosh^2\chi_0 - \left(\frac{\partial\varphi_0}{\partial\varphi'}\right)^2 \cosh^2\chi_0 \sin^2\theta_0 = -\cosh^2\chi'\sin^2\theta', \tag{34}$$

$$\frac{\partial\chi_0}{\partial\theta'}\frac{\partial\chi_0}{\partial\varphi'} - \frac{\partial\theta_0}{\partial\theta'}\frac{\partial\theta_0}{\partial\varphi'}\cosh^2\chi_0 - \frac{\partial\varphi_0}{\partial\theta'}\frac{\partial\varphi_0}{\partial\varphi'}\cosh^2\chi_0\sin^2\theta_0 = 0. \tag{35}$$

The remaining g'_{12}, g'_{13} give no additional condition. Since $\omega_1 = 1$ and we consider as identical transformations differing in the higher order of (26)–(29), we conclude that the group of asymptotic symmetries is isomorphic to the group of transformations $(\chi_0, \theta_0, \varphi_0) \to (\chi', \theta', \varphi')$ which satisfy (30)–(35). It contains the "translation" subgroup $\chi_0 = \chi' + \text{const}$, θ_0, φ_0 unchanged and the "pure rotation" subgroup $\chi_0 = \chi'$, $\theta_0 = \theta_0(\theta', \varphi')$, $\varphi_0 = \varphi_0(\theta', \varphi')$ satisfying (33)–(35). In the general case we can prove after some calculations that

$$\left(\frac{\partial\chi_0}{\partial\chi'}\right)^2 - \left(\frac{\partial\chi_0}{\partial\theta'}\right)^2 \cosh^{-2}\chi' - \left(\frac{\partial\chi_0}{\partial\varphi'}\right)^2 \cosh^{-2}\chi'\sin^{-2}\theta' = 1, \tag{36}$$

which means that the two-dimensional surface $\chi_0(\chi', \theta', \varphi') = \text{const}$ is a spacelike surface on $\mathscr{S}$. Furthermore, the Jacobian of the transformation $(\theta_0, \varphi_0) \to (\theta', \varphi')$ is

$$J(\theta_0, \varphi_0; \theta', \varphi') = \pm \frac{\partial\chi_0}{\partial\chi'} \frac{\cosh^2\chi'\sin\theta'}{\cosh^2\chi_0\sin\theta_0}, \tag{37}$$

which is a generalization of a known formula.[17,18]

Another way to define the group of asymptotic symmetries is to consider the points at spatial infinity as forming a manifold on its own right ("detached" from the original physical space-time). As we have shown in the previous section, this is a three-dimensional manifold with metric given by

$$g_{ij} = \mathrm{diag}[1, -\cosh^2\chi, -\cosh^2\chi\sin^2\theta] \tag{38}$$

with $i, j = 0,2,3$. Thus we can define the group of asymptotic symmetries as the group of transformations $(\chi, \theta, \varphi) \to (\chi', \theta', \varphi')$ which preserve the form (38). Calculating explicitly g'_{ij} we find again the conditions (30)-(35) with χ_0, θ_0, φ_0 replaced by χ, θ, φ. Thus this second definition is equivalent to the first.

Finally, we can define the group of asymptotic symmetries as the group generated by the asymptotic Killing vectors in the unphysical space–time. According to this definition we consider the spatial boundary $\mathscr{S}$ as a submanifold of $\hat{\mathscr{M}}$ (see the Appendix). A vector field ξ^μ on $\hat{\mathscr{M}}$ is an asymptotic Killing vector field of $\mathscr{M}$ iff it satisfies the conditions

$$q_\mu{}^\lambda q_\nu{}^\rho \hat{\nabla}_{(\lambda}\xi_{\rho)} \mathrel{\widehat{=}} 0, \quad n_\mu \xi^\mu \mathrel{\widehat{=}} 0. \tag{39}$$

These relations imply that the restriction of ξ^μ to $\mathscr{S}$ is a Killing vector field for the intrinsic geometry of $\mathscr{S}$. But the Killing vector fields on $\mathscr{S}$ generate the group of the previous definition. Hence this definition is equivalent to the previous two.

Thus all the previous definitions give a unique and unambiguous result: The group of asymptotic symmetries is the group of isometries of $\mathscr{S}$. This group is finite-dimensional and does not contain any supertranslations. It is isomorphic to the Lorentz group. It should be noted that Geroch has obtained the same group (although somehow artificially), while Ashtekar and Hansen have obtained a larger (infinite-dimensional) group.

The knowledge gained from the study of the asymptotic symmetries is useful in attacking some related and otherwise inaccessible problems. Thus, the fact that all allowed transformations (26)–(29) have $\omega_1 = 1$ implies that although the scalar field Ω is not unique (there is "gauge" freedom), two possible choices Ω and Ω' satisfy the condition $\Omega = \Omega' + O_2$ (or $\Omega/\Omega' \mathrel{\widehat{=}} 1$). This property should be regarded as a direct consequence of the second of conditions (12) (or of the existence of an invertible metric on $\mathscr{S}$). A related problem is that of the uniqueness of the completion. Although this problem has not yet been defined clearly (and will not be examined here), the above considerations indicate that a theorem similar to the corresponding theorem for null infinity will be very useful and not difficult to prove.

5. ASYMPTOTICALLY FLAT SPACE–TIMES

In Sec. 3 we defined the concept of almost asymptotically flat space–times. Why "almost"? The main reasons are two. First, our experience from null infinity have taught us to be very careful about the precise conditions to be imposed on the space–time. It is easy to impose conditions obviously too strong, e.g., $R_{\mu\nu} = 0$ on some open neighborhood of $\mathscr{S}$, or perhaps too weak, e.g., no additional condition. The difficulty lies in finding the conditions with the right strength. Conditions too strong will destroy interesting properties of the space–time and eliminate space–times which should be considered asymptotically flat. Conditions to weak will give uninteresting structure and include unwanted space–times. Second, the structure at spatial infinity has not been analyzed sufficiently to allow us to determine the additional (if any) delicate conditions needed. Up to now we have imposed conditions on the geometrical fields. However, the conditions for asymptotic flatness at spatial infinity will probably include some restrictions on the physical fields, as the case is at null infinity. Or the situation will be completely different. To solve this problem we need to know first the consequences of the conditions imposed, namely the properties of AAFS. This will be done in this section. Second, a complete study of the physical fields is needed. The consequences of conditions such as $R_{\mu\nu} \mathrel{\widehat{=}} 0$, $R_{\mu\nu} = O_n$, $R^{\mu\nu}n_\mu = 0$, $\Omega^{-4}T_{\mu\nu} = 0$, $\widetilde{C}^{\lambda\mu\rho\nu} \mathrel{\widehat{=}} 0$, $\hat{\nabla}_{[\lambda} R_{\mu]\nu} = O_n$, etc., should be studied as well as their interrelationships. This second step will be considered in a future paper.

In what follows we will prove that an AAFS satisfies the conditions of the two other four-dimensional formulations. Hence no condition of those formulations is needed. Furthermore, we will establish some properties of the physical fields which will exclude some of the candidates as conditions of asymptotic flatness.

The conditions of Ashtekar and Hansen[3,7] are given in Sec. 1. Their formulation is based on the conformal transformation $\tilde{\tilde{g}}_{\mu\nu} = \Omega_0^2 g_{\mu\nu}$, where Ω_0 behaves as ω^2 near i^0. Hence in coordinates χ, ω, θ, φ we have $\Omega_0 = \omega^2$ and from (25)

$$\tilde{\tilde{g}}_{\mu\nu} = \begin{bmatrix} \omega^2 + O_3 & O_2 & O_3 & O_3 \\ O_2 & -1 + O_1 & O_2 & O_2 \\ O_3 & O_2 & -\cosh^2\chi\cdot\omega^2 + O_3 & O_3 \\ O_3 & O_2 & O_3 & -\cosh^2\chi\sin^2\theta\cdot\omega^2 + O_1 \end{bmatrix}. \tag{40}$$

Condition (i) is obviously satisfied. Furthermore, (40) implies that all $\tilde{\tilde{g}}_{\mu\nu,\rho}$ exist. However, $\tilde{\tilde{g}}_{11,1} = O_0$ which means that the zero order term of $\tilde{\tilde{g}}_{11,1}$ depends in general on χ, θ, φ at i^0, i.e., $\tilde{\tilde{g}}_{11,1}$ is not continuous at i^0. Hence $\tilde{\tilde{g}}_{\mu\nu,\rho}$ is $C^{>0}$ and condition (ii) is satisfied. Finally we find $\tilde{\tilde{\nabla}}_\mu \Omega_0 = 2\omega\delta_\lambda^1$, $\tilde{\tilde{\nabla}}_\mu \tilde{\tilde{\nabla}}_\nu \Omega_0 = 2\delta_\mu{}^1\delta_\nu{}^1 - 2\omega\tilde{\tilde{\Gamma}}{}^1_{\mu\nu}$, which satisfy condition (iii).

The conditions of Sommers' formulation are also given in Sec. 1. Obviously conditions (i) and (ii) are satisfied if we take $\Omega \equiv \Sigma$. The verification of conditions (iii)–(v) require some long but straightforward calculations. From the physical metric (25) we calculate the intrinsic metric $g_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu$ and the extrinsic curvature $p_{\mu\nu} = q_\mu{}^\lambda q_\nu{}^\rho \nabla_\lambda n_\rho$ of a hypersurface $\Omega = \mathrm{const}$. Thus we find (α, β, γ, δ are as in Theorem 1)

$$\Omega^2 q_{\mu\nu} = \Omega p_{\mu\nu} + O_1 = \begin{bmatrix} 1+O_1 & \alpha+O_1 & O_1 & O_1 \\ \alpha+O_1 & 1+\beta+O_1 & \gamma+O_1 & \delta+O_1 \\ O_1 & \gamma+O_1 & -\cosh^2\chi + O_1 & O_1 \\ O_1 & \delta+O_1 & O_1 & -\cosh^2\chi\sin^2\theta + O_1 \end{bmatrix}. \tag{41}$$

Hence on $\mathscr{S}$ we have $\Omega^2 q_{\mu\nu} \hat{=} \Omega p_{\mu\nu}$ and Sommers' condition (iii) is satisfied. Also condition (iv) is satisfied since the unit normal n^μ to the hypersurface $\Omega =$ const is continuously extended to $\omega = 0$ and coincides with the normal to $\mathscr{S}$. Finally, as it will be proven below, condition (v) is satisfied. Hence, *every AAFS is asymptotically flat in the sense of Ashtekar and Hansen and Sommers.*

It must be clear by now that an advantage of the present formulation is the fact that calculations near the boundary can be carried out easily without any worry about differentiability class (as in the Ashtekar–Hansen formulation) or convergence (as in the Sommers formulation). In particular the order of any gravitational tensor field near $\mathscr{S}$ can be found in a straightforward way from the physical metric while in the previous formulations we had to consider direction-dependent limits[3] (as for $\Omega^{1/2}\tilde{\tilde{R}}_{\lambda\mu\rho\nu}$ and $\Omega_0^{1/2}\tilde{\tilde{C}}_{\lambda\mu\rho}{}^{\nu}$) or we could not find it at all and it had to be imposed on the space–time by an additional condition [compare condition (v) in Sommers' formulation].

In what follows we establish some asymptotic properties of the gravitational field of an AAFS. In addition to their own significance, these properties help us to eliminate some of the candidates as conditions for the definition of asymptotically flat space–times. All the calculations are long but straightforward and are carried out in a coordinate system $(\chi, \omega, \theta, \varphi)$ in which the physical metric is given by (25). First we find the Christofell symbols $\Gamma_{\lambda\mu\nu}$ and $\Gamma^{\lambda}{}_{\mu\nu}$. Then we calculate the components of the Riemann tensor $R_{\lambda\mu\rho\nu}$, the Ricci tensor $R_{\mu\nu}$ and $R_\mu{}^\nu$, the Ricci scalar R, the Weyl tensor $C_{\lambda\mu\rho\nu}$ and $C_{\lambda\mu}{}^{\rho\nu}$ and the dual of the Weyl tensor $C^*_{\lambda\mu\rho\nu} = \frac{1}{2}\epsilon_{\lambda\mu\zeta\eta}\, C_{\rho\nu}{}^{\zeta\eta}$.. We find that the order of the various tensor components in the above coordinate system is given concisely by the relations

$$R_{\lambda\mu\rho\nu} = O_{-1-n}, \quad R_{\mu\nu} = O_{1-n}, \quad R_\mu{}^\nu = O_{3+m-n},$$
$$R = O_3, \tag{42}$$

$$C_{\lambda\mu\rho\nu} = O_{-1-n}, \quad C_{\lambda\mu}{}^{\rho\nu} = O_{3+m-n},$$
$$C^*_{\lambda\mu\rho\nu} = O_{-1-n}, \tag{43}$$

where m (n) is the number of upper (lower) indices which are equal to 1 (e.g., $C_{01}{}^{23} = O_{3+0-1} = O_2$). Further calculations using (42) and (43) give for the contravariant components

$$R^{\lambda\mu\rho\nu} = O_{7+m}, \quad R^{\mu\nu} = O_{5+m}, \quad C^{\lambda\mu\rho\nu} = O_{7+m}. \tag{44}$$

From these relations and since the Weyl tensor of the conformal metric $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ is $\tilde{C}^{\lambda\mu\rho\nu} = \Omega^{-6} C^{\lambda\mu\rho\nu}$ we have the following theorem.

Theorem 3: For any almost asymptotically flat (at spatial infinity) space–time we have (on $\mathscr{S}$)

$$\tilde{C}^{\lambda\mu\rho\nu} \hat{=} 0, \quad \Omega^{-1}\tilde{C}^{\lambda\mu\rho\nu}\Omega_{;\mu} \hat{=} 0,$$
$$\Omega^{-2}\tilde{C}^{\lambda\mu\rho\nu}\Omega_{;\mu}\Omega_{;\nu} \hat{=} 0. \tag{45}$$

The above properties correspond to the property "$\tilde{C}_{\lambda\mu\rho\nu} = 0$ on $\mathscr{I}$" ($\tilde{C}_{\lambda\mu\rho\nu} = 0$ and $\tilde{C}^{\lambda\mu\rho\nu} = 0$ are equivalent at null infinity). They are, however, "heavier" properties than that at null infinity, and they are consequences of the form of the metric (while at null infinity $\tilde{C}_{\lambda\mu\rho\nu} \hat{=} 0$ had to be imposed[10]). Physically this can be explained perhaps as follows: The dynamical processes that are taking place inside the space–time (a) do not affect the physical fields at spatial infinity and, if once $\tilde{C}^{\lambda\mu\rho\nu} \hat{=} 0$, then $\tilde{C}^{\lambda\mu\rho\nu} \hat{=} 0$ forever, while they (b) do affect the physical fields at null infinity and the condition $\tilde{C}^{\lambda\mu\rho\nu} = 0$ on $\mathscr{I}$ has to be always imposed.

Further calculations of the "electric" and "magnetic" parts of the Weyl tensor

$$E_{\mu\nu} = C_{\lambda\mu\rho\nu} n^\lambda n^\rho, \quad B_{\mu\nu} = C^*_{\lambda\mu\rho\nu} n^\lambda n^\rho \tag{46}$$

give $E_{\mu\nu} = O_1$ and $B_{\mu\nu} = O_1$ near $\mathscr{S}$. Hence we have the following theorem:

Theorem 4: For any almost asymptotically flat (at spatial infinity) space–time we have (on $\mathscr{S}$)

$$E_{\mu\nu} \hat{=} 0, \quad \mathrm{E}_{\mu\nu} \hat{=} 0. \tag{47}$$

An immediate consequence of this theorem is that $\Omega^{-1}E_{\mu\nu}$ and $\Omega^{-1}B_{\mu\nu}$ induce on $\mathscr{S}$ two (symmetric and trace-free) tensor fields which represent the gravitational field and satisfy Sommers condition (v). Thus the ADM 4-momentum can be defined. But angular momentum cannot be defined since it requires $\Omega^{-2}B_{\mu\nu}$ to have a smooth extension to $\mathscr{S}$. Hence, for a definition of angular momentum to be possible, an additional condition appears to be necessary on the space–time. At this point the situation is as in the Ashtekar–Hansen and Sommers formulations. Important questions arise.[19] What are the weakest conditions which guarantee existence of angular momentum at spatial infinity? What are the weakest conditions which guarantee that the 4-momentum and the angular momentum are independent of the cross section of integration of $\mathscr{S}$? How these conditions are related (i.e., they are stronger or weaker) to the conditions $R_\mu{}^\nu = O_n$, $\nabla_\mu E^{\mu\nu} = 0$, $\nabla_\mu B^{\mu\nu} = 0$, $\nabla_{[\lambda} E_{\mu]\nu} = 0$, etc.? Is it necessary to impose any condition or is it possible to extract the information contained in the angular momentum without demanding $B_{\mu\nu} = O_2$? Or is perhaps the best choice not to define at all angular momentum at spatial infinity as the case is (up to now) at null infinity? Are more than one degrees of asymptotic flatness useful (AAFS and AFS) or we must just drop the word "almost" from the definition of AAFS?

Since these questions cannot be answered before a thorough study of the physical fields (which will be the subject of a future paper), it does not seem appropriate to define at this stage the concept of an asymptotically flat space–time (AFS). However, the following definition seems reasonable:

Definition: A space–time is *asymptotically Minkowskian (or flat) and empty at spatial infinity* (AFES) iff it is almost asymptotically flat and $R_{\mu\nu} = 0$ on some open neighborhood of $\mathscr{I}$.

Thus an AFS is expected to be a concept between an AAFS and an AFES. Every AFES will be an AFS.

6. GENERAL REMARKS

In this paper a *new* formulation has been presented for the study of asymptotic structure at spatial infinity. The reasons justifying the need for the new formulation have been presented in Sec. 1, while its advantages are scattered in Secs. 2–5. The evidence in its favor can be summarized as "right strength" and "smooth operation". Its "right strength" is supported by the following arguments: It gives the Ashtekar–Hansen formulation and the Sommers formulation (Sec. 5) and encompasses the Schwarzschild, Weyl, and Kerr space–times, as it can be proved easily. These space–time will be guiding examples in the study of the physical fields. Finally, the "smooth operation" of the new formulation is manifested in the applicability of tensor calculus and local differential geometry on an open neighborhood of $\mathscr{I}$.

It should be emphasized, however, that although the foundations of the new formulation seem strong enough, there are still a lot of problems to be solved before we have a complete and satisfactory picture of asymptotic structure. Beyond the study of the physical fields we have to answer the following questions for spatial infinity:

(a) Is symptotic flatness stable? The answer seems to be positive, since disturbances from the interior of the space–time do not reach $\mathscr{I}$. However, a better formulation of the question (e.g., as in Ref. 5, p. 74) and more precise answers are desirable.

(b) What are the implications of the existence of Killing vector fields on $\mathscr{M}$ for the asymptotic geometry?[5] Are there any theorems similar to those for null infinity?[20]

(c) Can multipole moments[21] be defined on $\mathscr{I}$ for gravitational fields which have no Killing vector fields?

(d) Is there a corresponding "peeling theorem"[16] for spatial infinity?

(e) Is $\mathscr{I}$ unique?[5] What can be said for two different completions of the same asymptotic region?[7]

(f) What are the implications of asymptotic flatness for the evolution of initial-data sets? Does it guarantee a "sufficient number"[3,4] of them?

In a more general framework the problem of asymptotic flatness of timelike infinity should be solved and a unified formulation of timelike, null, and spatial infinity should be presented.It appears that all the above problems can now be solved. Our experience from similar studies for null infinity and from the previous formulations of Geroch, Ashtekar, Hansen, and Sommers is invaluable.However, until such a program is completed, our understanding of asymptotic structure will be incomplete.

ACKNOWLEDGMENTS

Most of this work has been carried out at the Institute of Astronomy of the University of Cambridge, where the author was Senior Visiting Fellow, and it has been partially supported by the UK Science Research Council. The hospitality and support of the members and the staff of the Institute of Astronomy is gratefully acknowledged.

APPENDIX: SYMMETRIES OF A SUBSPACE

Let $(\mathscr{M},g)$ be a space–time and $\mathscr{H}$ a three-dimensional hypersurface assumed for simplicity to be timelike. Independently of whether or not the space–time admits a Killing vector field, the submanifold $\mathscr{H}$ (considered as a manifold on its own right) can have an intrinsic Killing vector field. If n_μ is the unit vector normal to $\mathscr{H}$, $q_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu$ the induced metric of $\mathscr{H}$ and $q_\mu{}^\nu = \delta_\mu{}^\nu + n_\mu n^\nu$ the projection operator, we can easily prove the following:

Lemma: Let $T_{\mu\nu}$ be a tensor field on $\mathscr{M}$. Then $q_\mu{}^\lambda q_\nu{}^\rho T_{\lambda\rho} = 0$ *on* $\mathscr{H}$, iff in a coordinate system x^μ in which $\mathscr{H}$ is represented by $x^1 = 0$ we have $T_{ij} = 0$ on $\mathscr{H}$ $(\mu, \nu = 0,1,2,3,\ i,j = 0,2,3)$.

For an intrinsic Killing vector field of $\mathscr{H}$ we have the following:

Theorem: Considered as a manifold, $\mathscr{H}$ admits a Killing vector field iff there is a vector field on $\mathscr{M}$ such that $q_\mu{}^\lambda q_\nu{}^\rho \nabla_{(\lambda}\xi_{\rho)} = 0$ and $n_\mu \xi^\mu = 0$ *on* $\mathscr{H}$.

This theorem can be proved easily if we use the previous lemma. In a coordinate system in which $\mathscr{H}$ is $x^1 = 0$ we have $\nabla_{(i}\xi_{j)} = 0, \xi^1 = 0$ on $\mathscr{H}$. These reduce to and can be obtained from the Killing equation on $\mathscr{H}$ with respect to its 3-metric. The condition $n_\mu \xi^\mu = 0$ (which is usually ommited in the literature) indicates how to construct from the three-dimensional vector field ξ^i on $\mathscr{H}$ the four-dimensional vector field ξ^μ on $\mathscr{M}$ and vice versa.

If $(\mathscr{M}, g)$ is an almost asymptotically flat space–time, then a vector field ξ^μ on $\mathscr{M}$ is an *asyptotic Killing vector field* (at spatial infinity) iff it has a smooth extension to the boundary $\mathscr{I}$ and this extension gives a Killing vector field on $\mathscr{I}$ (considered as a submanifold of $\hat{\mathscr{M}}$).

The boundary $\mathscr{I}$ of an AAFS admits six linearly independent Killing vector fields (the maximum number, which correspond to the six possible Lorentz rotations of Minkowski's space–time).

[1] S. Persides, J. Math. Phys. **21**, 135 (1980).

[2] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1973).

[3] A. Ashtekar and R.O. Hansen, J. Math, Phys. **19**, 1542 (1978).

[4] R. Geroch, J. Math. Phys. **13**, 956 (1972).

[5] R. Geroch in *Asymptotic Structure of Space-Time*, edited by F.P. Esposito and L. Witten (Plenum, New York, 1977).

[6] R. Arnowitt, S. Desser, and C.W. Misner, Phys. Rev. **117**, 1595 (1960); **118**, 1100 (1960); **121**, 1556 (1961); **122**, 997 (1961); the article in *Gravitation, An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

[7] A. Ashtekar, "Asymptotic Structure of the Gravitational Field at Spatial Infinity" (Preprint).

[8] P. Sommers, J. Math. Phys. **19**, 549 (1978).

[9] The extrinsic curvature as it is taken here differs in sign with that of Ref. 8.

[10] S. Persides, "A Definition of Asymptotically Minkowskian Space-Times", J. Math. Phys. (to appear).

[11] R. Penrose, Proc. Roy. Soc. A **284**, 159 (1965).

[12] The way asymptotic simplicity was introduced originally (see Ref. 11) required later the introduction of weak asymptotic simplicity. It seems, however, preferable to call asymptotically simple a space–time which in the previous terminology we would have called weakly asymptotically simple. Thus in the new terminology Minkowski's and Schwarzschild's space–times are asymptotically simple. When we are dealing with asymptotic structure, we are not interested in the topology of the interior of the space–time.

[13] Throughout the paper we assume that $\mathscr{M}$, $\hat{\mathscr{M}}$, $\mathbf{g}$, $\hat{\mathbf{g}}$, Ω are C^∞ to simplify the presentation. The whole discussion, however, would go through if $\mathscr{M}$ and $\hat{\mathscr{M}}$ are C^4, while $\mathbf{g}$, $\hat{\mathbf{g}}$, and Ω are C^3.

[14] It is implicitly assumed that $\mathscr{S}$ is covered exactly by χ, θ, φ with values in the intervals specified at the end of Sec. 2.

[15] The term "Minkowskian" is considered equivalent to "flat, four-dimensional, with signature $+ - - -$". Hence, as long as we refer to space–time, "Minkowskian" and "flat" are equivalent terms.

[16] H. Bondi, M.G.J. van der Burg, and A.W.K. Metzner, Proc. Roy. Soc. A **269**, 21 (1962); R.K. Sachs, Proc. Roy. Soc. A **270**, 103 (1962).

[17] R.K. Sachs, Phys. Rev. **128**, 2851 (1962).

[18] S. Persides, "Energy and Momentum in General Relativity," Gen. Rel Grav. (to appear).

[19] A. Ashtekar and A. Magnon-Ashtekar, "On Conserved Quantities in General Relativity" (preprint).

[20] A. Ashtekar and B. Xanthopoulos, "Isometries Compatible with Asymptotic Flatness at Null Infinity: A Complete Description" (preprint).

[21] R.O. Hansen, J. Math. Phys. **15**, 46 (1976).

Dust cylinders in static spacetimes

J. D. Lathrop[a)]
Department of Astronomy, The Pennsylvania State University, University Park, Pennsylvania 16802
M. S. Orsene
Department of Physics and Astronomy, Williams College, Williamstown, Massachusetts 01267

(Received 4 May 1979; accepted for publication 30 August 1979)

We argue that the work of Vishveshwara and Winicour suggests that there is an upper limit to the linear mass densities of infinite cylinders in static, cylindrically symmetric spacetimes. We investigate the class of static cylinders having two oppositely directed circular flows of dust, and show that the above conjecture is true for these cylinders.

1. INTRODUCTION

Vishveshwara and Winicour[1] have presented a class of differentially rotating dust cylinders that includes the van Stockum universe[2,3] as a special case. The spacetimes of these cylinders admit closed timelike lines (CTL) when the mass M in a length Δz along the cylinder axis satisfies the inequality $M > \frac{1}{4}\Delta z$. Thus the criterion for the existence of CTL in the case of a rotating cylinder resembles the criterion for the existence of a horizon in the Schwarzschild case (but see also Tipler[4] and Charlton[5]). Physically, the idea suggests itself that the spacetime of an infinite, rotating cylinder with a large linear mass density does not have an horizon because of the relativistic rotational motion of such a cylinder, which leads to a spacetime with CTL instead.

It is clear that the *static*, cylindrically symmetric spacetimes of cylinders having vanishing angular momentum do not admit CTL. But at the same time, the exterior spacetimes[6] of such cylinders do not have horizons. In light of the discussion above, it is natural to conjecture that the reason for this is that there is an upper limit, given by $M \leqslant \frac{1}{4}\Delta z$, to the allowed linear mass densities of these cylinders. Below, we show that this conjecture is in fact true for all static dust cylinders having two oppositely directed circular flows of dust.

In Sec. 2 we present a general discussion of these cylinders. We find that the opposing rotations, which in general are differential, have the same angular frequency and that the two components of dust have equal proper densities. We reduce the interior field equations to a simple form, solve them for the particular case that the angular frequency is constant, and present the general exterior solution and the Newtonian analog. In Sec. 3 we show that the linear mass density has the upper limit quoted above.

2. THE STATIC DUST CYLINDERS

We choose the fundamental form[7]

$$ds^2 = -F dt^2 + H(dr^2 + dz^2) + L d\varphi^2,$$

where F, H, and L are functions of r only. For the geodesics of circular orbits ($\dot r = \dot z = 0$; a dot denotes d/ds) we then find

$$\dot t = (F - L\omega^2)^{-1/2}, \qquad \dot\varphi = \pm\,\omega(F - L\omega^2)^{-1/2}, \tag{1}$$

where

$$\omega = |\dot\varphi|/\dot t = (F'/L')^{1/2}. \tag{2}$$

A prime indicates d/dr. Thus the two components of dust rotate with angular frequencies $+\omega(r)$ and $-\omega(r)$.

Let ρ^+ and ρ^- be the proper densities of the two components of dust. One of the field equations immediately yields $\rho^+ = \rho^-$. It is then easy to verify that $-8\pi(T^0_0 + T^\varphi_\varphi - T) = R^0_0 + R^\varphi_\varphi$ vanishes in our coordinates, so that without loss of generality we may impose the coordinate condition[7]

$$L = r^2/F. \tag{3}$$

The metric then has the Levi-Civita form[6] in both the interior and the exterior.

From the field equations and Eqs. (2) and (3) we find

$$(1/r)[r(H'/H)]' = -8\pi\rho H, \tag{4}$$

$$H'/H = -(F'/F)(1 - rF'/2F), \tag{5}$$

$$\omega^2 r^2/F^2 = (rF'/2F)(1 - rF'/2F)^{-1}, \tag{6}$$

where $\rho = \rho^+ + \rho^- = 2\rho^+$. This completes the development of the interior solution for the general case. To proceed to specific interior solutions, one of the functions ρ, H, F, and ω must be supplied; Eqs. (4)–(6) then determine the other three. Physically acceptable solutions are of course subject to the usual conditions, such as $\rho \geqslant 0$, etc.

As an example, let us consider the case $\omega = \omega_0$, a constant. From Eq. (6) we find $F = \frac{1}{2}[1 + (1 + 4\omega_0^2 r^2)^{1/2}]$, and Eqs. (5) and (4) then yield $H = (1 + 4\omega_0^2 r^2)^{-1/4}$ and $2\pi\rho = \omega_0^2(1 + 4\omega_0^2 r^2)^{-7/4}$.

For the exteriors of the cylinders, we have $\rho = 0$. Equations (4) and (5) then give the results of Levi–Civita,

$$F = F_R (r/R)^\alpha, \quad H = H_R (r/R)^{-\beta} \tag{7}$$

and Eq. (6) becomes

$$\omega r = \omega_R R\,(r/R)^\alpha. \tag{8}$$

Here $\alpha = RF'_R/F_R$, $\beta = \alpha - \frac{1}{2}\alpha^2$, and the notation f_R means $f(r)$ evaluated at the boundary $r = R$ of the cylinder.

We note that in the Newtonian limit ($F \simeq 1 + 2\phi$, where ϕ is the Newtonian potential) Eqs. (4)–(8) give $H \simeq 1 - 2\phi$ and the correct Newtonian results

[a)]On leave from the Department of Physics and Astronomy, Williams College, Williamstown, Massachusetts 01267.

$$(1/r)(r\phi')' \simeq 4\pi\rho, \quad \omega^2 r \simeq \phi,$$

and, in the exterior,

$$\alpha \simeq 2\omega_R^2 R^2 \ll 1,$$

$$\phi \simeq \phi_R + (\alpha/2)\ln(r/R), \quad \omega r \simeq \omega_R R.$$

3. THE LINEAR MASS DENSITY

Owing to the fact that the spacetimes of these cylinders are not asymptotically flat, there is a certain freedom in the definition of the linear mass density: There are many definitions that agree in the Newtonian limit but differ when the field is strong. We use the formula for the linear mass density given by Vishveshwara and Winicour,[1] which leads to the results described in Sec. 1.

This definition is based on time-translation and rotational killing vectors (T^a and Φ^a, respectively) for the exterior. In terms of these vectors, the mass M in a length Δz along the axis of a cylinder is defined in the general case to be

$$M = -\tfrac{1}{2}\tau^{-1}(\lambda_{11}\lambda_{00,\tau} - \lambda_{01}\lambda_{01,\tau})\Delta z, \tag{9}$$

where

$$\lambda_{00} = T^a T_a, \quad \lambda_{11} = \Phi^a \Phi_a, \quad \lambda_{01} = T^a \Phi_a,$$

$$\tau^2 = 2\lambda_{01}^2 - 2\lambda_{00}\lambda_{11}.$$

In the present case, we have $T^a = (1,0,0,0)$ and $\Phi^a = (0,0,0,1)$. Then Eq. (9) becomes

$$M = \tfrac{1}{4}\alpha\Delta z. \tag{10}$$

We now show that $0 \leqslant \alpha \leqslant 1$. The lower limit follows at once from Eq. (6), which for $r = R$ gives

$$\alpha = 2\omega_R^2 R^2/(F_R^2 + \omega_R^2 R^2). \tag{11}$$

To obtain the upper limit, we calculate the Lorentz contraction factor $\gamma = -t^a u_a$ for the dust (u^a is the four velocity of the dust) relative to an observer whose four velocity t^a is obtained by parallel transport along a geodesic in the space $t = \text{const}$ from the four velocity of an observer at rest on the axis. We find $t^a = (F^{-1/2},0,0,0)$, so that from Eqs. (1) and (3)

$$\gamma = (1 - \omega^2 r^2/F^2)^{-1/2}$$

for both components of the dust. Thus $F_R > \omega_R R$, so that Eq. (11) gives $\alpha \leqslant 1$ and, from Eq. (10),

$$M \leqslant \tfrac{1}{4}\Delta z.$$

[1] C.V. Vishveshwara and J. Winicour, J. Math Phys. **18**, 1280 (1977).
[2] W. van Stockum, Proc. Roy. Soc. Edinburgh **57**, 135 (1937).
[3] F.J. Tipler, Phys. Rev. D **9**, 2203 (1974).
[4] F.J. Tipler, Phys. Rev. Lett. **37**, 879 (1976).
[5] N.J. Charlton, J. Phys. A **11**, 2207 (1978).
[6] T. Levi-Civita, Rend. Acc. Lincei **28**, 101 (1919).
[7] See, e.g., J.L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), p. 309 ff.

On algebraically special space–times with nontwisting rays

Garry Ludwig
Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

(Received 22 May 1979; accepted for publication 21 August 1979)

With the aid of the Newman–Penrose formalism and Penrose's conformal technique Einstein's gravitational field equations are first solved exactly as far as is possible for arbitrary sources. It is assumed, however, that space–time is algebraically special with hypersurface-orthogonal geodesic and shear-free rays. Special cases are considered. Next, the asymptotic behavior of the components of the metric tensor, the Weyl tensor, the Ricci tensor and the spin coefficients is determined in a suitable frame. Einstein–Maxwell space–times with the above properties are treated in some detail.

1. INTRODUCTION

Many years ago Newman and Penrose[1] and Newman and Unti[2] investigated the asymptotic properties of the metric tensor, the Weyl tensor, and the spin coefficients in a suitable frame in empty asymptotically flat space–time. Later, Kozarzewski[3] and Exton *et al.*[4] did the same for an Einstein–Maxwell space–time. The present author[5] generalized their results to space–times with arbitrary Ricci tensors, using Penrose's conformal technique[6,7] as the main tool. In this paper we shall employ the same method in an attempt to find exact algebraically special solutions with geodesic and shear-free hypersurface-orthogonal rays but for arbitrary Ricci tensors. A similar attempt[8] for the twisting case (but with some restrictions on the Ricci tensor) led to results which had previously been obtained by Lind[9] and Trim and Wainwright[10,11] for a more restrictive class of Ricci tensors. Explicit solutions are hard to get. Instead, one obtains reduced equations which, in general, are few in number but quite difficult to solve. In the Robinson–Trautman[12] case, however, actual explicit solutions have been found.

We shall also derive the asymptotic properties of the metric tensor, the spin coefficients, and other variables for the special case considered here, namely that of an algebraically special space–time whose repeated principal null direction is geodesic, shear-free, and twist-free. The results of Ref. 5 cannot be adapted since they were based on a more or less specific congruence of null geodesics (due to the choice $\dot{P}=0$). This congruence is not necessarily the one that is shear-free and whose tangent vector field is the repeated principal null vector field for the Weyl tensor. Moreover, in the Einstein–Maxwell case many of the leading terms vanish. It is, therefore, instructive to obtain higher-order terms.

In Secs. 2 and 3 we go a fair distance towards finding algebraically special space–times whose repeated principal null vectors are at each point tangential to a congruence of shear-free and hypersurface-orthogonal null geodesics and whose Ricci tensor is arbitrary (in particular the component Φ_{00} need not vinish). The reduced equations we obtain simplify considerably if space–time is assumed to be type III (or N), i.e., if the Weyl tensor component Ψ_2 is assumed to vanish in addition to Ψ_0 and Ψ_1. They simplify even more, regardless of whether Ψ_2 vanishes or not, if the Ricci tensor component Φ_{00} is put equal to zero. This is done in Sec. 4. For this special case the equations can, with some effort, also be derived from those of the twisting case[8] by putting the twist equal to zero. We specialize further and take the source to be a Maxwell field one of whose principal null directions is the repeated principal null direction of the Weyl tensor. The reduced equations now become independent of the radial coordinate. In Sec. 5 the approximate results are obtained from the exact results of Secs. 2 and 3. In Sec. 6 we investigate the asymptotic behavior of an Einstein–Maxwell field with the properties described above. In particular, we show that if the component ϕ_0 of the Maxwell field does not vanish identically, i.e., if the Maxwell field is not "aligned" with the repeated principal null direction of the Weyl tensor, then the leading terms in the expansions of ϕ_1 and ϕ_2 vanish. Instead of having $\phi_1=O(r^{-2})$ and $\phi_2=O(r^{-1})$ we have $\phi_1=O(r^{-3})$ and $\phi_2=O(r^{-2})$, where r is the radial coordinate. The Appendix consists of a collection of formulas used for this paper but discussed previously.[1,5]

The space–times under consideration are not necessarily asymptotically flat in the sense of Penrose, but along each of the geodesics they behave as if they were. That is, as in Ref. 8, we drop all gobal requirements on future null infinity $\mathscr{I}^+$ (such as the requirement that the topology of $\mathscr{I}^+$ be that of a cylinder) and keep only local ones.

The notation used will be the same as in Refs. 5 and 8. In particular, careted quantities refer to the rescaled space–time $\hat{M}$, those without carets refer to the physical space–time M. Superscripts on a careted variable denote the appropriate coefficient in the expansion of that variable in powers of the conformal factor Ω; e.g., $\hat{B}=\hat{B}^0+\hat{B}^{(1)}\Omega+\hat{B}^{(2)}\Omega^2+0(\Omega^3)$. Similarly, superscripts on an uncareted variable refer to the expansion of that variable in powers of r^{-1}, where r is the radial coordinate. The usual symbols are used for the spin coefficients and other NP (Newman–Penrose[1]) quantities. ∇ stands for $\partial/\partial x^3+i(\partial/\partial x^4)$. Equation (NP4.2b), for example, refers to Eq. (4.2b) of Ref. 1. For a detailed discussion of the conformal technique as applied here the reader is referred to Ref. 5.

2. NP QUANTITIES IN RESCALED SPACE TIME

In Ref. 5 coordinates (u,Ω,x^3,x^4) and a frame were chosen in the unphysical (rescaled) space time $\hat{M}$ such that

 0022-2488/80/010154-07$01.00

$$\hat{\kappa} = \hat{\pi} = \hat{\epsilon} = \hat{\rho} = \hat{\tau} - (\hat{\bar{\alpha}} + \hat{\beta}) = 0 \tag{2.1}$$

identically, and such that on $\mathscr{I}^+$

$$\begin{aligned} \hat{U} = \hat{X}^i &= \hat{\omega} = \hat{\Psi}_0 = \hat{\Psi}_1 = \hat{\Psi}_2 = \hat{\Psi}_3 = \hat{\Psi}_4 \\ &= \hat{\tau} = \hat{\lambda} = \hat{\nu} = 0 , \end{aligned}$$

$$\begin{aligned} &\hat{f} = -1, \quad \hat{\xi}^3 = -i\hat{\xi}^4 = P(u,x^3,x^4), \\ &\hat{\alpha} = -\hat{\beta} = \tfrac{1}{2}\bar{\nabla}P, \quad \hat{\gamma} = -\tfrac{1}{2}\hat{\mu} = \tfrac{1}{2}\dot{P}P^{-1} . \end{aligned} \tag{2.2}$$

The coordinate lines along which u, x^3, and x^4 are constant are hypersurface-orthogonal null geodesics.

Actually, this was done only for P independent of the u coordinate, i.e., for $\dot{P}=0$, but the proof given generalizes quite readily and will be omitted. We cannot choose $\dot{P}$ to vanish here since this condition determines to a large extent the null geodesics of our coordinate system, and these null geodesics may not be the ones that are shear-free and that correspond to the repeated principal null direction of the Weyl tensor. The freedom left in the choice of frame and coordinates is given by the Newman–Unti group[13,8] (and reduces to the BMS group for $\dot{P}=0$).

Since we are aiming for algebraically special solutions with shear-free geodesic rays we shall assume that σ and Ψ_1 vanish. Both assumptions are necessary since there is no Goldberg–Sachs theorem when the Ricci tensor is arbitrary. The vanishing of Ψ_0 follows from Eq. (NP4.2b). Thus, in addition to Eq. (2.1), we have

$$\hat{\sigma} = \hat{\Psi}_0 = \hat{\Psi}_1 = 0 \tag{2.3}$$

identically in $\hat{M}$. When we substitute conditions (2.1)–(2.3) into Eqs. (NP4.2) and into the metric equations (A1) we obtain the following preliminary result:

$$\begin{aligned} \hat{\kappa} = \hat{\sigma} &= \hat{\rho} = \hat{\tau} = \hat{\pi} = \hat{\epsilon} = \hat{\lambda} = \hat{X}^i \\ &= \hat{\Psi}_0 = \hat{\Psi}_1 = \hat{\Phi}_{00} = \hat{\Phi}_{01} = \hat{\Phi}_{02} = 0 , \\ -\hat{\beta} &= \hat{\alpha} = \tfrac{1}{2}\bar{\nabla}P, \\ \hat{\mu} &= -\dot{P}P^{-1}, \quad \hat{\xi}^3 = -i\hat{\xi}^4 = P(u,x^3,x^4), \\ \hat{\Lambda} &= -\tfrac{1}{2}\hat{\Psi}_2, \quad \hat{\Phi}_{11} = \tfrac{1}{2}K + \tfrac{3}{2}\hat{\Psi}_2 \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} &\hat{\Psi}_3 = \bar{\hat{\delta}}\hat{\gamma} - \tfrac{1}{2}P\bar{\nabla}(\dot{P}P^{-1}), \quad \hat{\Psi}_4 = \bar{\hat{\delta}}\hat{\nu} + (\bar{\nabla}P)\hat{\nu}, \\ &\hat{\Phi}_{12} = P\nabla(\dot{P}P^{-1}) + \bar{\hat{\Psi}}_3 , \\ &\hat{\Phi}_{22} = \hat{\delta}\hat{\nu} + (\ln P)_{,11} - (\dot{P}P^{-1})^2 + 2\hat{\gamma}\dot{P}P^{-1} - (\nabla P)\hat{\nu} , \end{aligned} \tag{2.5}$$

as well as

$$\begin{aligned} &\hat{D}\hat{\gamma} = \tfrac{1}{2}K + 3\hat{\Psi}_2, \quad \hat{D}\hat{\nu} = 2\bar{\hat{\delta}}\hat{\gamma} , \\ &\hat{D}\hat{\omega} = \hat{\delta}\hat{f}, \quad \hat{D}\hat{U} = \hat{\Delta}\hat{f} - 2\hat{f}\hat{\gamma} , \end{aligned} \tag{2.6}$$

where $K = 2P^2\nabla\bar{\nabla}\ln P$ is the Gaussian curvature of the cut $u=$ constant on $\mathscr{I}^+$. The Bianchi identities[14] yield nothing further.

The first of Eqs. (A2) relates the metric variable $\hat{f}(u,\Omega,x^3,x^4)$ to Φ_{00} as follows:

$$\hat{f} = -\left[1 + 2\int_0^{\Omega} \Phi_{00}\Omega^{-3} d\Omega\right]^{1/2} \tag{2.7}$$

Let us also define a variable $T(\Omega,u,x^3,x^4)$ by

$$T = -\int_0^{\Omega} \hat{f}^{-1} d\Omega \tag{2.8}$$

and note that $\hat{D} = \hat{f}(\partial/\partial\Omega) = -\partial/\partial T$. We proceed by integrating Equations (2.6) and substituting the results into Eqs. (2.5). The NP quantities in the rescaled space–time $\hat{M}$ are, therefore, given by Eqs. (2.4) and by Eqs. (2.9) below:

$$\begin{aligned} \hat{\gamma} &= \tfrac{1}{2}\dot{P}P^{-1} - \tfrac{1}{2}KT - 3S, \\ \hat{\nu} &= -PT\bar{\nabla}(\dot{P}P^{-1}) + \tfrac{1}{2}T^2P\bar{\nabla}K + 6P\bar{\nabla}'R , \\ \hat{\omega} &= \hat{f}P\nabla T, \quad \hat{U} = \hat{f}\,[\dot{T} + \dot{P}P^{-1}T - \tfrac{1}{2}KT^2 - 6R\,] , \\ \hat{\Psi}_3 &= -\tfrac{1}{2}P(\bar{\nabla}K)T - 3P\bar{\nabla}'S , \\ \hat{\Psi}_4 &= -T\bar{\nabla}[P^2\bar{\nabla}(\dot{P}P^{-1})] + \tfrac{1}{2}T^2\bar{\nabla}(P^2\bar{\nabla}K) \\ &\quad + 12P(\bar{\nabla}P)\bar{\nabla}'R + 6P^2\bar{\nabla}'\bar{\nabla}'R , \\ \hat{\Phi}_{12} &= P\nabla(\dot{P}P^{-1}) - \tfrac{1}{2}P(\nabla K)T - 3P\nabla'S , \\ \hat{\Phi}_{22} &= (\ln P)_{,11} - \dot{P}P^{-1}KT - 6\dot{P}P^{-1}S \\ &\quad - P^2T\nabla\bar{\nabla}(\dot{P}P^{-1}) + \tfrac{1}{2}T^2P^2\nabla\bar{\nabla}K + 6P^2\nabla'\bar{\nabla}'R , \end{aligned} \tag{2.9}$$

where

$$S = \int_0^T \hat{\Psi}_2\, dT, \quad R = \int_0^T S\, dT ,$$

and

$$\nabla'S(u,T,x^3,x^4) = \frac{\partial S}{\partial x^3} + i\frac{\partial S}{\partial x^4} .$$

The variables P and $\hat{\Psi}_2$ are so far arbitrary real variables. Equations (A2) will relate them to the source variables.

3. NP QUANTITIES IN PHYSICAL SPACE TIME

In Ref. 5 a specific convention was adopted to relate a frame in $\hat{M}$ to one in M and the transformation rules for the NP quantities were given. For quick reference these rules are once more listed in the Appendix. With the aid of Eqs. (A2)–(A6) we obtain the NP quantities in M from Eqs. (2.4) and (2.9). They are

$$\begin{aligned} &\kappa = \sigma = \epsilon = \lambda = \Psi_0 = \Psi_1 = X^i = 0 , \\ &\rho = \Omega\hat{f}, \quad \tau = -\bar{\pi} = \hat{\omega}, \quad \alpha = \hat{\alpha}^0\Omega + \bar{\hat{\omega}}, \\ &\beta = -\hat{\bar{\alpha}}^0\Omega, \quad \mu = -\dot{P}P^{-1} - \Omega^{-1}\hat{U}, \\ &\gamma = \hat{\gamma} + \Omega^{-1}\hat{U}, \quad \nu = \hat{\nu}\Omega^{-1}, \quad \Psi_2 = \Omega^2\hat{\Psi}_2, \\ &\Psi_3 = \Omega\hat{\Psi}_3, \quad \Psi_4 = \hat{\Psi}_4 , \\ &\xi^3 = -i\xi^4 = P\Omega, \quad \omega = \Omega^{-1}\hat{f}^{-1}\left(\hat{\omega} - \hat{\xi}^i\frac{\partial\Omega}{\partial x^i}\right), \\ &U = \Omega^{-2}\hat{f}^{-1}\left(\hat{U} - \frac{\partial\Omega}{\partial u}\right), \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} \Phi_{01} &= \Omega^2\hat{D}\hat{\omega}, \quad \Phi_{02} = \Omega\,[\nabla(P\hat{\omega}) + \hat{\omega}\hat{f}^{-1}\hat{D}\hat{\omega}] , \\ \Phi_{11} &= \tfrac{1}{2}\Omega^2(K + 3\hat{\Psi}_2) + \tfrac{1}{2}\Omega\,[\hat{D}\hat{U} + P^2\bar{\nabla}(\hat{\omega}P^{-1}) \\ &\quad + \bar{\hat{\omega}}\hat{f}^{-1}\hat{D}\hat{\omega} + \dot{P}P^{-1}\hat{f}\,] , \\ \Phi_{12} &= \Omega\hat{\Phi}_{12} + \dot{P}P^{-1}\hat{\omega} + P\nabla\hat{U} + \hat{\omega}\hat{f}^{-1}\hat{D}\hat{U} , \\ \Phi_{22} &= \hat{\Phi}_{22} + \Omega^{-1}[\dot{\hat{U}} + \hat{U}\hat{f}^{-1}\hat{D}\hat{U} - \hat{\nu}\hat{\omega} - \bar{\hat{\nu}}\bar{\hat{\omega}} + 2\hat{\gamma}\hat{U}\,] , \\ \Lambda &= -\tfrac{1}{2}\hat{\Psi}_2\Omega^2\hat{f}\hat{U} - \hat{\omega}\bar{\hat{\omega}} + \tfrac{1}{2}\Omega\,[-\hat{D}\hat{U} + P^2\bar{\nabla}(\hat{\omega}P^{-1}) \\ &\quad + \bar{\hat{\omega}}\hat{f}^{-1}\hat{D}\hat{\omega} + \dot{P}P^{-1}\hat{f}\,] , \end{aligned} \tag{3.2}$$

where the quantities $\hat{\omega}$, $\hat{U}$, $\hat{\nu}$, etc., are given by Eqs. (2.9). Once P, $\hat{\Psi}_2$, and Φ_{00} are determined as functions of the coordinates by solving the reduced equations (3.2) together with the source equations (e.g., Maxwell's equations if the source is an electromagnetic field), Eqs. (3.1), (3.2), and the source variables will constitute a solution to Einstein's equations. The contravariant components of the metric tensor will be given by[2]

$$g^{11}=g^{13}=g^{14}=0,\quad g^{12}=1,\quad g^{22}=2(U-\omega\bar{\omega}),$$
$$g^{2i}=X^i-(\xi^i\bar{\omega}+\bar{\xi}^i\omega),\quad g^{ij}=-(\xi^i\bar{\xi}^j+\xi^j\bar{\xi}^i)\quad (i,j=3,4).$$

Of course, the conformal factor Ω must be expressed in terms of the radial coordinate r. The equations

$$\frac{\partial\Omega}{\partial r}\,(=D\Omega=\Omega^2\hat{D}\Omega)=\Omega^2\hat{f} \tag{3.3}$$

and

$$\Phi_{00}\,(=\Omega^3\hat{D}\hat{f}=\Omega D\hat{f})=\Omega\frac{\partial\hat{f}}{\partial r}$$

imply that

$$\Phi_{00}(r,u,x^3x^4)=\frac{1}{\Omega}\frac{\partial^2\Omega}{\partial r^2}-\frac{2}{\Omega^2}\left(\frac{\partial\Omega}{\partial r}\right)^2. \tag{3.4}$$

Therefore, once Φ_{00} is known, $\Omega\,(u,r,x^3,x^4)$ is found as a solution to Eq. (3.4) with boundary condition $\lim_{r\to\infty}\Omega=0$. The variables $\hat{f}$ and $\hat{T}$ are then found from Eqs. (2.7) and (2.8), respectively.

The reduced equations (3.2) involve integrals of $\hat{\Psi}_2$. But by suitable differentiation they can be changed quite readily to purely differential equations. However, if we restrict ourselves to type III (or N) space–times then $\hat{\Psi}_2$ vanishes and the equations are purely differential as they stand.

Since the spin coefficient π does not vanish one might want to change to a parallelly propagated frame by means of a null rotation using Eqs. (A.3) of Ref. 5. This will be done when approximate solutions are considered in Sec. 5.

4. $\Phi_{00}=\Lambda=0$

When the component Φ_{00} of the Ricci tensor is chosen to be zero the results of the previous section simplify considerably. The metric variables $\hat{f}$ and $\hat{\omega}$ become -1 and 0, respectively, and $T=\Omega=r^{-1}$. Differentiating the last of Eqs. (3.2) twice with respect to Ω changes it into the following differential equation for $\hat{\Psi}_2$:

$$\Omega^2\frac{\partial^2\hat{\Psi}_2}{\partial\Omega^2}-2\Omega\frac{\partial\hat{\Psi}_2}{\partial\Omega}+2\hat{\Psi}_2=-2\frac{\partial^2\Lambda}{\partial\Omega^2}. \tag{4.1}$$

If, for simplicity, we demand that Λ be zero, then the solution of Eq. (4.1) is given by

$$\hat{\Psi}_2=\hat{\Psi}_2^{(1)}\Omega+\hat{\Psi}_2^{(2)}\Omega^2,$$

where $\hat{\Psi}_2^{(1)}$ and $\hat{\Psi}_2^{(2)}$ are "constants" of integration. The NP variables for $\hat{M}$ now simplify considerably and are given by Eqs. (2.4) and by

$$\begin{aligned}
&\hat{\omega}=0,\quad \hat{f}=-1,\\
&\hat{U}=-\dot{P}P^{-1}\Omega+\tfrac{1}{2}K\Omega^2+\hat{\Psi}_2^{(1)}\Omega^3+\tfrac{1}{2}\hat{\Psi}_2^{(2)}\Omega^4,\\
&\hat{\gamma}=\tfrac{1}{2}\dot{P}P^{-1}-\tfrac{1}{2}K\Omega-\tfrac{3}{2}\hat{\Psi}_2^{(1)}\Omega^2-\hat{\Psi}_2^{(2)}\Omega^3,\\
&\hat{\nu}=-P\bar{\nabla}(\dot{P}P^{-1})\Omega+\tfrac{1}{2}P(\bar{\nabla}K)\Omega^2+P\bar{\nabla}\hat{\Psi}_2^{(1)}\Omega^3+\tfrac{1}{2}P\bar{\nabla}\hat{\Psi}_2^{(2)}\Omega^4,\\
&\hat{\Psi}_2=\hat{\Psi}_2^{(1)}\Omega+\hat{\Psi}_2^{(2)}\Omega^2,\\
&\hat{\Psi}_3=-\tfrac{1}{2}P(\bar{\nabla}K)\Omega-\tfrac{3}{2}P\bar{\nabla}\hat{\Psi}_2^{(1)}\Omega^2-P\bar{\nabla}\hat{\Psi}_2^{(2)}\Omega^3,\\
&\hat{\Psi}_4=-\bar{\nabla}[P^2\bar{\nabla}(\dot{P}P^{-1})]\Omega+\tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}K)\Omega^2\\
&\qquad+\bar{\nabla}(P^2\bar{\nabla}\hat{\Psi}_2^{(1)})\Omega^3+\tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}\hat{\Psi}_2^{(2)})\Omega^4,\\
&\hat{\Phi}_{12}=P\nabla(\dot{P}P^{-1})-\tfrac{1}{2}P(\nabla K)\Omega\\
&\qquad-\tfrac{3}{2}P\nabla\hat{\Psi}_2^{(1)}\Omega^2-P\nabla\hat{\Psi}_2^{(2)}\Omega^3,\\
&\hat{\Phi}_{22}=(\ln P)_{,11}-[\dot{P}P^{-1}K+P^2\nabla\bar{\nabla}(\dot{P}P^{-1})]\Omega\\
&\qquad+(\tfrac{1}{2}P^2\nabla\bar{\nabla}K-3\dot{P}P^{-1}\hat{\Psi}_2^{(1)})\Omega^2\\
&\qquad+(P^2\nabla\bar{\nabla}\hat{\Psi}_2^{(1)}-2\dot{P}P^{-1}\hat{\Psi}_2^{(2)})\Omega^3\\
&\qquad+\tfrac{1}{2}P^2\nabla\bar{\nabla}\hat{\Psi}_2^{(2)}\Omega^4.
\end{aligned} \tag{4.2}$$

The NP variables in the actual space–time M become

$$\begin{aligned}
&\kappa=\sigma=\epsilon=\pi=\tau=\lambda=\omega=X^i\\
&\quad=\Psi_0=\Psi_1=\Lambda=\Phi_{00}=\Phi_{01}=\Phi_{02}=0,\\
&\rho=-r^{-1},\quad \alpha=-\bar{\beta}=\tfrac{1}{2}(\bar{\nabla}P)r^{-1},\\
&\mu=-\tfrac{1}{2}Kr^{-1}-\hat{\Psi}_2^{(1)}r^{-2}-\tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-3},\\
&\gamma=-\tfrac{1}{2}\dot{P}P^{-1}-\tfrac{1}{2}\hat{\Psi}_2^{(1)}r^{-2}-\tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-3},\\
&\nu=-P\bar{\nabla}(\dot{P}P^{-1})+\tfrac{1}{2}P(\bar{\nabla}K)r^{-1}\\
&\qquad+P\bar{\nabla}\hat{\Psi}_2^{(1)}r^{-2}+\tfrac{1}{2}P\bar{\nabla}\hat{\Psi}_2^{(2)}r^{-3},\\
&\xi^3=-i\xi^4=Pr^{-1},\\
&U=r\dot{P}P^{-1}-\tfrac{1}{2}K-\hat{\Psi}_2^{(1)}r^{-1}-\tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-2},\\
&\Psi_2=\hat{\Psi}_2^{(1)}r^{-3}+\hat{\Psi}_2^{(2)}r^{-4}=\bar{\Psi}_2,\\
&\Psi_3=-\tfrac{1}{2}P(\bar{\nabla}K)r^{-2}-\tfrac{3}{2}P\bar{\nabla}\hat{\Psi}_2^{(1)}r^{-3}-P\bar{\nabla}\hat{\Psi}_2^{(2)}r^{-4},\\
&\Psi_4=-\bar{\nabla}(P^2\bar{\nabla}(\dot{P}P^{-1}))r^{-1}+\tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}K)r^{-2}\\
&\qquad+\bar{\nabla}(P^2\bar{\nabla}\hat{\Psi}_2^{(1)})r^{-3}+\tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}\hat{\Psi}_2^{(2)})r^{-4},
\end{aligned} \tag{4.3}$$

and

$$\begin{aligned}
&\Phi_{11}=\tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-4},\\
&\Phi_{12}=-\tfrac{1}{2}P\nabla\hat{\Psi}_2^{(1)}r^{-3}-\tfrac{1}{2}P\nabla\hat{\Psi}_2^{(2)}r^{-4},\\
&\Phi_{22}=r^{-2}(\tfrac{1}{2}P^2\nabla\bar{\nabla}K-3\dot{P}P^{-1}\hat{\Psi}_2^{(1)}+\hat{\dot{\Psi}}{}_2^{(1)})\\
&\qquad+r^{-3}(P^2\nabla\bar{\nabla}\hat{\Psi}_2^{(1)}-2\dot{P}P^{-1}\hat{\Psi}_2^{(2)}+\tfrac{1}{2}\hat{\dot{\Psi}}{}_2^{(2)})\\
&\qquad+r^{-4}(\tfrac{1}{2}P^2\nabla\bar{\nabla}\hat{\Psi}_2^{(2)}).
\end{aligned} \tag{4.4}$$

Note that the frame is parallelly propagated and that there is no need for a null rotation. With some effort Eqs. (4.3) and (4.4) could also have been obtained by specializing the results of Ref. 8 to the twist-free case.

If the source is a Maxwell field satisfying the source-free Maxwell equations, Eqs. (A8), then the ϕ_0 component of the field must vanish in order for Φ_{00} to be zero. That is, the repeated principal null direction of the space time must also be a principal null direction of the Maxwell field. Taking Eqs. (2.4) into account the conformally transformed Maxwell equations become

$$\begin{aligned}
&\hat{D}\hat{\phi}_1=0,\quad \hat{D}\hat{\phi}_2=\hat{\bar{\delta}}\hat{\phi}_1,\quad \hat{\delta}\hat{\phi}_1=0,\\
&\hat{\delta}\hat{\phi}_2=\hat{\Delta}\hat{\phi}_1-2\dot{P}P^{-1}\hat{\phi}_1+(\nabla P)\hat{\phi}_2.
\end{aligned} \tag{4.5}$$

The solution to the first two equations is

$$\hat{\phi}_1 = \hat{\phi}_1^0, \quad \hat{\phi}_2 = \hat{\phi}_2^0 - P(\bar{\nabla}\hat{\phi}_1^0)\Omega,$$

so that, according to Eqs. (A7),

$$\phi_1 = \hat{\phi}_1^0 r^{-2}, \quad \phi_2 = \hat{\phi}_2^0 r^{-1} - P(\bar{\nabla}\hat{\phi}_1^0)r^{-2}. \tag{4.6}$$

Hence

$$\begin{aligned}
&\Phi_{00} = \Phi_{01} = \Phi_{02} = \Lambda = 0, \\
&\Phi_{11} = |\hat{\phi}_1^0|^2 r^{-4}, \quad \Phi_{12} = \hat{\phi}_1^0 \bar{\hat{\phi}}_2^0 r^{-3} - \hat{\phi}_1^0 P(\nabla\bar{\hat{\phi}}_1^0)r^{-4}, \\
&\Phi_{22} = |\hat{\phi}_2^0|^2 r^{-2} - (\bar{\hat{\phi}}_2^0 P\bar{\nabla}\hat{\phi}_1^0 + \hat{\phi}_2^0 P\nabla\bar{\hat{\phi}}_1^0)r^{-3} \\
&\qquad + P^2|\nabla\bar{\hat{\phi}}_1^0|^2 r^{-4}.
\end{aligned} \tag{4.7}$$

The last two of Eqs. (4.5) and comparison of Eqs. (4.4) with (4.7) yield

$$\begin{aligned}
&\nabla\hat{\phi}_1^0 = 0, \quad \hat{\Psi}_2^{(2)} = 2|\hat{\phi}_1^0|^2, \quad P\nabla\hat{\Psi}_2^{(1)} = -2\hat{\phi}_1^0\bar{\hat{\phi}}_2^0, \\
&\dot{\hat{\phi}}_1 = 2\dot{P}P^{-1}\hat{\phi}_1^0 + P^2\nabla(P^{-1}\hat{\phi}_2^0), \\
&\dot{\hat{\Psi}}_2^{(1)} = 3\dot{P}P^{-1}\hat{\Psi}_2^{(1)} - \tfrac{1}{2}P^2\nabla\bar{\nabla}K + |\hat{\phi}_2^0|^2.
\end{aligned} \tag{4.8}$$

These equations must be solved for the variables P, $\hat{\Psi}_2^{(1)}$, $\hat{\Psi}_2^{(2)}$, ϕ_1^0, and ϕ_2^0 before Eqs. (4.3) and (4.6) constitute an actual solution. If we put ϕ_1^0 and ϕ_2^0 equal to zero then our solution reduces to that of Robinson and Trautman.[12] For this particular case Eqs. (4.8) were analyzed in some detail by Foster and Newman.[15]

5. ASYMPTOTIC BEHAVIOR

We wish to determine the asymptotic behavior of the NP quantities, i.e., of the metric tensor, the Weyl tensor, the Ricci tensor, and the spin coefficients. Space–time is assumed to be algebraically special with repeated principal null vectors that are tangent to hypersurface-orthogonal and shear-free null geodesics. The Ricci tensor is left as general as is compatible with the "almost asymptotic flatness" of the space–time. We cannot simply put $\kappa = \sigma = \Psi_0 = \Psi_1 = 0$ in the results of Ref. 5 for reasons explained in the Introduction. However, we need not start from scratch since most of the work has been done in Secs. 2 and 3. The "solution" obtained there is, unfortunately, very complicated (except in special circumstances). Therefore, it is instructive to make power series expansions in order to see the asymptotic behavior of the NP quantities.

Accordingly, let us start by expanding each variable in Eqs. (2.7)–(2.9) as a power series in the conformal factor Ω. A straightforward calculation leads to the following expressions for the careted NP quantities:

$$\begin{aligned}
\hat{\gamma} &= \tfrac{1}{2}\dot{P}P^{-1} - \tfrac{1}{2}K\Omega - \tfrac{3}{2}\hat{\Psi}_2^{(1)}\Omega^2 \\
&\quad - (\hat{\Psi}_2^{(2)} + \tfrac{1}{6}K\hat{f}^{(2)})\Omega^3 + O(\Omega^4), \\
\hat{\nu} &= -P\bar{\nabla}(\dot{P}P^{-1})\Omega + \tfrac{1}{2}P\bar{\nabla}K\Omega^2 \\
&\quad + \left[P\bar{\nabla}\hat{\Psi}_2^{(1)} - \tfrac{1}{3}P\hat{f}^{(2)}\bar{\nabla}(\dot{P}P^{-1})\right]\Omega^3 + O(\Omega^4), \\
\hat{\omega} &= -\tfrac{1}{3}P\nabla\hat{f}^{(2)}\Omega^3 - \tfrac{1}{4}P\nabla\hat{f}^{(3)}\Omega^4 + O(\Omega^5), \\
\hat{U} &= -\dot{P}P^{-1}\Omega + \tfrac{1}{2}K\Omega^2 \\
&\quad + (\hat{\Psi}_2^{(1)} - \tfrac{1}{3}\dot{\hat{f}}^{(2)} + \tfrac{2}{3}\hat{f}^{(2)}\dot{P}P^{-1})\Omega^3 \\
&\quad + (\tfrac{1}{2}\hat{\Psi}_2^{(2)} - \tfrac{1}{4}\dot{\hat{f}}^{(3)} - \tfrac{1}{6}K\hat{f}^{(2)} + \tfrac{3}{4}\dot{P}P^{-1}\hat{f}^{(3)})\Omega^4 \\
&\quad + O(\Omega^5), \\
T &= \Omega + \tfrac{1}{3}\hat{f}^{(2)}\Omega^3 + \tfrac{1}{4}\hat{f}^{(3)}\Omega^4 + O(\Omega^5), \\
\hat{\Psi}_2 &= \hat{\Psi}_2^{(1)}\Omega + \hat{\Psi}_2^{(2)}\Omega^2 + O(\Omega^3), \\
\hat{\Psi}_3 &= -\tfrac{1}{2}P\bar{\nabla}K\Omega - \tfrac{3}{2}P\bar{\nabla}\hat{\Psi}_2^{(1)}\Omega^2 + O(\Omega^3), \\
\hat{\Psi}_4 &= -\bar{\nabla}[P^2\bar{\nabla}(\dot{P}P^{-1})]\Omega \\
&\quad + \tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}K)\Omega^2 + O(\Omega^3), \\
\hat{\Phi}_{12} &= P\nabla(\dot{P}P^{-1}) - \tfrac{1}{2}P\nabla K\Omega - \tfrac{3}{2}P\nabla\hat{\Psi}_2^{(1)}\Omega^2 + O(\Omega^3), \\
\hat{\Phi}_{22} &= (\ln P)_{,11} - [P^2\nabla\bar{\nabla}(\dot{P}P^{-1}) + K\dot{P}P^{-1}]\Omega \\
&\quad + (\tfrac{1}{2}P^2\nabla\bar{\nabla}K - 3\dot{P}P^{-1}\hat{\Psi}_2^{(1)})\Omega^2 + O(\Omega^3), \\
\hat{f} &= -1 - \tfrac{1}{2}\Phi_{00}^{(4)}\Omega^2 - \tfrac{1}{3}\Phi_{00}^{(5)}\Omega^3 + O(\Omega^4).
\end{aligned} \tag{5.1}$$

Higher-order terms could easily be calculated but they become rather lengthy in general. The remaining careted variables are given by Eq. (2.4).

Equation (3.3) determines Ω as a function of (r,u,x^3,x^4):

$$\Omega = r^{-1} - \hat{f}^{(2)}r^{-3} - \tfrac{1}{2}\hat{f}^{(3)}r^{-4} + O(r^{-5}). \tag{5.2}$$

The uncareted variables in Eqs. (3.1) and (3.2) may now be written as power series in r^{-1}. Since the spin coefficient π does not vanish we transform to a parallelly propagated frame, where it does. The appropriate null rotation parameter is given by

$$c = \tfrac{1}{6}P\bar{\nabla}\hat{f}^{(2)}r^{-2} + \tfrac{1}{12}P\bar{\nabla}\hat{f}^{(3)}r^{-3} + O(r^{-4}). \tag{5.3}$$

The transformation laws for this null rotation can be found in the Appendix of Ref. 5 for all quantities except the metric variables. For the latter they are

$$\begin{aligned}
&\xi^{i\prime} = \xi^i, \quad \omega' = \omega + \bar{c}, \\
&X^{i\prime} = X^i + c\xi^i + \bar{c}\bar{\xi}^i, \\
&U' = U + c\omega + \bar{c}\bar{\omega} - c\bar{c}.
\end{aligned}$$

In this new frame the uncareted variables have the following expansions:

$$\begin{aligned}
\kappa &= \sigma = \epsilon = \pi = \Psi_0 = \Psi_1 = 0, \\
\rho &= -r^{-1} + 2\hat{f}^{(2)}r^{-3} + \tfrac{3}{2}\hat{f}^{(3)}r^{-4} + O(r^{-5}), \\
\tau &= -\tfrac{1}{2}P\nabla\hat{f}^{(2)}r^{-3} - \tfrac{1}{3}P\nabla\hat{f}^{(3)}r^{-4} + O(r^{-5}), \\
\alpha &= \tfrac{1}{2}\bar{\nabla}Pr^{-1} - \tfrac{1}{2}[\bar{\nabla}(P\hat{f}^{(2)})]r^{-3} + O(r^{-4}), \\
\beta &= -\tfrac{1}{2}(\nabla P)[r^{-1} - \hat{f}^{(2)}r^{-3} + O(r^{-4})], \\
\lambda &= \tfrac{1}{6}\bar{\nabla}(P^2\bar{\nabla}\hat{f}^{(2)})r^{-3} + O(r^{-4}), \\
\mu &= -\tfrac{1}{2}Kr^{-1} - (\hat{\Psi}_2^{(1)} - \tfrac{1}{3}\dot{\hat{f}}^{(2)} + \tfrac{2}{3}\hat{f}^{(2)}\dot{P}P^{-1})r^{-2} \\
&\quad + (\tfrac{2}{3}K\hat{f}^{(2)} - \tfrac{1}{2}\hat{\Psi}_2^{(2)} + \tfrac{1}{4}\dot{\hat{f}}^{(3)} - \tfrac{3}{4}\dot{P}P^{-1}\hat{f}^{(3)} \\
&\quad + \tfrac{1}{6}P^2\nabla\bar{\nabla}\hat{f}^{(2)})r^{-3} + O(r^{-4}), \\
\gamma &= -\tfrac{1}{2}\dot{P}P^{-1} + (\tfrac{2}{3}\dot{P}P^{-1}\hat{f}^{(2)} - \tfrac{1}{3}\dot{\hat{f}}^{(2)} - \tfrac{1}{2}\hat{\Psi}_2^{(1)})r^{-2} \\
&\quad + \left[\tfrac{3}{4}\dot{P}P^{-1}\hat{f}^{(3)} - \tfrac{1}{3}K\hat{f}^{(2)} - \tfrac{1}{4}\dot{\hat{f}}^{(3)} - \tfrac{1}{2}\hat{\Psi}_2^{(2)} \right. \\
&\quad \left. + \tfrac{1}{12}P(\bar{\nabla}P)\nabla\hat{f}^{(2)} - \tfrac{1}{12}P(\nabla P)\bar{\nabla}\hat{f}^{(2)}\right]r^{-3} \\
&\quad + O(r^{-4}), \\
\nu &= -P\bar{\nabla}(\dot{P}P^{-1}) + \tfrac{1}{2}P\bar{\nabla}Kr^{-1} \\
&\quad + \left[P\bar{\nabla}\hat{\Psi}_2^{(1)} - \tfrac{1}{3}P\bar{\nabla}(\dot{P}P^{-1}\hat{f}^{(2)}) + \tfrac{1}{6}P\bar{\nabla}\dot{\hat{f}}^{(2)}\right]r^{-2} \\
&\quad + O(r^{-3}), \\
\xi^i &= \hat{\xi}^{i0}[r^{-1} - \hat{f}^{(2)}r^{-3} + O(r^{-4})], \\
X^i &= \left[\tfrac{1}{6}P\hat{\xi}^{i0}\bar{\nabla}\hat{f}^{(2)} + \tfrac{1}{6}P\bar{\hat{\xi}}^{i0}\nabla\hat{f}^{(2)}\right]r^{-3} + O(r^{-4}), \\
\omega &= -\tfrac{1}{2}P\nabla\hat{f}^{(2)}r^{-2} + O(r^{-3}), \\
U &= \dot{P}P^{-1}r - \tfrac{1}{2}K - (\hat{\Psi}_2^{(1)} + \tfrac{2}{3}\dot{\hat{f}}^{(2)} - \tfrac{4}{3}\dot{P}P^{-1}\hat{f}^{(2)})r^{-1}
\end{aligned} \tag{5.4}$$

$$
\begin{aligned}
&-(\tfrac{1}{2}\hat{\Psi}_2^{(2)} + \tfrac{1}{4}\hat{\dot{f}}^{(3)} + \tfrac{1}{3}K\hat{f}^{(2)} - \tfrac{3}{4}\dot{P}P^{-1}\hat{f}^{(3)})r^{-2} \\
&+ O(r^{-3}), \\
\Psi_2 &= \hat{\Psi}_2^{(1)}r^{-3} + \hat{\Psi}_2^{(2)}r^{-4} + O(r^{-5}), \\
\Psi_3 &= -\tfrac{1}{2}P\bar{\nabla}Kr^{-2} - \tfrac{3}{2}P\bar{\nabla}\hat{\Psi}_2^{(1)}r^{-3} + O(r^{-4}), \\
\Psi_4 &= -\bar{\nabla}(P^2\bar{\nabla}(\dot{P}P^{-1}))r^{-1} + \tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}K)r^{-2} \\
&\quad + O(r^{-3}),
\end{aligned}
$$

and

$$
\begin{aligned}
\Phi_{01} &= P\nabla\hat{f}^{(2)}r^{-4} + P\nabla\hat{f}^{(3)}r^{-5} + O(r^{-6}), \\
\Phi_{02} &= -\tfrac{1}{3}\nabla(P^2\nabla\hat{f}^{(2)})r^{-4} + O(r^{-5}), \\
\Phi_{11} &= (\tfrac{1}{2}\hat{\dot{f}}^{(2)} - \hat{f}^{(2)}\dot{P}P^{-1})r^{-3} \\
&\quad + (\tfrac{1}{2}\hat{\Psi}_2^{(2)} - \tfrac{1}{6}P^2\bar{\nabla}\nabla\hat{f}^{(2)} + \tfrac{1}{2}\hat{\dot{f}}^{(3)} - \tfrac{3}{2}\dot{P}P^{-1}\hat{f}^{(3)} \\
&\quad + \tfrac{5}{6}K\hat{f}^{(2)})r^{-4} + O(r^{-5}), \qquad (5.5) \\
\Phi_{12} &= [-\tfrac{1}{2}P\nabla\hat{\Psi}_2^{(1)} - \tfrac{1}{3}P\nabla\hat{\dot{f}}^{(2)} + \tfrac{2}{3}P\nabla(\dot{P}P^{-1}\hat{f}^{(2)})]r^{-3} \\
&\quad + O(r^{-4}), \\
\Phi_{22} &= [\tfrac{1}{2}P^2\nabla\bar{\nabla}K - 3\dot{P}P^{-1}\hat{\Psi}_2^{(1)} + \hat{\dot{\Psi}}_2^{(1)} - \tfrac{1}{3}\hat{\ddot{f}}^{(2)} \\
&\quad + \tfrac{5}{3}\hat{\dot{f}}^{(2)}\dot{P}P^{-1}] + \tfrac{2}{3}\hat{f}^{(2)}(\dot{P}P^{-1})_{,1} \\
&\quad - 2\hat{f}^{(2)}\dot{P}^2P^{-2}]r^{-2} + O(r^{-3}), \\
\Lambda &= (-\tfrac{1}{6}\hat{\dot{f}}^{(2)} + \tfrac{1}{3}\dot{P}P^{-1}\hat{f}^{(2)})r^{-3} \\
&\quad + (\tfrac{3}{4}\dot{P}P^{-1}\hat{f}^{(3)} - \tfrac{1}{4}\hat{\dot{f}}^{(3)} - \tfrac{1}{6}K\hat{f}^{(2)} \\
&\quad - \tfrac{1}{6}P^2\bar{\nabla}\nabla\hat{f}^{(2)})r^{-4} + O(r^{-5}),
\end{aligned}
$$

where $\hat{f}^{(2)} = -\tfrac{1}{2}\Phi_{00}^{(4)}$ and $\hat{f}^{(3)} = -\tfrac{1}{3}\Phi_{00}^{(5)}$. Equations (5.5) are the reduced equations. Their left-hand sides must be equated to the energy–momentum tensor and the resultant equations must be solved in conjuction with the source equations.

6. APPROXIMATE EINSTEIN–MAXWELL SOLUTIONS

In this section the source is assumed to be an electromagnetic field satisfying the source-free Maxwell equations (4.5). The physical and the unphysical field variables are related as follows:

$$
\phi_0 = \Omega^3\hat{\phi}_0, \quad \phi_1 = \Omega^2\hat{\phi}_1, \quad \phi_2 = \Omega\hat{\phi}_2. \qquad (6.1)
$$

Since the Ricci tensor is related to the Maxwell field by $\Phi_{\alpha\beta} = \phi_\alpha\bar{\phi}_\beta$ ($\alpha,\beta = 0,1,2$) we find quite easily that $\Phi_{00} = 0(\Omega^6)$ and hence that $\hat{f}^{(2)}$ and $\hat{f}^{(3)}$ vanish. Many terms in Eqs. (5.1)–(5.5) now disappear and it becomes easier to calculate higher-order terms. The results are

$$
\begin{aligned}
\hat{f} &= -1 - \tfrac{1}{4}\Phi_{00}^{(6)}\Omega^4 - \tfrac{1}{5}\Phi_{00}^{(7)}\Omega^5 + O(\Omega^6), \\
T &= \Omega + \tfrac{1}{5}\hat{f}^{(4)}\Omega^5 + \tfrac{1}{6}\hat{f}^{(5)}\Omega^6 + O(\Omega^7), \\
\Omega &= r^{-1} - \tfrac{1}{3}\hat{f}^{(4)}r^{-5} - \tfrac{1}{4}\hat{f}^{(5)}r^{-6} + O(r^{-7}), \\
c &= \tfrac{1}{20}P\bar{\nabla}\hat{f}^{(4)}r^{-4} + \tfrac{1}{30}P\bar{\nabla}\hat{f}^{(5)}r^{-5} + O(r^{-6}), \\
\kappa &= \sigma = \epsilon = \pi = \Psi_0 = \Psi_1 = \Lambda = 0, \\
\rho &= -r^{-1} - \tfrac{1}{3}\Phi_{00}^{(6)}r^{-5} - \tfrac{1}{4}\Phi_{00}^{(7)}r^{-6} + O(r^{-7}), \\
\alpha &= \tfrac{1}{2}(\bar{\nabla}P)r^{-1} - (\tfrac{1}{6}\hat{f}^{(4)}\bar{\nabla}P + \tfrac{1}{4}P\bar{\nabla}\hat{f}^{(4)})r^{-5} \\
&\quad - (\tfrac{1}{8}\hat{f}^{(5)}\bar{\nabla}P + \tfrac{1}{5}P\bar{\nabla}\hat{f}^{(5)})r^{-6} + O(r^{-7}), \\
\beta &= \tfrac{1}{2}(\nabla P)[-r^{-1} + \tfrac{1}{3}\hat{f}^{(4)}r^{-5} + \tfrac{1}{4}\hat{f}^{(5)}r^{-6} + O(r^{-7})], \\
\tau &= -\tfrac{1}{4}P\nabla\hat{f}^{(4)}r^{-5} - \tfrac{1}{5}P\nabla\hat{f}^{(5)}r^{-6} + O(r^{-7}), \\
\gamma &= -\tfrac{1}{2}(\dot{P}P^{-1}) - \tfrac{1}{2}\hat{\Psi}_2^{(1)}r^{-2} - \tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-3} \\
&\quad + [-\tfrac{9}{20}\hat{\Psi}_2^{(3)} - \tfrac{1}{5}\hat{\dot{f}}^{(4)} + \tfrac{4}{5}\dot{P}P^{-1}\hat{f}^{(4)}]r^{-4} + O(r^{-5}), \\
\mu &= -\tfrac{1}{2}Kr^{-1} - \hat{\Psi}_2^{(1)}r^{-2} - \tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-3} \\
&\quad - [\tfrac{3}{10}\hat{\Psi}_2^{(3)} - \tfrac{1}{5}\hat{\dot{f}}^{(4)} + \tfrac{4}{5}\dot{P}P^{-1}\hat{f}^{(4)}]r^{-4} + O(r^{-5}), \\
\lambda &= \tfrac{1}{20}\bar{\nabla}(P^2\bar{\nabla}\hat{f}^{(4)})r^{-5} + \tfrac{1}{30}\bar{\nabla}(P^2\bar{\nabla}\hat{f}^{(5)})r^{-6} + O(r^{-7}), \\
\nu &= -P\bar{\nabla}(\dot{P}P^{-1}) + \tfrac{1}{2}P(\bar{\nabla}K)r^{-1} + P\bar{\nabla}\hat{\Psi}_2^{(1)}r^{-2} \\
&\quad + \tfrac{1}{2}P\bar{\nabla}\hat{\Psi}_2^{(2)}r^{-3} + O(r^{-4}), \\
\Psi_2 &= \hat{\Psi}_2^{(1)}r^{-3} + \hat{\Psi}_2^{(2)}r^{-4} + \hat{\Psi}_2^{(3)}r^{-5} + O(r^{-6}), \\
\Psi_3 &= -\tfrac{1}{2}P(\bar{\nabla}K)r^{-2} - \tfrac{3}{2}P\bar{\nabla}\hat{\Psi}_2^{(1)}r^{-3} - P\bar{\nabla}\hat{\Psi}_2^{(2)}r^{-4} \\
&\quad - \tfrac{3}{4}P\bar{\nabla}\hat{\Psi}_2^{(3)}r^{-5} + O(r^{-6}), \\
\Psi_4 &= -\bar{\nabla}(P^2\bar{\nabla}(\dot{P}P^{-1}))r^{-1} + \tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}K)r^{-2} \qquad (6.2) \\
&\quad + \bar{\nabla}(P^2\bar{\nabla}\hat{\Psi}_2^{(1)})r^{-3} \\
&\quad + \tfrac{1}{2}\bar{\nabla}(P^2\bar{\nabla}\hat{\Psi}_2^{(2)})r^{-4} + O(r^{-5}), \\
\omega &= -\tfrac{1}{12}P\nabla\hat{f}^{(4)}r^{-4} - \tfrac{1}{20}P\nabla\hat{f}^{(5)}r^{-5} + O(r^{-6}), \\
\xi^i &= \hat{\xi}^{i0}[r^{-1} - \tfrac{1}{3}\hat{f}^{(4)}r^{-5} - \tfrac{1}{4}\hat{f}^{(5)}r^{-6} + O(r^{-7})] \\
&\qquad (i = 3,4), \\
X^i &= \tfrac{1}{20}(P\hat{\xi}^{i0}\bar{\nabla}\hat{f}^{(4)} + P\hat{\bar{\xi}}^{i0}\nabla\hat{f}^{(4)})r^{-5} \\
&\quad + \tfrac{1}{30}(P\hat{\xi}^{i0}\bar{\nabla}\hat{f}^{(5)} + P\hat{\bar{\xi}}^{i0}\nabla\hat{f}^{(5)})r^{-6} + O(r^{-7}), \\
U &= \dot{P}P^{-1}r - \tfrac{1}{2}K - \hat{\Psi}_2^{(1)}r^{-1} - \tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-2} \\
&\quad - (\tfrac{3}{10}\hat{\Psi}_2^{(3)} + \tfrac{2}{15}\hat{\dot{f}}^{(4)} - \tfrac{8}{15}\hat{f}^{(4)}\dot{P}P^{-1})r^{-3} + O(r^{-4}), \\
\Phi_{00} &= \Phi_{00}^{(6)}r^{-6} + \Phi_{00}^{(7)}r^{-7} + O(r^{-8}),
\end{aligned}
$$

and

$$
\begin{aligned}
\Phi_{01} &= P\nabla\hat{f}^{(4)}r^{-6} + P\nabla\hat{f}^{(5)}r^{-7} + O(r^{-8}), \\
\Phi_{02} &= -\tfrac{1}{5}\nabla(P^2\nabla\hat{f}^{(4)})r^{-6} - \tfrac{1}{6}\nabla(P^2\nabla\hat{f}^{(5)})r^{-7} + O(r^{-8}), \\
\Phi_{11} &= \tfrac{1}{2}\hat{\Psi}_2^{(2)}r^{-4} + (\tfrac{3}{4}\hat{\Psi}_2^{(3)} + \tfrac{1}{2}\hat{\dot{f}}^{(4)} - 2\hat{f}^{(4)}\dot{P}P^{-1})r^{-5} \\
&\quad + O(r^{-6}), \\
\Phi_{12} &= -\tfrac{1}{2}P\nabla\hat{\Psi}_2^{(1)}r^{-3} - \tfrac{1}{2}P\nabla\hat{\Psi}_2^{(2)}r^{-4} \\
&\quad + [-\tfrac{9}{20}P\nabla\hat{\Psi}_2^{(3)} - \tfrac{1}{5}P\nabla\hat{\dot{f}}^{(4)} + \tfrac{4}{5}P\nabla(\hat{f}^{(4)}\dot{P}P^{-1})]r^{-5} \\
&\quad + O(r^{-6}), \\
\Phi_{22} &= (\tfrac{1}{2}P^2\nabla\bar{\nabla}K - 3\dot{P}P^{-1}\hat{\Psi}_2^{(1)} + \dot{\Psi}_2^{(1)})r^{-2} \qquad (6.3) \\
&\quad + (P^2\nabla\bar{\nabla}\hat{\Psi}_2^{(1)} - 2\dot{P}P^{-1}\hat{\Psi}_2^{(2)} + \tfrac{1}{2}\hat{\dot{\Psi}}_2^{(2)})r^{-3} \\
&\quad + O(r^{-4}), \\
\Lambda &= -\tfrac{1}{20}r^{-5}(\hat{\Psi}_2^{(3)} + 6\hat{\dot{f}}^{(4)} - 24\dot{P}P^{-1}\hat{f}^{(4)}) + O(r^{-6}).
\end{aligned}
$$

Note that $\hat{f}^{(4)} = -\tfrac{1}{4}\Phi_{00}^{(6)}$ and $\hat{f}^{(5)} = -\tfrac{1}{5}\Phi_{00}^{(7)}$. If in Eq. (6.3), the parameter r^{-1} is replaced by the conformal factor Ω then, to the order shown, we get the correct expansions in terms of Ω.

If we now expand the first two careted Maxwell equations, Eqs. (4.5), and hence obtain the field variables in physical space–times by means of Eqs. (6.1), we find that

$$
\begin{aligned}
\phi_1 &= \hat{\phi}_1^0r^{-2} - P^2\bar{\nabla}(\hat{\phi}_0^0P^{-1})r^{-3} - \tfrac{1}{2}P^2\bar{\nabla}(\hat{\phi}_0^{(1)}P^{-1})r^{-4} \\
&\quad - \tfrac{1}{3}P^2\bar{\nabla}(\hat{\phi}_0^{(2)}P^{-1})r^{-5} + O(r^{-6}), \\
& \qquad (6.4) \\
\phi_2 &= \hat{\phi}_2^0r^{-1} - P(\bar{\nabla}\hat{\phi}_1^0)r^{-2} + \tfrac{1}{2}P\bar{\nabla}[P^2\bar{\nabla}(\hat{\phi}_0^0P^{-1})]r^{-3} \\
&\quad + \tfrac{1}{6}P\bar{\nabla}[P^2\bar{\nabla}(\hat{\phi}_0^{(1)}P^{-1})]r^{-4} + O(r^{-5}).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\phi_0 &= r^{-3}\hat\phi^0_0 + r^{-4}\hat\phi^{(1)}_0 + r^{-5}\hat\phi^{(2)}_0 + r^{-6}\hat\phi^{(3)}_0 \\
&\quad + r^{-7}(\hat\phi^{(4)}_0 - \hat\phi^0_0\hat f^{(4)}) + O(r^{-8}) .
\end{aligned}
$$

From the remaining two Maxwell equations we obtain

$$
\begin{aligned}
\hat{\dot\phi}_0 &= (P\nabla\hat\phi^0_1 + 2\dot P P^{-1}\hat\phi^0_0) + \Omega\,[3\dot P P^{-1}\hat\phi^{(1)}_0 \\
&\quad - K\hat\phi^0_0 - P\nabla[P^2\bar\nabla(\hat\phi^0_0 P^{-1})]] + O(\Omega^2) , \\
\hat{\dot\phi}_1 &= [2\dot P P^{-1}\hat\phi^0_1 + P^2\nabla(\hat\phi^0_2 P^{-1})] \\
&\quad - \Omega\,[P^2\nabla\bar\nabla\hat\phi^0_1 + 3\dot P P\bar\nabla(\hat\phi^0_0 P^{-1}) + P\bar\nabla(\dot P P^{-1})\hat\phi^0_0] \\
&\quad + O(\Omega^2) .
\end{aligned}
\tag{6.5}
$$

The equations $\Phi_{\alpha\beta} = \phi_\alpha\bar\phi_\beta$ $(\alpha, \beta = 0, 1, 2)$ lead to

$$
\hat f^{(4)} = -\tfrac{1}{4}\hat\phi^0_0\hat{\bar\phi}{}^0_0 , \quad \hat f^{(5)} = -\tfrac{1}{5}(\hat\phi^{(1)}_0\hat{\bar\phi}{}^0_0 + \hat\phi^0_0\hat{\bar\phi}{}^{(1)}_0) ,
$$

and

$$
\begin{aligned}
\Phi_{01} &= \hat\phi^0_0\hat{\bar\phi}{}^0_1\Omega^5 + [\hat\phi^{(1)}_0\hat{\bar\phi}{}^0_1 - \hat\phi^0_0 P^2\nabla(\hat{\bar\phi}{}^0_0 P^{-1})]\Omega^6 \\
&\quad + [\hat\phi^{(2)}_0\hat{\bar\phi}{}^0_1 - \hat\phi^{(1)}_0 P^2\nabla(\hat{\bar\phi}{}^0_0 P^{-1}) \\
&\quad - \tfrac{1}{2}P^2\hat\phi^0_0\nabla(\hat{\bar\phi}{}^{(1)}_0 P^{-1})]\Omega^7 + O(\Omega^8) , \\
\Phi_{02} &= \hat\phi^0_0\hat{\bar\phi}{}^0_2\Omega^4 + [\hat\phi^{(1)}_0\hat{\bar\phi}{}^0_2 - \hat\phi^0_0 P\nabla\hat{\bar\phi}{}^0_1]\Omega^5 \\
&\quad + [\hat\phi^{(2)}_0\hat{\bar\phi}{}^0_2 - \hat\phi^{(1)}_0 P\nabla\hat{\bar\phi}{}^0_1 + \tfrac{1}{2}P\hat\phi^0_0\nabla(P^2\nabla(\hat{\bar\phi}{}^0_0 P^{-1}))]\Omega^6 \\
&\quad + [\hat\phi^{(3)}_0\hat{\bar\phi}{}^0_2 - \hat\phi^{(2)}_0 P\nabla\hat{\bar\phi}{}^0_1 + \tfrac{1}{2}P\hat\phi^{(1)}_0\nabla(P^2\nabla(P^{-1}\bar\phi^0_0)) \\
&\quad + \tfrac{1}{6}P\hat\phi^0_0\nabla(P^2\nabla(\hat{\bar\phi}{}^{(1)}_0 P^{-1}))]\,\Omega^7 + O(\Omega^8) , \\
\Phi_{11} &= \hat\phi^0_1\hat{\bar\phi}{}^0_1\Omega^4 - [\hat\phi^0_1 P^2\nabla(P^{-1}\hat{\bar\phi}{}^0_0) \\
&\quad + \hat{\bar\phi}{}^0_1 P^2\bar\nabla(P^{-1}\hat\phi^0_0)]\Omega^5 + O(\Omega^6) , \\
\Phi_{12} &= \hat\phi^0_1\hat{\bar\phi}{}^0_2\Omega^3 - [\hat\phi^0_1 P\nabla\hat{\bar\phi}{}^0_1 + \hat{\bar\phi}{}^0_2 P^2\bar\nabla(\hat\phi^0_0 P^{-1})]\Omega^4 \\
&\quad + [\tfrac{1}{2}P\hat\phi^0_1\nabla(P^2\nabla(P^{-1}\hat{\bar\phi}{}^0_0)) + P^3\bar\nabla(P^{-1}\hat\phi^0_0)\nabla\hat{\bar\phi}{}^0_1 \\
&\quad - \tfrac{1}{2}P^2\hat{\bar\phi}{}^0_2\bar\nabla(\hat\phi^{(1)}_0 P^{-1})]\Omega^5 + O(\Omega^6) , \\
\Phi_{22} &= \hat\phi^0_2\hat{\bar\phi}{}^0_2\Omega^2 - [\hat\phi^0_2 P\nabla\hat{\bar\phi}{}^0_1 + \hat{\bar\phi}{}^0_2 P\bar\nabla\hat\phi^0_1]\Omega^3 + O(\Omega^4) .
\end{aligned}
\tag{6.6}
$$

Equations (6.6) must be equated to Eqs. (6.3) and the resultant equations must be solved together with $\Lambda = 0$ and Eqs. (6.5) for the remaining variables before Eqs. (6.2) and (6.4) constitute an actual solution.

Comparing the two expressions for each of Φ_{01} and Φ_{02} we see that $\hat\phi^0_0\hat{\bar\phi}{}^0_1 = 0$ and $\hat\phi^0_0\hat{\bar\phi}{}^0_2 = 0$. Therefore, if $\hat\phi^0_0 \neq 0$ then $\phi_1 = O(r^{-3})$ and $\phi_2 = O(r^{-2})$. This can be generalized to show that as long as ϕ_0 does not vanish identically the same result follows. To prove this let us assume that $\hat\phi_0 = \hat\phi^{(m)}_0\Omega^m + O(\Omega^{m+1})$ with $\hat\phi^{(m)}_0 \neq 0$ for some nonnegative integer m. Then we find from Eqs. (2.7) and (3.2) that

$$
\hat f = -1 - (4+2m)^{-1}\Omega^{4+2m}|\hat\phi^{(m)}_0|^2 + O(\Omega^{5+2m}) ,
$$

and

$$
\begin{aligned}
\Phi_{01} &= \Omega^2\hat D\hat\omega = -(5+2m)\hat\omega^{(5+2m)}\Omega^{6+2m} \\
&\qquad + O(\Omega^{7+2m}) , \\
\Phi_{02} &= \Omega\,(\hat\delta\hat\omega + 2\hat{\bar\alpha}{}^0\hat\omega) = \nabla(P\hat\omega^{(5+2m)})\Omega^{6+2m} \\
&\qquad + O(\Omega^{7+2m}) .
\end{aligned}
\tag{6.7}
$$

On the other hand,

$$
\begin{aligned}
\Phi_{01} &= \Omega^5\hat\phi_0\hat{\bar\phi}_1 = \Omega^{5+m}\hat\phi^{(m)}_0\hat{\bar\phi}{}^0_1 + O(\Omega^{6+m}) , \\
\Phi_{02} &= \Omega^4\hat\phi_0\hat{\bar\phi}_2 = \Omega^{4+m}\hat\phi^{(m)}_0\hat{\bar\phi}{}^0_2 + O(\Omega^{5+m}) .
\end{aligned}
$$

Comparison of Eqs. (6.7) and (6.8) leads to the desired conclusion.

ACKNOWLEDGMENT

This work was supported in part by a grant from the Natural Sciences/Engineering Research Council of Canada.

APPENDIX

For easy reference we collect here various formulas and equations that have been used in this paper.

The differential operators $\hat D$, $\hat\delta$, and $\hat\Delta$ are given in terms of the chosen frame and coordinate system by[5]

$$
\begin{aligned}
\hat D &= \hat f\frac{\partial}{\partial\Omega} , \quad \hat\delta = \hat\omega\frac{\partial}{\partial\Omega} + \hat\xi^i\frac{\partial}{\partial x^i} , \\
\hat\Delta &= \frac{\partial}{\partial u} + \hat U\frac{\partial}{\partial\Omega} + \hat X^i\frac{\partial}{\partial x^i} .
\end{aligned}
$$

Similarly, in the physical space–time

$$
\begin{aligned}
D &= \frac{\partial}{\partial r} , \quad \delta = \omega\frac{\partial}{\partial r} + \xi^i\frac{\partial}{\partial x^i} , \\
\Delta &= \frac{\partial}{\partial u} + U\frac{\partial}{\partial r} + X^i\frac{\partial}{\partial x^i} .
\end{aligned}
$$

The metric equations in $\hat M$ are[5]

$$
\begin{aligned}
&\hat D\hat\xi^i = \hat\sigma\hat{\bar\xi}{}^i , \quad \hat D\hat\omega = \hat\delta\hat f + \hat\sigma\hat{\bar\omega} - \hat\tau\hat f , \\
&\hat D\hat X^i = \hat\tau\hat{\bar\xi}{}^i + \hat{\bar\tau}\hat\xi^i , \\
&\hat D\hat U = \hat\Delta\hat f + \hat\tau\hat{\bar\omega} + \hat{\bar\tau}\hat\omega - \hat f(\hat\gamma + \hat{\bar\gamma}) , \\
&\hat\delta\hat X^i - \hat\Delta\hat\xi^i = (\hat\mu + \hat{\bar\gamma} - \hat\gamma)\hat\xi^i + \hat{\bar\lambda}\hat{\bar\xi}{}^i , \\
&\hat\delta\hat{\bar\xi}{}^i - \hat{\bar\delta}\hat\xi^i = (\hat{\bar\tau} - 2\hat\alpha)\hat\xi^i + (2\hat{\bar\alpha} - \hat\tau)\hat{\bar\xi}{}^i , \\
&\hat\delta\hat{\bar\omega} - \hat{\bar\delta}\hat\omega = (\hat{\bar\tau} - 2\hat\alpha)\hat\omega + (2\hat{\bar\alpha} - \hat\tau)\hat{\bar\omega} + \hat f(\hat\mu - \hat{\bar\mu}) , \\
&\hat\delta\hat U - \hat\Delta\hat\omega = (\hat\mu - \hat\gamma + \hat{\bar\gamma})\hat\omega + \hat{\bar\lambda}\hat{\bar\omega} - \hat{\bar\nu}\hat f .
\end{aligned}
\tag{A1}
$$

Next we list for various NP quantities the transformation laws between the physical space–time M and the rescaled space–time $\hat M$. They are, respectively, for the components of the Ricci tensor:

$$
\begin{aligned}
\Phi_{00} &= \Omega^4\hat\Phi_{00} + \Omega^3[\hat D\hat D\Omega - (\hat\epsilon + \hat{\bar\epsilon})\hat D\Omega + \hat\kappa\hat{\bar\delta}\Omega + \hat{\bar\kappa}\hat\delta\Omega\,] , \\
\Phi_{01} &= \Omega^3\hat\Phi_{01} + \Omega^2[\hat D\hat\delta\Omega - \hat{\bar\pi}\hat D\Omega + (\hat{\bar\epsilon} - \hat\epsilon)\hat\delta\Omega + \hat\kappa\hat\Delta\Omega\,] , \\
\Phi_{02} &= \Omega^2\hat\Phi_{02} + \Omega\,[\hat\delta\hat\delta\Omega - \hat{\bar\lambda}\hat D\Omega + (\hat{\bar\alpha} - \hat\beta)\hat\delta\Omega + \hat\sigma\hat\Delta\Omega\,] , \\
\Phi_{11} &= \Omega^2\hat\Phi_{11} + \tfrac{1}{2}\Omega\,[\hat D\hat\Delta\Omega + \hat{\bar\delta}\hat\delta\Omega - \hat{\bar\mu}\hat D\Omega - \hat{\bar\pi}\hat{\bar\delta}\Omega \\
&\qquad + (\hat{\bar\beta} - \hat\alpha - \hat\pi)\hat\delta\Omega + (\hat\rho + \hat\epsilon + \hat{\bar\epsilon})\hat\Delta\Omega\,] , \\
\Phi_{12} &= \Omega\hat\Phi_{12} + \hat\delta\hat\Delta\Omega - \hat\mu\hat\delta\Omega - \hat{\bar\lambda}\hat{\bar\delta}\Omega + (\hat{\bar\alpha} + \hat\beta)\hat\Delta\Omega , \\
\Phi_{22} &= \hat\Phi_{22} + \Omega^{-1}[\hat\Delta\hat\Delta\Omega - \hat\nu\hat\delta\Omega - \hat{\bar\nu}\hat{\bar\delta}\Omega + (\hat\gamma + \hat{\bar\gamma})\hat\Delta\Omega\,] , \\
\Lambda &= \Omega^2\hat\Lambda + (\hat D\Omega)(\hat\Delta\Omega) - (\hat\delta\Omega)(\hat{\bar\delta}\Omega) \\
&\qquad + \tfrac{1}{2}\Omega\,[\hat{\bar\delta}\hat\delta\Omega - \hat D\hat\Delta\Omega - \hat{\bar\mu}\hat D\Omega + (\hat\pi + \hat{\bar\beta} - \hat\alpha)\hat\delta\Omega \\
&\qquad + \hat{\bar\pi}\hat{\bar\delta}\Omega + (\hat\rho - \hat\epsilon - \hat{\bar\epsilon})\hat\Delta\Omega\,] ,
\end{aligned}
\tag{A2}
$$

for the components of the Weyl tensor:

$$
\begin{aligned}
&\hat\Psi_0 = \Omega^{-4}\Psi_0 , \quad \hat\Psi_1 = \Omega^{-3}\Psi_1 , \\
&\hat\Psi_2 = \Omega^{-2}\Psi_2 , \quad \hat\Psi_3 = \Omega^{-1}\Psi_3 , \quad \hat\Psi_4 = \Psi_4 ,
\end{aligned}
\tag{A3}
$$

for the spin coefficients:

$$
\begin{aligned}
&\hat\kappa = \Omega^{-3}\kappa , \quad \hat\sigma = \Omega^{-2}\sigma , \quad \hat\epsilon = \Omega^{-2}\epsilon , \\
&\hat\beta = \Omega^{-1}\beta , \quad \hat\lambda = \lambda , \quad \hat\nu = \Omega\nu , \\
&\hat\rho = \Omega^{-2}(\rho - D\ln\Omega) , \quad \hat\tau = \Omega^{-1}(\tau - \delta\ln\Omega) ,
\end{aligned}
\tag{A4}
$$

$$\hat{\pi} = \Omega^{-1}(\pi + \bar{\delta}\ln\Omega), \quad \hat{\alpha} = \Omega^{-1}(\alpha - \bar{\delta}\ln\Omega), \tag{A4}$$

$$\hat{\mu} = \mu + \Delta \ln\Omega, \quad \hat{\gamma} = \gamma - \Delta \ln\Omega,$$

for the differential operators:

$$\hat{D} = \Omega^{-2}D, \quad \hat{\delta} = \Omega^{-1}\delta, \quad \hat{\Delta} = \Delta, \tag{A5}$$

for the metric variables:

$$\xi^i = \Omega\hat{\xi}^i, \quad \omega = \Omega^{-1}\hat{f}^{-1}\left(\hat{\omega} - \hat{\xi}^i\frac{\partial\Omega}{\partial x^i}\right), \quad X^i = \hat{X}^i,$$

$$U = \Omega^{-2}\hat{f}^{-1}\left(\hat{U} - \frac{\partial\Omega}{\partial u} - \hat{X}^i\frac{\partial\Omega}{\partial x^i}\right), \tag{A6}$$

and for the components of the Maxwell field:

$$\phi_0 = \Omega^3\hat{\phi}_0, \quad \phi_1 = \Omega^2\hat{\phi}_1, \quad \phi_2 = \Omega\hat{\phi}_2. \tag{A7}$$

Finally, Maxwell's equations [Eqs. (NPA1)] take the following form in spin coefficient notation:

$$\begin{aligned} D\phi_1 - \bar{\delta}\phi_0 &= (\pi - 2\alpha)\phi_0 + 2\rho\phi_1 - \kappa\phi_2, \\ D\phi_2 - \bar{\delta}\phi_1 &= -\lambda\phi_0 + 2\pi\phi_1 + (\rho - 2\epsilon)\phi_2, \\ \delta\phi_1 - \Delta\phi_0 &= (\mu - 2\gamma)\phi_0 + 2\tau\phi_1 - \sigma\phi_2, \\ \delta\phi_2 - \Delta\phi_1 &= -\nu\phi_0 + 2\mu\phi_1 + (\tau - 2\beta)\phi_2. \end{aligned} \tag{A8}$$

[1]E.T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[2]E.T. Newman and T.W.J. Unti, J. Math. Phys. **3**, 891 (1962).
[3]B. Kozarzewski, Acta Phys. Polon. **27**, 775 (1965).
[4]A.R. Exton, E.T. Newman, and R. Penrose, J. Math. Phys. **10**, 1566 (1969).
[5]G. Ludwig, Gen. Rel. Grav. **7**, 293 (1976).
[6]R. Penrose, Proc. Roy. Soc. A **284**, 159 (1965).
[7]R. Penrose, in *Battelle Rencontre* (Benjamin, New York, 1968).
[8]G. Ludwig, Gen. Rel. Grav. **9**, 1009 (1978).
[9]R.W. Lind, Gen. Rel. Grav. **5**, 25 (1974).
[10]D.W. Trim and J. Wainwright, J. Math. Phys. **12**, 2494 (1971).
[11]D.W. Trim and J. Wainwright, J. Math. Phys. **15**, 535 (1974).
[12]I. Robinson and A. Trautman, Proc. Roy. Soc. A **265**, 463 (1962).
[13]B. Aronson, R. Lind, J. Messmer, and E.T. Newman, J. Math. Phys. **12**, 2462 (1971).
[14]F.A.E. Pirani, in *Lectures on General Relativity* (Prentice Hall, Englewood Cliffs, N.J., 1965).
[15]J. Foster and E.T. Newman, J. Math. Phys. **8**, 189 (1967).

The virial series of the ideal Bose gas

Robert M. Ziff
Department of Mechanical Engineering, State University of New York, Stony Brook, New York 11794
John M. Kincaid
Thermophysics Division, National Bureau of Standards, Washington, D.C. 20234

(Received 30 May 1979; accepted for publication 26 September 1979)

The radius of convergence ρ_R of the virial series of the d-dimensional ideal Bose gas is estimated by the method of Padé approximants, using at least thirty virial coefficients, which were numerically determined. A finite ρ_R is found for $d = 1$ and 2, even though no phase transition occurs for these d. For $d = 3$, ρ_R is consistent with Fuchs' analytical bounds, and for $d = 4$, 5, and 6, ρ_R is equal to the critical density ρ_c. These findings are supported by some analytic results for the equation of state.

I. INTRODUCTION

The equation of state of a gas may be defined parametrically by the two fugacity series,

$$p = \sum_{l=1}^{\infty} b_l z^l \tag{1}$$

and

$$\rho = \sum_{l=1}^{\infty} l b_l z^l , \tag{2}$$

or directly in powers of ρ as the virial series,

$$p = \sum_{n=1}^{\infty} B_n \rho^n . \tag{3}$$

Here p is the pressure divided by kT, ρ is the number density, z is the fugacity, $b_l = b_l(T)$ are the fugacity coefficients, $B_n = B_n(T)$ are the virial coefficients, T is the temperature, and k is Boltzmann's constant. For a given model, the series (1) and (2) have a certain radius of convergence z_R about the origin on the complex z plane and the series (3) has a certain radius of convergence ρ_R on the complex ρ plane. There has been a considerable amount of work studying these radii for general systems to see if ρ_R and $\rho(z_R)$ are related to each other and especially to see if either corresponds to the gas–liquid phase transition point.[1] In this paper we will investigate these questions for the specific model of a d-dimensional ideal Bose gas (IBG). Note that, because the particles in this model have no repulsive core, most of the results of the papers of Ref. 1 do not apply here.

For the IBG, it is convenient to make all quantities appearing in Eqs. (1)–(3) dimensionless by scaling p and ρ by λ^d, where $\lambda^2 \equiv h^2/2\pi mkT$, h is Planck's constant, and m is the boson mass. For example, ρ is now taken to be the actual number density multiplied by λ^d. Then the b_l are given by[2]

$$b_l = l^{-d/2-1} \tag{4}$$

for a d-dimensional IBG. Note that the temperature does not appear in the b_l because T has been combined into the dimensionless variables p and ρ. For $d > 2$, this model exhibits a phase transition (the Bose–Einstein condensation) at a density ρ_c defined by

$$\rho_c \equiv \sum_{l=1}^{\infty} l^{-d/2} \equiv \zeta(d/2), \tag{5}$$

such that for all $\rho > \rho_c$, the pressure is constant and equal to

$$p_c \equiv \zeta(1 + d/2). \tag{6}$$

The radii of convergence, z_R, of the fugacity series (1) and (2), with the b_l given by (4), are easily found to be unity for all d. For $d = 1$ and 2, $\rho(z_R)$ is infinite, and since $\rho(z)$ is a monotonically increasing function of z, the fugacity series describe the entire physical region $0 \leqslant \rho < \infty$. For all $d > 2$, where $\rho(z_R) = \rho_c$, the fugacity series describe precisely the one phase region $0 \leqslant \rho < \rho_c$, and for densities greater than ρ_c, one must refer to Eq. (6) for the pressure. Thus, in the fugacity series there is a direct connection between the limit of convergence and the occurrence of condensation.

The problem of finding ρ_R is much more difficult, because the virial coefficients B_n, although formally related to the b_l, cannot be written in closed form for this model (except for the special case $d = 2$). For $d = 3$, Fuchs has shown that[3]

$$12.57 \approx 4\pi \leqslant \rho_R \leqslant \zeta(3/2) + 8\pi \approx 27.72. \tag{7}$$

A striking consequence of the lower bound being much greater than $\rho_c = \zeta(3/2) \approx 2.612$ is that the virial series converges far beyond the transition point. Yet ρ_R is not known any more accurately than by these bounds, and apparently the problem has not been looked at for other d.

II. ANALYTIC RESULTS

For $d = 2$, the B_n may be found explicitly in the following way. According to Eqs. (1), (2), and (4), $dp/d\rho$ may be written as

$$\frac{dp}{d\rho} = \frac{dp}{dz} \Big/ \frac{d\rho}{dz} = \frac{g_1(z)}{g_0(z)} , \tag{8}$$

where we have introduced the so-called Bose functions $g_n(z)$ defined by

$$g_n(z) = \sum_{l=1}^{\infty} l^{-n} z^l . \tag{9}$$

When $n = 0$ and 1, this series may be written in closed form, so that

$$\rho = g_1(z) = -\log(1 - z) , \tag{10}$$

or $z = 1 - e^{-\rho}$, and

 0022-2488/80/010161-05$01.00

$$g_0(z) = z/(1-z) = e^p - 1. \tag{11}$$

Thus, Eq. (8) becomes

$$\frac{dp}{d\rho} = \frac{\rho}{e^\rho - 1}. \tag{12}$$

The right-hand side may be expanded in powers of ρ with the aid of the Bernoulli numbers, B_n^*,[4] and then integrated term by term, giving for the virial coefficients the following expression:

$$B_n = B_{n-1}^*/n!. \tag{13}$$

Thus $B_1 = 1$, $B_2 = -\frac{1}{4}$, and $B_{2n} = 0$ $(n = 2,3,4,\cdots)$. Making use of the expression of the B_n^* in terms of the ζ function, one may also express the virial coefficients as follows:

$$B_{2n+1} = \frac{(-1)^{n+1} 2\zeta(2n)}{(2\pi)^{2n}(2n+1)}, \quad n = 1,2,3,\cdots. \tag{14}$$

This formula provides an asymptotic expression for B_n since $\zeta(2n) \to 1$ as $n \to \infty$, and the ratio test shows that ρ_R is exactly 2π. This may seem surprising because the 2-d equation of state $p(\rho)$ shows no phase transition for $0 \leqslant \rho < \infty$; however, the latter implies only that $p(\rho)$ is analytic on the positive real axis in the complex ρ plane, and any singularities in p off that axis will make ρ_R finite. Indeed, the complete analytic structure of $p(\rho)$ may be deduced directly from Eq. (12), which shows that $dp/d\rho$ has simple poles on the imaginary axis at $\rho = \pm 2\pi i n$, $n = 1,2,3,\cdots$. Therefore, p has logarithmic poles at these same points. Closest to the origin are the pair of singularities at $\rho = \pm 2\pi i$, which make $\rho_R = 2\pi$, as found above.

Unfortunately, this trick used to find the B_n does not work for other dimensions, as z cannot in general be eliminated between $dp/d\rho$ and ρ. Still, the following interesting result can be proven: The functional dependence of $dp/d\rho$ upon ρ in d dimensions is the same as the functional dependence of p/ρ upon p in $(d-2)$ dimensions. In terms of the coefficients A_n in the expansion of $1/\rho$ in powers of p, as defined by the expression

$$\frac{1}{\rho} = \sum_{n=1}^{\infty} A_n p^{n-2}, \tag{15}$$

this result says that $A_n^{[d-2]} = nB_n^{[d]}$, where the bracketed superscript indicates the dimensionality. If, for example, the A_n of the two-dimensional gas could be found, then the virial coefficients of the four-dimensional gas would follow immediately. However we have not been able to find a closed expression for $A_n^{[2]}$, even though we know the result (12)!

Another useful property which can be derived is the behavior of $p(\rho)$ about ρ_c (for $d > 2$, of course)[5] making use of the known behavior of the Bose functions about $z = 1$. In terms of $\alpha \equiv -\log z$, those functions have the expansion[6]

$$g_n(e^{-\alpha}) = \begin{cases} \Gamma(1-n)\alpha^{n-1} + \sum_{k=0}^{\infty} \zeta(n-k)\frac{(-\alpha)^k}{k!}, & n \neq 1,2,3,\cdots, \\ \frac{(-\alpha)^{n-1}}{(n-1)!}\left[-\log\alpha + \sum_{m=1}^{n-1}\frac{1}{m}\right] + \sum_{\substack{k=0 \\ \neq n-1}}^{\infty} \zeta(n-k)\frac{(-\alpha)^k}{k!}, & n = 1,2,3,\cdots \end{cases} \tag{16}$$

which is valid for $|\alpha| < 2\pi$. In terms of α, p and ρ are given by

$$p = g_{1+d/2}(e^{-\alpha}), \tag{17}$$

$$\rho = g_{d/2}(e^{-\alpha}). \tag{18}$$

For $d = 3$, Eq. (16) gives

$$p_c - p = \alpha\zeta(3/2) - (4\pi^{1/2}/3)\alpha^{3/2} - (\alpha^2/2)\zeta(1/2) + \cdots, \tag{19}$$

$$\rho_c - \rho = 2\pi^{1/2}\alpha^{1/2} + \alpha\zeta(1/2) - (\alpha^2/2)\zeta(-1/2) + \cdots, \tag{20}$$

where $p_c = \zeta(5/2)$ and $\rho_c = \zeta(3/2)$, according to Eqs. (5) and (6). These series are written in order of decreasing importance for α small. Eliminating α between (19) and (20), we find

$$(p_c - p) = \frac{\zeta(3/2)}{4\pi}(\rho_c - \rho)^2 - \left(\frac{1}{6\pi} + \frac{\zeta(3/2)\zeta(1/2)}{8\pi^2}\right) \times (\rho_c - \rho)^3 + \cdots. \tag{21}$$

This series contains only integral powers of $(\rho_c - \rho)$ to all orders, as may be seen in the following way. The two series given by Eqs. (19) and (20) may be thought of as simple power series in $\alpha^{1/2}$, and successive powers of this parameter may be eliminated from the series of $(\rho_c - \rho)$ by subtracting the series $(\rho_c - \rho)$ raised to a successive integral power—by long division, so to speak. The first two terms of this procedure gives Eq. (21). The next term will be of order α^2, which can clearly be eliminated by raising $(\rho_c - \rho)$ to the fourth power. After each step, there will remain a series whose leading term will contain $\alpha^{1/2}$ raised to an integral power, which can always be eliminated by raising $(\rho_c - \rho)$ to an integral power. Thus, we conclude that the Taylor expansion of $p(\rho)$ about $\rho = \rho_c$ exists, and consequently $p(\rho)$ has no singularity at the transition point ρ_c.[7] This conclusion agrees with Fuchs' result that $p(\rho)$ is analytic about ρ_c. Furthermore, Eq. (21) provides the analytic continuation of $p(\rho)$ beyond $\rho = \rho_c$, and shows that the pressure decreases when $\rho > \rho_c$.

Next we will consider the case for $d = 5$, for which

$$p_c - p = \alpha\zeta(5/2) - (\alpha^2/2)\zeta(3/2) + (8\pi^{1/2}/15)\alpha^{5/2}\cdots, \tag{22}$$

$$\rho_c - \rho = \alpha\zeta(3/2) - (4\pi^{1/2}/3)\alpha^{3/2} - (\alpha^2/2)\zeta(1/2) + \cdots. \tag{23}$$

Eliminating α to second order gives the relation

$$(p_c - p) = \frac{\zeta(5/2)}{\zeta(3/2)}(\rho_c - \rho) + \frac{\zeta(5/2)}{\zeta(3/2)^{5/2}} \times \frac{4\pi^{1/2}}{3}(\rho_c - \rho)^{3/2} + \cdots, \tag{24}$$

in which the second term contains a fractional power of $(\rho_c - \rho)$, so that p is not real for $\rho > \rho_c$. On the complex ρ plane there is a singularity at $\rho = \rho_c$, and a branch cut must connect to it. Although $(p_c - p)$ and $(\rho_c - \rho)$ are both simple power series in α, as in the previous case, the elimination of α does not give a simple power series in $(\rho_c - \rho)$. The reason is that the leading term of Eq. (23) is of order α in-

TABLE I. The virial coefficients, B_n, for various dimensions, d.

n	$d=1$	$d=2$	$d=3$
1	1.0	1.0	1.0
2	−0.35355 33905 9327	−0.25	−0.17677 66952 9664
3	0.11509 98205 4025	0.02777 77777 7778	−0.00330 00598 1992
4	−0.03413 86050 9159	0.0	−0.00011 12893 2847
5	0.00900 77375 0204	−0.00027 77777 7778	−0.00000 35405 0410
6	−0.00199 91196 5573	0.0	−0.00000 00838 6347
7	0.00031 17106 1171	0.00000 47241 1187	−0.00000 00003 6621
8	0.00000 23432 7492	0.0	0.00000 00001 0281
9	−0.00002 66697 2643	−0.00000 00918 5773	0.00000 00000 0706
10	0.00001 31427 7163	0.0	0.00000 00000 0027
11	−0.00000 42506 4773	0.000000 00189 789	
12	0.00000 09409 2635	0.0	
13	−0.00000 00703 6085	−0.00000 00000 4065	
14	−0.00000 0060 16153	0.0	
15	0.00000 00404 0756	0.00000 00000 0089	
n	$d=4$	$d=5$	$d=6$
1	1.0	1.0	1.0
2	−0.125	−0.08838 83476 4832	−0.065
3	−0.01157 40740 7407	−0.01151 66866 0664	−0.00906 63580 2469
4	−0.00260 41666 6667	−0.00322 74884 4806	−0.00271 26736 1111
5	−0.00079 82253 0864	−0.00120 36191 8526	−0.00107 94758 4448
6	−0.00028 69405 8642	−0.00052 35046 3138	−0.00050 27505 1708
7	−0.00011 37836 8679	−0.00025 09020 5384	−0.00025 89549 2916
8	−0.00004 82418 1433	−0.00012 85954 8669	−0.00014 30949 9582
9	−0.00002 14753 7035	−0.00006 92363 9579	−0.00008 32913 7682
10	−0.00000 99191 5471	−0.00003 87103 4623	−0.00005 04630 4140
11	−0.00000 47175 0052	−0.00002 22987 4467	−0.00003 15630 3208
12	−0.00000 22971 2533	−0.00001 31597 3086	−0.00002 02606 0322
13	−0.00000 11405 1809	−0.00000 79234 6472	−0.00001 32888 6677
14	−0.00000 05755 8149	−0.00000 48518 3013	−0.00000 88762 2382
15	−0.00000 02945 4015	−0.00000 30140 0411	−0.00000 60219 0440

stead of $\alpha^{1/2}$, and the elimination of $\alpha^{3/2}$, for example, requires that Eq. (23) be raised to the $\frac{2}{3}$ power. In the same manner, it can be shown that $p(\rho)$ has a branch singularity at $\rho = \rho_c$ for all odd dimensionalities $d \geqslant 5$.

To study the even-dimensionality cases, it is necessary to use the second expression in Eq. (16) for $g_n(\alpha)$. For $d = 4$, it follows that

$$p_c - p = \alpha\zeta(2) + (\alpha^2/2)(\log\alpha - 3/2) + \cdots, \qquad (25)$$

$$\rho_c - \rho = -\alpha(\log\alpha - 1) - (\alpha^2/2)\,\zeta(0) + \cdots. \qquad (26)$$

Note that, even to lowest order in α, (25) and (26) imply a transcendental equation for $p(\rho)$. The first derivative,

$$\frac{dp}{d\rho} = \frac{\zeta(2)}{\rho'} \sim \frac{\zeta(2)}{-\log\alpha} \qquad (27)$$

goes to zero as $\alpha \to 0$, where $\rho' \equiv -\partial\rho/\partial\alpha$. (For $d = 3$, the first derivative is also zero at $\rho = \rho_c$.) However, the second derivative here is infinite, since

$$\frac{d^2p}{d\rho^2} = \frac{(\rho')^2 - \rho\rho''}{(\rho')^3} \sim \frac{\zeta(2)}{\alpha(\log\alpha)^3} \to \infty, \qquad (28)$$

as $\alpha \to 0$, and therefore $p(\rho)$ is singular at $\rho = \rho_c$. Likewise, one can show that for all even $d \geqslant 4$, some derivative of p will be infinite at ρ_c and therefore p will be singular there.

Because of the nature of the singularity in $p(\rho)$ at ρ_c for $d \geqslant 4$, the equation of state may not be analytically continued as a real function beyond the condensation point. Such a continuation of $p(\rho)$ is often identified with the existence of metastable states, and so it appears that there can be no metastability for these models. While there is an analytic continuation of the equation of state beyond ρ_c for $d = 3$, it represents unstable states because $dp/d\rho < 0$. It also does not lead to a connection of the gaseous and condensed phases by means of a Maxwell construction, in the way that the van der Waals equation does, for example, and therefore seems to have little physical significance.

The singularity of $p(\rho)$ at $\rho = \rho_c$ for all $d \geqslant 4$ also implies that the virial series cannot converge beyond $\rho = \rho_c$. Yet this places only an upper bound on ρ_R, since ρ_R may be less than ρ_c if $p(\rho)$ contains other singularities on the complex ρ plane, off the real axis, and closer to the origin. The results of a numerical search for these singularities will be given in the next section. The case of $d = 3$ is unusual in that there is no singularity in the equation of state at $\rho = \rho_c$ and therefore the condensation places no bound on ρ_R. In this case, as for $d = 1$ and 2, ρ_R will be determined by singularities off the real axis.

III. NUMERICAL ESTIMATES FOR THE RADIUS OF CONVERGENCE

To find ρ_R, we first determined the B_n by eliminating z from Eqs. (1) and (2). This was done for $d = 1$–6, typically to 25-place accuracy and to 30th order, with the precise degree of accuracy and order depending somewhat upon d. The first 15 of these are listed in Table I, to 14-place accuracy. Note

TABLE II. The location of the first singularity $p(\rho)$.

d	ρ_0	$\lvert\rho_0\rvert=\rho_R$
1	$(-2.279\pm0.015)\pm i(1.486\pm0.005)$	2.72 ± 0.01
2	$\pm 2\pi i$	2π
3	$(14.10\pm0.03)\pm i(12.07\pm0.01)$	18.56 ± 0.02
4	1.65 ± 0.03	$1.65\pm0.03\approx\zeta(2)$
5	1.341487	$1.341487\approx\zeta(5/2)$
6	1.202057	$1.202057\approx\zeta(3)$

that the results for $d=2$ agree exactly with (13); for $d=3$ the virial coefficients decrease most rapidly, indicating a large radius of convergence, and for $d=4$, 5, and 6, all B_n have the same sign (for $n>1$) and decrease monotonically as n increases. We then used Padé approximant methods to obtain estimates of the radius of convergence, ρ_R, of the series (3). Specifically, for a given series $f(x)=\Sigma_{i=0}^{\infty} f_i\, x^i$ we determined the coefficients of the polynomials $Q_M(x)$, $P_L(x)$, and $R_N(x)$ such that

$$Q_M(x)(df/dx)+P_L(x)f(x)+R_N(x) = O(x^{L+M+N+2}). \tag{29}$$

We obtained the $[N/L;M]$ integral approximant to $f(x)$ by integrating this differential equation.[8] Here

$$Q_M(x)=1+\sum_{i=1}^{M} q_i\, x^i,\quad P_L(x)=\sum_{i=0}^{L} p_i\, x^i,$$

and

$$R_N(x)=\sum_{i=1}^{N} r_i\, x^i. \tag{30}$$

When $M=-1$, $Q_M(x)\equiv 0$ so that one obtains the usual $[N/L]$ Padé approximants to $f(x)$. When $N=0$, $R_N(x)\equiv 0$ and one obtains the $[L/M]$ $d\log$ approximant to $f(x)$.

In this integral approximant method, ρ_R appears as the absolute value of the smallest root, ρ_0, of the polynomial Q_M (in the case $M\geqslant 0$) or P_L (in the case $M=-1$). A listing of our estimates for ρ_0 and ρ_R are given in Table II. The values were obtained by studying the integral approximants to p, p/ρ, and $dp/d\rho$. The apparent errors indicated for ρ_R, obtained by the method of Baker and Hunter,[9] reflect the amount by which the various approximants disagree. For $d=1$, we have a conjugate pair of singularities with negative real part, whose absolute value gives $\rho_R\approx 2.72$. For $d=2$, simple poles are found for $dp/d\rho$ at $\rho=\pm 2\pi i$, which agrees with our analysis in Sec. II. For $d=3$, a conjugate pair of singularities with positive real part are found, and they give $\rho_R\approx 18.6$—well within Fuchs' bounds, Eq. (7). For $d=4$, 5, and 6, the first singularity is on the real axis and coincides with the phase transitions at $\rho=\rho_c$. For $d=4$, the uncertainty of our estimate for ρ_R is rather large but is consistent with ρ_R being equal to $\rho_c=\zeta(2)=\pi^2/6\approx 1.644934$. The calculation places the singularity right on the real axis, in which case ρ_R must be equal to ρ_c, because on that axis $p(\rho)$ is analytic for $0\leqslant\rho<\rho_c$. For $d=5$, ρ_R was found to equal $\rho_c=\zeta(5/2)\approx 1.341487$ to seven digit accuracy, and the nature of the pole was found to be consistent with Eq. (25). Likewise, for $d=6$, ρ_R was found to equal ρ_c to high accuracy.

We have thus found, in the examples provided by the d-dimensional IBG, that there is a relation between ρ_R of the virial series and a phase transition in the equation of state, but that the relation is not a consistent one. For $d=1$, 2, and 3, there are singularities of $p(\rho)$ off the real axis which give a finite ρ_R, even though in the first two cases no phase transition exists, and in the third case the phase transition occurs for ρ much less than ρ_R. For $d=4$, 5, and 6, however, $\rho_R=\rho_c$, so that the virial expansion is valid in these cases for all $0<\rho<\rho_c$, and points to the Bose–Einstein condensation at ρ_c. Based upon the analytic structure of $p(\rho)$ discussed in Sec. II, we expect that this is true for all $d\geqslant 4$.

IV. RELATED MODELS

In this section we give examples of two systems whose fugacity series are similar to those of the IBG for $d=3$ and 5, yet whose virial series can be simply expressed and the radius of convergence can be easily determined. Firstly, we consider a system described by the elementary equation of state,

$$p=\rho-\tfrac{1}{2}\rho^2/(2\pi)^{1/2}, \tag{31}$$

which has a fugacity series with coefficients given by

$$b_l=(2\pi)^{1/2}\, l^{\,l-2}e^{-l}/l!. \tag{32}$$

These coefficients may be derived as follows. From Eq. (31) and the general relation $\rho=z\,dp/dz$, which may be deduced from Eqs. (1) and (2), it follows by integration that

$$z=c\rho e^{-\rho/\sqrt{2\pi}}, \tag{33}$$

where c is a constant. The power series $\rho(z)$ follows from the above by means of Lagrange's expansion,[10] and comparison with Eq. (2) shows that the b_l are given by Eq. (32), if c is taken to be $e/(2\pi)^{1/2}$. Note that changing c only scales the b_l and does not affect the equation of state.

The notable feature of this equation of state is its similarity to the 3-d IBG. In the Sterling approximation, b_l becomes exactly $l^{-5/2}$, the same as for the 3-d IBG. Although the virial series here contains just two terms, according to Eq. (31), it may be noted from Table I that only the first two terms of the virial series of the 3-d IBG are really important. The radius of convergence of the fugacity series is again $z_R=1$, which corresponds to $\rho=(2\pi)^{1/2}$. This density also corresponds to the maximum value of pressure, $p_c=(2\pi)^{1/2}/2$, and therefore we may identify the former with the critical density ρ_c. Beyond ρ_c, the pressure decreases and is unphysical. The radius of convergence of (31), which is clearly infinite, is much greater than ρ_c. These properties are all qualitatively the same as those of the 3-d IBG. The infinite value of ρ_R is perhaps a big difference, and evidently this example demonstrates that an apparently small change in the b_l may have a drastic effect upon ρ_R.

The equation of state (31) is of the same form as the van der Waals equation of state, in the limit that the core size goes to zero. Thus, for such a system, the b_l are given by an expression of the form of (32), and the fugacity series converges just up to the spinodal. Although for the complete van der Waals equation of state its does not seem possible to find closed expressions for the b_l by the procedure used above, it does appear that the fugacity series always converges up to the first spinodal. For the Bragg–Williams, or Husimi–Temperley,[11,12] equation of state, which is similar to the van der Waals one, and for which fairly simple expressions can be found for the b_l, Katsura has proven that the fugacity series converges exactly up to the spinodal.[12]

Secondly, a system analogous to the 5-d IBG may be constructed with the fugacity coefficients,

$$b_l = (2\pi)^{1/2}\, l^{\,l-3} e^{-l}/l!, \tag{34}$$

which becomes $l^{-7/2}$ in the Sterling approximation. These b_l imply the equation of state

$$\begin{aligned} p &= \rho/2 + ((2\pi)^{1/2}/6)\left[1-(1-\rho/\rho_c)^{3/2}\right] \\ &= \rho - 2\,(2\pi)^{1/2}\sum_{n=2}^{\infty}\frac{(2n-4)!}{n!(n-2)!}\left(\frac{\rho}{4\rho_c}\right)^n, \end{aligned} \tag{35}$$

where $\rho_c \equiv (\pi/2)^{1/2}$. This follows from (31)–(32) by virtue of the relation following Eq. (15), which applies here also. As for the 5-d IBG, z_R equals unity and corresponds to the density ρ_c where the pressure is at its maximum and has a $\frac{3}{2}$-power fractional singularity in $(\rho_c - \rho)$. For large n,

$$B_n \sim -(2n)^{-5/2}(1/\rho_c)^n, \tag{36}$$

and clearly $\rho_R = \rho_c$. Note that the B_n of the 5-d IBG, as given in Table I, also show the above asymptotic behavior for large n.

Thus, the behavior of each of these model systems is qualitatively the same as that of the corresponding IBG. In particular, for both 3-d models, ρ_R is much greater than ρ_c, and in both 5-d models, ρ_R and ρ_c are equal.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge helpful comments from G. Stell, B. Widom, and S. Merajver. We are also grateful to G.A. Baker, Jr., for making available to us information on integral approximant methods. R.Z. would like to thank the National Science Foundation, and the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for their support.

[1]See, for example, J. Groeneveld, Phys. Lett. **3**, 50 (1962); O. Penrose, J. Math. Phys. **4**, 1312 (1963); and J.L. Lebowitz and O. Penrose, J. Math. Phys. **5**, 841 (1964). These papers are reprinted in *The Equilibrium Theory of Classical Fluids*, edited by H.L. Frisch and J.L. Lebowitz (Benjamin, New York, 1964).

[2]See, for example, R.M. Ziff, G.E. Uhlenbeck, and M. Kac, Phys. Rep. C **32**, 169–248 (1977), Sec. 2.

[3]W.H.J. Fuchs, J. Rat. Mech. Anal. **4**, 647–52 (1955). Previous to Fuchs' work, weaker bounds on R were found by H.S. Shapiro, Phys. Rev. **99**, 1673–4 (1955), and B. Widom, Phys. Rev. **96**, 16–7 (1954). In this paper Widom gives an explicit expression of the B_n of the 3-d gas in the form of a determinant.

[4]See, for example, M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964), p. 804.

[5]The critical behavior of a d-dimensional gas has also been looked at by J.D. Gunton and M.J. Buckingham, Phys. Rev. **166**, 152 (1968); M.J. Cooper and M.S. Green, Phys. Rev. **176**, 302 (1968).

[6]J.E. Robinson, Phys. Rev. **83**, 678–9 (1951); also, Ref. 2, p. 216.

[7]This of course depends upon the radius of convergence of (21) being nonzero. To prove this, observe that (21) may be considered to be the result of inverting the series (20) to find $\alpha^{1/2}$ as a function of $(\rho_c - \rho)$, and then inserting this new series into Eq. (19) to eliminate all powers of $\alpha^{1/2}$. In principle it is possible to invert (20) because it contains a term linear in $\alpha^{1/2}$. The inverted series has no constant term because (20) has no constant term. When this inverted series is inserted into (19), we are guaranteed that the radius of convergence of the resulting series, (21), will be nonzero, because the constant term of the inverted series (which is zero) is less than the radius of convergence of (19). See, for instance, K. Knopp, *Infinite Sequences and Series* (Dover, New York, 1956), Chap. 4.

[8]D.L. Hunter and George A. Baker, Jr., Phys. Rev. B **19**, 3808 (1979); M.E. Fisher and H. Au-Yang, to be published in J. Phys. A (1979).

[9]D.L. Hunter and George A. Baker, Jr., Phys. Rev. B **7**, 3346 (1973); George A. Baker, Jr., and D.L. Hunter, Phys. Rev. B **7**, 3377 (1973).

[10]E.T. Whittaker and G.N. Watson, *Modern Analysis* (Cambridge University, London, 1927), 4th ed. 132–3; also Ref. 4, p. 14.

[11]S. Katsura, Prog. Theor. Phys. **13**, 571–86 (1955).

[12]H.N.V. Temperley, Proc. Phys. Soc. A **67**, 233–8 (1954).

N-dimensional instantons and monopoles

D. H. Tchrakian[a)]
Fachbereich Physik, Universität Kaiserslautern, D–6750 Kaiserslautern, Federal Republic of Germany

(Received 28 August 1978; accepted for publication 25 June 1979)

The possibility of finding solutions of the instanton and monopole types to gauge field theories on arbitrary even and odd dimensional Euclidean manifolds respectively is investigated. Suitable boundary conditions for both types are given, and new self-duality criteria are developed, for gauge field theories on N-dimensional manifolds ($N \geq 5$) which are also endowed with new Action and Lagrangian densities.

1. INTRODUCTION

It is our intention in this article to extend the notion of instanton and monopole solutions to the Euclidean gauge field system to manifolds of arbitrary dimension N.

From the point of view of application to physics, we hope that this may improve our insight into theories with finite action/topological invariants. It is also not inconceivable that field theories on higher dimensional manifolds may be of relevance to physical theories. On the other hand, it is hoped that the following development may be of some intrinsic interest.

In our presentation below we start in Sec. 2 with the definitions of the topological invariants, as integrals of functions of the curvature. For all even N-dimensional manifolds a Pontryagin number is defined in terms of the curvature only, while for all odd dimensional manifolds a topolgocal monopole charge is defined in terms of the curvature and the section (i.e., the Higgs field). The boundary conditions that those solutions must satisfy are stated, and for even N, analogs of the self-dual solutions of $N = 4$ are developed. To this end, an extension of the operation of duality is developed for even N-dimensional manifolds in Sec. 3. Section 4 deals with the Bianchi identities for N-dimensional manifolds, and finally in Sec. 5 we define an action density, leading to equations of motion which are then solved identically by virtue of the generalized Bianchi identities and extended criteria of self-duality for all even N, explicit consideration being given to $N = 6$. Also in Sec. 5, we have given a modified criterion of self-duality for odd $N = 3$, which agrees with the Bogomolny'i bound condition, and have generalized this to arbitrary odd N, again giving explicit consideration to $N = 5$.

The presentation throughout concentrates on the even $N = 6$ and odd $N = 5$ examples, but the extensions to arbitrary N are obvious and straightforward, although not trivial. To this end we have added an appendix dealing with the $N = 8$ case, thus indicating the way for further generalization.

In this paper we give no proofs of the existence of the types of solution considered, nor do we give any explicit solutions. The latter are under investigation at present.

2. TOPOLOGICAL INVARIANTS

A. Instantons: even N

Following the arguments of Belavin *et al.*,[1] we generalize the definition of the Pontryagin number for $N = 4$, to higher order (in the curvature) topological invariants,[2] corresponding to the solutions in question, for arbitrary even N, starting with $N = 4$:

$$q_4 = \frac{1}{16\pi^2} \int \mathrm{Tr}\epsilon_{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}\, d_4x, \tag{1}$$

$$q_6 = \frac{1}{D_G V_6} \int \mathrm{Tr}\epsilon_{\mu\nu\rho\sigma\tau\lambda} F_{\tau\lambda} F_{\rho\sigma} F_{\tau\lambda}\, d_6x, \tag{1'}$$

where $V_N = 2\pi^{N/2}/N\Gamma(N/2)$ is the volume of the N-dimensional unit sphere and D_G is a normalization factor depending on the gauge group and the representation of its algebra in which the curvature $F_{\mu\nu}$ takes its values.

That q_N will have an integral spectrum follows from the fact that the solutions we consider are required to satisfy the following boundary conditions at infinity[1]:

$$A_\mu(x) \underset{|x|\to\infty}{\longrightarrow} g^{-1}(x)\partial_\mu g(x), \tag{2}$$

where $A_\mu(x)$ is the connection which also takes its values in the algebra of the gauge group G. Then substituting (2) into

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu], \tag{3}$$

and applying similar arguments as in Ref. 1, we see that the quantities defined by Eqs. (1) must have integral spectrum.

B. Monopoles: odd N

We define the topological invariants, pertaining to the solutions in question, corresponding to the magnetic monopole charge for the Yang–Mills–Higgs system (A,ϕ), on arbitrary odd N-dimensional manifold, starting with $N = 3$, which is the 't Hooft–Polyakov[3] solution, and followed by $N = 5$,

$$\mu_3 = \frac{1}{4\pi} \int \mathrm{Tr}\epsilon_{ijk}\, \phi F_{jk}\, dS_i^{(2)}, \tag{4}$$

$$\mu_5 = \frac{1}{\Omega_s} \int \mathrm{Tr}\epsilon_{ijklm}\, \phi F_{jk} F_{lm}\, dS_i^{(4)}, \tag{4'}$$

where $\Omega_N = 2\pi^{N/2}/\Gamma(N/2)$ is the surface area of the N-dimensional unit sphere, and $dS^{(N-1)}$is the $(N-1)$-dimensional "area" element. Here the Higgs field ϕ takes its value in the algebra of G and hence the field ϕ_a ($a = 1,\ldots,n$ for an n parameter group G) is in the adjoint representation of G. We expect that the above constructions can be generalized to the case where ϕ_a is not in the adjoint representation of G, as has already been done for the $N = 3$ case by Schwarz[4] and the present author.[5]

 0022-2488/80/010166-04$01.00

That μ_N has an integral spectrum follows from the fact that the integrals (4), (4′), etc., are Kronecker integrals, subject to the following boundary conditions:

$$D_\mu \phi \underset{|x|\to\infty}{\longrightarrow} 0, \tag{5a}$$

$$\phi^2 \underset{|x|\to\infty}{\longrightarrow} \eta^2 1. \tag{5b}$$

The first condition (5a), is the so-called finite energy condition used for the $N=3$ monopole,[3] implying $\mathrm{Tr}\phi^2 \underset{|x|\to\infty}{\longrightarrow}$ const, while the second (5b), is a stronger condition for all gauge groups G larger than SU(2), and was found to be necessary in the $N=3$ case by Rawnsley and the present author.[6]

We end this section by extending the arguments made for $N=3$ in Ref. 6 to the $N=5$ case here. We define the "electromagnetic" field for $N=5$ to be

$$\mathscr{F}_{ijkl} = \mathrm{Tr}\phi \,|\, (F_{ij} + \tfrac{1}{4}[D_i\phi, D_j\phi\,]), (F_{kl} + \tfrac{1}{4}[D_k\phi, D_l\phi\,])| \,.$$

It is straightforward to verify, using (5b), that $\mathscr{F}_{ijkl}$ reduces to

$$\mathscr{F}_{ijkl} = \tfrac{1}{2}\,\mathrm{Tr}\phi\partial_i\phi\partial_j\phi\partial_k\phi\partial_l\phi \; + \text{total divergence terms,}$$

such that the "monopole" charge of this "electromagnetic" field is

$$\frac{1}{(4\pi)^2}\int \mathscr{F}_{ijkl}\epsilon_{ijklm}\, dS_m, \tag{4″}$$

where, in the absence of any contribution from the total divergence terms in $\mathscr{F}_{ijkl}$, (4″) reduces to a Kronecker integral. But comparison of (4″) with (4′) shows that these two integrals are equal for large surfaces on which condition (5a) is valid, justifying our assertion that integrals (4), (4′), etc., are Kronecker integrals.

3. DUALITY OPERATIONS

We define the operation of dual on the curvature as a set of conjugations whose repeated application results in the identity operation.

This set of operations is defined for even dimensional manifolds. We shall consider below the cases of $N=6$ and $N=8$ since these two cases are typical of all the others, namely, those of odd $N/2$ and even $N/2$, respectively ($N \geqslant 4$).

For $N=6$ we start by introducing the notation

$$F_{\mu\nu\rho\sigma} = \{F_{\mu\nu}, F_{\rho\sigma}\} - \{F_{\mu\rho}, F_{\nu\sigma}\} - \{F_{\mu\sigma}, F_{\rho\sigma}\}, \tag{6}$$

and then define the first dual as

$${}^{(1)}F_{\mu\nu} = \frac{i}{4!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda}F_{\rho\sigma\tau\lambda}, \tag{7}$$

and the second dual is

$${}^{(2)}F_{\mu\nu\rho\sigma} = -\,\frac{i}{2!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda}F_{\tau\lambda}. \tag{7′}$$

Repeated application of these conjugations then yield, respectively,

$${}^{(2)}({}^{(1)}F)_{\mu\nu\rho\sigma} = -\,\frac{i}{2!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda}\;{}^{(1)}F_{\tau\nu} = F_{\mu\nu\rho\sigma}, \tag{8}$$

$${}^{(1)}({}^{(2)}F)_{\mu\nu} = \frac{i}{4!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda}\;{}^{(2)}F_{\rho\sigma\tau\lambda} = F_{\mu\nu}. \tag{8′}$$

Similarly, for $N=8$, using in addition to (6) the notation

$$F_{\mu\nu\rho\sigma\tau\lambda} = \{F_{\mu\nu}, F_{\rho\sigma\tau\lambda}\} - \{F_{\mu\rho}, F_{\nu\sigma\tau\lambda}\} - \{F_{\mu\sigma}, F_{\rho\sigma\tau\lambda}\} - \{F_{\mu\tau}, F_{\rho\sigma\mu\lambda}\} - \{F_{\mu\lambda}, F_{\rho\sigma\tau\lambda}\}, \tag{9}$$

we have the following duality operations:

$${}^{(1)}F_{\mu\nu} = \frac{1}{6!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda\kappa\eta}F_{\rho\sigma\tau\lambda\kappa\eta}, \tag{10}$$

$${}^{(2)}F_{\mu\nu\rho\sigma} = \frac{1}{4!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda\kappa\eta}F_{\tau\lambda\kappa\eta}, \tag{10′}$$

$${}^{(3)}F_{\mu\nu\rho\sigma\tau\lambda} = \frac{1}{2!}\,\epsilon_{\mu\nu\sigma\rho\tau\lambda\kappa\eta}F_{\kappa\eta}. \tag{10″}$$

Then these conjugations close under the following repeated applications:

$${}^{(3)}({}^{(1)}F)_{\mu\nu\rho\sigma\tau\lambda} = \frac{1}{2!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda\kappa\eta}\;{}^{(1)}F_{\kappa\eta} = F_{\mu\nu\rho\sigma\tau\lambda}, \tag{11}$$

$${}^{(1)}({}^{(3)}F)_{\mu\nu} = \frac{1}{6!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda\kappa\eta}\;{}^{(3)}F_{\rho\sigma\tau\lambda\kappa\eta} = F_{\mu\nu}, \tag{11′}$$

$${}^{(2)}({}^{(2)}F)_{\mu\nu\rho\sigma} = \frac{1}{4!}\,\epsilon_{\mu\nu\rho\sigma\tau\lambda\kappa\eta}\;{}^{(2)}F_{\tau\lambda\kappa\eta} = F_{\mu\nu\rho\sigma}. \tag{11″}$$

The case for arbitrary N needs no further comment, except to note that $\pm i$ in the definitions of these conjugations occur only for odd $N/2$ and not for even $N/2$.

4. BIANCHI IDENTITIES

The Bianchi identities of an $N=3$ or 4 gauge theory arise from the Jacobi identity for the covariant derivatives which can also be expressed as

$$\epsilon_{\mu\lambda\rho\sigma}[D_\lambda, [D_\rho, D_\sigma\,]]f = 0. \tag{12}$$

For an N-dimensional manifold, (12) can similarly be expressed as

$$\epsilon_{\mu_1\mu_2\cdots\mu_{N-3}\lambda\rho\sigma}[D_\lambda, [D_\rho, D_\sigma\,]]f = 0. \tag{12′}$$

It is now obvious from (12′) that for $n>4$, there are further identities of the sort

$$\epsilon_{\mu_1\mu_2\cdots\mu_{N-5}\lambda\tau\rho\sigma}[D_\lambda, \{[D_\tau, D_\nu\,], [D_\rho, D_\sigma\,]\}]f = 0 \tag{12″}$$

and so forth, until only one index remains free for even N, and no free indices for odd N.

We illustrate the above procedure explicitly with the $N=6$ example, as that will be applied in the next section. To this end we recall the definitions of curvature and covariant derivative,

$$[D_\mu, D_\nu\,]f = F_{\mu\nu}f, \tag{13}$$

$$D_\mu f = \partial_\mu f + A_\mu f, \tag{14}$$

and in particular, when f is in the adjoint representation of the algebra of G, then (14) is expressed as

$$D_\mu f = \partial_\mu f + [A_\mu, f\,]. \tag{14′}$$

Using (14′) in (12′) and (12″) we get, respectively,

$$\epsilon_{\mu\nu\rho\sigma\tau\lambda}D_\sigma F_{\tau\lambda} = 0 \quad \text{or} \quad D_\sigma\;{}^{(2)}F_{\mu\nu\rho\sigma} = 0, \tag{15}$$

$$\epsilon_{\mu\nu\rho\sigma\tau\lambda}\mathscr{D}_\nu F_{\rho\sigma\tau\lambda} = 0 \quad \text{or} \quad \mathscr{D}_\nu\;{}^{(1)}F_{\mu\nu} = 0, \tag{15′}$$

where the lengthened derivative operator $\mathscr{D}_\mu$ is defined by

$$\mathscr{D}_\mu\,\{F_{\rho\sigma}, F_{\tau\lambda}\} = \partial_\mu\,\{F_{\rho\sigma}, F_{\tau\lambda}\} + [A_\mu, \{F_{\rho\sigma}, F_{\tau\lambda}\}], \tag{16}$$

which looks very similar to the covariant derivative (14′) but differs from it in that the anticommutator $\{F_{\rho\sigma},F_{\tau\lambda}\}$ takes values inside as well as outside the algebra of G.

It is easy to see how larger families of Bianchi identities will arise with increasing N, and furthermore, unlike our considerations in Sec. 3, the results here are valid both for even and odd N-dimensional manifolds. For example, the identities (15) and (15′) also hold for $N = 5$. Further, it is clear that for $N = 7,8$ the following identitites hold:

$$D_\sigma\, {}^{(3)}F_{\mu\nu\rho\sigma\tau\lambda} = 0,$$

$$\mathscr{D}_\sigma\, {}^{(2)}F_{\mu\nu\rho\sigma} = 0,$$

$$\mathscr{D}_\sigma\, {}^{(1)}F_{\mu\nu} = 0.$$

5. SELF-DUALITY

For the special cases $N = 3$ and $N = 4$, respectively, the nature of the solutions of the equations of motion are very different. In the $N = 3$ case only solutions with spherical symmetry are proved to exist[7] and of these only a special case solution is found explicitly[8] by integrating the equations of motion. In contrast to this, for the $N = 4$ case the (second order differential) equations of motion are solved by first order equations of self-duality,[9] explicit solutions to which are known,[1,10,11] and in particular for $G = SU(2)$ all solutions are known.[12]

It is our purpose in this section to seek similar self-duality criteria that solve the equations of motion for all cases of even N. As in the above, we consider the $N = 6$ example explicitly, whence the extension to arbitrary even N cases follows straightforwardly.

Our procedure is essentially that of Ref. 1. We start by considering one of the following inequalities:

$$\operatorname{Tr}\int (F_{\mu\nu} - {}^{(1)}F_{\mu\nu})^2\, d_6x \geqslant 0, \tag{17}$$

$$\operatorname{Tr}\int (F_{\mu\nu\rho\sigma} - {}^{(2)}F_{\mu\nu\rho\sigma})^2\, d_6x \geqslant 0. \tag{17'}$$

Let us consider (17), since (17′) will result in exactly the same conclusions. It follows that

$$\operatorname{Tr}\int \left(F^2_{\mu\nu} + \frac{2}{4!}F^2_{\mu\nu\rho\sigma}\right) d_4x \geqslant \frac{2\pi^3 D_G}{4!}\, q_6. \tag{18}$$

Had we chosen the action density[13] to be

$$\mathscr{S} = \operatorname{Tr}\left(F^2_{\mu\nu} + \frac{2}{4!}F^2_{\mu\nu\rho\sigma}\right), \tag{19}$$

and had we also required that

$$F_{\mu\nu} = {}^{(1)}F_{\mu\nu}, \tag{20}$$

then the inequality (18) would have become an equality, and such solutions, corresponding to (20) would be finite action instanton solutions endowed with a topological invariant q_6.

Before proceeding, we note that as a consequence of (20),

$$F_{\mu\nu\rho\sigma} = {}^{(2)}F_{\mu\nu\rho\sigma}, \tag{20'}$$

which is the condition relevant to the inequality (17′).

Conditions (20) and (20′) are the promised extensions to $N = 6$ of the $N = 4$ criterion of self-duality of Ref. 1.

Finally, we check that (20) does indeed solve the equations of motion of the gauge field. From the action density (19) follow these equations of motion:

$$D_\mu F_{\mu\nu} + \tfrac{1}{2}\mathscr{D}_\mu \{F_{\mu\nu\rho\sigma},F_{\rho\sigma}\} = 0. \tag{21}$$

Using (20′) and (20), in that order, in (21) we find

$$D_\mu F_{\mu\nu} = 0. \tag{21'}$$

We now notice that it follows from (20) and (7) that (21′) *is* the Bianchi identity (15′), and thus the equations of motion (21) are identically satisfied by virtue of conditions (20) and (21′).

We note that the self-duality conditions (20) and (20′) lead to a trivial solution with vanishing curvature for $G = SU(2)$. It was also found that vanishing curvature solutions are implied by these extended self-duality conditions for $N = 6$, $G = SO(6)$. Thus the interesting gauge structures are those of $SU(n)$ with $n \geqslant 3$.

As a final remark we note that the procedure in this section is extendable to arbitrary even N, and in particular $N = 8$ is considered in detail in the Appendix.

We end this section by applying the above arguments, as far as possible, to the odd N cases. We start by the $N = 3$ example and consider the following inequality:

$$\operatorname{Tr}\int (\tfrac{1}{2}\epsilon_{ijk}F_{jk} - D_i\phi)^2\, d_3x \geqslant 0, \tag{22}$$

which is obviously going to play the role that integral (17) played for even N. It follows that

$$\operatorname{Tr}\int \left[\tfrac{1}{4}F^2_{ij} + \tfrac{1}{2}(D_i\phi)^2\right] d_3x \geqslant \operatorname{Tr}\int \epsilon_{ijk}(D_i\phi)\, F_{jk}\, d_3x, \tag{23}$$

where the left-hand side of the inequality (23) is nothing but the energy of the 't Hooft–Polyakov[3] with vanishing potential.[8] The right-hand side is equal to 2π times the magnetic charge. This follows by noticing that the integrand on the right-hand side of (23) is a total divergence by virtue of the Bianchi identity,

$$\epsilon_{ijk}D_iF_{jk} = 0,$$

whence, by converting the volume integral into a surface integral, we obtain the integral (4), which is our definition for the magnetic charge. We can then rewrite (23) in the following form,

$$\int \mathscr{L}\, d_3x \geqslant 2\pi\mu_3, \tag{23'}$$

and if we had required the following "self-duality" conditions,

$$\tfrac{1}{2}\epsilon_{ijk}F_{jk} = D_i\phi, \tag{24}$$

then the inequality (23′) would become an equality and the minimum thus attained. It is easy to check that (24) solves the equations of motion of the 't Hooft–Polyakov system. We recognize that (24) is the same condition that Bogomolny'i[14] requires for the minimal value of the energy to be attained.

We now proceed to the $N = 5$ case by considering the following inequality,

$$\operatorname{Tr}\int \left(\frac{1}{4!}\epsilon_{ijklm}F_{jklm} - D_i\phi\right)^2 d_5x \geqslant 0, \tag{25}$$

which by use of the Bianchi identity,

$$\epsilon_{ijklm} D_i F_{jklm} = 0, \tag{26}$$

and the definition (4′) of μ_5, can be reexpressed as

$$\mathrm{Tr} \int \mathscr{L}\, d_5 x \geqslant 2\pi^2 \mu_5, \tag{25′}$$

where we have taken the Lagrangian $\mathscr{L}$ to be

$$\mathscr{L} = \mathrm{Tr}\big[\tfrac{1}{8} F_{ijkl}^2 + (D_i\phi)^2\big]. \tag{27}$$

It is now clear that the minimum in this case will be attained if we require the self-duality conditions

$$\frac{1}{4!}\epsilon_{ijklm} F_{jklm} = D_i\phi. \tag{28}$$

It is also easy to verify that (28), together with the Bianchi identities, actually solves the Euler–Lagrange equations corresponding to the Lagrangian (27).

Furthermore, it is now clear what the self-duality conditions for higher N should be; for example, for $N = 7$ they are

$$\frac{1}{6!}\epsilon_{ij_1k_1j_2j_3k_3} F_{j_1k_1j_2k_2j_3k_3} = D_i\phi. \tag{28′}$$

We end this section by remarking that, for dimensions higher than $N = 3$, the gauge group must be larger than SU(2), for in that case one can see by taking the trace of (28) that this leads to the Higgs field being covariantly constant *everywhere*, and solutions of this type are not of the (finite-energy) soliton type. This is similar to the (even) $N = 6$ case where $G = \mathrm{SU}(2)$ led to another trivial solution where the curvature vanished.

ACKNOWLEDGMENTS

I have benefited from discussions and correspondence with Professor W. Rühl, Professor H.R. Petry, and Professor J.H. Rawnsley. I am also grateful to Professor Rühl for his kind hospitality at Kaiserslautern and to Professor Doebner (Clausthal/Zellerfeld) and Professor Jan S. Nilsson (Göetborg) for enabling me to expose this material in its preliminary stages at their seminars.

APPENDIX

The new self-duality criteria introduced in Sec. 5, for gauge theories on $N = 5$ and $N = 6$ dimensional manifolds, were said to be readily generalizable to arbitrary N. It is our aim in this Appendix to indicate this procedure by considering the $N = 8$ case in some detail.

For $N = 8$, the topological invariant, which is the next order higher than (1′) is the following[2]:

$$q_8 = \frac{1}{D_G V_8}\int \epsilon_{\mu\nu\rho\sigma\tau\lambda\kappa\eta}\big[\mathrm{Tr}(F_{\mu\nu}F_{\rho\sigma}F_{\tau\lambda}F_{\kappa\eta}) - \tfrac{1}{2}\mathrm{Tr}(F_{\mu\nu}F_{\rho\sigma})\cdot\mathrm{Tr}(F_{\tau\lambda}F_{\kappa\eta})\big]\, d_8 x, \tag{A1}$$

and if we seek finite action instanton solutions endowed with this topolgoical invariant, then we must choose a gauge theory whose action density is determined by consideration of the following inequality[13]:

$$\int \mathrm{Tr}[F(2) - {}^{(1)}F(2)]^2\, dx + \tfrac{1}{2}\int [\mathrm{Tr}F(4) + \mathrm{Tr}\,{}^{(2)}F(4)]^2\, dx \geqslant 0, \tag{A2}$$

where F(2) denotes the 2-form $F_{\mu\nu}$ and the 4- and 6- forms $F(4)$ and $F(6)$ are given respectively by (6) and (9), and the dual forms ${}^{(1)}F(2)$, ${}^{(2)}F(4)$, and ${}^{(3)}F(6)$ by (10), (10′), and (10″).

The expansion of (A2) gives, using (A1)

$$\int \{\mathrm{Tr}[F(2)^2 + {}^{(1)}F(2)^2] + 2[\mathrm{Tr}F(4)]^2\}\, dx \geqslant 2D_G V_8 q_8. \tag{A3}$$

It is now clear that if we choose the action density for this $N = 8$ gauge theory to be

$$\mathscr{S} = \mathrm{Tr}[F(2)^2 + {}^{(1)}F(2)^2] + 2[\mathrm{Tr}F(4)]^2, \tag{A4}$$

then the minimum will have been attained, and the corresponding equations of motion, in the sense of Sec. 5, will be satisfied with finite total action, if the following self-duality conditions are satisfied:

$$F(2) = {}^{(1)}F(2) \quad [\text{or } F(6) = {}^{(3)}F(6)], \tag{A5a}$$

$$F(4) = -{}^{(2)}F(4). \tag{A5b}$$

[1] A.A. Belavin, A.M. Polyakov, A.S. Schwarz, and Yu.S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[2] For $N = 2k$, k = even integer, these invariants are related to Pontryagin class, while for k = odd integer > 2 they are related to Chern class. See, for example, S. Kobayashi and K. Nomizu, *Foundation of Differential Geometry* (Interscience, New York, 1969), pp. 305–14.
[3] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974); A.M. Polyakov, JETP Lett **20**, 1940 (1974).
[4] A.S. Schwarz, Nucl. Phys. B **112**, 358 (1976).
[5] D.H. Tchrakian, Kaiserslautern preprint, 1978.
[6] J.H. Rawnsley and D.H. Tchrakian, J. Phys. A **10**, L88 (1977).
[7] Yu.S. Tyupkin, V.A. Fateyev, and A.S. Schwarz, Theor. Math. Phys. **26**, 397 (1977); Teor. Mat. Fiz. **26**, 270 (1976).
[8] M. Prasad and C. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).
[9] There is as yet no proof that all instanton solutions are self-dual.
[10] R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
[11] D.H. Tchrakian and J.H. Rawnsley, J. Phys. A **10**, L195 (1977).
[12] M.F. Atiyah, V.G. Drinfeld, N. Hitchin, and Yu.I. Manin, Phys. Lett. A **65**, 185 (1978).
[13] If we worked with (17′), $\mathscr{S}$ would then be a numerical multiple of (19), and the inequality (18) would not be disturbed.
[14] E.B. Bogomolny'i, Sov. J. Nucl. Phys, **24**, 449 (1976).

Calculations of Dirac-spinor amplitudes by means of complex Lorentz-transformations and trace calculus

J. O. Eeg

Institute of Physics, University of Oslo, Norway
(Received 3 August 1978; revised manuscript received 5 January 1979)

A method for expressing spinor amplitudes $M = \bar{u}(p'\sigma')\Gamma_M u(p\sigma)$ in a formal covariant way and calculating them by trace calculations is described. By means of complex Lorentz transformations, an expression for $u(p\sigma)\bar{u}(p'\sigma')$ in terms of Dirac γ matrices, four vectors, and the complex Lorentz transformation coefficients is obtained. M can then be written as a trace of γ matrices similar to the expression for $\sum_{\rm pol}|M|^2$. The method is easily extended to cases when higher spin spinors and matrices are involved.

I. INTRODUCTION

As is well known, the square modulus of an amplitude written in terms of Dirac spinors and matrices can be given as a trace of γ-matrix products. Fearing and Silbar[1] have pointed out that one often wants an expression for the amplitude itself which does not involve spinors and γ matrices. They have invented a method (hereafter called the FS method) to calculate the amplitude times some complex number by means of trace calculus. (This complex number is expressed in terms of appropriate Dirac spinors).

In this paper we present a method which enables us to express the amplitude itself in terms of a trace of γ matrices. Because we have to deal with complex Lorentz transformations, our method is probably more of formal than great practical interest in the Dirac case. However, when higher spin spinors are involved, our method may be an alternative. It is also possible to combine our method with the FS method.

II. TRACE CALCULUS IN THE SPIN-$\frac{1}{2}$ CASE

The amplitude (or part of the amplitude) can be written in the form

$$M \equiv M(p'\sigma';p\sigma) = \bar{u}(p'\sigma')\Gamma_M u(p\sigma), \tag{1}$$

where $u(p\sigma)$ is a Dirac spinor corresponding to 4-momentum p and spin quantum number $\sigma = \pm\frac{1}{2}$. $\bar{u} = u^\dagger\gamma_0$ is the adjoint of u and Γ_M involves products of Dirac γ_μ matrices (sometimes contracted by 4-vectors). The squared modulus of the amplitude is then

$$|M(p'\sigma';p\sigma)|^2 = \mathrm{Tr}\{[u(p'\sigma')\bar{u}(p'\sigma')]\Gamma_M[u(p\sigma)\bar{u}(p\sigma)]\bar{\Gamma}_M\},$$

where $\bar{\Gamma}_M \equiv \gamma_0\Gamma_M^\dagger\gamma_0$. The calculation is then straightforward, using[2]

$$u(p\sigma)\bar{u}(p\sigma) = \frac{\gamma\cdot p + m}{2m}\,\frac{1}{2}(1 + 2\sigma\gamma_5\gamma\cdot n_{(3)}). \tag{2}$$

$n_{(3)\mu} \equiv n_{(3)}(p)_\mu$ is the covariant spin vector obtained by boosting up the unit vector $n_{(3)}^{(0)} = (0,\mathbf{n}_{(3)}^{(0)})$:

$$n_{(3)}(p)_\mu = L_\mu{}^\nu(p)n_{(3)\nu}^{(0)}, \quad p_\mu = L_\mu{}^\nu(p)p_\nu^{(0)}, \quad p^{(0)} = (m,\mathbf{0}). \tag{3}$$

$L(p)$ is the Lorentz transformation from the rest frame to the frame where the particle has momentum p. For later use we introduce $\mathbf{n}_{(1)}^{(0)}$ and $\mathbf{n}_{(2)}^{(0)}$ such that $\mathbf{n}_{(l)}^{(0)}$, $l = 1,2,3$, constitutes a right-handed system of unit vectors, and we define the combinations

$$\mathbf{n}_{(2\sigma)}^{(0)} \equiv \tfrac{1}{2}(\mathbf{n}_{(1)}^{(0)} + 2i\sigma\,\mathbf{n}_{(2)}^{(0)}) \tag{4a}$$

We also define $n_{(l)}(p)$ for $l = 2\sigma = \pm 1$ similar to $n_{(3)}(p)$ in Eq. (3), and we have

$$p\cdot n_{(l)}(p) = 0, \quad [n_{(l)}(p)]^2 = -1, \quad p^2 = m^2, \quad p'^2 = m'^2. \tag{4b}$$

Fearing and Silbar[1] multiply the amplitude M by a complex number S^*, where

$$S \equiv S(p'\sigma';p\sigma) = \bar{u}(p'\sigma')u(p\sigma). \tag{5}$$

Using (2), S^*M can be obtained by trace calculus:

$$[S(p'\sigma';p\sigma)]^*M(p'\sigma';p\sigma) = \mathrm{Tr}\{[u(p'\sigma')\bar{u}(p'\sigma')]\Gamma_M[u(p\sigma)\bar{u}(p\sigma)]\}, \tag{6}$$

which means that S^*M can be written in terms of p, p', $2\sigma n_{(3)}(p)$, and $2\sigma' n_{(3)}(p')$.

Our method is based on the fact that M can be written as a trace directly:

$$M(p'\sigma';p\sigma) = \mathrm{Tr}[\Gamma_M H(p\sigma;p'\sigma')], \tag{7}$$

where

$$H(p\sigma,p'\sigma') \equiv u(p\sigma)\bar{u}(p'\sigma'). \tag{8}$$

Moreover, we have

$$\mathrm{Tr}[H(p\sigma;p'\sigma')] = S(p'\sigma';p\sigma). \tag{9}$$

In the next section we show that the matrix H defined in (8) can be written in the form:

$$\begin{aligned} H(p\sigma;p'\sigma') &= [(\gamma\cdot p + m)/2m](\gamma_\alpha h_{(+)}^\alpha + \gamma_5\gamma_\alpha h_{(-)}^\alpha) \\ &= (\gamma_\alpha h_{(+)}^\alpha + \gamma_5\gamma_\alpha h_{(-)}^\alpha)(\gamma\cdot p' + m')/2m', \end{aligned} \tag{10}$$

where $h_{(+)}^\alpha$ and $h_{(-)}^\alpha$ are coefficients depending on the spin quantization in rest frame and p, p' through real and complex Lorentz-transformation coefficients (details are given in Appendices A and B). Consequently, using (10), M can be calculated directly [see Eq. (7)].

 0022-2488/80/010170-05$01.00

III. COMPLEX LORENTZ TRANSFORMATION

In the Kramer–Weyl representation, the Dirac matrices are

$$\gamma_\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \tilde\sigma_\mu & 0 \end{pmatrix}, \tag{11}$$

where $\sigma_\mu = (1,\boldsymbol\sigma)$, $\tilde\sigma_\mu = (1, -\boldsymbol\sigma)$. an ordinary physical proper Lorentz transformation $L(A)$ is generated by an element $A \in \mathrm{SL}(2C)$ (i.e., A is a complex 2×2 matrix with $\det A = 1$). A complex Lorentz transformation is generated by *two* elements $A,B \in \mathrm{SL}(2C)$ in the following way[3,4]:

$$A\sigma_\mu B^\dagger = \sigma_\nu L^\nu{}_\mu(A,B), \quad B^{-1\dagger}\tilde\sigma_\mu A^{-1} = \tilde\sigma_\nu L^\nu{}_\mu(A,B), \tag{12}$$

which gives the following transformation for the γ_μ's:

$$\begin{aligned} \Lambda(A)\gamma_\mu \Lambda^{-1}(B) \\ = \tfrac12(1+\gamma_5)\gamma_\nu L^\nu{}_\mu(A,B) + \tfrac12(1-\gamma_5)\gamma_\nu L^\nu{}_\mu(B,A), \end{aligned} \tag{13}$$

where

$$\Lambda(A) = \begin{pmatrix} A & 0 \\ 0 & A^{-1\dagger} \end{pmatrix}, \quad \gamma_5 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{14}$$

From (12) and $\mathrm{Tr}(\tilde\sigma_\mu\sigma_\nu) = 2g_{\mu\nu}$ one obtains the coefficients:

$$L_{\mu\nu}(A,B) = \tfrac12 \mathrm{Tr}(\tilde\sigma_\mu A\sigma_\nu B^\dagger). \tag{15}$$

For $B = A$ we obtain a physical Lorentz transformation. For boosts corresponding to momenta p and p', we define

$$L_{\mu\nu}(p,p') \equiv L_{\mu\nu}(A(p),A(p')), \quad A(p) = \exp[\tfrac12\omega\hat{\mathbf p}\cdot\boldsymbol\sigma], \tag{16a}$$

where $\sinh\omega = (|\mathbf p|/m)$ and $\hat{\mathbf p} \equiv (1/|\mathbf p|)\mathbf p$. Moreover, we define the combinations

$$L^\pm_{\mu\nu} \equiv L^\pm_{\mu\nu}(p,p') \equiv \tfrac12[L_{\mu\nu}(p,p') \pm L_{\mu\nu}(p',p)]. \tag{16b}$$

$L^+_{\mu\nu}$ corresponds to the real part (but not physical!) and $L^-_{\mu\nu}$ to the imaginary part of the transformations. For $\mathbf p = \mathbf p' = \mathbf 0$ we obtain

$$\begin{aligned} H(0\sigma;0\sigma) &= u(0\sigma)\bar u(0\sigma) \\ &= \tfrac12(\gamma\cdot\xi^{(0)} + 1)\tfrac12(\gamma\cdot\xi^{(0)} + 2\sigma\gamma_5\gamma\cdot n^{(0)}_{(3)}), \end{aligned} \tag{17a}$$

$$\begin{aligned} H(0\sigma;0-\sigma) &= u(0\sigma)\bar u(0-\sigma) \\ &= \tfrac12(\gamma\cdot\xi^{(0)} + 1)\tfrac12\gamma_5\gamma\cdot n^{(0)}_{(2\sigma)}, \end{aligned} \tag{17b}$$

where $\xi^{(0)} \equiv (1,\mathbf 0) = (1/m)p^{(0)} = (1/m')p'^{(0)}$; the $n_{(l)}$'s are given by Eqs. (3), (4). $H(p\sigma;p'\sigma')$ is now obtained from (17), (13), and $u(p) = \Lambda(p)u(0\sigma)$, $\Lambda(p) \equiv \Lambda(A(p))$, as

$$H(p\sigma;p'\sigma') = \Lambda(p)H(0\sigma;0\sigma')\Lambda^{-1}(p'). \tag{18}$$

This means that the coefficients in (10) are given by

$$2h_{(\pm)\alpha} = \delta_{\sigma'\sigma} L_\alpha^{(\pm)\beta}\xi^{(0)}_\beta + c(l)L_\alpha^{(\mp)\beta} n^{(0)}_{(l)\beta}. \tag{19}$$

[$l = 3$ for $\sigma' = \sigma$ and $l = 2\sigma$ for $\sigma' = -\sigma$. Moreover, $c(3) = 2\sigma$, $c(2\sigma) = 1$.] More details about $h_{(\pm)\alpha}$ are given in Appendix B. Explicit expressions for M for the cases $\Gamma_M = 1$ (i.e., $M = S$), $\Gamma_M = \gamma_5$, $\Gamma_M = \gamma_\mu$, and $\Gamma_M = \gamma_5\gamma_\mu$ are given in Appendix C. If we combine our method with the FS method, we find alternative expressions for $h^\mu_{(\pm)}$ which are not given in terms of complex Lorentz transformations as in (19). These expressions are given in the last part of Appendix C, and can for general Γ_M be used in (7) instead of (19) and (B1)–(B6).

IV. THE SPIN-$\frac32$ CASE

If $M(p'\sigma';p\sigma)$ is the amplitude for some process where a spin-$\frac32$ particle is produced, the Dirac spinor $u(p'\sigma')$ in (1) has to be replaced by a Rarita–Schwinger[5] spinor $u_\mu(p'\sigma')$ or a Joos–Weinberg (JW)[6] spinor $u^{[3/2]}(p'\sigma')$.[7] The RS and JW adjoint spinors are related by[8]

$$\bar u_\mu(p'\sigma') = \bar u^{[3/2]}(p'\sigma')\Sigma^{[3/2,1/2]}_{\mu\nu}p'^\nu\gamma_5 i/m', \tag{20a}$$

$$\bar u^{[3/2]}(p'\sigma') = (1/2im')\bar u^\mu(p'\sigma')\gamma_5\Sigma^{[1/2,3/2]}_{\mu\alpha}p'^\alpha, \tag{20b}$$

where $\Sigma^{[1/2,3/2]}_{\mu\nu} = -\Sigma^{[1/2,3/2]}_{\nu\mu}$ are a set of nonsquare covariant transforming matrices which is a generalization of $\frac12 i[\gamma_\mu,\gamma_\nu]$ in the Dirac case. Let χ_σ be a spin-$\frac12$ rest spinor (i.e., Pauli spinor) and $\chi^{[3/2]}_{\sigma'}$ a spin-$\frac32$ rest spinor. One obtains

$$\begin{aligned} \chi_\sigma(\chi^{[3/2]}_{\sigma'})^\dagger = a(\sigma,\sigma')\boldsymbol\sigma^{[1/2,3/2]}\cdot\mathbf n^{(0)}_{(l)} \\ + b(\sigma,\sigma')K^{[1/2,3/2]}_{km} n^{(0)}_{(l_1)k} n^{(0)}_{(l_2)m}, \end{aligned} \tag{21}$$

where $\boldsymbol\sigma^{[1/2,3/2]}$ are vector operator (dipole transition)—and $K^{[1/2,3/2]}_{km}$ tensor operator (quadrupole transition)—matrices to operate between spin-$\frac12$ and spin-$\frac32$ rest spinors (further details on spin-$\frac32$ matrices are given in Appendix D). The $\mathbf n^{(0)}$'s in (21) are the same as in (2)–(4); their l indices take the values 3 and 2σ depending on σ and σ', and $a(\sigma,\sigma')$, $b(\sigma,\sigma')$ are numerical coefficients (details are given in Appendix E).

Using (21), an expression similar to Eq.(17) can be found for $u(0\sigma)\bar u_\mu(0\sigma')$ and $u(0\sigma)\bar u^{[3/2]}(0\sigma')$, and by means of complex Lorentz transformations we obtain for the RS case

$$\begin{aligned} H_\mu&(p\sigma;p'\sigma') \\ &= u(p\sigma)\bar u_\mu(p'\sigma') \\ &= [(\gamma\cdot p + m)/2m][\gamma_\alpha h_{(+)}{}^\alpha{}_\mu + \gamma_5\gamma_\alpha h_{(-)}{}^\alpha{}_\mu] \\ &= [\gamma_\alpha h_{(+)}{}^\alpha{}_\mu + \gamma_5\gamma_\alpha h_{(-)}{}^\alpha{}_\mu][(\gamma\cdot p' + m')/2m'], \end{aligned} \tag{22}$$

where $h_{(\pm)}{}^\alpha{}_\mu$ are given in Appendix E. For the JW case we obtain

$$\begin{aligned} H^{[1/2,3/2]}&(p\sigma;p'\sigma') \equiv u(p\sigma)\bar u^{[3/2]}(p'\sigma') \\ &= \frac{\gamma\cdot p + m}{2m}(\gamma^{[1/2,3/2]}_{\mu\nu\alpha} g^{\mu\nu\alpha}_{(+)} + \gamma_5\gamma^{[1/2,3/2]}_{\mu\nu\alpha} g^{\mu\nu\alpha}_{(-)}) \\ &= (\gamma^{[1/2,3/2]}_{\mu\nu\alpha} g^{\mu\nu\alpha}_{(+)} + \gamma_5\gamma^{[1/2,3/2]}_{\mu\nu\alpha} g^{\mu\nu\alpha}_{(-)}) \\ &\quad \times \frac{\gamma^{[3/2]}_{\tau\rho\lambda} p'^\tau p'^\rho p'^\lambda + (m')^3}{2(m')^3}, \end{aligned} \tag{23}$$

where $\gamma^{[1/2,3/2]}_{\mu\nu\alpha}$ and $\gamma^{[3/2]}_{\tau\rho\lambda}$ are generalized Dirac matrices (see Appendix D) and

$$g^{\mu\nu\alpha}_{(\pm)} = (1/2m')p'^\nu h^{\alpha\mu}_{(\mp)}. \tag{24}$$

V. CONCLUSION

The advantage of the method given here (over FS) is that we take the trace of fewer γ matrices. This can be seen if we compare Eqs. (2) and (6) with Eqs. (7) and (10). But in the Dirac case this is not of much practical interest because Dirac γ matrices are easy to handle and the right-hand side of (6) involves at most only two γ matrices more than the right-hand side of (7). The FS method is therefore probably preferable in most practical calculations. For higher spin, however, $[u(p\sigma)\bar{u}(p\sigma)]$ is much more complicated than for spin $\frac{1}{2}$, while H_μ in (22), say, is not too complicated. Moreover, our method can be a helpful device in studying the amplitude in some cases. For instance in the Breit frame of p and p' (where $\mathbf{p}' = -\mathbf{p}$), the complex Lorentz-transformation coefficients $L^{(\pm)}_{\mu\nu}$ is very simple.[4] It should also be noted that the complex Lorentz-transformation equations (13) and (A1) can be used to write the amplitude as a product of Pauli matrices sandwiched between spin-$\frac{1}{2}$ rest spinors (i.e., Pauli spinors. This can also be done for higher spin.[4])

The results for $S(p'\sigma';p\sigma)$ and similar quantities defined in Appendix C [see (C1)–(C6)] will give the same answer as in the FS method. The connection between the two methods (see Appendix C) also gives the combined method mentioned at the end of Sec. III. This combined method may be an attractive alternative for some cases.

ACKNOWLEDGMENT

The authors thanks P.A. Amundsen for useful comments.

APPENDIX A

The complex Lorentz transformation coefficients can be calculated directly from (15) with $A = \exp(\frac{1}{2}\omega\hat{\mathbf{p}}\cdot\boldsymbol{\sigma})$ $B = \exp(\frac{1}{2}\omega'\hat{\mathbf{p}}'\cdot\boldsymbol{\sigma})$, $|\hat{\mathbf{p}}| = |\hat{\mathbf{p}}'| = 1$. But in our case it seems more favorable to calculate $L^{(\pm)}_{\mu\nu}$ directly. combining (13) and (16), we obtain

$$\Lambda(p)\gamma_\mu \Lambda^{-1}(p') = \gamma_\nu L^{(+)\nu}{}_\mu + \gamma_5\gamma_\nu L^{(-)\nu}{}_\mu, \quad \text{(A1)}$$

which implies

$$L^{(\pm)}_{\mu\nu} = \tfrac{1}{4}\,\mathrm{Tr}[\gamma^{(\pm)}\gamma_\mu \Lambda(p)\gamma_\nu \Lambda^{-1}(p')] \quad \text{(A2)}$$

where $\gamma^{(+)}\equiv 1$ and $\gamma^{(-)}\equiv -\gamma_5$. Using $\Lambda(p) = \cosh(\omega/2) + \boldsymbol{\alpha}\cdot\hat{\mathbf{p}}\sinh(\omega/2)$ and $\boldsymbol{\gamma} = \boldsymbol{\alpha}\gamma_2 = -\gamma_0\boldsymbol{\alpha}$, we obtain

$$\Lambda(p) = \gamma\cdot v\gamma_0, \quad v\equiv(\cosh(\omega/2),\hat{\mathbf{p}}\sinh(\omega/2)), \quad \text{(A3a)}$$

and similarly

$$\Lambda^{-1}(p') = \gamma_0\,\gamma\cdot v', \quad v'\equiv(\cosh(\omega'/2),\hat{\mathbf{p}}'\sinh(\omega'/2)). \quad \text{(A3b)}$$

Let $a = (a_0, \mathbf{a})$ be an arbitrary quantity with four components. Using $\gamma_0\gamma_\mu\gamma_0 = \tilde{\gamma}_\mu\equiv(\gamma_0 - \boldsymbol{\gamma})$, we obtain $[\tilde{a}\equiv(a_0, -\mathbf{a})]$

$$L^{(+)}_{\mu\nu}a^\nu = v_\mu\tilde{a}\cdot v' - \tilde{a}_\mu(v\cdot v') + v'_\mu\tilde{a}\cdot v, \quad \text{(A4)}$$

$$L^{(-)}_{\mu\nu}a^\nu = i\epsilon_{\mu\nu\alpha\beta}\,\tilde{a}^\nu v^\alpha v'^\beta. \quad \text{(A5)}$$

Note that

$$v^2 = 1, \quad p\cdot v = mv\cdot\xi^{(0)}, \quad mv\cdot\tilde{v} = p\cdot\xi^{(0)}. \quad \text{(A6)}$$

Using (A2)–(A3), the coefficients $L^{(\pm)}_{\mu\nu}$ can also be written

$$L^{(\pm)}_{\mu\nu} = \tfrac{1}{4}\,\mathrm{Tr}(\gamma^{(\pm)}\gamma_\mu\,\gamma\cdot v\gamma\cdot\xi^{(0)}\,\gamma_\nu\,\gamma\cdot\xi^{(0)}\,\gamma\cdot v'), \quad \text{(A7)}$$

which gives explicitly

$$L^{(+)}_{\mu\nu} = 2\xi^{(0)}_\nu(v'_\mu\, v\cdot\xi^{(0)} - \xi^{(0)}_\mu\, v\cdot v' + v_\mu\, v'\cdot\xi^{(0)}) - (\xi^{(0)})^2(v'_\mu\, v_\nu - g_{\mu\nu}\, v'\cdot v + v_\mu\, v'_\nu), \quad \text{(A8)}$$

$$L^{(-)}_{\mu\nu} = i\epsilon_{\mu\nu\alpha\beta}\,v^\alpha v'^\beta[2\xi^{(0)}_\nu\xi^{(0)\lambda} - g_\nu{}^\lambda(\xi^{(0)})^2], \quad \text{(A9)}$$

which is in agreement with (A4)–(A5) because $\xi^{(0)} = (1,\mathbf{0})$.

APPENDIX B

Define

$$N^{(\pm)}_{(l)\mu}\equiv L^{(\pm)}{}_\mu{}^\nu\, n^{(0)}_{(l)\nu}, \quad \xi^{(\pm)}_\mu\equiv L^{(\pm)}{}_\mu{}^\nu\,\xi^{(0)}_\nu, \quad \text{(B1)}$$

where the index l can take the values 3 and $2\sigma = \pm 1$, as in (2)–(4). then, using $\tilde{n}^{(0)} = -n^{(0)}$, $\tilde{\xi}^{(0)} = +\xi^{(0)}$, and $\tilde{a}\cdot\tilde{b} = a\cdot b$, we obtain

$$N^{(+)}_{(l)\mu} = v\cdot v'\, n^{(0)}_{(l)\mu} - (v'\cdot n^{(0)}_{(l)})\, v_\mu - (v\cdot n^{(0)}_{(l)})\, v'_\mu, \quad \text{(B2)}$$

$$N^{(-)}_{(l)\mu} = -i\epsilon_{\mu\alpha\beta\lambda}\, v^\alpha v'^\beta\, n^{(0)\lambda}_{(l)}, \quad \text{(B3)}$$

$$\xi^{(+)}_\mu = v_\mu(\xi^{(0)}\cdot v') - v\cdot v'\xi^{(0)}_\mu + v'_\mu(\xi^{(0)}\cdot v), \quad \text{(B4)}$$

$$\xi^{(-)}_\mu = i\epsilon_{\mu\nu\alpha\beta}\, v^\nu v'^\alpha\, \xi^{(0)\beta}. \quad \text{(B5)}$$

The coefficients in (10) are then given by [see (19)]

$$2h^\alpha_{(\pm)} = \delta_{\sigma'\sigma}\,\xi^{(\pm)\alpha} + c(l)N^{(\mp)\alpha}_{(l)}. \quad \text{(B6)}$$

Note that $\xi^{(+)}_\mu\to(1/m)p_\mu$ and $\xi^{(-)}_\mu\to 0$ when $p'\to p$. Moreover,

$$(1/m)p\cdot\xi^{(+)} = (1/m')p'\cdot\xi^{(+)} = v\cdot v', \qquad p\cdot\xi^{(-)} = p'\cdot\xi^{(-)} = 0, \quad \text{(B7)}$$

$$(1/m)p\cdot N^{(-)} = -(1/m')p'\cdot N^{(-)} = i\mathbf{n}^{(0)}_{(l)}\cdot[\mathbf{v}'\times\mathbf{v}], \quad \text{(B8)}$$

$$p\cdot n^{(0)}_{(l)} = 2(v\cdot p)(v\cdot n^{(0)}_{(l)}). \quad \text{(B9)}$$

APPENDIX C

Using (7) and (10), one obtains

$$S(p'\sigma';p\sigma)\equiv\bar{u}(p'\sigma')u(p\sigma) = (2/m)p\cdot h_{(+)} = (2/m')p'\cdot h_{(+)}, \quad \text{(C1)}$$

$$P(p'\sigma';p\sigma)\equiv\bar{u}(p'\sigma')\gamma_5 u(p\sigma) = (-2/m)p\cdot h_{(-)} = (2/m')p'\cdot h_{(-)}, \quad \text{(C2)}$$

$$V_\mu(p'\sigma';p\sigma)\equiv\bar{u}(p'\sigma')\gamma_\mu u(p\sigma) = 2h_{(+)\mu}, \quad \text{(C3)}$$

$$A_\mu(p'\sigma';p\sigma)\equiv\bar{u}(p'\sigma')\gamma_5\gamma_\mu u(p\sigma) = -2h_{(-)\mu}, \quad \text{(C4)}$$

where the $h_{(\pm)\mu}$'s are given by (B2)–(B6). Similarly, in the spin-$\frac{3}{2}$ case we have

$$S_\mu(p'\sigma';p\sigma)\equiv\bar{u}_\mu(p'\sigma')u(p\sigma) = (2/m)p_\alpha h_{(+)}{}^\alpha{}_\mu = (2/m')p'_\alpha h_{(+)}{}^\alpha{}_\mu, \quad \text{(C5)}$$

$$P_\mu(p'\sigma';p\sigma)\equiv\bar{u}_\mu(p'\sigma')\gamma_5 u(p\sigma) = (-2/m)p_\alpha h_{(+)}{}^\alpha{}_\mu = (2/m')p'_\alpha h_{(+)}{}^\alpha{}_\mu, \quad \text{(C6)}$$

$$V_{\alpha\mu}(p'\sigma';p\sigma)\equiv\bar{u}_\alpha(p'\sigma')\gamma_\mu u(p\sigma) = 2h_{(+)\mu\alpha}, \quad \text{(C7)}$$

$$A_{\alpha\mu}(p'\sigma';p\sigma)\equiv\bar{u}_\alpha(p'\sigma')\gamma_5\gamma_\mu u(p\sigma) = 2h_{(-)\mu\alpha}. \quad \text{(C8)}$$

From (C1) and Eqs. (5)–(10) we obtain

$$|S(p'\sigma';p\sigma)|^2 = \left(\frac{2p\cdot h_{(+)}}{m}\right)\left(\frac{2p'\cdot h^*_{(+)}}{m'}\right) = \mathrm{Tr}\{[u(p'\sigma')\bar{u}(p'\sigma')][u(p\sigma)\bar{u}(p\sigma)]\}. \quad \text{(C9a)}$$

Using (2) and the notation $s_\mu \equiv s(p,\sigma)_\mu \equiv 2\sigma n_{(3)}(p)_\mu$, we obtain

$$16(p\cdot h_{(+)})(p'\cdot h^*_{(+)})$$
$$= p\cdot p' + (p\cdot s')(p'\cdot s) - (mm' + p\cdot p')(s\cdot s'). \qquad \text{(C9b)}$$

To show (C9b) explicitly and directly from (B1)–(B8) and (3) is tedious because $h^\mu_{(+)}$ involves the four component objects v and v'. These objects are not 4-vectors and satisfy certain formulas together with $\xi^{(0)}$ and the 4-vectors p and p'. We also have to use relations like $\mathbf{n}^{(0)}_{(2\sigma)}\cdot[\mathbf{n}^{(0)}_{(2\sigma)}]^* = 1 - (\mathbf{n}^{(0)}_{(3)})^2$. Similar to (C9) we obtain

$$2[S(p'\sigma';p\sigma)]^* h_{(+)}{}^\mu$$
$$= \mathrm{Tr}\{[u(p'\sigma')\bar{u}(p'\sigma')]\gamma^\mu[u(p\sigma)\bar{u}(p\sigma)]\}$$
$$= (4mm')^{-1}[i\epsilon^{\mu\alpha\beta\rho} p'^\alpha p^\beta (s-s')^\sigma$$
$$+ (1 - s'\cdot s)(mp'^\mu + m'p^\mu) + m'(p\cdot s')s^\mu + m(p'\cdot s)s'^\mu], \qquad \text{(C10)}$$

$$-2[S(p'\sigma';p\sigma)]^* h_{(-)}{}^\mu$$
$$= \mathrm{Tr}\{[u(p'\sigma')\bar{u}(p'\sigma')]\gamma_5\, \gamma^\mu[u(p\sigma)\bar{u}(p\sigma)]\}$$
$$= (4mm')^{-1}[p'\cdot s\, p^\mu + p\cdot s' p'^\mu - (p'\cdot p + m'm)(s+s')^\mu$$
$$+ \epsilon^{\mu\alpha\sigma\rho}(mp' + m'p)^\alpha s'^\sigma s^\rho], \qquad \text{(C11)}$$

$$\pm 2[S_\alpha(p'\sigma';p\sigma)]^*\, h_{(\pm)\beta\mu}$$
$$= \mathrm{Tr}\{[u_\alpha(p'\sigma')\bar{u}_\beta(p'\sigma')]\gamma_\pm\, \gamma_\mu[u(p\sigma)\bar{u}(p\sigma)]\}, \qquad \text{(C12)}$$

where $\gamma_+ \equiv 1$ and $\gamma_- \equiv \gamma_5$.

APPENDIX D

The JW equation[4,6,7] for spin $\frac{3}{2}$ is

$$(\gamma^{[3/2]}_{\mu\nu\sigma} i^3\, \partial^\mu\, \partial^\nu\, \partial^\sigma - m^3)\psi^{[3/2]}(x) = 0 \qquad \text{(D1)}$$

where $\psi^{[3/2]}(x)$ is a $[2\cdot(2\cdot\frac{3}{2}+1)]$-component spinor field and

$$\gamma^{[3/2]}_{\mu\nu\sigma} = \begin{pmatrix} 0 & t^{[3/2]}_{\mu\nu\sigma} \\ \tilde{t}^{[3/2]}_{\mu\nu\sigma} & 0 \end{pmatrix}, \qquad \text{(D2)}$$

similar to (11). We have $t^{[3/2]}_{000} = 1$ and $t^{[3/2]}_{k00} = \frac{2}{3}s^{[3/2]}_k$, where $s^{[3/2]}_k$ are the spin matrices for spin $\frac{3}{2}$. Defining

$$\beta^{[3/2]} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{\alpha}^{[3/2]} = \frac{2}{3}\begin{pmatrix} \mathbf{s}^{[3/2]} & 0 \\ 0 & -\mathbf{s}^{[3/2]} \end{pmatrix}, \qquad \text{(D3)}$$

we have

$$\gamma^{[3/2]}_{000} = \beta^{[3/2]}, \quad \gamma^{[3/2]}_{k00} = \alpha^{[3/2]}_k \beta^{[3/2]}. \qquad \text{(D4)}$$

Further details are given in Refs. 4,6,7. The nonsquare matrices $\Sigma^{[1/2,3/2]}_{\mu\nu}$ in (20) are given by

$$\Sigma^{[1/2,3/2]}_{kl} = \epsilon_{kln}\Sigma^{[1/2,3/2]}_n,$$
$$\Sigma^{[1/2,3/2]}_{k0} = i\alpha^{[1/2,3/2]}_k = i\gamma_5\Sigma^{[1/2,3/2]}_k, \qquad \text{(D5)}$$
$$\boldsymbol{\Sigma}^{[1/2,3/2]} = \begin{pmatrix} \boldsymbol{\sigma}^{[1/2,3/2]} & 0 \\ 0 & \boldsymbol{\sigma}^{[1/2,3/2]} \end{pmatrix},$$
$$\Sigma^{[3/2,1/2]}_{\mu\nu} = \beta^{[3/2]}(\Sigma^{[1/2,3/2]}_{\mu\nu})^\dagger \beta^{[1/2]}. \qquad \text{(D6)}$$

The dipole-transition matrices $\boldsymbol{\sigma}^{[1/2,3/2]} = (\boldsymbol{\sigma}^{[3/2,1/2]})^\dagger$ occurring in (21) and (D6) are given by

$$\sigma^{[3/2,1/2]}_1 = \frac{1}{\sqrt{6}}\begin{pmatrix} -\sqrt{3} & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & \sqrt{3} \end{pmatrix},$$

$$\sigma^{[3/2,1/2]}_2 = \frac{i}{\sqrt{6}}\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & \sqrt{3} \end{pmatrix},$$

$$\sigma^{[3/2,1/2]}_3 = \sqrt{2/3}\begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}. \qquad \text{(D7)}$$

and the quadrupole transition matrices in (21) are given by

$$K^{[3/2,1/2]}_{kl} \equiv \tfrac{1}{2}(\sigma^{[3/2,1/2]}_k\, \sigma_l + \sigma^{[3/2,1/2]}_l \sigma_k),$$
$$K^{[3/2,1/2]}_{ll} = 0, \quad K^{[1/2,3/2]}_{kl} \equiv (K^{[3/2,1/2]}_{kl})^\dagger, \qquad \text{(D8)}$$

where the $\boldsymbol{\sigma}$ are the ordinary Pauli matrices. The matrices in (D7) and (D8) satisfy a certain algebra together with the spin matrices for spin $\frac{1}{2}$ and spin $\frac{3}{2}$, for example:

$$\sigma^{[1/2,3/2]}_k\, \sigma^{[3/2,1/2]}_l = \tfrac{2}{3}\,\delta_{kl} - \tfrac{1}{3}i\epsilon_{klm}\sigma_m. \qquad \text{(D9)}$$

Further details are given in Ref. 8. The nonsquare $\gamma^{[1/2,3/2]}$-matrices in (23) are given by

$$\gamma^{[1/2,3/2]}_{\mu\nu\alpha} = i\gamma_\alpha \Sigma^{[1/2,3/2]}_{\mu\nu}. \qquad \text{(D10)}$$

APPENDIX E

The coefficients a and b and the l indices in (21) are given by

$$\sigma' = \sigma: \quad a = \tfrac{1}{2}\sqrt{3/2}, \quad b = 2\sigma\cdot a, \quad l = l_1 = l_2 = 3,$$
$$\sigma' = 3\sigma: \quad a = -(3/\sqrt{2}),$$
$$b = \tfrac{1}{3}\cdot 2\sigma\cdot a, \quad l = l_2 = -2\sigma, \quad l_1 = 3, \qquad \text{(E1)}$$
$$\sigma' = -\sigma: \quad a = (\sqrt{6}/4)\cdot 2\sigma,$$
$$b = 2\cdot 2\sigma\cdot a, \quad l = l_2 = 2\sigma, \quad l_1 = 3,$$
$$\sigma' = -3\sigma: \quad a = 0, \quad b = 2\sigma\cdot\sqrt{2}, \quad l_1 = l_2 = 2\sigma,$$

Using (20), (21), (A1), (D9), one obtains

$$h_{(\pm)}{}^\alpha{}_\mu = \xi^{(\pm)\alpha}\tfrac{2}{3}\, a(\sigma,\sigma')n^{(0)}_{(l)\mu} + F^{(\mp)\alpha}{}_\mu, \qquad \text{(E2)}$$

where

$$F^{(\mp)\alpha}{}_\mu = L^{(\mp)\alpha}{}_\nu F^{(0)\nu}{}_\mu,$$
$$F^{(0)}_{00} = F^{(0)}_{0k} = F^{(0)}_{k0} = 0, \qquad \text{(E3)}$$

$$F^{(0)}_{mk} = \tfrac{1}{3}ia(\sigma,\sigma')\epsilon_{mkr}\, n^{(0)}_{(l)r} + b(\sigma,\sigma')$$
$$\times[\tfrac{1}{2}\delta_{ks}\,\delta_{rm} + \tfrac{1}{2}\delta_{kr}\,\delta_{sm} - \tfrac{1}{3}\delta_{sr}\,\delta_{km}]n^{(0)}_{(l_1)r}\, n^{(0)}_{(l_2)s}. \qquad \text{(E4)}$$

[1]H.W. Fearing and R.R. Silbar, Phys. Rev. D 6, 471 (1972).

[2]We use a metric where $a\cdot b = a_0 b_0 - \mathbf{a}\cdot\mathbf{b}$ and the following normalization of Dirac Spinors: $\bar{u}(p\sigma')u(p\sigma) = \delta_{\sigma'\sigma}$.

[3]Streater and Wightman, *PCT, Spin, Statistics and All That* (Benjamin, New York, 1964).
[4]J.O. Eeg, Phys. Norvegica **6**, 153 (1972).
[5]W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).
[6]H. Joos, Fortschr. Physik **10**, 65 (1962); S. Weinberg, Phys. Rev. **133**, B1318 (1964).
[7]This equation is studied by S. Sankaranarayanan and R.H. Good Jr., Nuovo Cimento **36**, 1303 (1965), D. Shay, H.S. Song, and R.H. Good Jr., Suppl. Nuovo Cimento **3**, 455 (1965), and in Ref. 4, which also have further references.
[8]J.O. Eeg, Lett. Nuovo Cimento **13**, 14 (1975) and Phys. Norvegica **8**, 137 (1976).

On the stable analytic continuation with rational functions

I. Sabba Stefanescu[a)]

Institut für Theoretische Kernphysik, Universität Karlsruhe, Germany
(Received 31 October 1978)

We discuss the use of rational approximants in the performance of stable analytic extrapolation from interior points of the analyticity domain to other interior points. We show that the instability of analytic extrapolation and the presence of noise sets an upper bound to the number of parameters that can be used in the solution. We generalize this result to other classes of functions which are used to fit experimental data and present a number of practical examples in form factor analysis.

1. INTRODUCTION

The process of data analysis in high energy physics involves analytic continuation as an essential step. The problem of analytic continuation is known to be mathematically ill-posed; by this one means that, although the analytic extension of a function defined on a certain piece of curve C in the complex plane is uniquely determined, the extensions of functions differing on C by any amount, no matter how small, can be arbitrarily different, at points lying outside C.

Consequently, the problem of the analytic extrapolation of a function given with errors on C—the "data function"—is, strictly speaking, meaningless. However, if one knows that the data refers to a function—the "true function"—which is holomorphic in some domain D of the complex plane, it is not absurd to try to devise a method of continuation, enjoying the property of stability: this means that its results, dependent on both the data function and the size of its errors, should converge in some sense to the "true" continuation, as the errors shrink to zero on C around the true function. Such stable analytic extrapolation procedures have been developed lately.[1–5]

Their prescription for the extrapolation is not uniquely defined. The function which is taken as the extrapolate of the data, and for which stability can be proved, obeys an extremal property: usually it is that analytic function for which some norm $\|\cdot\|_{\partial D}$ on the boundary ∂D of D attains its minimal value that is still consistent with the errors ϵ of the data function on C. Crudely speaking, in order that it converge to the true result, the extrapolate should not oscillate in D more than is strictly required by the data.

There is much arbitrariness allowed in the precise definition of this extremal property. By changing the type of the norm on ∂D, we change the extrapolation, but we still get a convergent result in the limit of zero errors. The rate of convergence changes, however, in an uncontrollable way. If we had some theoretical information about an upper bound for such a norm on ∂D, we could estimate, in principle at least, this convergence rate. Such information is in practice usually absent.

To summarize, analytic extrapolation methods require in general the extremal element f_e of the problem:

$$\inf\{\|f\|_{\partial D} \mid \|f-h\|_c \leqslant \epsilon\}, \tag{1.1}$$

where $\|f-h\|_c$ is some way of measuring the goodness of the fit to the data on C (this is usually fixed by statistical considerations to be a χ^2 expression). The choices of the norm $\|\cdot\|_{\partial D}$ are at present done in such a way that the computational algorithm for the extrapolation be maximally simple or that the aesthetic quality of the extrapolate come closer to the standards of some physicist.

The methods developed in Refs. 1–5 have been extensively applied in practical analysis.[5–9]

Despite the concern with aesthetics, the functions in terms of which the extrapolate is obtained have only a remote connection to the models in use in high energy physics, or even with standard approximation techniques, like the Padé one. Conversely, functions which fulfill the requirements of analyticity, but singled out by models, rather than by the mathematical techniques of Refs. 1–5 are used extensively in practice to fit data, without much regard for the stability problem. The excuse is sometimes that the simplicity of physics should select the solution with the smallest number of parameters. The most notable example is offered by rational approximants, which fall completely outside the scope of Refs. 1–5, despite their eminent physical relevance.

The stability of rational approximants under noise has been studied to the knowledge of the author, only "experimentally", in connection with the Padé technique.[10] The results are given as criteria fixing the numbers of terms which one must use in the solution in order to avoid fitting the noise and the development of instabilities. There is, however, no theoretical justification for these rules.

In this paper, we wish to ask the following question: Is it possible to choose the norm $\| \ \|_{\partial D}$ in such a way that the extremal element of (1.1) is naturally expressed in terms of rational functions? We shall prove that this is indeed the case, and show that it is possible to set an upper bound to the number of terms needed in the extrapolate f_e. This bound is different from the simple one given by the number of experimental points used in the fit, exists even for a continuum distribution of data, and is related to a peculiar property of the norm we choose. The estimates we give for it are rather coarse, but we try to make it plausible by means of practical

[a)]Supported by Deutsche Forschungsgemeinschaft, Bonn-Bad Godesberg.

examples that, by further refinement, it can be made quite low. The existence of this bound shows that we are able to produce, for any distribution of data, discrete or continuous, a sequence of rational approximants to it, whose poles lie outside a known holomorphy domain D and which tend to the data pointwise inside D.

In Sec. 2 of this paper, we discuss unstable problems in a general, slightly geometrical setting and analyze the relation to their reformulation, Eq. (1.1). Most of the results of Sec. 2 are known, but an attempt is made to generalization. In Sec. 3, we show that certain choices of the stabilizing norms do lead to stable extrapolates which contain a finite number of parameters, in particular to rational functions.

In Sec. 4 we derive an upper bound for the number of parameters that is needed in the stable extrapolation. The bound depends on the data function itself and on the noise. The procedure can be generalized to the other stabilizing norms, discussed in Sec. 3. In Sec. 5, we discuss applications to the analytic extrapolation in form factor problems (of the nucleon and pion). The conclusions contain some general remarks on the use of analytic extrapolation techniques in practice. The appendices gather some relevant numerical results.

All the discussion below refers to the extrapolation from interior points of the analyticity domain D, lying along a bounded data region C, to other points interior to D or to the boundary ∂D of D, as is the sitation in form factor problems or phase shift analyses.

2. UNSTABLE PROBLEMS AND "FLATTENED" BODIES

We wish now to review unstable problems in more accurate terms and introduce some notation.

A. Description of the data

We denote by $H(D)$ the vector space of functions holomorphic in the domain D. The function of interest, $f_0(z) \in H(D)$, is measured at N points z_i along the curve C, with errors $\sigma(z_i)$ and the results are the data function $h(z_i)$. We judge if a certain function $f(z) \in H(D)$ is in reasonable agreement with $h(z)$ on C, by computing

$$\chi_\rho^2(f-h) \equiv \sum_{i=1}^{N} \frac{|f(z_i)-h(z_i)|^2}{\sigma^2(z_i)} \equiv \int_C |f(z)-h(z)|^2 d\rho(z) \tag{2.1}$$

and then comparing $\chi_\rho^2(f-h)$ with a certain value χ_0^2, which is supposed to be known in advance. In (2.1), $d\rho(z)$ is a discrete measure with support concentrated at a finite number of points.

We can generalize (2.1) to the case of an arbitrary positive measure $d\rho(z)$ on C, and a continuous data distribution $h(z)$. Many of the results that follow are independent of whether $d\rho(z)$ has discrete or continuous support. However, true analytic extrapolation can be performed only off an infinite number of points, and we still have to see what we mean by this when we use only a finite number. For continuous data distributions we shall assume that $\int_C d\rho(z)$ is finite and that $h \in L^2(\rho)$.

It is natural to regard χ_ρ^2 as a norm in a Hilbert space $L^2(\rho)$ of complex square integrable functions defined on C.

We wish to study the behavior of approximants in the limit of "errors going to zero." This means the following: For any ϵ, there exists a data function h_ϵ which is such that $\chi_\rho^2(h_\epsilon - f_0) \leqslant \chi_0^2\epsilon$; the approximants $f_\epsilon \in H(D)$ satisfy $\chi_\rho^2(f_\epsilon - h_\epsilon) \leqslant \chi_0^2\epsilon$, and clearly $\chi_\rho^2(f_\epsilon - f_0) \to 0$, when $\epsilon \to 0$. We notice that, with this definition, we keep the relative magnitude of the errors at different points unchanged, and let only an over-all scale factor go to zero.

B. Stabilizing bodies

If $d\rho(z)$ is not of finite type, χ_ρ^2 can be regarded as a norm in $H(D)$; indeed $\chi_\rho^2(f) = 0 \Rightarrow f = 0$ for $f \in H(D)$ by the uniqueness of analytic continuation. We call this normed space H_{χ^2}. Clearly, convergence in H_{χ^2} cannot be used to set up reasonable approximants to functions in $H(D)$: Two functions in $H(D)$ which are close to each other according to χ_ρ^2 may be very different at points lying outside C. This is just the statement of the instability of analytic continuation; a possible proof is given in Ref. 11, Appendix A. We conclude the open sets in H_{χ^2} are "too coarse."

We are interested in definitions of convergence which require an approximation of the limit function also at points of D, lying outside C. For the moment we refer loosely to such an interesting convergence as an "i-convergence." Only those convergences are really interesting which are strictly stronger than the χ_ρ^2 one: $f_n \xrightarrow{i} f \Rightarrow \chi_\rho^2(f_n - f) \to 0$, but not conversely.

We say that a set $\mathscr{S} \subset H_{\chi^2}$ is a stabilizer with respect to a certain i-convergence if it is closed with respect to χ_ρ^2 and any sequence $f_n \in \mathscr{S}$ which converges in the sense of χ_ρ^2 also converges according to i. This means that, by intersecting with $\mathscr{S}$, we can throw out enough elements in each open set of H_{χ^2}, so that χ_ρ^2 convergence becomes equivalent to an interesting i-convergence. Practically we know then how to find i-approximants to the true function f_0, if we know that $f_0 \in \mathscr{S}$: any function $f_\epsilon \in \mathscr{S}$, which satisfies

$$\chi_\rho^2(f_\epsilon - h_\epsilon) \leqslant \chi_0^2 \epsilon \tag{2.2}$$

is such an i-approximant. Indeed, since $f_0 \in \mathscr{S}$, we are sure that such an f_ϵ exists; we deduce $\chi_\rho^2(f_\epsilon - f_0) \leqslant 2\chi_0^2\epsilon$ and so $\lim_{\epsilon\to 0}(i) f_\epsilon = f_0$.

We now wish to see how one obtains such stabilizing sets: A sufficient characterization is given by Tykhonov.[12]

Theorem 2.1: If $\mathscr{S}$ is compact with respect to i-convergence, then it is also a stabilizer with respect to it.

Proof: Clearly, $\mathscr{S}$ is closed with respect to χ_ρ^2 convergence, because it is closed with respect to i. We can assume $0 \in \mathscr{S}$. Let then $f_n \in \mathscr{S}$ be a sequence of elements which converges with respect to χ_ρ^2 to an element $f \in \mathscr{S}$ and has a unique limit. From f_n we can extract an i-converging subsequence, to $f' \in \mathscr{S}$. Then, also $\chi_\rho^2(f'_n - f') \to 0$. Since $\chi_\rho^2(f'_n - f) \to 0$, we conclude $\chi_\rho^2(f' - f) = 0$ and, because χ_ρ^2 is a norm, that $f' = f$ and $f'_n \xrightarrow{i} f$. This is true for any converging subsequence, and we conclude that $f_n \xrightarrow{i} f$.

We shall assume from now on that the set $\mathscr{S}$ is also convex. This is always the case in practical applications, since convexity mirrors our ignorance best.

C. Stabilizing levers. Critical values

We have shown how we can find approximants to the true function, in the sense of a quite freely defined *i*-convergence. This works only if the true function is known to belong to the stabilizing body $\mathscr{S}$. In many cases, we do not know this, but we can almost always assert that the true function can be brought inside the body $M\mathscr{S}$, $M>0$, for a sufficiently large value of M. We call M a stabilizing lever.

We wish to show now how, even if we do not know this value of M, we can nevertheless construct an *i*-convergent sequence of approximants to the true function.

To this end, for a given M and a data function h, define $\chi^2_{\rho\,\min}(M,h)$ to be the smallest distance from h to $M\mathscr{S}$. It is relevant that:

Theorem 2.2: There exists a unique function in $M\mathscr{S}$ for which the smallest distance is attained.

Proof: This is just the statement that the minimal distance to a closed convex set in the Hilbert space $L^2(\rho)$ is always attained by some element.[13]

Clearly, $\chi^2_{\min}(M)$ is a monotonically decreasing function of M. For errors of magnitude ϵ we define $M_0(\epsilon,h)$ to be the smallest value of M for which $\chi^2_{\rho\,\min}(M_0,h)=\chi^2_0\epsilon$ (if this value exists) and we call the corresponding extremal function $f_{M_0(\epsilon);h}$. M_0 is called the critical value of the stabilizing lever.

Consider now a sequence of data functions h_ϵ which approximate the true function f_0 increasingly well in χ^2_ρ norm:

$$\chi^2_\rho(h_\epsilon-f_0)\leqslant\chi^2_0\epsilon \tag{2.2'}$$

and let M_{true} be such that $f_0\in M_{\text{true}}\mathscr{S}$. Then we are sure that $M_0(h_\epsilon;\epsilon)$ exists and $M_0(h_\epsilon;\epsilon)\leqslant M_{\text{true}}$. Consequently $f_{M_0(\epsilon);h_\epsilon}\in M_{\text{true}}\mathscr{S}$ for all ϵ. With the help of Theorem 2.1, we have now proven the following.[14]

Theorem 2.3: The functions $f_{M_0(\epsilon);h_\epsilon}$ converge (i) to f_0, as $\epsilon\to 0$. This statement is, in our opinion, important, because no knowledge of M_{true} is assumed and we still get convergence. We do not know, however, how good this convergence is (unless we know M_{true}). Even if M_{true} is known, we shall see that the computation of M_0 and of the corresponding extremal function $f_{M_0}\equiv f_{M_0(\epsilon);h_\epsilon}$ serves estimating the quality of the extrapolation. We conclude that the problem of analytic continuation is turned into the well-defined task of finding M_0 and f_{M_0} [cf. Eq. (1.1)].

D. Finite numbers of points; geometrical interpretation

Assume now that $d\rho(z)$ is of finite type, with support at $\{z_i\}_{i=0}^N$. It is now natural to define the analytic extrapolation of data given at these points as the function $f_{M_0(\epsilon);h_\epsilon^{(N)}}$ for a certain choice of the stabilizing set $\mathscr{S}$ (convex and *i*-compact). The existence of at least one such function is ensured by Theorem 2.2, if $f_0\in M\mathscr{S}$ for some M.

Now, the limit $f_{M_0(\epsilon);h_\epsilon^{(N)}}\to f_0$ means that we must allow both $\epsilon\to 0$ and $N\to\infty$. If only $\epsilon\to 0$, the functions $f^N_{M_0(\epsilon);h_\epsilon}$ do not have as limit the true function, but the function with the smallest M, which assumes a given set of values at the N points z_i.

If ϵ is sufficiently small and N is sufficiently large, the functions $f_{M_0(\epsilon);h_\epsilon^{(N)}}$ can (i) approximate f_0 as well as one wishes. To see this, let $\{z_n\}_{n=0}^\infty$ be a sequence of points on C, and let $f^{(N)}_{\min}$ be the function with the smallest $M\equiv M^{(N)}_{\min}$ which assumes the values $f_0(z_1),\dots,f_0(z_N)$ at the first N points $z_1,\dots,z_N$. If N is sufficiently large, $f^{(N)}_{\min}$ is arbitrarily close (i) to f_0. This follows since $f^{(N)}_{\min}\subset M_{\text{true}}\mathscr{S}$ for all N. On the other hand, at fixed N, the function $f_{M_0(\epsilon);h_\epsilon^{(N)}}$ is arbitrarily close to $f^{(N)}_{\min}$ if ϵ is sufficient small. This proves our assertion.

When $d\rho(z)$ is of finite type, we can picture the M_0 problem in $L^2(\rho)$ as follows: Along the N axes of $L^2(\rho)$, we measure the values of holomorphic functions at $z_1,z_2,\dots,z_N$. The set of admissible values according to statistics is a sphere (or ellipsoid) centered around the point of coordinates $h_1,h_2,\dots,h_N$ (and semiaxes $\sigma_1,\sigma_2,\dots,\sigma_N$). It is important that the noise is isotropic with respect to the axes [Eq. (2.1)].

The set of values assumed at $z_1,z_2,\dots,z_N$ by the functions belonging to the stabilizing set $\mathscr{S}$ makes up a closed convex body $\mathscr{S}_N$. This set is dilated, or contracted by a factor M, until it just touches the sphere $\chi^2_\rho=\chi^2_0$; the corresponding value of M is M_0.[15]

It is helpful to notice that the instability of analytic continuation is synonimous with the fact that the body $\mathscr{S}_N$ is "flattened," i.e., it has the property that, for any $r>0$, there exists a number N_r or dimensions (an index N_r in the sequence $\{z_n\}_{n=0}^\infty$), so that $\mathscr{S}_N$ contains no sphere of radius r for $N>N_r$.

To see this, consider the set $\mathscr{S}'$ of sequences $\{f(z_n)\}_{n=0}^\infty$ of values assumed by functions in $\mathscr{S}$ at the points $\{z_n\}_{n=0}^\infty$, and for some sequence $\{f_1(z_n)\}_{n=0}^\infty\in\mathscr{S}'$ the "sphere" of sequences $\{h_i\}_{i=0}^\infty$, so that $\Sigma_{i=1}^\infty[f_1(z_i)-h_i]^2/\sigma_i^2<r^2$. The instability of analytic continuation means that, for any r, there are points $\{h_i\}_{i=0}^\infty$ in this sphere, so that there is no $f\in\mathscr{S}$ with $f(z_i)=h_i$, for all i. If, on the other hand, $\mathscr{S}_N$ did contain a sphere of radius r, for any N, we could contradict this statement.

We conclude there are directions in $L^2(\rho)$ along which $\mathscr{S}_N$ gets increasingly flattened, as N increases. We notice then that the experimental point $h_1,h_2,\dots,h_N$ which is the noise-affected "measurement" of a point in $\mathscr{S}_N$, has almost no chance to lie inside $\mathscr{S}_N$; the isotropic noise throws it, with a probability going to 1 as $N\to\infty$ outside $\mathscr{S}_N$. We expect therefore that the minimal distance from h to $\mathscr{S}_N$ is (almost) always nonzero, even for quite large values of the scale factor M (see also Sec. 2E).

The simplest example of such a flattened body is an ellipsoid

$$\sum_{i=1}^N\lambda_i y_i^2=M^2 \tag{2.3}$$

$[y_i=f(z_i)]$, with λ_i rapidly increasing with i. Its surface has regions of large curvature (almost "edges") at the ends of the big semiaxes and "flat" regions at the ends of the small semiaxes. We expect then the following effect: There is a much larger probability for the function which realizes the smallest χ^2 in $\mathscr{S}_N$ to lie near the region of "edges" than on the flat sides. We expect then that the extremal function can be very well approximated by a few of its components, lying on the big semiaxes of (2.3).

In this paper, we shall describe certain stabilizing sets $\mathscr{S}$, whose images $\mathscr{S}_N$ do not have just regions of high curvature, but well-defined "edges," that is, points on their surface at which a continuum of planes is tangent to $\mathscr{S}_N$. We shall show that these "edges" can be described by a smaller number of parameters than is needed by the whole $N-1$ dimensional surface. As a consequence, we expect the exact extremal function (rather than approximants of it) to have, in these situations, a simple structure, if the data function is sufficiently simple. This is, in the opinion of the author, the basic reason why rational functions appear as solutions to problems like (1.1); we shall deal with this in Sec. 4.

E. Examples

We now discuss some examples, just to show that generality has been profitable. Let us namely see how we can choose $\mathscr{S}$ (the "i-convergence") so that, for any $z\in D$, the sequence of values of the approximants $f_\epsilon(z)$ converges to the value of the true function, $f_0(z)$.

$e1$: There are many ways in which we can identify this requirement with that of weak convergence in a certain normed linear subspace of $H(D)$. Indeed, if D is mapped onto the unit disk, for any H^p subspace of $H(D)$,[13,16] $p\geqslant 1$, the integrals

$$f(z)=\int_{\partial D}\frac{f(z')dz'}{z'-z} \tag{2.4}$$

are well defined for any z inside D, and represent bounded linear functionals on H^p, $p\geqslant 1$. So, it is sufficient to take the (i) compact set of Theorem 2.1 as any weakly compact set of $H^p, p\geqslant 1$. For $1<p<\infty$ any bounded set in H^p norm has this property (by the Alaoglu theorem[13] and the Riesz representation of linear functionals[17]). For $p=\infty$, any bounded set contains a subsequence which converges pointwise on compact subsets of D, by Montel's principle.[13] So, any set $\mathscr{S}$ of the type

$$\|f\|_p\leqslant M,\quad 1<p\leqslant\infty, \tag{2.5}$$

can be taken as a stabilizing set. These stabilizing sets have been used in practice extensively, particularly the H^2 and H^∞ ones.[1,3,4,5,15]

$e2$: Instead of resorting to the Cauchy representation, we can use the Poisson one (Schwartz–Villat[13]), by which we regard the value at an interior point as a continuous linear functional over some normed space of real functions defined on the boundary:

$$f(z)=\frac{1}{2\pi}\oint_{\partial D}\frac{\zeta+z}{\zeta-z}\,d\sigma(\zeta), \tag{2.6}$$

where $\sigma(\zeta)$ is a function with bounded variation. We choose then weakly compact sets in such spaces of functions. In (2.6) ζ maps the domain D onto the unit disk.

If $f(z)$ is real analytic, and the boundary ∂D consists of a cut, we have a Cauchy formula, similar to (2.6), in terms of the imaginary part $d\sigma(\zeta)$:

$$f(z)=\frac{1}{\pi}\int_{\text{cut}}\frac{d\sigma(\zeta)}{\zeta-z},\quad \zeta \text{ real}. \tag{2.7}$$

If $d\sigma(\zeta)=\hat\sigma(\zeta)d\zeta$ and $\hat\sigma(\zeta)$ belongs to some $L^p(\partial D)$, $1<p\leqslant\infty$, any bounded set in such an L^p is weakly compact and stabilizes the extrapolation. So, the set $\mathscr{S}$ can be chosen of the form $\|\hat\sigma(\zeta)\|_p\leqslant M$.

It is interesting to consider some special cases:

(a) A bounded set in the space of functions of normalized bounded variation on ∂D,

$$\int_{\partial D}|d\sigma(\zeta)|\leqslant M. \tag{2.8}$$

The space of functions of normalized bounded variation is dual to that of continuous functions defined on ∂D. The set (2.8) is then weak-*-compact (by Helly's theorem or, equivalently, by Alaoglu's theorem) and again serves our purposes as stabilizer (i.e., any bounded sequence $\{\mu_n\}$ contains a subsequence $\{\mu_n'\}$ so that $\int f d\mu_n'$ converges, for any continuous function f on ∂D). The set $\mathscr{S}$ described by (2.8) is the one which produces rational functions as elements of minimal distance to an exterior point h.

(b) A closed subset of (2.8) is that of bounded positive measures. It is the only one which occurs sometimes more or less naturally as a stabilizing condition, because of positivity requirements on the amplitude. This set is of interest in the discussion of the extrapolation from interior points of D, under an L^∞ stabilizing condition.

(c) We shall also use (Sec. 5) the condition

$$-\hat\sigma_0(\zeta)\leqslant\hat\sigma(\zeta)\leqslant\hat\sigma_0(\zeta), \tag{2.9}$$

which is a closed subset of (2.8) [$\hat\sigma_0(\zeta)$ is a given function].

We might be interested in producing approximants which converge even on the boundary ∂D of D to the value of the true function (extrapolation to the cut). For $z\in\partial D$, the functional (2.6) is unbounded on the sets described above. It is nevertheless bounded with respect to the uniform norm of $\hat\sigma(\zeta)(d\sigma(\zeta)\equiv\hat\sigma(\zeta)d\zeta)$ on the subset of functions which are Hölder continuous of some index $\alpha\geqslant\alpha_0>0$ on ∂D. According to the Arzela–Ascoli theorem, the set of functions $\hat\sigma$ for which $\sup_{z_1,z_2\in\partial D}|\hat\sigma(z_1)-\hat\sigma(z_2)|/|z_1-z_2|^{\alpha_0}$ is bounded, is compact in the uniform norm. Then, according to Theorem 2.1, within the set of functions for which this quantity is bounded by some number, the extrapolation to the cut is stable.[18] We discuss an application of this to the pion form factor in Sec. 5D ($\alpha=1$).

It is interesting to see the relation of these facts to the reproducing kernel methods for analytic continuation, introduced in Ref. 19. To this end, let us notice that stability on the cut is expressed, e.g., by the statement $\chi_p^2(f_\epsilon-f_0)\to 0$ implies $\sup_{z\in\partial D}|f_\epsilon(z)-f_0(z)|\to 0$. Consequently, according to Theorem 2.1 we should be able to isolate a compact subset in the set A of analytic functions continuous on the boundary. The following statement is then relevant:

Let L be a Hilbert space of functions analytic in $|z|<1$, whose norm we denote by $\|\cdot\|_1$, and which admits of a reproducing kernel $H(z,w)$. Assume the latter is such that, for any z, $|z|=1$, $H(z,w)$ is continuous with respect to w on $|w|=1$. Then a bounded set with respect to $\|\ \|_1$ is compact with respect to uniform norm on $|z|=1$.

This is a consequence of the Arzela–Ascoli theorem. To see this, it suffices to show that the set of functions f, $\|f\|_1 \leqslant K$ is uniformly bounded and equicontinuous on $|z| = 1$. Uniform boundedness follows from the boundedness of $H(z,z)$:

$$|f(z)|^2 = |(H(z,w),f(w))_1|^2 \leqslant \|f(w)\|_1^2 H(z,z) \leqslant K^2 M. \tag{2.10}$$

Equicontinuity follows from

$$\begin{aligned}|f(z_1) - f(z_2)|^2 &= |(H(z_1,w) - H(z_2,w_1), f(w))_1|^2 \\ &\leqslant K^2 \|H(z_1,w) - H(z_2,w)\|_1^2 \\ &= K^2(H(z_1,z_1) - H(z_1,z_2) - H(z_2,z_1) \\ &\quad + H(z_2,z_2)) \leqslant K^2|H(z_1,z_1) - H(z_1,z_2)| \\ &\quad + K^2|H(z_2,z_1) - H(z_2,z_2)|\,.\end{aligned} \tag{2.11}$$

We notice that the degree of continuity of f is dictated by that of $H(z,w)$. Equations (2.10) and (2.11) prove the announced statement.

Reproducing kernels with various degrees of smoothness on the boundary are discussed in detail in Ref. 19. In particular, the author introduces there the two parameter hypergeometric kernel

$$H(z_1,z_2) = \sum_{n=2}^{\infty} \frac{\Gamma(2\nu + b + 1)\Gamma(b+n)}{\Gamma(2\nu + b + n + 1)\Gamma(b)} (z_1^* z_2)^n\,, \tag{2.12}$$

which is sufficient by flexible to accommodate many hypotheses concerning the behavior of the function on the boundary. Practical applications are discussed in, e.g., Refs. 6, 20, 21. The analysis of Ref. 7 uses a space of functions corresponding approximately to the choice: $\nu = 1, b = 0$ in (2.12). More exactly it reads for this case:

$$H(z_1,z_2) = \sum_{n=0}^{\infty} \frac{1}{(n+1)^3} (z_1^* z_2)^2\,. \tag{2.13}$$

If R_n denotes the coefficients of $(z_1^* z_2)$ in (2.12), (2.13), the norms for which the kernels (2.12), (2.13) are reproducing are $\|f\|_1 = \Sigma_{n=0}^{\infty} c_n^2 R_n$, where c_n are the Taylor coefficients of f.

*e*3: After having described some ways in which an interesting convergence can be chosen, we now turn to some practical examples concerning the flattening of the stabilizing body, discussed in Sec. 2D.

"Flattening" is equivalent to the statement that the measurements of the values of an analytic function inside its holomorphy domain have a large chance to produce results which are inconsistent with analyticity and a "reasonable, physical" value of the bound M for a large enough number of points. The analysis of the data in the spacelike region of the pion form factor offers such situations, since a value of the bound M [L^∞ norm, cf. Eq. (2.5)] in the timelike region is approximately known.

In Ref. 22, the authors obtained bounds, by means of analyticity, for the value of the pion form factor at various spacelike points from a timelike bound and knowledge of a few, "experimental" spacelike values. Their results got inconsistent with the measured values at those points, if they took into account more than three spacelike values. This is not due to any systematic experimental errors, but just shows the "flattening" effect of the timelike bound.

Another example of how sensitive a property analyticity and boundedness together represent can be obtained by computing, for the pion form factor, for instance, the minimal timelike bound for which the centers of the experimental points are the trace of a function analytic in the cut plane. This is done in Appendix A, for an increasing number of such points taken from actual experiments. By the same mechanism, the computation of the minimal value of the hadronic contribution to the muon magnetic moment leads to values 10^4 larger than the true one, in the situation when errors are not taken into account (see Ref. 23 for more details).

We emphasize that these statements concerning "flattening" are strictly equivalent to the statement of the instability of analytic continuation, and contain no new information. In the opinion of the author, they can help, however, specify in geometrical language the problem of analytic extrapolation.

3. DESCRIPTION OF THE EXTREMAL FUNCTIONS

We now turn to the problem of actually finding $\chi^2_{\min}$ and the corresponding unique extremal function for some stabilizing bodies $\mathscr{S}$.

We shall deal only with the stabilizers described in Sec. 2, e2, e3, i.e., with sets of holomorphic functions described by

$$f(z) = \int_I k(z,w)d\sigma(w)\,, \tag{3.1}$$

where $k(z,w)$ is the Poisson, Cauchy, or a related kernel (see Sec. 5C). The values of the function $f(z)$ on C, which are measured, will be assumed to be real, as is the case in practice. The function $k(z,w)$ is analytic in w, in a neighborhood of the domain where the integral is performed, and assumes for $z \in C$ real values as a function of w in the integration domain. The stabilizers of Sec. 2, *e*1, particularly the L^2 ones, have already received much attention.[1,4,15] We shall use the notation $(\ ,\ )_p$ for the scalar product defined by χ_p^2, Eq. (2.1). We shall assume that I is compact, although this is by no means essential. A well-known characterization of the extremal element $\bar{f}(z)$ in $\mathscr{S}$ is given by the following:

Theorem 3.1: For any $f \in \mathscr{S}$

$$(h - \bar{f}, \bar{f})_p \geqslant (h - \bar{f}, f)_p\,. \tag{3.2}$$

Proof: This follows by writing

$$\frac{\partial}{\partial \lambda} \|h - \lambda f - (1-\lambda)\bar{f}\|_p^2 \geqslant 0 \tag{3.3}$$

at $\lambda = 0$, for any $f \in \mathscr{S}$. We shall denote for convenience

$$n \equiv \frac{h - \bar{f}}{\chi_p(h - \bar{f})}\,. \tag{3.4}$$

It is the unit normal to the plane tangent to $\mathscr{S}_N$ at the extremal point $\bar{f}$.

We can now describe the points of minimal distance from $\mathscr{S}$ to h, when we choose $\mathscr{S}$ to be generated by the set (2.8).

Theorem 3.2: The function f generated according to (3.1) by $d\sigma(w)$ obeying (2.8) on which the minimal distance is attained is of the form

$$\bar{f}(z) = \sum \lambda_i \, k(z,w_i)\,, \tag{3.5}$$

where the sum extends over a finite number of terms and

$$\sum |\lambda_i| = M \tag{3.6}$$

Proof: We notice first that, as a corollary of (3.2), if we choose

$$d\sigma(w) = \pm M\delta(w - w_0)dw, \quad w_0 \in I, \tag{3.7}$$

we have the inequality

$$(n,\bar{f}) \geqslant |(n,k(\cdot,w_0))_\rho| M \tag{3.8}$$

for all w_0 in the domain of integration in (3.1).

We have then the following chain of inequalities, using (3.1) and (3.8):

$$(n,\bar{f})_\rho = \left(n, \int_I k(z,w)d\bar{\sigma}(w)\right)_\rho$$

$$= \int_I (d\bar{\sigma}(w)(n,k(\cdot,w))_\rho \leqslant \int_I |d\bar{\sigma}(w)|\,|(n,k(\cdot,w))_\rho| \tag{3.9a}$$

$$\leqslant \int_I |d\bar{\sigma}(w)| \frac{(n,\bar{f})_\rho}{M} \tag{3.9b}$$

$$\leqslant (n,\bar{f})_\rho \tag{3.9c}$$

We conclude that inequalities (3.9a), (3.9b), and (3.9c) must turn into equalities. Inequality (3.9a) requires

$$\operatorname{sgn}[d\bar{\sigma}(w)] = \operatorname{sgn}(n,k(\cdot,w))_\rho \tag{3.10}$$

for w in the domain of integration. Inequality (3.9b) requires

$$|(n,k(\cdot,w))_\rho| = (n,\bar{f}/M)_\rho \tag{3.11}$$

at all points of the domain of integration where $d\bar{\sigma}(w) \neq 0$.

But $(n,k(\cdot,w))_\rho$ is an analytic function of w, which is real in the compact domain of integration. It cannot assume either of the values $\pm(n,\bar{f}/M)$ more than a finite number of times. We conclude

$$d\bar{\sigma}(w) = \sum \lambda_i \, \delta(w - w_i)|dw|\,, \tag{3.12}$$

which shows that the extremal function is indeed of the form (3.5). Finally, inequality (3.9c) requires condition (3.6).

By using the same reasoning, we can also get the following results:

(a) For the body $\mathscr{S}_+$, generated by functions with positive real (or imaginary) parts:

Theorem 3.3: The function $\bar{f}$, generated via (3.1) by $d\sigma(w)$ obeying

$$d\sigma(w) \geqslant 0, \quad w \in I \tag{3.13}$$

on which the minimal distance to $\mathscr{S}_+$ is attained is of the form

$$\bar{f} = \sum \lambda_i k(z,w_i)\,, \quad \lambda_i > 0, \tag{3.14}$$

where the sum extends over a finite number of terms.

Proof: For this situation, Theorem 3.1 implies both

$$(n,\bar{f}) = 0 \tag{3.15}$$

and

$$(n,k(\cdot,w_0))_\rho \leqslant 0, \quad w_0 \in I. \tag{3.16}$$

Indeed, according to Theorem 3.1, we have both

$$(n,\bar{f}) \geqslant \lambda(n,\bar{f}) \tag{3.17}$$

for all positive λ, which implies (3.15) and

$$(n,\bar{f})_\rho \geqslant \lambda(n,k(\cdot,w_0))_\rho \tag{3.18}$$

for all positive λ, by choosing

$$d\sigma(w) = \lambda\delta(w - w_0), \quad \lambda > 0. \tag{3.19}$$

Eq. (3.18) implies (3.16). The chain of inequalities analogous to (3.9) is then

$$0 = (n,\bar{f}) = \int_I d\bar{\sigma}(w)(n,k(\cdot,w))_\rho \leqslant 0. \tag{3.20}$$

Equality in the last step is possible only if $(n,k(\cdot,w)) = 0$ inside supp$d\bar{\sigma}(w)$. Again, analyticity in w of $(n,k(\cdot,w))$ implies the discreteness and finiteness of the support, and so the form (3.14) for the extremal function.

(b) For the body $\mathscr{S}$ generated by functions with bounded real (or imaginary parts):

Theorem 3.4: The function $\bar{f}$, generated via (3.1) by $d\sigma(w)$ obeying Eq. (2.9), for which the minimal distance to $\mathscr{S}$ is attained is of the form

$$\bar{f}(z) = \sum_i (-1)^i \int_{i-1}^{x_i} k(z,w)d\sigma_0(w), \quad x_0 = a, x_N = b \tag{3.21}$$

where the sum extends over a finite number of terms.

Proof: By means of Theorem 3.1 we have the following chain of inequalities:

$$(n,\bar{f})_\rho = \int_I d\bar{\sigma}(w)(n,k(\cdot,w))_\rho$$

$$\leqslant \sum_i \left[\int_{x_i}^{x_i} d\sigma_0(w)(n,k(\cdot,w))_{\rho^+}\right.$$

$$\left. - \int_{x_i}^{x_i} d\sigma_0(w)(n,k(\cdot,w))_{\rho^-}\right]$$

$$\leqslant (n,\bar{f})_\rho\,. \tag{3.22}$$

In Eq. (3.22), the points x_i are the roots in w of $(n,k(\cdot,w))_\rho$, which are situated on a, b. Equality can be obtained in (3.22) if the extremal measure $d\bar{\sigma}(w)$ is alternatively equal to $\pm d\sigma_0(w)$, where $(n,k(\cdot,w))_\pm$ is positive/negative. Analyticity in w of $(n,k(\cdot,w))$ implies as before that it can change sign only a finite number of times on a,b, therefore, that the sum must stop after a finite number of terms.

We now interpret these results: We see that the functions which minimize χ_0^2 must be found among a restricted, finite-dimensional class of possibilites, depending on the stabilizing body $\mathscr{S}$. Since $k(z,w)$ can be a rational function, we see that the expressions (3.5), (3,14) are rational. The stable analytic extrapolation is to be found among these extremal functions, for that special choice of the scale factor M of $\mathscr{S}$ for which $\chi^2_{\min}(M) = \chi_0^2$ (cf. Sec. 2). So, we get a sequence of rational functions which converges to the true function in the limit of zero errors. In Eq. (3.14), the rational function has positive residua, but the stabilizing condition under which this happens is likely to be valid only under special circumstances.

It is important to notice the formal independence of the solution from the special form assumed by $\rho(z)$. In particular, even if the number of experimental points is infinite, but the errors are finite, the stable analytic extrapolation under condition (2.8) is a rational function.

The results of this section are a generalization of those obtained by Raszillier and Schmidt, for the case of a finite number of points and an L^∞ bound. They may *a priori* be regarded as rather weak, because we do not yet know any criterion for fixing the number of terms occuring in these sums. In practice, these methods seem, however, to be surprisingly efficient, in that the number of terms (parameters) needed by the extremal function appears to very small. In the next section, we shall show how one can understand this occurrence, in that we shall place bounds on the number of terms needed in these sums, as a function of h and the errors. The practical results are discussed in Sec. 5.

4. THE SIMPLIFYING PROPERTIES OF NOISE

A. Edges and simple formulas

In this paragraph we discuss shortly an analytic continuation problem of the L^2 type, in order to show that the flattening of the stabilizing body implies the existence of good low-dimensional approximants to the exact minimal function (for "reasonable" data functions). Plausible geometric reasons for this have already been given at the end of Sec. 2D. We then show that the stabilizing bodies studied in Sec. 3 have, indeed, well-defined edges, which can be described by "few" parameters; we expect, therefore, that, in these situations, the exact extremal function can assume a simple form, not just approximants of it.

The results presented in this paragraph are not new; those concerning the L^2 problem can be found in Refs. 1, 15 or in Ref. 24. Those concerning the description of the stabilizing bodies for a finite number of points can be abstracted from books on the theory of moments.[25] The author thinks, nevertheless, that discussing them helps clarifying the role played by the theorems of the next paragraph concerning the number of parameters which appear in the extremal function (for the stabilizing sets described in Sec. 3).

$a1$: Let $x_1, x_2, \ldots, x_N$ be N points, $-1 < x_1, \ldots, x_N < 1$, where the values of a function $f_0(z)$ is measured with results $h_1, h_2, \ldots, h_N$ and equal errors σ. We wish to find among all functions f satisfying

$$\oint_{|z|=1} |f(z)|^2 |dz| \leqslant M^2 \tag{4.1}$$

that one which realizes $\inf \chi^2(M)$. We assume that M is such that $\chi^2_{\min}(M) \neq 0$. It is well known[24,1,26] that the solution of this problem is to be found among the linear combinations of functions

$$f(z) = \sum_{i=1}^{N} y_i \, H(z,x_i), \tag{4.2}$$

where $H(z,x)$ is the reproducing kernel of the space of analytic functions with norm (4.1):

$$H(z,x) = 1/2\pi(1 - z\bar{x}).$$

The function χ^2 can be written as a quadratic form of the parameters y:

$$\chi^2 = \sum_{i=1}^{N} \frac{(\sum_{j=1}^{N} y_j H(x_j, x_i) - h_i)^2}{\sigma^2} \tag{4.3}$$

The matrix of this quadratic form is $(1/\sigma^2)\mathbb{H}^2$, where $\mathbb{H}_{ij} = H(x_i, x_j)$. The stabilizing condition (4.1) can be written as

$$\sum_i y_i y_j H(x_i, x_j) \leqslant M^2, \tag{4.4}$$

where we have used the fact that H is a reproducing kernel. We now go over to a basis where the matrix $\mathbb{H}$ is reduced to unity, so that the form (4.3) reads

$$\chi^2 = \frac{1}{\sigma^2} \sum (\tilde{y}_i - \tilde{h}_i)^2 \tag{4.5}$$

and the form (4.4) is

$$\sum \tilde{y}_i^2 \frac{1}{\kappa_i} \equiv \sum \tilde{y}_i^2 \lambda_i \leqslant M^2. \tag{4.6}$$

In (4.5), (4.6), $\tilde{y}_i$, $\tilde{h}_i$ are appropriate linear combinations of the y_i, h_i, and the κ_i are the eigenvalues of $\mathbb{H}$. It is clear that the eigenvalues of $\mathbb{H}$ are rapidly decreasing, since two columns of $\mathbb{H}$, $H(x_i, x_j)$, $H(x_i, x_{j+1})$ are almost linearly dependent, if N is sufficiently large. In the limit of a continuous distribution of data, the eigenvalues decrease exponentially.[15]

The extremal values of $\chi^2, \tilde{y}_i$ are obtained by Lagrange multipliers and they read:

$$\tilde{y}_i = \frac{h_i}{1 + \mu \lambda_i \sigma^2}, \tag{4.7}$$

$$\chi^2_{\min} = \mu^2 \sigma^4 \sum \frac{\chi_i^2 h_i^2}{(1 + \mu \sigma^2 \lambda_i)^2}, \tag{4.8}$$

and μ is the solution of

$$\sum_{i=1}^{N} \frac{\lambda_i h_i^2}{(1 + \mu \sigma^2 \lambda_i)^2} = M^2. \tag{4.9}$$

We now make the following remarks:

(1) We expect that, for large i, $h_i/\lambda_i \approx 0$, since λ_i increases (almost exponentially) with i. So, "reasonable" estimates of the extremal function $\bar{f}(z) = \Sigma \tilde{y}_i \varphi_i(z)$ [the $\varphi_i(z)$ are combinations of the $H(z,x_i)$] can be obtained by truncating the sum at that n, for which $\lambda_n \sim 1/\mu\sigma^2$. We see that the extremal function can be better approximated by a given number of terms if μ is large.

(2) At fixed h, the value of μ decreases with increasing M. This follows because the left-hand side of (4.9) is monotonously decreasing as a function of μ, for $\mu > 0$ and one must choose the root with $\mu > 0$ (see Ref. 24, p. 204 and Ref. 15). We conclude the extremal function can be approximated more efficiently by a given number of terms for small values of M than for high ones. Since $\chi^2_{\min}(M)$ is a monotonously decreasing function, we can read off (4.7)–(4.9) the following statement:

Given $\delta > 0$, $\chi_0^2 > 0$, there exists a number $n_0(h, \sigma, \chi_0^2, \delta)$, so that if $\chi^2_{\min} \geqslant \chi_0^2$, the solution of the problem (4.3), (4.4) can be approximated within δ by a linear combination of the first n_0 eigenfunctions.

Explicitly, to find n_0, we proceed as follows: From $\chi^2_{\min} \geqslant \chi_0^2$, we deduce via (4.8) that $\mu \geqslant \mu_0(h,\sigma,\chi_0^2)$. We choose then $n_0(h,\sigma,\chi_0^2,\delta)$ so that $\Sigma_{i=n_0}^{\infty} h_i^2/(1+\mu_0\lambda_i\sigma_2)^2 \leqslant \delta$.

This statement is rather useless for the L^2 problem, since we know its exact solution (4.7). We shall prove, however, an analog of it for the stabilizing bodies of Sec. 3 for which we do not know the exact solution. However, there we shall show this not for approximants, but rather for the extremal function itself.

*a*2. We now show that the stabilizing bodies described in Theorems 3.2–3.4 are such that there exist points on their surface at which they admit of more than one tangent plane. We consider only the case of a finite number N of experimental points, so that the problem can be visualized in R^N.

1. We consider first the set $\mathscr{S}$, given by Eq. (2.8): $\int_I |d\sigma(w)| \leqslant M$. A plane PH in $R^N(\chi^2)$ is described by a set of numbers $(n_1, n_2, \dots, n_N)$ and a constant α:

$$(n,y)_p = \sum_{i=1}^{N} \frac{n_i y_i}{\sigma_i^2} = \alpha; \quad \|n\|_p = 1. \tag{4.10}$$

The plane $PH(n;\alpha)$ is tangent to $\mathscr{S}_N$ if, for all $c \in \mathscr{S}_N$, $c = (c_1, c_2, \cdots)$,

$$|(n,c)_p| = \left| \sum_{i=1}^{n} \frac{n_i c_i}{\sigma_i^2} \right| \leqslant \alpha \tag{4.11}$$

and there exists a point $\bar{c} \in \mathscr{S}_N$ for which equality is obtained. Using the expression (3.1) for $c \in \mathscr{S}_N$, we associate with each vector n the "polynomial"

$$P_N(n,w) = (n, k(\cdot,w))_p = \sum_{i=1}^{N} n_i \frac{1}{\sigma_i^2} k(z_i,w) \tag{4.12}$$

and notice that if the plane $PH(n,\alpha)$ is tangent to $\mathscr{S}_N$, then, for all $d\sigma(w)$ obeying (2.8),

$$\left| \int_I P_N(n,w) d\sigma(w) \right| \leqslant \alpha. \tag{4.13}$$

Then, from the reasoning of Theorem 3.2, we deduce that a family of planes given by $n = (n_1, \dots, n_N)$ attains its extreme value α on $\mathscr{S}_N$ at a point given by a linear combination $\Sigma_{i=1}^{p} \lambda_i k(z_l, w_i)$, $p \leqslant N$, $\Sigma_{i=1}^{p} |\lambda_i| = M$, which contains p terms, if and only if the corresponding polynomial $P_N(n,w)$ satisfies

$$|P_N(n;w)| \leqslant \alpha/M \tag{4.14}$$

and

$$|P_N(n,w_j)| = \alpha/M \tag{4.15}$$

at p points $w_j \in I$ at least.

Now, we see that a point $c \in \mathscr{S}_N$ which is generated by one δ function only, via (3.1), and which lies on the boundary has its tangent plane associated with a polynomial which reaches its maximum modulus once along I. Then there is really no need to prove that if N is sufficiently large, one can add to $P(w)$ a small amount of some other polynomial of degree N, without disturbing either its norm (maximum modulus), or the point where it attains it. In other words, we can generate a continuum of planes tangent to $\mathscr{S}_N$ at the point c. This is the property by which we characterize an edge. We see therefore that we can talk now about the "dimensionality" of an edge, as the number of parameters which occur in a finite linear combination of the type (3.5). We also conclude that the minimum is attained on a p-dimensional edge only if the analytic function (4.12) has a sufficient number of oscillations.

Completely analogous results hold for the bodies described by Theorems 3.2 and 3.3. We just state them:

2. A plane $(n_1, \dots, n_N)$ is tangent to the body $\mathscr{S}_N$ of moments generated by (3.1) with the positivity condition (3.13) at a point given by the sum (3.14) with p terms if and only if the polynomial $P_N(n,w) = \Sigma_{i=1}^{N} (n_i/\sigma_1^2) k(z_i;w)$, $\Sigma n_i^2/\sigma_i^2 = 1$ has at least p double roots at the points w_i, and is positive on I $(w_i \in I)$.

This is a well-known result in the theory of moments.

3. A plane $(n_1, \dots, n_N)$ is tangent to the body $\mathscr{S}_N$ generated via Eq. (3.1) by $d\sigma(w)$ obeying Eq. (2.9) at a point $\bar{c}$ of coordinates $\bar{c}_i$, given by Eq. (3.21), where the sum extends over p terms if and only if the "polynomial" $P_N(n,w)$ Eq. (4.12) is such that it has p simple roots at $x_j \in I$ and no others.

In both these latter cases, we see that, if the minimum is attained on points consisting of $p < N/2$ (N, respectively) terms, we can vary the polynomial $P_N(n,w)$ (and so the unit vector n) by small amounts without violating the conditions it has to obey in order to describe a tangent plane. So, we can again safely talk about edges and their dimensionality.

In Fig. 1 we show as an example the body $\mathscr{S}_2$ generated via (3.1) by $d\sigma(w)$ obeying (2.8) and see that indeed it has two sharp corners at A and B.

Intuitively, from Fig. 1 we expect now that there are "large" sets of points in R^N for which the minimal distance to $\mathscr{S}_N$ is attained on a k-dimensional edge, with $k < N-1$. In Fig. 1, the minimum distance from (h_1,h_2) to $\mathscr{S}_2$ is attained on a one-dimensional boundary; however, for all points lying in the hatched domain (at corner B), the minimum distance to $\mathscr{S}_2$ is attained on the zero-dimensional corner B. (The same holds for the reflected domain near A.)

Ideally, given h,σ,M we would like to know the dimension of that edge (possibly $N-1$) on which the minimum of

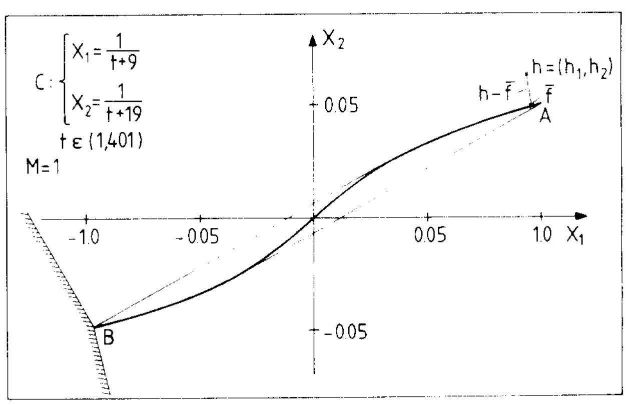


FIG. 1. The body $\mathscr{S}_2$ generated by the values at two points $a_1 = -9$, $a_2 = -19$ of the possible form factors with absolute area M of the spectral function less than 1 [Eq (3.1)]. $\mathscr{S}_2$ is the convex hull of the curve C, generated by $x = 1/(t+9)$, $y = 1/(t+19)$ as t varies over the cut (1.401). For all points (h_1,h_2) lying the hatched area, the exact minimum of the distance to $\mathscr{S}_2$ is attained at B.

χ_ρ^2 is attained. We cannot do this, but shall instead prove in the next paragraph a statement analogous to the one at the end of Sec. 4, a1, concerning the correlation between the dimension of the edge and $\chi_{\min}^2$ at given h,σ. This is sufficient to specify the number of parameters appearing in the stable extrapolate.

We emphasize that the statements of the present paragraph will not play any direct role in the proofs of the next section. They just might help making their content geometrically more intuitive.

B. Bounds on the number of parameters of the stable extrapolate

The result of this chapter is essentially the following: Given h,σ and $\chi_0^2>0$ we construct a number $p_0(h,\sigma,\chi_0^2)$ so that, if the number of terms of the extremal function (3.5), (3.14), (3.21) is bigger than p_0, then $\chi_{\min}^2 \leqslant \chi_0^2$. We know on the other hand, that a stable analytic extrapolation is given by the function f_{M_0} (Theorem 2.3). This function is such that $\chi_{\min}^2=\chi_0^2$, a certain fixed number. As a consequence of the results of this and the preceding paragraphs, we see that this function must be looked for in a well-defined parameter space, with a bounded dimensionality. This makes its search in principle feasible on a computer. It is *a priori* surprising that the solution of such an extremal problem can be expressed in closed form; this corresponds to the picture of the preceding paragraph concerning the "edges" of the stabilizing bodies.

The bound $p_0(h,\sigma,\chi_0^2)$ we obtain may or may not be interesting, according to whether it is less than or bigger than N, the number of experimental points. We point out, however, that this number is always finite, even for a continuous distribution of data, and it is easy to give examples where, for a large enough number N of points, it is sensibly smaller than N.

The procedure we adopt is the following: We write

$$\chi_{\min}^2=(n,h-\bar{f})_\rho=(n,h)_\rho-(n,\bar{f})_\rho \leqslant (n,h)_\rho\,. \quad (4.16)$$

The last inequality follows from Theorem 3.1, since $0\in\mathscr{S}$ and then it is true that $(n,\bar{f})_\rho \geqslant 0$.

We prove then that $(n,h)_\rho$ is bounded by a number which goes to zero as the number of parameters of the extremal function f increases. To show this, we use the fact that this number of parameters is related to the number of zeros or of oscillations within a certain interval of the analytic function of w, $(n,k(\cdot,w))$, and that the magnitude of this function is controlled by the maximum modulus principle.

We shall prove this for the situation which is of interest for the extrapolation with rational functions, and choose $k(z,w)$ to be the Cauchy kernel. We can get almost identical results for the other situations.

Theorem 4.1: Let $k_c(z,w)$ be the Cauchy kernel. Let p be the number of terms in the expression (3.5) for the extremal function of (2.8). Then there exist constants C,b such that $(0<b<1)$

$$|(n,k_c(\cdot,w))_\rho| < Cb^p \quad (4.17)$$

for w inside the domain of integration I [some part of the real axis; see Eq. (2.7)].

Proof: The function $(n,k_c(\cdot,w))_\rho$ is analytic in w outside the data region $[a,b]$ and goes to zero as $w\to\infty$. If the extremal function contains p terms, there exist p real points $x_1,x_2,\ldots,x_p$, where the function $(n,k_c(\cdot,w))_\rho$ assumes the same absolute value, say $|\alpha|$ (cf. Theorem 3.2). There are then at least $p/2$ values, where $(n,k(\cdot,w))_\rho$ assumes the value α.

Consider then a bounded curve $\mathscr{C}$ which surrounds the data region $[a,b]$ and avoids the integration region I. Then, for $w\in\mathscr{C}$,

$$|(n,k_c(\cdot,w))_\rho| \leqslant \|n\|_\rho \|k_c(z,w)\|_\rho \leqslant \sup_{w\in\mathscr{C}}\|k_c(\cdot,w)\|_\rho \equiv C_1\,. \quad (4.18)$$

Let us now map $\mathscr{C}$ onto the unit disk, so that I comes inside a region $[-\beta,\beta]$ along the real axis, $|\beta|<1$, and the point at infinity comes, e.g., at β. The (at least) $p/2$ zeros of $(n,k(\cdot,w))_\rho-\alpha$ come at $x_1',x_2',\ldots,x_N'$ inside this interval. We use then the maximum modulus principle to get

$$|(n,k_c(\cdot,w))_\rho-\alpha| \leqslant C_1\prod_{i=1}^{p/2}\left|\frac{z-x_i}{1-x_iz}\right| \leqslant C_1\left(\frac{2\beta}{1+\beta^2}\right)^{p/2}. \quad (4.19)$$

But, at β, $(n,k_c(\cdot,w))_\rho=0$ and we get

$$|\alpha| \leqslant C_2\left(\frac{2\beta}{1+\beta^2}\right)^{p/2} \quad (4.20)$$

We obtain inequality (4.17) by letting $C=2C_2$, $b=(2\beta/1+\beta^2)^{1/2}$.

We now estimate the quantity $(n,h)_\rho$. To this end we use:

Theorem 4.2: For any $\epsilon>0$, there exists a $\delta>0$, so that $|(n,h)_\rho|\leqslant\epsilon$ as soon as $|(n,k(\cdot,w))_\rho|\leqslant\delta$

Proof: We notice to this end that the operator A: $n(x)\mapsto(n,k(\cdot,w))_\rho$ is compact and one-to-one on its domain of values. The last property follows from the uniqueness of analytic continuation. As a compact operator, it transforms any weakly converging subsequence in $L^2(\rho)$ into a strongly converging one on I (e.g., in the uniform norm in I). A sequence $\{n_k\}$ such that $|(n_k,k(\cdot,w))_\rho|\to 0$ uniformly on I must then converge weakly to zero in $L^2(\rho)$. In particular $|(n_k\ h)_\rho|\to 0$ since $h\in L^2(\rho)$.

Notice that the proof is independent of whether $\rho(x)$ has discrete or continuous support.

We see thus that, if we know that $\chi_{\min}^2\geqslant\chi_0^2$, it is sufficient to choose the number of terms p_0 in the sum (3.7) as

$$p_0=\frac{\ln(\delta(\chi_0^2/2)/C)}{\ln b}\,. \quad (4.21)$$

where $\delta(\chi_0^2/2)$ is the value of δ corresponding to $\epsilon=\chi_0^2/2$ by Theorem 4.2.

With this we have shown that a stable analytic extrapolation can be obtained by minimizing χ^2 in a p_0-dimensional space. Recall that $\ln b$ is negative and therefore p_0 decreases as χ_0^2 increases, as expected.

We now wish to describe more closely how we can get the function $\delta(\epsilon)$ of Theorem 4.2. Clearly, we get a majorization similar to Theorem 4.1 for any L^2 norm on the integration interval I (e.g., by integrating the uniform bound, Eq. (4.17), in the conformally mapped variable, Theorem 4.1). We write this as follows:

$$\|An\|_I^2 = \|(n,k(\cdot,w)\|_I^2 \leqslant C_0 b^{2p}. \tag{4.22}$$

The operator A^+A maps then $L^2(\rho)$ into itself, is compact and self-adjoint. It has then a set of eigenfunctions $f_i(x)$ which is complete in $L^2(\rho)$ and a corresponding set of eigenvalues λ_i which accumulates to zero. Consider then an approximation to $h(x)$ of order $\epsilon/2$ by linear combinations of the first N_0 eigenfunctions

$$\left\|h(x) - \sum_{i=1}^{N_0} h_i f_i(x)\right\|_\rho^2$$
$$\equiv \|h(x) - h_{N_0}(x)\|_\rho^2 \equiv \|h'_{N_0}(x)\|_\rho^2 \leqslant \epsilon^2/4. \tag{4.23}$$

Let then n_{N_0} be the component of $n(x)$ on the N_0-dimensional subspace spanned by these eigenfunctions. If we choose then

$$\delta = \lambda_{N_0}\epsilon^2/4\|h_{N_0}\|^2, \tag{4.24}$$

we get

$$\|n_{N_0}\|_\rho^2 \leqslant \frac{1}{\lambda_{N_0}}\|An_{N_0}\|_I^2 \leqslant \frac{1}{\lambda_{N_0}}\|An\|_I^2 \leqslant \frac{\epsilon^2}{4\|h_{N_0}\|_\rho^2}. \tag{4.25}$$

Consequently,

$$\begin{aligned}|(n,h)|_\rho &\leqslant |(n_{N_0},h_{N_0})|_\rho + |(n'_{N_0},h'_{N_0})|_\rho \\ &\leqslant \|n_{N_0}\|_\rho\|h_{N_0}\|_\rho + \|n'_{N_0}\|_\rho\|h'_{N_0}\|_\rho \\ &\leqslant \frac{\epsilon}{2\|h_{N_0}\|_\rho}\|h_{N_0}\|_\rho + \frac{\epsilon}{2} = \epsilon.\end{aligned} \tag{4.26}$$

We have in Eq. (4.24) an explicit expression for $\delta(\epsilon)$ of Eq. (4.21). This estimate can be refined by using Lagrange multiplier techniques.

If λ_{N_0} decreases exponentially with N_0, $\lambda_{N_0} \sim a^{N_0}$, $|a| < 1$, we see that $p_0 \propto N_0$, i.e., the number of coefficients in the extrapolate is proportional to the number of eigenfunctions required to fit $h(x)$ or to the "structure that is seen in the data."

We also expect in this picture that small changes of the errors [or of the measure $\rho(x)$] do not change the number of parameters needed to attain the minimum. This is indeed the case, as one persuades oneself by means of some epsilontics.

We now describe the practical procedure that follows from these considerations, for the determination of the stable extrapolate. We assume that the data are not consistent with zero, so that, for $M = 0$, $\chi^2_{\min}(M) > \chi_0^2$. From formula (4.21), we can compute $p_0(\chi_0^2 - \epsilon, h, \sigma)$, for some $\epsilon > 0$; then we know that, for small enough M, $\chi^2_{\min}(M)$ is attained on a p_0-dimensional subspace. We assume then we can vary M in small enough steps, so that we can correctly approximate that value M_0 for which $\chi^2_{\min}(M_0) = \chi_0^2$. The corresponding function is the stable extrapolate.

Unfortunately, the estimates the author can get for p_0, although considerably smaller than the number of experimental points N, still lie sensibly above the number of parameters that the computer seems to require for the exact minimum. The author believes, however, that the existence of the bound p_0 makes it plausible enough that the rejection of the extra parameters at the minimum is exact and that further work will improve the estimate of p_0. In the next section we shall describe various applications of the statements of Secs. 3 and 4, and we shall constantly assume that we can fix p_0 by simply choosing a sufficiently small value of M and minimizing χ^2 for an increasing number of parameters. The rejection of the unnecessary ones is quite sharp, and we can then fix p_0. We follow then this number by continuity as we increase the value of M.

5. APPLICATIONS

A. Nucleon electromagnetic form factors and rational functions

We have checked this "edge effect," i.a while performing an analysis of the nucleon electromagnetic form factors.[27] We assumed, in accordance to physical models, a simple pole ansatz, like (3.5) for the spectral functions of the Dirac and Pauli form factors $F_{1s}, F_{1v}, F_{2s}, F_{2v}$, where s and v stand for isoscalar and isovector, respectively, and required that the total area under them be bounded by a constant A:

$$\sum_{i=1}^{2}\frac{1}{\kappa_i}\int_{4m_\pi^2}^{\infty}|F_{1s} + F_{1v}|dt \leqslant A. \tag{5.1}$$

(the k's are normalization constants, chosen for scale reasons). We know already (Sec. 3) that the ansatz, with suitably chosen parameters, minimizes χ^2 among all other functions with area bounded by A. The number of poles required to reach $\chi^2_{\min}$ is found by inspection, using the minimization program MINUIT, with its variable metric part. As announced, for small values of A, we do not see the slightest decrease of $\chi^2_{\min}$ if one adds more poles, after a certain critical number is reached. Figure 2 shows a curve $\chi^2_{\min}(A)$ as obtained with 10 and 15 parameters, respectively, for a certain range of A. The positions of the parameters at the minimum point obtained with 10 of them free are taken as starting points for the minimization with 15 parameters. As the allowed area A increases, the difference between the minima of the two situations increases. One expects this since, if A be-

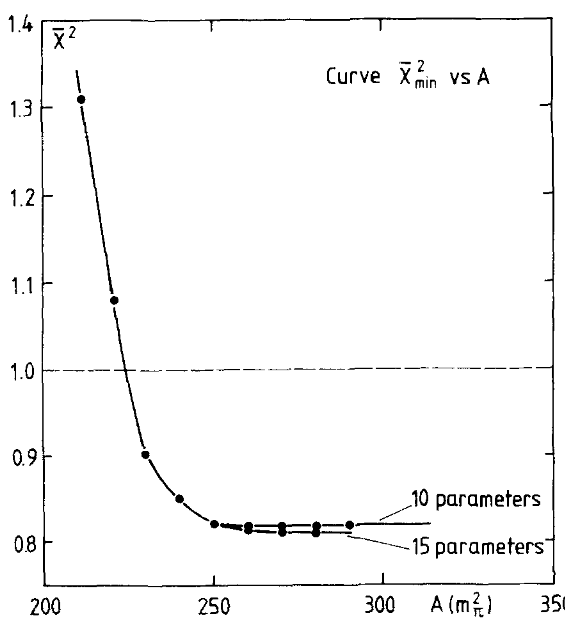


FIG. 2. Curve $\chi^2_{\min}$ vs A discussed in Sec. 5A.

comes so large that $\chi^2_{\min}(A) = 0$, one needs, in general, $N-1$ nontrivial parameters. The numerical results are displayed in Appendix B.

The results of Secs. 3 and 4 have a rather surprising consequence: Consider a form factor whose spectral function has an area bounded by A and a set of spacelike data generated by it. Let now these data be affected by noise and try to fit them with functions of area bounded by A. The best fit is not in general the original form factor, but rather a function consisting of simple poles and, with the above argument, really only few of them.

B. Positivity and rational functions

We have seen in Sec. 2 that positivity (of the real or imaginary part) can work as a stabilizing condition and that the stable analytic extrapolation is obtained by means of a sum of pole terms with positive residua. The theorems of Sec. 4 go through word for word for the case when the imaginary part is positive.

This is the situation which occurred in Ref. 28 in a search for the best $KN\Lambda$ coupling constant compatible with analyticity and the positivity of the forward KN amplitude. The authors noticed that the best fit was obtained when the imaginary part was replaced by just one pole. This is the effect described by Theorem 4.2.

Nothing is changed if we replace the positivity condition by

$$\mathrm{Im}[f(t)] \geqslant -\mathscr{L} k_0(t) \tag{5.2}$$

for t on the cut, where $k_0(t)$ is a known positive function.

Using the fact that $\chi^2_{\min}$ is obtained with a finite number of terms, it is possible to obtain on the computer "rigorous results" concerning the functions under study. For instance it is possible to show that the spectral function of the proton Pauli form factor must get positive somewhere on the cut (presumably beyond the $N\bar{N}$ threshold), if consistency is to be achieved with the data and the ρ peak of Ref. 29 within $\chi^2 \leqslant 1$. (We assume again that we can follow the exact minimum on the computer, according to the discussion at the end of Sec. 4B.)

It has been shown in a recent paper[30] that a unique determination of phase shifts, which relies only on the assumptions of axiomatic field theory, can be obtained by computing first the amplitude in the strip $0 \leqslant t \leqslant 4m_\pi^2$, where positivity of the imaginary part holds. As a consequence of the latter, the number of zeros in the complex s plane at fixed real t is bounded from above. In Ref. 30 it is shown that their position in the complex s plane is fixed by unitarity at threshold. This might not be a practically effective way, but for the present purpose let us imagine that one can find other methods to fix them. Then, one is left with the problem of extrapolating the modulus $m(t)$ of the amplitude to $0 \leqslant t \leqslant 4m_\pi^2$. It appears natural to stabilize this extrapolation by means of an L^∞ function bound on the cuts at fixed energy

$$|m(t)| \leqslant MB(t), \tag{5.3}$$

where $B(t)$ is taken from a Regge model. Using an outer function, we can reduce this condition in a standard manner[3] to

$$|\tilde{m}(t)| \leqslant 1. \tag{5.4}$$

It is remarkable that this stabilizing condition leads to the same geometry as the conditions on the spectral function (2.8), (2.9), or (3.13). Theorems analogous to 4.1, 4.2 can also be proven in this case. However, the extremal function is no longer a linear combination of terms, but rather a finite Blaschke product. The proofs are somewhat more involved than those in this paper and will be presented elsewhere.[31]

In this paper we just point out that we can also transform (with some loss of information) the boundedness condition to one of positivity of the real part of an analytic function, by using the Cayley transform

$$C_m = (1-\tilde{m})/(1+\tilde{m}). \tag{5.5}$$

Theorems 4.1 and 4.2 apply then also to this case; there are some minor differences in the proof of Theorem 4.1: The analyticity domain of the function $(n, k_p(\cdot, w))$, where k_p is the Poisson kernel, is now the w plane cut along the data region and along its reflection across the unit circle. The curve $\mathscr{C}$ consists then of two pieces (surrounding the two cuts); the Blaschke factors must be replaced in (4.19) by functions which have modulus 1 on the curve $\mathscr{C}$ and a zero on the unit circle. The factor b of (4.17) is the maximum of such a function on the unit circle. There is no other change.

C. Extrapolation to the cut. The pion form factor in the timelike region

We have shown (Sec. 2) that a bound on the derivative of the spectral function stabilizes the extrapolation to the cut. In the dispersion relation for the pion form factor

$$F_\pi(t) = \frac{1}{\pi}\int_{4m_\pi^2}^{\infty} \frac{\mathrm{Im}[f(t')dt']}{t'-t} \tag{5.6}$$

we can integrate under this assumption once by parts, to get

$$F_\pi(t) = \frac{1}{\pi}\int_{4m_\pi^2}^{\infty} \ln\left(\frac{t'-t}{t'-4m_\pi^2}\right)\mathrm{Im}[f'(t')]dt', \tag{5.7}$$

where we impose

$$|\mathrm{Im}[f'(t')]| \leqslant DB(t') \tag{5.8}$$

[$B(t)$ is a given function on the cut]. We can then apply Theorem 3.4 to find the expression for the extremal function, and then the reasonings of Theorems 4.1, 4.2 to limit the numbers of terms. This was done in Ref. 32, where we tried to reconstruct the spectral function of the cut in the region of ρ' (1600), by using $B(t) = 1/(t+2)^2$. If the extremal steplike functions [Eq. (3.21)] for the derivative are integrated once, one obtains a spectral function which consists of some modulated "triangular resonances." Fits were performed with three and five parameters for the spectral function and no improvement was observed when the number of parameters increased (see Fig. 2 of Ref. 32). The unpleasantly looking cusps[32] disappear if condition (5.8) refers to a line on the second sheet, rather than to the physical cut.

D. M_0 Curves

The simplicity of the extremal function for the situations above can be used in drawing efficiently M_0-type graphs, to determine the best value of some functional over a set of analytic functions, or for the location of poles or zeros of the latter. We recall briefly the basic idea of the method[3,11,15,33]: We wish that a certain functional F be able to assume the values F_0 in the set $\mathscr{S}$ of functions which obey some experimental constraints and an inequality like (2.8), (2.9), (3.13), with unknown value of the stabilizing lever, say generically M. This is impossible unless M is large enough, namely larger than a certain critical value, M_0 (F_0). For all values of M larger than $M_0(F_0)$, there exist functions realizing the value F_0, and complying with all the other constraints. We now draw a curve $M_0(F_0)$; if it turns out that there are values of F_0 for which $M_0(F_0)$ must get exceedingly large (that is, only "wild" functions realize the value F_0), we can exclude these from the possible values of F. It is plausible[33] to regard the minimum of this curve as really "the most probable" value of F. If the errors are sufficiently small or the functional F is luckily chosen (function values at points near the data region, etc.) these curves are quite spectacular.

Clearly, the precise shape of the curve and its minimum depends on the stabilizing restriction is use. It is remarkable that this is so only to a rather small extent and presumably the reason lies in the very ill-posed nature of the analytic extrapolation problem.

Let us notice the following: If we know a value M of the stabilizing lever, then the allowed values of F_0 are those for which $M_0(F_0) \leqslant M$. The possibility of computing exactly the value M_0 (cf. Sec. 4) solves then the complete extrapolation problem, in the sense of Ref. 3 (cf. Sec. 2C).

To get the critical value $M_0(F_0)$ of the stabilizing lever M, for the situations described in this paper, one draws curves $\chi^2_{\min}(M)$, for various values of M and an increasing number of parameters in formulas (3.5), (3.14), or (3.21), respectively. As in Fig. 2 one notices at fixed M a quick cutoff in the number of parameters needed, or a "freezing" of the additional ones as we try to put more in to bring $\chi^2_{\min}$ further down. The present data on the nucleon form factors allows best fits, in the sense of Theorems 4.1, 4.2 with not more than three free poles, for each form factor.

As an example, we present the determination of the ω residue, by using the area A of Eq. (5.1) as a stabilizing condition. We draw a graph for the minimal value of A versus possible values of the ω residue (Fig. 3). The curve is rather flat, which shows that the determination is affected by rather large errors. There are different interpretations of the spectral function beyond ω, according to the region of the graph. The minimum implies a coupling of the ϕ to nucleon–antinucleon, which satisfies SU(3) (together with r_ω), but violates Zweig's rule.[27] The interpretation of SU(3) constraints appears, however, to be ambiguous. A smaller coupling of ω requires a spectral function showing essentially an $\omega' \approx 1200$ MeV, and almost no ϕ coupling, which is in accordance with Zweig's rule.

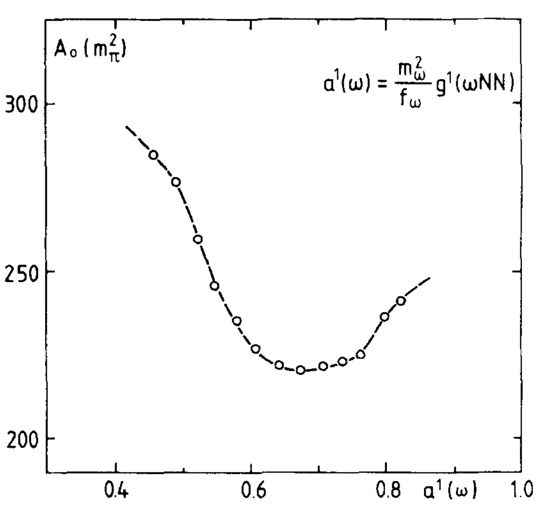


FIG. 3. Determination of the ω residue, with condition (5.1) as a stabilizing requirement.

Curves similar to the $M_0(F_0)$ ones can be obtained for the best value f_0 of the spectral function of the pion form factor at a point on the cut; one must draw there graphs $D_0(f_0)$, where D is the scale factor, Eq. (5.8) for a bound on the derivative. Reference 32 shows that the errors of the extrapolation in the neighborhood of the ρ' (1600) peak are really big.

6. CONCLUSIONS

a. We have shown that certain classes of functions, which are suitable as physical models, in particular rational approximants, can be brought in the frame of stable analytic extrapolation methods. This can be done in such a way that they are extremal functions for the transformed problem of finding the best extrapolation [Eq. (1.1)]. In other words, consideration of other functions, belonging to the same stabilizing class is of no effect in finding $\chi^2_{\min}$.

It is in the opinion of the author remarkable that the number of parameters needed in such fits can be sometimes limited nontrivially from above, once the data function and the errors are given. This gives a sound meaning to the sentence "this data is described by this many parameters," with the provision that one must specify the stabilizing principle which is being used.

We dwell again shortly on the meaning of the stable extrapolations: We see that there are two ways of truncating a series of terms, like, say (3.5) when performing a fit: (a) to use the smallest possible number of terms, for which essentially $\chi^2 \approx 1$ (this can be done with varying amounts of sophistication); (b) to find the minimal area A, Eqs. (5.1) (or a minimal value of the stabilizing lever, in general) for which $\chi^2 \approx 1$ and use this "critical" function for the fit. Is there any difference between these two procedures? As long as both produce a function with the desired analytic properties, nobody can tell which is closer to the true one. Simplicity would rule in favor of procedure (a), since in general, it contains less parameters than (b).

What distinguishes (a) from (b) is really the virtual process of letting the errors $\sigma(x)$ in (2.1) go to zero: namely, if we perform by both methods a sequence of fits $\{f_1^a, f_2^a, \ldots, f_n^a\}$, $\{f_1^b, f_2^b, \ldots, f_n^b \cdots\}$, respectively, at errors ϵ_n going to zero, then only the second sequence is guaranteed to tend to the true function. Although all the members of the first one have the correct analytic properties, the sequence does not, in general, tend to a function with the desired structure. (This is just the statement of the instability of analytic continuation.)

So methods (b), "stable methods," are in this respect, systematic as opposed to (a).

One expects the difference between (a) and (b) to show up if the errors are very small, so that the limiting process is almost carried out. In practice, this is seldom observed, the reason being presumably that analytic extrapolation is such an ill-posed problem. Small differences can, however, be seen in the analysis of nucleon electromagnetic form factors.

b.The results of this paper can presumably be generalized to other sets of functions of physical interest. In other words, there exist many times a stabilizing assumption within which the functions used for the fit are even the best ones, (the stable extrapolation) in the sense of problem (1.1). One must decide then whether this assumption is tenable or has some interest.

For instance, one can accomodate presumably second sheet (Breit–Wigner) poles by using first for clarity a mapping onto the unit disk D and then the representation of the functions analytic in D,

$$f(z) = \iint_{CD} \frac{d\mu(z')}{z' - z}, \tag{6.1}$$

where $\mu(z')$ is an arbitrary complex function of bounded variation of z' extending over CD. The extremal points of the stabilizing condition

$$\iint_{CD} \rho(z')|d\mu(z')| \leqslant S \quad \rho \geqslant 0, \tag{6.2}$$

are just second sheet poles, i.e., poles placed outside D. The well-known argument of Callucci–Fonda–Ghirardi[34] requires then huge values of S in (6.2). On the other hand, Breit–Wigner poles are functions preferred by the fit under condition (6.2). The use of rational functions[35] in amplitude analysis has many advantages, simplicity being one of the most important.

ACKNOWLEDGMENTS

I wish to express my gratitude to Professor E. Thomas, who had the kindness to discuss in detail with me the mathematical part of this paper, related to positive rational functions. The whole mathematical presentation was much influenced by the discussions I had with him. I am grateful to Professor D. Atkinson for his interest in the results of this paper and for his encouragement. I have drawn much benefit from previous work with Professor S. Ciulli, to whom I owe my concern with questions of analytic extrapolation. In the work on form factors, from which examples were taken, I benefited much from discussions with Professor G. Höhler and Professor W. Schmidt. I owe many thanks to Dr. C.B. Lang for his continued interest in these questions and many useful remarks. I wish to thank the referee for pointing out to me that a previous version of the paper lead to quite a number of confusions and for suggesting some rewriting.

APPENDIX A

We wish to verify the consistency of the "exact" spacelike "experimental" points $h(t_i)$ on the pion form factor with analyticity and a physically plausible bound on the modulus in the timelike region. We use the bound of Kiehlmann and Schmidt, Ref. 9, which goes along the upper end of the errors bars of the ρ peak and continues by a constant to infinity, and a Breit–Wigner formula down to threshold. Call it $m(\theta)$, where a mapping of the cut t plane onto the unit disk has already been carried out. Dividing off the outer function

$$C(z) = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} \ln m(\theta)\, d\theta\right], \tag{A1}$$

we reduce the problem to the one of verifying the consistency of some values $w(z_i)$ of modulus less than one

$$w(z_i) = h(t_i)/C(t_i), \quad z_i = z(t_i) \tag{A2}$$

with an analytic function of modulus bounded by one in the unit disk.

This is done by means of the Riesz representation theorem for positive polynomials and by means of the Cayley transform relation between the functions of modulus less than one and those of positive real part in the unit disk. Any book on the theory of moments would give details about this.[25] The conclusion is that the values are consistent with analyticity and a modulus less than one on the unit circle if and only if the quadratic form in ξ_i

$$\sum_{i,l} \frac{1 - w(z_i)w(z_l)}{1 - z_i z_l} \xi_i \xi_l \tag{A3}$$

is positive definite.

We recast this statement in a new form, as follows: It is first equivalent to the positivity of all the eigenvalues of the matrix

$$T = A - \Lambda A \Lambda \tag{A4}$$

where

$$A_{ij} = 1/(1 - z_i z_j) \tag{A5}$$

$$\Lambda_{ij} = h(t_i)/C(z_i)\delta_{ij}\,. \tag{A6}$$

Now, the matrix A is positive definite, since the function equal to zero throughout is consistent with a modulus less than one. So is $\Lambda A \Lambda$ since it differs from A by a similarity transformation. We conclude that (a) we can bring them both simultaneously to canonical form, by a real transformation and (b) the roots of the equation in λ

$$\det(\Lambda A \Lambda - \lambda A) = 0 \tag{A7}$$

TABLE I. Numerical demonstration of the flattened shape of the body $\mathscr{S}_N$ generated in $\mathbf{R}^N$ by values $f(t_i)$ at N points $-t_i$ of a holomorphic function with modulus bounded by $m(t)$ (see Appendix A) in the t plane, cut along $(4m_\pi^2, \infty)$. The point with coordinate h_i (the "measured" values) lies outside of $\mathscr{S}_N$ and the last column shows the dilation factor M needed to absorb h inside $M\mathscr{S}_N$.

N	$-t$	h	C	M
1	−0.067	1.100	1.359	0.8094
2	3.987	0.102	0.661	0.8703
3	1.069	0.320	0.983	0.8952
4	1.982	0.200	0.825	0.9004
5	0.795	0.378	1.057	2.0222
6	1.204	0.282	0.953	102.83
7	0.294	0.612	1.255	703.51
8	0.176	0.784	1.318	1834.3
9	0.620	0.449	1.115	5576.4

TABLE II. Numerical appearance of the "edge effect." The parameters are those of formula (B1). Two runs are being compared: with 10 and 15 parameters, respectively. The underlined numbers are kept fixed in the run with fewer parameters. As the area increases, the edge effect should get less visible with the same number of parameters. Notice that those pole positions are most mobile, which have the smallest couplings. The author suggests that the minimum is in both cases attained on the same function (at lest for $A = 230$) and that the small variations are due to computer inaccuracies. The suggestion is based on Theorem 4.2.

A	n		$t_{\omega'}$	$t_{\omega''}$	$t_{\omega'''}$	$t_{\rho'}$	$t_{\rho''}$	$t_{\rho'''}$	$t_{\rho^{\mathrm{IV}}}$
230	10		54.919	$\underline{100.0}$	139.79	76.705	$\underline{90.734}$	396.11	1527.75
	15		54.911	76.250	139.53	76.736	92.799	395.40	1533.62
240	10		52.542	$\underline{100.0}$	172.60	78.564	$\underline{90.734}$	396.02	2330.99
	15		52.518	84.38	168.71	78.421	84.380	411.82	2476.00
		$a^{1,2}_{\omega}$	$a^{1,2}_{\omega'}$	$a^{1,2}_{\omega''}$	$a^{1,2}_{\omega'''}$	$a^{1,2}_{\rho'}$	$a^{1,2}_{\rho''}$	$a^{1,2}_{\rho'''}$	$a^{1,2}_{\rho^{\mathrm{IV}}}$
230	10	35.12	30.974	$\underline{0.1}$	− 7.409	0.997	$\underline{0.09}$	− 15.45	4.568
	15	35.12	30.971	0.02	− 7.395	0.982	0.009	− 15.82	4.585
240	10	36.27	31.104	$\underline{0.1}$	− 9.320	0.824	$\underline{0.09}$	14.26	4.405
	15	36.27	31.090	0.92	− 8.941	0.031	0.928	14.82	4.777
230	10	4.594	4.832	—	− 0.238	103.55	$\underline{0.09}$	22.261	− 1.222
	15	4.657	4.942	—	− 0.279	103.52	0.18	22.381	− 1.226
240	10	4.940	5.149	—	− 0.224	107.39	$\underline{0.09}$	28.632	− 4.662
	15	4.720	4.726	—	− 0.135	107.04	0.09	29.775	− 5.199

must all be positive and less than one. The fact that they must be less than one is nontrivial. We conclude that the condition

$$\|A^{-1}\Lambda A \Lambda\| \leqslant 1 \tag{A8}$$

is necessary and sufficient for consistency. If the boundedness by 1 is replaced by M, we just write M^2 on the right-hand side of (A8). We compute this norm for the data described at the beginning (see Table I).

The last column shows the square root of the norm of (A8) after the point in the corresponding line was introduced in the quadratic form (A3). The value of M is a measure of the "squashness" of the stabilizing body $\mathscr{S}_N$, generated in R^N by the functions of modulus less than one along the unit circle. M shows the amount with which S_N has to be dilated in order to absorb the experimental point of coordinates h_i.

APPENDIX B

We fit the four nucleon form factors at once. The position of the ω pole is fixed in F_{1s} and F_{2s}. The same position of the poles is assumed in both Dirac and Pauli form factors, which agrees with physics but not exactly with the theory of Sec. 2. This is, however, harmless, since unnecessary poles have vanishing couplings. Superconvergence of F_{2p}, F_{2n}, G_{Mp} is enforced and limits, together with normalization, the number of parameters. Table II contains the parameters of the formula

$$F_{is(v)} = \sum_k \frac{a'_{\omega^{(k)}(\rho^{(k)})}}{t - t_{\omega^{(k)}(\rho^{(k)})}}. \tag{B1}$$

The units are m_π^2 for A, a's and t's.

[1]R.E. Cutkosky, Ann. Phys. (N.Y.) **54**, 110 (1969).
[2]E. Pietarinen, Nuovo Cimento A **12**, 522 (1972); Nucl. Phys. B **55**, 541 (1973).
[3]S. Ciulli and G. Nenciu, J. Math. Phys. **14**, 1675 (1973).
[4]P. Presnajder and J. Pisut, Nuovo Cimento A **3**, 603 (1971).
[5]G. Nenciu, I. Raszillier, W. Schmidt, and H. Schneider, Nucl. Phys. B **55**, 106 (1973).
[6]D.P. Hodgkinson *et al.*, Eds., *Proceedings of the Topical Conferences of Baryon Resonance* (Ross & Saxon, Oxford, 1976); R.E. Cutkosky *et al.*, Nucl. Phys. B **102**, 139 (1976).
[7]E. Pietarinen, Nucl. Phys. B **107**, 21 (1976).
[8]J. Lichard, preprint, CERN TH 1977.
[9]H.D. Kiehlmann and W. Schmidt, Nucl. Phys. B **94**, 67 (1975).
[10]J.L. Basdevant, Fortsch. Physik **20**, 283 (1972).
[11]S. Ciulli, C. Pomponiu, and I.S. Stefanescu, Phys. Rep. C **17**, 135 (1975).
[12]A.N. Tykhonov, *Méthodes de resolution des problemes mal posés* (Mir Publ., Moscow, 1975).
[13]W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1974), p. 83.
[14]S. Ciulli and G. Nenciu, Nuovo Cimento **8**, 735 (1972).
[15]I.S. Stefanescu, Nucl. Phys. B **56**, 287 (1973).
[16]P.L. Duren, *Theory of H^p Spaces* (Academic, New York, 1970).
[17]W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973).
[18]S. Ciulli and G. Nenciu, Commun. Math. Phys. **26**, 237 (1972).
[19]R.E. Cutkosky, J. Math. Phys. **14**, 1231 (1973).
[20]S.C. Cheung, Preprint C00-3066-1, Carnegie Mellon (1971).
[21]Y.A. Chao *et al.*, Phys. Lett. B **57**, 150 (1975).
[22]I. Raszillier and W. Schmidt, Nucl. Phys. B **55**, 106 (1973).
[23]G. Nenciu and I. Raszillier, Nuovo Cimento A **11**, 319 (1972).
[24]H. Meschkowski, *Hilbertsche Räume mit Kernfunktion* (Springer-Verlag, Berlin, 1962), p. 115.
[25]T. Shohat and T. Tamarkin, *The Problem of Moments* (Mathematical Surveys, Providence, R.I., 1943); M.G. Krein and A.A. Nudelman, *The Theory of Markov Moments and Extremal Problems* (in Russian) (Nauka, Moscow, 1973).
[26]S. Okubo, J. Math. Phys. **15**, 963 (1974).
[27]G. Höhler, E. Pietarinen, I.S. Stefanescu, F. Borkowski, G.G. Simon, V.H. Walther, and R.D. Wendling, Nucl. Phys. B **114**, 505 (1976).
[28]N. Sznajder-Hald, J.L. Petersen, and G. Nenciu, Nucl. Phys. B **59**, 93 (1973).
[29]G. Höhler and E. Pietarinen, Nucl. Phys. B **95**, 210 (1975).
[30]H. Burkhardt and A. Martin, Nuovo Cimento A **29**, 141 (1975).
[31]I.S. Stefanescu, in preparation.
[32]C.B. Lang and I.S. Stefanescu, Phys. Lett. B **58**, 450 (1975).
[33]I. Caprini, S. Ciulli, C. Pomponiu, and I.S. Stefanescu, Phys. Rev. D **5**, 1658 (1972); I. Caprini, Lett. Nuovo Cimento **5**, 443 (1972); Nucl. Phys. B **68**, 561 (1974).
[34]G. Callucci, L. Fonda, and G.C. Ghirardi, Phys. Rev. **166**, 1729 (1968).
[35]W. Langbein, Nucl. Phys. B **94**, 519 (1975).

Spherical integral representations for finite velocity conduction of an abruptly activated current through a crystal

Lim Chee-Seng

Department of Mathematics, University of Singapore, Singapore
(Received 7 July 1978)

A source current is abruptly switched on within a crystalline medium with unaligned sets of permittivity and permeability principal axes. The current then propagates with finite velocity along an arbitrarily oriented conductor. The problem incorporates a partial zero condition and is tackled by Fourier analysis. The general solution for a scalar field is derived as a combination of four spherical integrals over the unit sphere. Two of these are time dependent and could represent the switch-on effect. Their time independent versions relative to the traveling current edge constitute both remaining spherical integrals. These must therefore participate in any eventual steady state and possibly contribute to Čerenkov radiation; their combination is expressible as a single spherical integral directly associated with a basic differential operator. A steady state is found to develop inside an expanding ellipsoid which retreats from the current edge.

1. INTRODUCTION

Suppose an electric current of density $\mathbf{J}$ is radiating within a crystalline medium having constant permittivity and permeability matrices $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$, respectively. These matrices are real, symmetric, and positive definite. Corresponding sets of principal axes are generally unaligned. The electric and magnetic field vectors $\mathbf{E}$ and $\mathbf{H}$ satisfy

$$\nabla\times\mathbf{E}+\mathbf{H}_t\boldsymbol{\mu}=0,\quad \nabla\times\mathbf{H}-\mathbf{E}_t\boldsymbol{\epsilon}=\mathbf{J}, \tag{1.1}$$

relative to the $\mathbf{x}=(x_1,x_2,x_3)$ frame, say. Eliminating $\mathbf{H}$, we have

$$\mathbf{E}_{tt}\boldsymbol{\epsilon}+\nabla\times[(\nabla\times\mathbf{E})\boldsymbol{\mu}^{-1}]=-\mathbf{J}_t\,. \tag{1.2}$$

In this paper, we select a singular source current that is suddenly switched on at instant $t=0$. It then shoots out from the point $\mathbf{x}=\mathbf{0}$ with a constant finite speed v along an infinitesimally thin conductor which is arbitrarily oriented, say, along the positive ζ_1 direction of some fixed orthogonal frame $\boldsymbol{\zeta}=(\zeta_1,\zeta_2,\zeta_3)$. This is coorigined with the $\mathbf{x}$ frame and is related to it by a rotational transformation $\boldsymbol{\zeta}=\mathbf{xR}$ with modulus $\det\mathbf{R}=1$. Thus, in particular,

$$\mathbf{J}=\hat{\mathbf{v}}H(t)[H(\zeta_1)-H(\zeta_1-vt)]\,\delta(\zeta_2)\delta(\zeta_3), \tag{1.3}$$

where $\hat{\mathbf{v}}$ is the unit vector along the (positive ζ_1) direction of current flow and $H(t)$ denotes the Heaviside unit function. Define $\mathbf{v}=v\hat{\mathbf{v}}$, the current flow velocity. Now

$$\delta(\zeta_1-vt)\delta(\zeta_2)\delta(\zeta_3)=\delta\{(\mathbf{x}-\mathbf{v}t)\mathbf{R}\}=\delta(\mathbf{x}-\mathbf{v}t) \tag{1.4}$$

since the Jacobian $[\partial(\mathbf{x}-\mathbf{v}t)\mathbf{R}/\partial\mathbf{x}]_{\mathbf{x}=\mathbf{v}t}\equiv\det\mathbf{R}^T\equiv1$, the superscript T indicating a transposition. Hence, from (1.3),

$$\mathbf{J}_t=\mathbf{v}H(t)\delta(\mathbf{x}-\mathbf{v}t), \tag{1.5}$$

a traveling point source abruptly activated at $t=0$. It specifies the inhomogeneous term of (1.2). In the initial value problem, one makes the logical assumption that prior to source activation,

$$\mathbf{E}\equiv\mathbf{0}\equiv\mathbf{H},\quad \text{during } t<0. \tag{1.6}$$

Our immediate objective is the formulation of certain integrals, especially integrals over the unit sphere. These will be applied in a subsequent paper [1] to a crystal with compounded uniaxiality in relation to a certain combination of a nondiagonal $\boldsymbol{\epsilon}$ with a nondiagonal $\boldsymbol{\mu}$. No uniaxiality of any kind is employed throughout the present investigation. Handelsman and Lewis [2] have established the asymptotic theory of Čerenkov radiation for a dispersive medium via a system of integrodifferential equations. These lead to other integral forms that are analytically tractable for certain crystalline and optically active media. Integral forms over the Minkowski 4-space have also been studied by Johannsen, [3] with special emphasis on isotropic and uniaxial media, as well as an ionized gas.

2. MATRIX TRANSFORMATIONS

Before tackling (1.2), we appeal to the following (see Ref. 4):

(i) the existence of a real symmetric nonsingular matrix $\boldsymbol{\mu}^{1/2}$ satisfying $\boldsymbol{\mu}^{1/2}\boldsymbol{\mu}^{1/2}=\boldsymbol{\mu}$ (*NB* the determinant $\det\boldsymbol{\mu}>0$; the inverse $(\boldsymbol{\mu}^{1/2})^{-1}=\boldsymbol{\mu}^{-1/2}$ is symmetric);

(ii) the rule $(\boldsymbol{\alpha}\mathbf{A}^T)\times(\boldsymbol{\beta}\mathbf{A}^T)=(\boldsymbol{\alpha}\times\boldsymbol{\beta})\mathbf{A}^{-1}\det\mathbf{A}$ (2.1)

for all 1×3 matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and any nonsingular 3×3 matrix $\mathbf{A}$ with its transpose indicated by the superscript T;

(iii) the definition of the real matrix

$$\boldsymbol{\tau}=\boldsymbol{\mu}^{-1/2}\boldsymbol{\epsilon}\boldsymbol{\mu}^{-1/2}, \tag{2.2}$$

which turns out to be symmetric and positive definite.

Let

$$\mathbf{y}=(y_1,y_2,y_3)=(\det\boldsymbol{\mu})^{1/2}\mathbf{x}\boldsymbol{\mu}^{-1/2}. \tag{2.3}$$

Then in $\mathbf{y}$-space, the gradient operator becomes

$$\nabla_y=\left(\frac{\partial}{\partial y_1},\frac{\partial}{\partial y_2},\frac{\partial}{\partial y_3}\right)=(\det\boldsymbol{\mu})^{-1/2}\nabla\boldsymbol{\mu}^{1/2}. \tag{2.4}$$

If the row vector $\mathbf{F}$ is defined as

$$\mathbf{F}=(\det\boldsymbol{\mu})^{-1/2}\mathbf{E}\boldsymbol{\mu}^{1/2}, \tag{2.5}$$

 0022-2488/80/010189-06$01.00

it follows from (2.1) and (2.4) that

$$\nabla \times \mathbf{E} = (\det \boldsymbol{\mu})^{1/2} (\nabla_y \times \mathbf{F}) \boldsymbol{\mu}^{1/2}. \tag{2.6}$$

Repeating this principle with regard to (1.2) and accounting for (2.2), we arrive at

$$\mathbf{F}_{tt} \boldsymbol{\tau} + \nabla_y \times (\nabla_y \times \mathbf{F}) = -(\det \boldsymbol{\mu})^{-1/2} \mathbf{J}_t \boldsymbol{\mu}^{-1/2}. \tag{2.7}$$

To incorporate (1.5), we first define

$$\mathbf{w} = (w_1, w_2, w_3) = (\det \boldsymbol{\mu})^{1/2} \mathbf{v} \boldsymbol{\mu}^{-1/2}. \tag{2.8}$$

We can then show that the Jacobian modulus

$$\left| \frac{\partial(\mathbf{x} - \mathbf{v}t)}{\partial \mathbf{y}} \right|_{\mathbf{y} = \mathbf{w}t} = (\det \boldsymbol{\mu})^{-1}, \tag{2.9}$$

in which case,

$$\delta(\mathbf{x} - \mathbf{v}t) = (\det \boldsymbol{\mu}) \delta(\mathbf{y} - \mathbf{w}t). \tag{2.10}$$

Let, next,

$$\mathbf{S}(\nabla_y) = \begin{pmatrix} 0 & -\dfrac{\partial}{\partial y_3} & \dfrac{\partial}{\partial y_2} \\ \dfrac{\partial}{\partial y_3} & 0 & -\dfrac{\partial}{\partial y_1} \\ -\dfrac{\partial}{\partial y_2} & \dfrac{\partial}{\partial y_1} & 0 \end{pmatrix}, \tag{2.11}$$

a skew symmetric matrix operator with the property

$$\det \mathbf{S}(\nabla_y) \equiv 0. \tag{2.12}$$

It is then seen that

$$[\nabla_y \times (\nabla_y \times \mathbf{F})]^T = \mathbf{S}^2(\nabla_y) \mathbf{F}^T. \tag{2.13}$$

By reconsidering the square bracketed quantity in terms of vectorial operations, we have

$$\mathbf{S}^2(\nabla_y) = \nabla_y^T \nabla_y - \mathbf{I} \nabla_y \nabla_y^T, \tag{2.14}$$

$\mathbf{I}$ being the 3×3 identity matrix.

Applying (1.5), (2.8), (2.10), and (2.13) to (2.7), we get

$$\mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right) \mathbf{F}^T = -\boldsymbol{\tau}^{-1} \mathbf{w}^T H(t) \delta(\mathbf{y} - \mathbf{w}t), \tag{2.15}$$

where the 3×3 matrix operator $\mathbf{Q}$ is defined as

$$\mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right) = \mathbf{I} \frac{\partial^2}{\partial t^2} + \mathbf{P}(\nabla_y), \tag{2.16}$$

with

$$\mathbf{P}(\nabla_y) = \boldsymbol{\tau}^{-1} \mathbf{S}^2(\nabla_y). \tag{2.17}$$

Note the property

$$\det \mathbf{Q}(0, \nabla_y) = \det \mathbf{P}(\nabla_y) \equiv 0,$$

which follows in view of (2.17) and (2.12). Hence, from (2.16). $\partial^2/\partial t^2$ is a factor of

$$\det \mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right) = \frac{\partial^2}{\partial t^2} L\left(\frac{\partial}{\partial t}, \nabla_y\right), \tag{2.18}$$

say. This involves the characteristic polynomial of $\mathbf{P}(\nabla_y)$ and is of order six in $\partial/\partial t$. Thus, the linear scalar operator (cf. Ref. 5).

$$L\left(\frac{\partial}{\partial t}, \nabla_y\right) = \frac{\partial^4}{\partial t^4} + \frac{\partial^2}{\partial t^2} \operatorname{tr}\mathbf{P}(\nabla_y) + \tfrac{1}{2} [\operatorname{tr}\mathbf{P}(\nabla_y)]^2 - \tfrac{1}{2} \operatorname{tr}\mathbf{P}^2(\nabla_y), \tag{2.19}$$

with tr denoting the trace. Observe, via (2.11) and (2.17), a second degree homogeneity of $\mathbf{P}(\nabla_y)$, from which it follows that for any constant X,

$$\operatorname{tr}\mathbf{P}(X\nabla_y) = X^2 \operatorname{tr}\mathbf{P}(\nabla_y), \quad \operatorname{tr}\mathbf{P}^2(X\nabla_y) = X^4 \operatorname{tr}\mathbf{P}^2(\nabla_y); \tag{2.20}$$

hence

$$L\left(X \frac{\partial}{\partial t}, X\nabla_y\right) = X^4 L\left(\frac{\partial}{\partial t}, \nabla_y\right). \tag{2.21}$$

Consider the scalar function ϕ governed by

$$L\left(\frac{\partial}{\partial t}, \nabla_y\right) \phi = H(t) \delta(\mathbf{y} - \mathbf{w}t). \tag{2.22}$$

Accounting then for (2.4), (2.5), and (2.8), the algebraic resolution of (2.15) via (2.18) leads to

$$\mathbf{E}_{tt} = -(\det \boldsymbol{\mu}) \mathbf{v} \boldsymbol{\mu}^{-1/2} \boldsymbol{\tau}^{-1} \times \operatorname{adj}^T \mathbf{Q}\left\{\frac{\partial}{\partial t}, (\det \boldsymbol{\mu})^{-1/2} \nabla \boldsymbol{\mu}^{1/2}\right\} \phi \boldsymbol{\mu}^{-1/2}, \tag{2.23}$$

a second derivative in the original $\mathbf{x}$-space; here adj^T denotes the transposed adjoint of the $\mathbf{Q}$-matrix. Condition (1.6) implies that $\mathbf{E}_{tt} \equiv \mathbf{0}$ before the source is activated. This is guaranteed by postulating

$$\phi \equiv 0 \quad \text{over } t < 0. \tag{2.24}$$

3. INVERSION

Relative to a translated frame defined by

$$\mathbf{s} = (s_1, s_2, s_3) = \mathbf{y} - \mathbf{w}t, \tag{3.1}$$

with corresponding gradient operator

$\nabla_s = (\partial/\partial s_1, \partial/\partial s_2, \partial/\partial s_3)$, (2.22) becomes

$$L\left(\frac{\partial}{\partial t} - \mathbf{w} \cdot \nabla_s, \nabla_s\right) \phi = H(t) \delta(\mathbf{s}). \tag{3.2}$$

We introduce the Fourier transform $\mathrm{FT}[\phi]$ whose inverse is

$$\phi = \iiint_{R_3} \exp(i\boldsymbol{\alpha} \cdot \mathbf{s})\, d\boldsymbol{\alpha} \int_{-\infty + ic}^{\infty + ic} \mathrm{FT}[\phi] \exp(-i\sigma t)\, d\sigma. \tag{3.3}$$

For the horizontal complex integral path $(-\infty + ic, \infty + ic)$, we choose $c > 0$. We shall see that this choice satisfies the condition (2.24). The outer triple integral with respect to $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ ranges over the infinite three-dimensional space $R_3: -\infty < \alpha_\nu < \infty (\nu = 1,2,3)$, equivalently, $0 \leqslant |\boldsymbol{\alpha}| < \infty$, with the unit position vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3) = \boldsymbol{\alpha}|\boldsymbol{\alpha}|^{-1} \in \Omega_3$, the three dimensional unit sphere. Thus,

$$\iiint_{R_3} \cdots d\boldsymbol{\alpha} = \int_{-\infty}^{\infty} d\alpha_1 \int_{-\infty}^{\infty} d\alpha_2 \int_{-\infty}^{\infty} \cdots d\alpha_3 = \int_{\Omega_3} d\Omega \int_0^{\infty} \cdots \boldsymbol{\alpha}^2\, d|\boldsymbol{\alpha}|, \tag{3.4}$$

$d\Omega$ being a surface element of Ω_3. Fourier transformation of (3.2), with the homogeneity rule (2.21) incorporated, yields

$$L(-\sigma - \boldsymbol{\alpha} \cdot \mathbf{w}, \boldsymbol{\alpha})\, \mathrm{FT}[\phi] = i(2\pi)^{-4} \sigma^{-1}. \tag{3.5}$$

Regarding the L function in (3.5), we deduce from

(2.19) that its primary version

$$L(-\sigma,\boldsymbol{\alpha}) = [\sigma^2 - \sigma_+^2(\boldsymbol{\alpha})]\,[\sigma^2 - \sigma_-^2(\boldsymbol{\alpha})], \tag{3.6}$$

where, accounting for (2.20), we can express

$$\sigma_\pm(\boldsymbol{\alpha}) = |\boldsymbol{\alpha}|\sigma_\pm(\boldsymbol{\xi}), \tag{3.7}$$

with

$$\sigma_\pm(\boldsymbol{\xi}) = \{-\tfrac{1}{2}\mathrm{tr}\mathbf{P}(\boldsymbol{\xi}) \pm \tfrac{1}{2}\{[\mathrm{tr}\mathbf{P}(\boldsymbol{\xi})]^2 - 2[(\mathrm{tr}\mathbf{P}(\boldsymbol{\xi}))^2 - \mathrm{tr}\mathbf{P}^2(\boldsymbol{\xi})]\}^{1/2}\}^{1/2}. \tag{3.8}$$

By the principal axes theorem, there is no loss of generality in assuming a diagonalized $\boldsymbol{\tau}$-matrix [Ref. 4, (3.45)]:

$$\boldsymbol{\tau} = \begin{pmatrix} \lambda_1^{-2} & 0 & 0 \\ 0 & \lambda_2^{-2} & 0 \\ 0 & 0 & \lambda_3^{-2} \end{pmatrix}. \tag{3.9}$$

However, $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ are generally nondiagonal. Positive definiteness of $\boldsymbol{\tau}$ demands that its eigenvalues $\lambda_j^{-2} (j = 1,2,3)$ are all positive. Thus, from (2.14) and (2.17), one derives

$$-\mathrm{tr}\mathbf{P}(\boldsymbol{\xi}) = \lambda_1^2(\xi_2^2 + \xi_3^2) + \lambda_2^2(\xi_3^2 + \xi_1^2) + \lambda_3^2(\xi_1^2 + \xi_2^2) > 0. \tag{3.10}$$

Furthermore, assuming, again without sacrificing generality, that

$$\lambda_1 \geqslant \lambda_2 > \lambda_3 > 0, \tag{3.11}$$

one can establish after some extensive computations that

$$\begin{aligned} &2\mathrm{tr}\mathbf{P}^2(\boldsymbol{\xi}) - [\mathrm{tr}\mathbf{P}(\boldsymbol{\xi})]^2 \\ &= [(|\xi_1|\,|\lambda_2^2 - \lambda_3^2|^{1/2} - |\xi_3|\,|\lambda_1^2 - \lambda_2^2|^{1/2})^2 + \xi_2^2(\lambda_1^2 - \lambda_3^2)] \\ &\quad\times[(|\xi_1|\,|\lambda_2^2 - \lambda_3^2|^{1/2} + |\xi_3|\,|\lambda_1^2 - \lambda_2^2|^{1/2})^2 + \xi_2^2(\lambda_1^2 - \lambda_3^2)]; \end{aligned} \tag{3.12}$$

also,

$$\begin{aligned} &[\mathrm{tr}\mathbf{P}(\boldsymbol{\xi})]^2 - \mathrm{tr}\mathbf{P}^2(\boldsymbol{\xi}) \\ &= 2(\xi_1^2\lambda_2^2\lambda_3^2 + \xi_2^2\lambda_3^2\lambda_1^2 + \xi_3^2\lambda_1^2\lambda_2^2) > 0. \end{aligned} \tag{3.13}$$

Thus,

$$0 \leqslant [\mathrm{tr}\mathbf{P}(\boldsymbol{\xi})]^2 - 2[(\mathrm{tr}\mathbf{P}(\boldsymbol{\xi}))^2 - \mathrm{tr}\mathbf{P}^2(\boldsymbol{\xi})] < [\mathrm{tr}\mathbf{P}(\boldsymbol{\xi})]^2. \tag{3.14}$$

By (3.10), then, (3.8) discloses $\sigma_\pm(\boldsymbol{\xi})$ to be real and positive, while

$$\sigma_+(\boldsymbol{\xi}) \not\equiv \sigma_-(\boldsymbol{\xi}) \quad \text{over } \Omega_3. \tag{3.15}$$

Nevertheless, it is seen from (3.8) and (3.12) that $\sigma_+(\boldsymbol{\xi})$ and $\sigma_-(\boldsymbol{\xi})$ do coincide discretely, viz., along the four separate directions

$$\begin{aligned} \boldsymbol{\xi} = &\pm |\lambda_1^2 - \lambda_3^2|^{-1/2}(|\lambda_1^2 - \lambda_2^2|^{1/2}, 0, |\lambda_2^2 - \lambda_3^2|^{1/2}), \\ &\pm |\lambda_1^2 - \lambda_3^2|^{-1/2} \\ &\times(|\lambda_1^2 - \lambda_2^2|^{1/2}, 0, -|\lambda_2^2 - \lambda_3^2|^{1/2}). \end{aligned} \tag{3.16}$$

Let

$$\omega_\nu = \omega_\nu(\boldsymbol{\alpha}) = \begin{cases} 0, & \nu = 1, \quad (3.17) \\ \sigma_+(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\cdot\mathbf{w}, & \nu = 2, \quad (3.18) \\ -\sigma_+(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\cdot\mathbf{w}, & \nu = 3, \quad (3.19) \\ \sigma_-(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\cdot\mathbf{w}, & \nu = 4, \quad (3.20) \\ -\sigma_-(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\cdot\mathbf{w}, & \nu = 5. \quad (3.21) \end{cases}$$

Then by (3.3), (3.5), and (3.6),

$$\phi = \frac{1}{(2\pi)^3}\iiint_{R_3} \exp(i\boldsymbol{\alpha}\cdot\mathbf{s})\,d\boldsymbol{\alpha} \times \frac{i}{2\pi}\int_{-\infty+ic}^{\infty+ic} \frac{\exp(-i\sigma t)}{\Pi_{\nu=1}^5\,\sigma - \omega_\nu(\boldsymbol{\alpha})}\,d\sigma. \tag{3.22}$$

The inner σ integral must be considered for each $|\boldsymbol{\alpha}| \in (0,\infty)$, in which case, we shall see that it approaches a singular behavior as $|\boldsymbol{\alpha}| \to 0$. Such a singularity ultimately contributes to an integration with respect to $|\boldsymbol{\alpha}|$ in accordance with (3.4). The σ integrand is meromorphic with, normally, five real poles at $\sigma = \omega_\nu(\boldsymbol{\alpha})$ $(\nu = 1,\dots,5)$. If $t < 0$, the path $(-\infty + ic, \infty + ic)$ may be closed by an infinite semicircle, necessarily drawn into $\mathrm{Im}\sigma > c$, and along which the extended integral vanishes because conditions of Jordan's lemma are satisfied. Due to the choice $c > 0$, all five poles occur outside the closed contour constructed, so that (2.24) indeed holds by virtue of Cauchy's theorem.

Henceforth, we assume that $t > 0$, whence, in order that the extended integral along a closing infinite semicircle vanishes, this semicircle must, by Jordan's lemma, be described within $\mathrm{Im}\sigma < c$. The completed closed contour is anticlockwise and circumscribes all five poles. These are simple for almost every $\boldsymbol{\xi} \in \Omega_3$ and $\forall |\boldsymbol{\alpha}| \in (0,\infty)$. Discrete exceptions are (i) the four directions of (3.16) along each of which ω_2 meets ω_4 while ω_3 meets ω_5 to form two separate poles, both of order two provided

$$\pm\,\boldsymbol{\xi}\cdot\mathbf{w} \neq \sigma_+(\boldsymbol{\xi}) \quad [= \sigma_-(\boldsymbol{\xi})] \tag{3.23}$$

along the particular $\boldsymbol{\xi}$ direction; (ii) other $\boldsymbol{\xi}$ directions satisfying

$$\boldsymbol{\xi}\cdot\mathbf{w} = \sigma_+(\boldsymbol{\xi}), \quad -\sigma_+(\boldsymbol{\xi}), \quad \sigma_-(\boldsymbol{\xi}), \quad -\sigma_-(\boldsymbol{\xi}), \tag{3.24}$$

i.e., where ω_ν $(\nu = 2, 3, 4, \text{ or } 5)$ meets $\omega_1 \equiv 0$ to form a double pole at $\sigma = 0$. Should (3.23) be violated by a coincidence of any of the $\boldsymbol{\xi}$ directions for (3.24) with one of the $\boldsymbol{\xi}$ directions of (3.16), then we have a triple pole at $\sigma = 0$ and a separate double pole at $\sigma = \omega_2\,(=\omega_4)$ or at $\sigma = \omega_3\,(=\omega_5)$.

It is established in the appendix that, whatever the order of each pole along any specified $\boldsymbol{\xi}$ direction, a uniformly consistent value for $\int_{-\infty+ic}^{\infty+ic}$ can be determined on the basis that every pole is simple $\forall\boldsymbol{\xi} \in \Omega_3$. The key factor is, of course, the fact that each pole is simple almost everywhere on Ω_3, whereupon residue theory yields

$$\frac{i}{2\pi}\int_{-\infty+ic}^{\infty+ic} \frac{\exp(-i\sigma t)}{\Pi_{\nu=1}^5\,\sigma - \omega_\nu(\boldsymbol{\alpha})}\,d\sigma = \sum_{j=1}^{5} \frac{\exp[-i\omega_j(\boldsymbol{\alpha})\,t]}{\Pi_{\nu(\neq j)}\,\omega_j(\boldsymbol{\alpha}) - \omega_\nu(\boldsymbol{\alpha})}. \tag{3.25}$$

Observe that this result confirms, in view of (3.7) and the forms (3.17)–(3.21), an inverse quartic singular behavior near $|\boldsymbol{\alpha}| = 0$.

4. THE SPHERICAL INTEGRALS

Accounting for (3.1), (3.17)–(3.21), it can be shown from (3.22) and (3.25) that

$$\phi = 2^{-1}(2\pi)^{-3} \sum_{\nu=\pm} \iiint_{R_3} \frac{\exp(i\boldsymbol{\alpha}\cdot\mathbf{s}) - \exp\{i[\boldsymbol{\alpha}\cdot\mathbf{y} - \sigma_\nu(\boldsymbol{\alpha})\,t\,]\}}{\sigma_\nu(\boldsymbol{\alpha})[\sigma_\nu^2(\boldsymbol{\alpha}) - \sigma_{-\nu}^2(\boldsymbol{\alpha})]\;[\boldsymbol{\alpha}\cdot\mathbf{w} - \sigma_\nu(\boldsymbol{\alpha})]}\,d\boldsymbol{\alpha}$$
$$- 2^{-1}(2\pi)^{-3} \sum_{\nu=\pm} \iiint_{R_3} \frac{\exp(i\boldsymbol{\alpha}\cdot\mathbf{s}) - \exp\{i[\boldsymbol{\alpha}\cdot\mathbf{y} + \sigma_\nu(\boldsymbol{\alpha})\,t\,]\}}{\sigma_\nu(\boldsymbol{\alpha})[\sigma_\nu^2(\boldsymbol{\alpha}) - \sigma_{-\nu}^2(\boldsymbol{\alpha})]\;[\boldsymbol{\alpha}\cdot\mathbf{w} + \sigma_\nu(\boldsymbol{\alpha})]}\,d\boldsymbol{\alpha}, \tag{4.1}$$

where $\sigma_{-\nu} = \sigma_{\mp}$ when $\nu = \pm$, over which the summation runs. Now (3.7), (3.8), (3.10), and (3.13) reveal that $\sigma_\nu(\boldsymbol{\alpha}) = \sigma_\nu(-\boldsymbol{\alpha})$. Consequently, we can show that if

$$\phi_\nu(\mathbf{y},t) = (2\pi)^{-3} \iiint_{R_3} \frac{\cos[\boldsymbol{\alpha}\cdot\mathbf{y} - \sigma_\nu(\boldsymbol{\alpha})\,t\,]\,d\boldsymbol{\alpha}}{\sigma_\nu(\boldsymbol{\alpha})[\sigma_\nu^2(\boldsymbol{\alpha}) - \sigma_{-\nu}^2(\boldsymbol{\alpha})]\;[\sigma_\nu(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\cdot\mathbf{w}]}, \tag{4.2}$$

then

$$\phi = \sum_{\nu=\pm} \phi_\nu(\mathbf{y},t) - \phi_\nu(\mathbf{s},0). \tag{4.3}$$

Within the context of generalized functions,[6]

$$\int_0^\infty \alpha^{-2}\cos(\alpha X)\,d\alpha = -\tfrac{1}{2}\pi|X| \tag{4.4}$$

for any real scalar X. This rule can be applied with (3.4) and (3.7) to (4.2) to secure the contribution imparted by the singularity at $|\boldsymbol{\alpha}| = 0$ arising from (3.25). Thus, recollecting (2.3) and (2.8), we eventually establish in the original $\mathbf{x}$-space,

$$\phi_\nu(\mathbf{y},t) = \frac{1}{(4\pi)^2} \iint_{\Omega_3} \frac{|(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{x}^T - \sigma_\nu(\boldsymbol{\xi})t\,|\,d\Omega}{\sigma_\nu(\boldsymbol{\xi})[\sigma_\nu^2(\boldsymbol{\xi}) - \sigma_{-\nu}^2(\boldsymbol{\xi})]\;[(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{v}^T - \sigma_\nu(\boldsymbol{\xi})]}. \tag{4.5}$$

Moreover, if

$$\mathbf{r} = (r_1, r_2, r_3) = \mathbf{x} - \mathbf{v}t, \tag{4.6}$$

i.e., the position vector relative to the advancing current edge, then

$$\phi_\nu(\mathbf{s},0) = \frac{1}{(4\pi)^2} \iint_{\Omega_3} \times \frac{|(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{r}^T|\,d\Omega}{\sigma_\nu(\boldsymbol{\xi})[\sigma_\nu^2(\boldsymbol{\xi}) - \sigma_{-\nu}^2(\boldsymbol{\xi})]\;[(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{v}^T - \sigma_\nu(\boldsymbol{\xi})]}, \tag{4.7}$$

which is time independent in $\mathbf{r}$-space. As a consequence of the coincidental poles phenomena discussed earlier, the integrands represented in (4.5) and (4.7) are singular on Ω_3 at each of the four $\boldsymbol{\xi}$ values of (3.16), as well as at any $\boldsymbol{\xi}$ value satisfying

$$(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{v}^T = \sigma_\nu(\boldsymbol{\xi}). \tag{4.8}$$

Cauchy principal value interpretations are therefore applicable to the spherical integrals of (4.5) and (4.7). In terms of these spherical integrals, a solution for the scalar field ϕ as expressed by (4.3) is now formally complete. The solution for the vector field $\mathbf{E}_u$ is then evaluated from (2.23). Observe that as $t \to 0_+$, $\phi_\nu(\mathbf{y},t) \to \phi_\nu(\mathbf{s},0)$, and so $\phi \to 0$; coupled with hypothesis (2.24), this indicates a continuity across $t = 0$.

A. A steady state

Since $\sigma_-(\boldsymbol{\xi}) > 0$, it achieves a positive minimum on Ω_3. Suppose

$$(\det\boldsymbol{\mu})^{1/2}|\mathbf{x}\boldsymbol{\mu}^{-1/2}| < t \min_{\Omega_3} \sigma_-(\boldsymbol{\xi}). \tag{4.9}$$

This describes a domain originating at $\mathbf{x} = \mathbf{0}$ with the current activation when $t = 0$. It thereafter evolves with time about its origin which becomes increasingly separated from the subsequent current edge at $\mathbf{r} = \mathbf{0}$. We shall demonstrate that $\mathbf{E}_u$ attains a steady state relative to this edge and inside the specified domain.

By (3.8) and (3.14), $\sigma_+(\boldsymbol{\xi}) \geqslant \sigma_-(\boldsymbol{\xi})$. So (4.9) implies that $\forall \boldsymbol{\xi} \in \Omega_3$,

$$(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{x}^T < \sigma_\nu(\boldsymbol{\xi})\,t \quad (\nu = \pm). \tag{4.10}$$

Now, in view of (2.4), (2.14) and (2.17), the 3×3 $\mathbf{Q}$ operator of (2.16) is homogeneous of degree two in $\partial/\partial t$, $\partial/\partial x_j$ ($j = 1,2,3$). Hence, its cofactors and therefore its transposed adjoint

$$\operatorname{adj}^T\mathbf{Q}\left\{\frac{\partial}{\partial t}, (\det\boldsymbol{\mu})^{-1/2}\nabla\boldsymbol{\mu}^{1/2}\right\} = \operatorname{adj}^T\mathbf{Q}\left\{\frac{\partial}{\partial t} - \mathbf{v}\cdot\nabla_r\,, (\det\boldsymbol{\mu})^{-1/2}\nabla_r\boldsymbol{\mu}^{1/2}\right\} \tag{4.11}$$

are each homogeneous of degree four in $\partial/\partial t$, $\partial/\partial x_j$ ($j = 1,2,3$). Here, $\nabla_r = (\partial/\partial r_1, \partial/\partial r_2, \partial/\partial r_3)$. So on Ω_3, the combination

$$|(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{x}^T - \sigma_\nu(\boldsymbol{\xi})t\,|$$

being linear in x_1, x_2, x_3 and t and smooth under (4.10), vanishes identically when acted upon by the $\operatorname{adj}^T\mathbf{Q}$ operator of (4.11). Applying this notion to (4.5) and recalling the time independence of $\phi_\nu(\mathbf{s},0)$ in $\mathbf{r}$-space, we have from (2.23), (4.3), and (4.11),

$$\mathbf{E}_{tt} = (\det\boldsymbol{\mu})\mathbf{v}\boldsymbol{\mu}^{-1/2}\tau^{-1}\mathrm{adj}^T\mathbf{Q}\{-\mathbf{v}\cdot\nabla_r, (\det\boldsymbol{\mu})^{-1/2}\nabla_r\boldsymbol{\mu}^{1/2}\} \times \sum_{\nu=\pm}\phi_\nu(\mathbf{s},0)\boldsymbol{\mu}^{-1/2}, \quad (4.12)$$

which is generally nontrivial, but is time independent relative to the $\mathbf{r}$-frame. In this respect, a steady state develops within the domain defined by (4.9), or in particular, at any given $\mathbf{x}$ position after a sufficiently long period.

The domain of (4.9) can be geometrically identified. First we apply the principal axes theorem to the symmetric matrix $\boldsymbol{\mu}$ to get the diagonal form

$$\mathbf{N}^{-1}\boldsymbol{\mu}\mathbf{N} = \begin{pmatrix}\mu_1 & 0 & 0\\ 0 & \mu_2 & 0\\ 0 & 0 & \mu_3\end{pmatrix}, \quad (4.13)$$

for a specific orthogonal matrix $\mathbf{N}$, i.e., $\mathbf{N}^{-1} = \mathbf{N}^T$. So

$$\boldsymbol{\mu}^{-1} = \mathbf{N}\begin{pmatrix}\mu_1^{-1} & 0 & 0\\ 0 & \mu_2^{-1} & 0\\ 0 & 0 & \mu_3^{-1}\end{pmatrix}\mathbf{N}^T. \quad (4.14)$$

Let $\mathbf{x}\mathbf{N} = \mathbf{X} = (X_1, X_2, X_3)$, i.e., the x_j axes and X_j axes are separated by a pure rotation. Then

$$|\mathbf{x}\boldsymbol{\mu}^{-1/2}|^2 = \mathbf{x}\boldsymbol{\mu}^{-1}\mathbf{x}^T = \sum_{j=1,2,3}\mu_j^{-1}X_j^2. \quad (4.15)$$

Since $\boldsymbol{\mu}$ is positive definite, its eigenvalues μ_j $(j = 1,2,3)$ introduced in (4.13) are positive. Consequently, the steady state supporting domain expressed by (4.9) is the interior of an ellipsoid with equation

$$(\det\boldsymbol{\mu})\sum_{j=1,2,3}\mu_j^{-1}X_j^2 = t^2\Big[\min_{\Omega_3}\sigma_-(\boldsymbol{\xi})\Big]^2. \quad (4.16)$$

This ellipsoid expands about its center at $\mathbf{x} = \mathbf{0}$ and retreats in relation to the current edge at $\mathbf{r} = \mathbf{0}$. As (4.9) merely represents a sufficiency criterion, the particular steady state of $\mathbf{E}_{tt}$ may well extend to a larger domain containing the present ellipsoid.

B. The resultant $\Sigma_{\nu=\pm}\phi_\nu(\mathbf{s},0)$

The resultant $\Sigma_{\nu=\pm}\phi_\nu(\mathbf{s},0)$ is crucially involved in (4.12) as well as in the main result (4.3). Its two constituents are individually represented by (4.7). However, a compact spherical integral form for this resultant exists in terms of an algebraic transform of the L operator defined by (2.19). Such a form can be extracted in the first instance from (3.25). Alternatively, it is easy enough to obtain from (4.2). Thus, via the σ_ν symmetry,

$$\phi_\nu(\mathbf{s},0) = \frac{1}{(2\pi)^3}\iiint_{R_3}\frac{\cos(\boldsymbol{\alpha}\cdot\mathbf{s})\,d\boldsymbol{\alpha}}{[\sigma_\nu^2(\boldsymbol{\alpha}) - \sigma_{-\nu}^2(\boldsymbol{\alpha})]\,[\sigma_\nu^2(\boldsymbol{\alpha}) - (\boldsymbol{\alpha}\cdot\mathbf{w})^2]}; \quad (4.17)$$

$$= \frac{1}{(4\pi)^2}\iint_{\Omega_3}\frac{|\boldsymbol{\xi}\cdot\mathbf{s}|\,d\Omega}{[\sigma_\nu^2(\boldsymbol{\xi}) - \sigma_{-\nu}^2(\boldsymbol{\xi})]\,[(\boldsymbol{\xi}\cdot\mathbf{w})^2 - \sigma_\nu^2(\boldsymbol{\xi})]}, \quad (4.18)$$

which follows from (4.17) via (3.7) and (4.4), and represents an alternative to (4.7). Accounting for (3.6), then,

$$\sum_{\nu=\pm}\phi_n(\mathbf{s},0) = \frac{(\det\boldsymbol{\mu})^{1/2}}{(4\pi)^2}\iint_{\Omega_3}\frac{|\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{r}^T|\,d\Omega}{L\{-(\det\boldsymbol{\mu})^{1/2}\boldsymbol{\xi}\boldsymbol{\mu}^{-1/2}\mathbf{v}^T,\boldsymbol{\xi}\}}, \quad (4.19)$$

the desired form. As this persists indefinitely in r-space, it could be interpreted as a possible Čerenkov radiation function. The complementary resultant $\Sigma_{\nu=\pm}\phi_\nu(\mathbf{y},t)$, which cannot contribute to $\mathbf{E}_{tt}$ in the steady state, would then represent the switch-on effect of the source current.

APPENDIX

By the reasoning accorded to (3.22), a typical real pole of its σ integrand at $\omega_j = \omega_j(\boldsymbol{\alpha})$ can, $\forall|\boldsymbol{\alpha}|\in(0,\infty)$, attain an order of $m_j = 1$, 2, or 3 depending on the direction of $\boldsymbol{\xi} = \boldsymbol{\alpha}|\boldsymbol{\alpha}|^{-1}\in\Omega_3$. However, the subsequent result (3.25) is formulated on the basis that each $m_j = 1$, this being indeed the case almost everywhere on Ω_3. Should $m_j > 1$ along some isolated $\boldsymbol{\xi}$ direction, the individual residue contribution from ω_j to (3.25) fails since ω_j equals ω_ν for at least one $\nu\neq j$; in this event, the residue contribution from the particular ω_ν $(=\omega_j)$ likewise fails. Nonetheless, this need not imply the failure of the resultant $\Sigma_{j=1}^5$. The objective in this appendix is to demonstrate that such a sum is, in fact, uniformly valid $\forall\boldsymbol{\xi}\in\Omega_3$.

When $t > 0$, the fundamentally correct residue representation for the σ integral of (3.22) at any $\boldsymbol{\xi}\in\Omega_3$ should be

$$\frac{i}{2\pi}\int_{-\infty+ic}^{\infty+ic}\frac{\exp(-i\sigma t)}{\Pi_{\nu=1}^5\,\sigma - \omega_\nu(\boldsymbol{\alpha})}\,d\sigma = \sum_j \mathrm{res}(\omega_j, m_j), \quad (\mathrm{A1})$$

where

$$\mathrm{res}(\omega_j, m_j) = \frac{1}{(m_j-1)!}\left[\left(\frac{\partial}{\partial\sigma}\right)^{m_j-1}\frac{\exp(-i\sigma t)}{\Pi_{\nu:\omega_\nu\neq\omega_j}\sigma - \omega_\nu}\right]_{\sigma=\omega_j}, \quad (\mathrm{A2})$$

which expresses a finite residue at the σ pole ω_j of order $m_j = m_j(\boldsymbol{\xi})$. The number of such residues encountered in (A1) is at most five (achieved only if every $m_j = 1$). The product $\Pi_\nu:\omega_\nu\neq\omega_j$ in (A2) involves 5-m_j factors. Should every $m_j = 1$ $(j = 1,\dots,5)$, (A1) and (3.25) become identical expressions.

A σ pole of order 2 occurs along any of the four $\boldsymbol{\xi}$ directions of (3.16) or when any of the four relations in (3.24) holds along $\boldsymbol{\xi}$ directions distinct from those of (3.16). Precisely, it occurs when two normally separate poles ω_j and ω_k of orders $m_j = 1$, $m_k = 1$ meet for some instantaneous $\boldsymbol{\xi}$ value, say, $\boldsymbol{\eta}\in\Omega_3$. If $\boldsymbol{\xi}$ is sufficiently near $\boldsymbol{\eta}$ for each $|\boldsymbol{\alpha}|\in(0,\infty)$, then $\omega_j\approx\omega_k$, so that

$$\mathrm{res}(\omega_j,1) = \frac{1}{\omega_j-\omega_k}\,\frac{\exp(-i\omega_j t)}{\Pi_{\nu(\neq j,k)}\,\omega_j - \omega_\nu} = \frac{1}{\omega_j-\omega_k}\,\frac{\exp(-i\omega_k t)}{\Pi_{\nu(\neq j,k)}\,\omega_k - \omega_\nu} \quad (\mathrm{A3})$$

$$+\frac{\partial}{\partial\omega_k}\frac{\exp(-i\omega_k t)}{\Pi_{\nu(\neq j,k)}\omega_k-\omega_\nu}+O(\omega_j-\omega_k), \quad (A4)$$

since, on the right side of (A3), the factor accompanying $(\omega_j-\omega_k)^{-1}$ is an analytic function of ω_j within some sufficiently small neighborhood of ω_k. The leading term in (A4) is $-\operatorname{res}(\omega_k,1)$, which is now near singular. However, in the limit as $\xi\to\eta$, the combination

$$\lim_{\xi\to\eta}[\operatorname{res}(\omega_j,1)+\operatorname{res}(\omega_k,1)] = \frac{\partial}{\partial\omega_k}\frac{\exp(-i\omega_k t)}{\Pi_{\nu(\neq j,k)}\omega_k-\omega_\nu}\bigg|_{\xi=\eta}, \quad (A5)$$

$$= \operatorname{res}(\omega_k,2)|_{\xi=\eta}\equiv\operatorname{res}(\omega_j,2)|_{\xi=\eta}, \quad (A6)$$

a bounded residue whose boundedness evidently follows from the mutual cancellation of a singular pair. Once ξ departs from η, $\operatorname{res}(\omega_j,2)|_{\xi=\eta}$ resolves spontaneously into the original constituents $\operatorname{res}(\omega_j,1)$ and $\operatorname{res}(\omega_k,1)$. The transition and recovery in the Σ_j composition of (3.25) as ξ crosses η is clearly consistent with (A1). Thus, (3.25) is uniformly valid $\forall\xi\in\Omega_3$, at least for σ poles whose orders never exceed two.

Reconsider (3.16). Suppose along any of the four specified ξ directions now denoted by η, say, $\sigma_+(\eta)=\eta\cdot\mathbf{w}$ (or $-\eta\cdot\mathbf{w}$). Let $\xi\to\eta$. Then two separate simple poles ω_j and ω_k, with $j=2$, $k=4$ (or $j=3$, $k=5$) converge towards the simple pole $\omega_1\equiv0$ to form, in the limit, a triple pole at $\sigma=0$. Meanwhile, the two remaining simple poles ω_3 and ω_5 (or ω_1 and ω_2) approach one another to produce, when $\xi=\eta$, a consistent net contribution associated with a double pole as already demonstrated. Suppose

$$A_l=\frac{1}{l!}\left[\frac{\partial^l}{\partial\sigma^l}\frac{\exp(-i\sigma t)}{\Pi_{\nu(\neq j,k,1)}\sigma-\omega_\nu}\right]_{\sigma=0} \quad (l=0,1,\cdots), \quad (A7)$$

a finite coefficient. Then

$$\operatorname{res}(\omega_j,1)=\frac{\exp(-i\omega_j t)}{\omega_j(\omega_j-\omega_k)\Pi_{\nu(\neq j,k,l)}\omega_j-\omega_\nu}, \quad (A8)$$

$$=(\omega_j-\omega_k)^{-1}(A_0\omega_j^{-1}+A_1+A_2\omega_j+A_3\omega_j^2+A_4\omega_j^3+\cdots) \quad (A9)$$

when ξ is sufficiently near η for each $|\alpha|\in(0,\infty)$. A corresponding representation holds for $\operatorname{res}(\omega_k,1)$, whereupon, when $\xi\sim\eta$,

$$\operatorname{res}(\omega_j,1)+\operatorname{res}(\omega_k,1) = -A_0\omega_j^{-1}\omega_k^{-1}+A_2+A_3(\omega_j+\omega_k)+A_4(\omega_j^2+\omega_j\,\omega_k+\omega_k^2)+\cdots. \quad (A10)$$

But

$$A_0\omega_j^{-1}\omega_k^{-1}=\operatorname{res}(\omega_1\equiv0,1), \quad (A11)$$

which is presently near singular. From (A10) and (A11), however,

$$\lim_{\xi\to\eta}[\operatorname{res}(\omega_j,1)+\operatorname{res}(\omega_k,1)+\operatorname{res}(\omega_1,1)] = \lim_{\xi\to\eta}A_2=\operatorname{res}(\omega_j,3)|_{\xi=\eta}\equiv\operatorname{res}(\omega_k,3)|_{\xi=\eta}\equiv\operatorname{res}(\omega_1,3)|_{\xi=\eta}, \quad (A12)$$

a finite value resulting from the combination of three singular quantities whose singular parts annihilate each other. Again (3.25) remains consistent with (A1), in particular, throughout an ξ neighborhood of η, if $j=2$ and $k=4$, the state of affairs is as follows:

$$\frac{i}{2\pi}\int_{-\infty+ic}^{\infty+ic} = \sum_{j=1}^{5}\operatorname{res}(\omega_j,1) \text{ outside the neighborhood,} \quad (A13)$$

$$\sim\operatorname{res}(\omega_1\equiv0,3)|_{\xi=\eta}+\operatorname{res}(\omega_3\equiv\omega_5,2)|_{\xi=\eta} \text{ inside the neighborhood.} \quad (A14)$$

Since no m_j exceeds 3 over Ω_3, the establishment for uniform validity of (3.25) is now complete.

ACKNOWLEDGMENTS

This investigation was originally prompted by conversations with Professor Kurt Suchy and Dr. Helmut Hebenstreit, to whom I am indebted. I would also like to thank them, as well as Frau Dr. Suchy, Jörg Thiel, Katja Havranek, Theresia Bernstein, and all other members of the Institut für Theoretische Physik der Universität Düsseldorf for the very warm hospitality received over the short tenure of a visiting appointment.

[1] L. Chee-Seng, J. Math. Phys. **21**, 195 (1980).
[2] R. Handelsman and R.M. Lewis, J. Math. Phys. 7, 1982 (1966).
[3] G. Johannsen, J. Math. Phys. **13**, 78 (1972).
[4] M. Kline, and I.W. Kay, *Electromagnetic Theory and Geometrical Optics* (Interscience, New York, 1965), Sec. 3.
[5] D.T. Finkbeiner, *Matrices and Linear Transformations* (Tuttle, Tokyo, 1960), Sec. 8.4.
[6] M.J. Lighthill, *Fourier Analysis and Generalized Functions* (Cambridge U.P., Cambridge, 1958).

Certain relativistic phenomena in crystal optics

Lim Chee-Seng

Department of Mathematics, University of Singapore, Singapore

(Received 12 July 1978; revised manuscript received 6 November 1978)

Relativistic unsteady phenomena are established for a crystalline medium with unaligned sets of permittivity and permeability principal axes, but incorporating a compounded uniaxiality about some nonprincipal direction. All effects originate from a suddenly activated, arbitrarily oriented, maintained line current conducted with a finite velocity $\mathbf{v}$. Integral representations studied in another paper (Chee-Seng) are applied. The original coordinate system is subjected to a series of rotational and translational, scaled and unscaled transformations. No specific coordinate frame is strictly adhered to. Instead, it is often expedient and advantageous to exploit several reference frames simultaneously in the course of the analysis and interpretations. The electric field is directly related to a net scalar field Δ involving another scalar Ψ and its complement $\bar{\Psi}$ which can be deduced from Ψ; Ψ and $\bar{\Psi}$ are associated with two expanding, inclined ellipsoidal wavefronts ξ and $\bar{\xi}$; these are cocentered at the current origin and touch each other twice along the uniaxis. Elsewhere, ξ leads $\bar{\xi}$. For a source current faster than ξ:$\mathbf{v}t \in \operatorname{ext}\xi$, $\Psi \not\equiv 0$ within a finite but growing "ice-cream cone" domain, its nontrivial composition being $\chi^{-1/2}$ inside ξ and $2\chi^{-1/2}$ inside part of a tangent cone from the advancing current edge $\mathbf{v}t$ to, and terminating at, ξ; the function χ vanishes along such a tangent cone. Alternatively, for a source current slower than ξ:$\mathbf{v}t \in \operatorname{int}\xi$, if $\mathbf{v}t$ is avoided, $\chi > 0$ everywhere, while $\Psi = \chi^{-1/2}$ inside ξ but vanishes identically outside ξ. However, the crucial scalar field Δ depends on three separate current–velocity regimes. Over a slow regime: $\mathbf{v}t \in \operatorname{int}\bar{\xi}$, Δ is nontrivial inside ξ wherein it is discontinuous across $\bar{\xi}$. Over an intermediate regime: $\mathbf{v}t \in \operatorname{int}\xi \cap \operatorname{ext}\bar{\xi}$, Δ takes four distinct forms on 12 adjacent domains bounded by ξ, $\bar{\xi}$ and a double-conical tangent surface linking $\mathbf{v}t$ to $\bar{\xi}$. But for a fast regime: $\mathbf{v}t \in \operatorname{ext}\xi$, Δ assumes six distinct forms on 18 adjacent domains bounded by ξ, $\bar{\xi}$ plus two double-conical tangent surfaces, convertexed at $\mathbf{v}t$, to both ξ and $\bar{\xi}$. Singularities are normally confined to these boundaries. Relative to a moving frame, χ is time-independent. Nevertheless, Ψ and, consequently, Δ evolve unsteadily, principally because of transitions across the expanding ellipsoids ξ and $\bar{\xi}$ which also acquire a relative retreat from the current edge $\mathbf{v}t$. An evolution scheme is discussed in detail. This produces, among other things, a steady state corollary which, in turn, covers Čerenkov radiation. A quadrical symmetry exists with respect to a family $\{Q_v\}$ of constant χ-surfaces. These are quadric surfaces cocentered at $\mathbf{v}t$ and having principal axes inclined to those of ξ (and $\bar{\xi}$). Their interactions with ξ are closely examined. If $\mathbf{v}t \in \operatorname{ext}\xi$, each Q_v is a hyperboloid of two sheets which are asymptotic to the double-conical tangent surface connecting ξ to $\mathbf{v}t$; Ψ can become nontrivial on only one sheet, viz., that which is approached by ξ as the latter retreats from $\mathbf{v}t$; eventually, two permanent intersections, one following the other, occur along two expanding and travelling parallel plane circuits. But if $\mathbf{v}t \in \operatorname{int}\xi$, each Q_v is an ellipsoid inside which ξ initially evolves until an encounter occurs, first as a point contact which immediately grows into a plane circuit; as this traverses Q_v, it expands and then contracts to a diametrically opposite point where contact breaks off. Finally, an elliptical axisymmetry about a principal direction of $\{Q_v\}$ is demonstrated. Corresponding behaviors hold in relation to $\bar{\xi}$.

1. INTRODUCTION

If a source current is conducted inside a crystal with a speed comparable to the wave speeds of the crystalline structure, relativistic steady state effects, which may include Čerenkov effects, are ultimately detectable within a moving frame. This is possible on the hypothesis that, since its activation, the current possesses a monotonous history over an indefinite period preceding observations. However, the evolution process throughout any finite period following a sudden activation is unsteady, primarily because of discontinuities across the fundamental wavefronts (i.e., those associated with a stationary localized impulse) which have not yet reached infinity prior to a steady state.

The present paper focuses on such an unsteady development within a crystalline medium with unaligned permittivi-

 0022-2488/80/010195-16$01.00

ty and permeability principal axes. However, a fundamental compounded uniaxiality prevails in the sense that a certain compounded permittivity-permeability matrix possesses a double eigenvalue. Nonetheless, the uniaxiality does not generally occur about any of the six principal directions serving as those of two inclined ellipsoidal wavefronts. Furthermore, it is inevitably modified by the finite velocity of the current flow which is arbitrarily oriented. Unless this orientation is along the uniaxis, it normally disrupts the uniaxiality.

Besieris[1] has tackled the problem of a moving crystalline medium with aligned pairs of principal permittivity and permeability axes. Both anisotropies are uniaxial about a common principal direction. Along this direction, the medium moves uniformly, thus preserving the uniaxial symmetry. Radiation originates from an impulsive stationary point source in the form of a longitudinally as well as a transversely placed magnetic dipole. Besieris' investigation covers various speed ranges for the medium relative to the appropriate directional velocities of the fundamental wavefronts; these are spheroidal. Čerenkov-type phenomena can be found inside circular cones. In two preceding papers, Besieris[2] and Besieris and Compton[3] determined the Green's function for a moving isotropic conducting medium; other investigators who worked along similar lines include Lee and Papas,[4,5] Compton and Tai,[6] Tai,[7–9] Compton,[10] Chen and Yen,[11] Kalafus,[12] Johannsen,[13,14] Solimeno.[15]

An analysis of Čerenkov radiation within a crystal has also been made by Majumdar and Pal[16] via a Lorentz frame, relative to which the medium propagates past a stationary charged particle. Both sets of principal axes are again aligned. To achieve uniaxiality, two of the permittivity–permeability ratios were assumed equal. Crystalline properties therefore parallel those of Besieris.[1] In a subsequent paper,[17] Majumdar and Pal extended their Čerenkov analysis to the case of a biaxial crystal and obtained exact results when the medium travels along a principal direction. In both papers, the radiated energy was considered. This was dealt with in greater detail by Majumdar[18] who additionally formulated (i) a generalized uniaxiality criterion for aligned as well as unaligned permittivity and permeability principal axes and (ii) a theorem for deducing double anisotropy results from those for electric anisotropy only via direct substitutions. Sastry[19] also studied Čerenkov radiation within a doubly anisotropic crystal along corresponding lines, but more comprehensively. Results were derived for the radiation cones and energy loss in a generalized uniaxial medium (cf. Ref. 20).

In an earlier paper on the crystalline field equations, Lewandowski[21] established the same uniaxiality criterion as that of Majumdar and Pal.[16] Other aspects of the moving uniaxial medium already examined include wave reflection and transmission (Lee and Lo[22]), and field quantization (Kong[23]). Another paper of Kong[24] concentrates on an optical approach from a bianisotropic theory. From a more general but formal consideration of the moving medium, Johannsen[25] made a brief application to the uniaxial case (see also Handelsman and Lewis[26]). Finally, related papers on the uniaxial medium have also been published by Clemmow,[27] Felsen,[28,29] Mei[30] and Lu and Mei.[31]

2. COMPOUNDED UNIAXIALITY

Consider a crystalline medium with constant permittivity and permeability matrices $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$; these are real, symmetric and positive definite. Within this medium, radiation from an electric current of density $\mathbf{J}$ is governed by the Maxwell's equations

$$\nabla\times\mathbf{E}+\mathbf{H}_t\boldsymbol{\mu}=\mathbf{0},\quad \nabla\times\mathbf{H}-\mathbf{E}_t\boldsymbol{\epsilon}=\mathbf{J},$$

relative to the $\mathbf{x}=(x_1,x_2,x_3)$ frame; $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic field vectors. Following Chee-Seng,[32] we again select a singular source current that is suddenly switched on at instant $t=0$. It then shoots out from the point $\mathbf{x}=\mathbf{0}$ with a constant velocity $\mathbf{v}$ along an infinitesimally thin conductor which is arbitrarily oriented, say, along the positive ζ_1-direction of some fixed orthogonal frame $\boldsymbol{\zeta}=(\zeta_1,\zeta_2,\zeta_3)$. This is cooriginal with the $\mathbf{x}$ frame and is related to it by a rotational transformation $\boldsymbol{\zeta}=\mathbf{x}\mathbf{R}$ with modulus $\det\mathbf{R}=1$. Thus, in particular,

$$\mathbf{J}=\mathbf{v}|\mathbf{v}|^{-1}H(t)[H(\zeta_1)-H(\zeta_1-|\mathbf{v}|t)]\delta(\zeta_2)\delta(\zeta_3),\quad (2.1)$$

where H and δ denote, respectively, the Heaviside unit function and the Dirac delta function. Since the induced electric and magnetic fields cannot precede current activation, therefore

$$\mathbf{E}\equiv\mathbf{0}\equiv\mathbf{H}\quad \text{over } t<0. \quad (2.2)$$

According to Kline and Kay,[33] there exists a nonsingular real symmetric matrix $\boldsymbol{\mu}^{1/2}$ whereby $\boldsymbol{\mu}=\boldsymbol{\mu}^{1/2}\boldsymbol{\mu}^{1/2}$. Clearly, the inverse $(\boldsymbol{\mu}^{1/2})^{-1}=\boldsymbol{\mu}^{-1/2}$ is also symmetric. Furthermore, the compounded permittivity-permeability matrix

$$\boldsymbol{\tau}=\boldsymbol{\mu}^{-1/2}\boldsymbol{\epsilon}\boldsymbol{\mu}^{-1/2} \quad (2.3)$$

is symmetric and positive definite, and so possesses only positive eigenvalues $\lambda_1^{-2},\lambda_2^{-2},\lambda_3^{-2}$, say. There is no loss of generality in assuming the diagonal form

$$\boldsymbol{\tau}=\begin{pmatrix}\lambda_1^{-2} & 0 & 0\\ 0 & \lambda_2^{-2} & 0\\ 0 & 0 & \lambda_3^{-2}\end{pmatrix}, \quad (2.4)$$

since this is attainable with an appropriate orientation of coordinate axes. However, $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ are generally nondiagonal. Corresponding permittivity and permeability principal axes are normally unaligned.

The previous paper[32] deals with the derivation of certain triple integrals and their reductions to surface integrals over the unit sphere. That paper covers a situation where all three eigenvalues of the compounded matrix $\boldsymbol{\tau}$ are distinct, but does allow the possibility of there being two coincidental eigenvalues, viz. $\lambda_1\geqslant\lambda_2>\lambda_3$. Coincidence of two such eigenvalues corresponds to a *compounded uniaxiality*. A comprehensive study of this compounded uniaxial case forms the objective of the present paper which will employ as a basis the triple integrals [refer ahead to (2.18)] obtained in the earlier paper.

Let us define the vectors[32]

$$\mathbf{y} = (y_1, y_2, y_3) = (\det\boldsymbol{\mu})^{1/2}\mathbf{x}\boldsymbol{\mu}^{-1/2}, \tag{2.5}$$

$$\mathbf{w} = (w_1, w_2, w_3) = (\det\boldsymbol{\mu})^{1/2}\mathbf{v}\boldsymbol{\mu}^{-1/2}, \tag{2.6}$$

$$\mathbf{s} = (s_1, s_2, s_3) = \mathbf{y} - \mathbf{w}t, \tag{2.7}$$

and, with $\mathbf{I}$ denoting the identity matrix and T denoting the transpose, the following matrix and scalar operators

$$\mathbf{P}(\nabla_y) \equiv \boldsymbol{\tau}^{-1}(\nabla_y^T\nabla_y - \mathbf{I}\nabla_y\nabla_y^T), \tag{2.8}$$

$$\mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right) \equiv \mathbf{I}\frac{\partial^2}{\partial t^2} + \mathbf{P}(\nabla_y), \tag{2.9}$$

$$L\left(\frac{\partial}{\partial t}, \nabla_y\right) \equiv \frac{\partial^4}{\partial t^4} + \frac{\partial^2}{\partial t^2}\operatorname{tr}\mathbf{P}(\nabla_y) + \tfrac{1}{2}[\operatorname{tr}\mathbf{P}(\nabla_y)]^2 - \tfrac{1}{2}\operatorname{tr}\mathbf{P}^2(\nabla_y), \tag{2.10}$$

where $\nabla_y = (\partial/\partial y_1, \partial/\partial y_2, \partial/\partial y_3)$, the gradient operator in $\mathbf{y}$-space. The operators $\mathbf{Q}$ and L are related by

$$\det\mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right) = \frac{\partial^2}{\partial t^2} L\left(\frac{\partial}{\partial t}, \nabla_y\right).$$

Let

$$\sigma_{\pm}(\boldsymbol{\alpha}) = \{-\tfrac{1}{2}\operatorname{tr}\mathbf{P}(\boldsymbol{\alpha}) \pm \tfrac{1}{2}[2\operatorname{tr}\mathbf{P}^2(\boldsymbol{\alpha}) - (\operatorname{tr}\mathbf{P}(\boldsymbol{\alpha}))^2]^{1/2}\}^{1/2}, \tag{2.11}$$

where the real vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$. (Actually, $\sigma_+(\boldsymbol{\alpha})$, $-\sigma_+(\boldsymbol{\alpha})$, $\sigma_-(\boldsymbol{\alpha})$ and $-\sigma_-(\boldsymbol{\alpha})$ are the four σ zeros of $L(-\sigma, \boldsymbol{\alpha})$, an algebraic transform of the fourth-order L operator.) In terms of the eigenvalues of $\boldsymbol{\tau}$,

$$\begin{aligned}&-\operatorname{tr}\mathbf{P}(\boldsymbol{\alpha})\\ &= \lambda_1^2(\alpha_2^2 + \alpha_3^2) + \lambda_2^2(\alpha_3^2 + \alpha_1^2) + \lambda_3^2(\alpha_1^2 + \alpha_2^2)\\ &\quad > 0,\end{aligned} \tag{2.12}$$

while

$$\begin{aligned}&2\operatorname{tr}\mathbf{P}^2(\boldsymbol{\alpha}) - [\operatorname{tr}\mathbf{P}(\boldsymbol{\alpha})]^2\\ &\quad = [(|\alpha_1|\,|\lambda_2^2 - \lambda_3^2|^{1/2} - |\alpha_3|\,|\lambda_1^2 - \lambda_2^2|^{1/2})^2 + \alpha_2^2(\lambda_1^2 - \lambda_3^2)]\\ &\quad\times[(|\alpha_1|\,|\lambda_2^2 - \lambda_3^2|^{1/2} + |\alpha_3|\,|\lambda_1^2 - \lambda_2^2|^{1/2})^2 + \alpha_2^2(\lambda_1^2 - \lambda_3^2)].\end{aligned} \tag{2.13}$$

In the procedure adopted by Chee-Seng (Ref. 32, Secs. 1–4), the Maxwell's equations incorporating the source current of (2.1) are first combined into the single vector equation

$$\mathbf{E}_{tt}\boldsymbol{\epsilon} + \nabla\times[(\nabla\times\mathbf{E})\boldsymbol{\mu}^{-1}] = -\mathbf{J}_t = -\mathbf{v}H(t)\delta(\mathbf{x} - \mathbf{v}t),$$

which is then used to formulate the relationship

$$\mathbf{E}_{tt} = -(\det\boldsymbol{\mu})\mathbf{v}\boldsymbol{\mu}^{-1/2}\boldsymbol{\tau}^{-1} \times \operatorname{adj}^T\mathbf{Q}\left\{\frac{\partial}{\partial t}, (\det\boldsymbol{\mu})^{-1/2}\right\}\nabla\boldsymbol{\mu}^{1/2}\phi\boldsymbol{\mu}^{-1/2} \tag{2.14}$$

between the electric field $\mathbf{E}$ and a scalar field ϕ. The latter satisfies

$$L\left(\frac{\partial}{\partial t}, \nabla_y\right)\phi = H(t)\delta(\mathbf{y} - \mathbf{w}t) \tag{2.15}$$

and, in consistency with (2.2),

$$\phi \equiv 0 \quad \text{over } t < 0. \tag{2.16}$$

Fourier transformation of (2.15) and subsequent inversion for ϕ, with (2.16) accommodated, then produces

$$\phi = \sum_{\nu = \pm} \phi_\nu(\mathbf{y}, t) - \phi_\nu(\mathbf{s}, 0), \tag{2.17}$$

where

$$\phi_\nu(\mathbf{y}, t) = (2\pi)^{-3} \times \iiint_{R_3} \frac{\cos[\boldsymbol{\alpha}\cdot\mathbf{y} - \sigma_\nu(\boldsymbol{\alpha})t]\,d\boldsymbol{\alpha}}{\sigma_\nu(\boldsymbol{\alpha})[\sigma_\nu^2(\boldsymbol{\alpha}) - \sigma_{-\nu}^2(\boldsymbol{\alpha})][\sigma_\nu(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\cdot\mathbf{w}]}, \tag{2.18}$$

with $\sigma_{-\nu} = \sigma_{\mp}$ $(\nu = \pm)$; the triple integral with respect to $\boldsymbol{\alpha}$ ranges over the entire three dimensional space R_3. Formulas (2.17) and (2.18) are results from the preceding paper that together play a key role in the present analysis. They hold throughout $t > 0$. Henceforth, unless otherwise specified we assume that $t > 0$.

Hereafter, we adhere strictly to the compounded uniaxial system wherein

$$\lambda_1 = \lambda_2 = \lambda, \quad \text{say, with } \lambda > \lambda_3 > 0. \tag{2.19}$$

This corresponds to a fundamental wave configuration that is uniaxial about the y_3 axis which passes through the point $(0,0,1)\boldsymbol{\mu}^{1/2}$ in $\mathbf{x}$-space. Permittivity and permeability principal axes remain normally unaligned.

Incorporating (2.19), (2.11)–(2.13) simplify to give

$$\sigma_+(\boldsymbol{\alpha}) = \lambda\,|\boldsymbol{\alpha}|, \tag{2.20}$$

$$\sigma_-(\boldsymbol{\alpha}) = [\lambda_3^2(\alpha_1^2 + \alpha_2^2) + \lambda^2\alpha_3^2]^{1/2}. \tag{2.21}$$

Now, consider the integrand in (2.18). By (2.20) and (2.21), the factor

$$\sigma_\nu^2(\boldsymbol{\alpha}) - \sigma_{-\nu}^2(\boldsymbol{\alpha}) = \pm(\lambda^2 - \lambda_3^2)(\alpha_1^2 + \alpha_2^2) \quad (\nu = \pm).$$

Its vanishment admits two complex integrand singularities at $\alpha_1 = \pm i|\alpha_2|$ (or, alternatively, at $\alpha_2 = \pm i|\alpha_1|$). To avoid encountering both these singularities, we operate on ϕ_+ of (2.18) with the two-dimensional Laplacian

$$\nabla_y^{*^2} = \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial y_2^2} \equiv \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} \tag{2.22}$$

to get, via (2.20) and (2.21),

$$4\pi\lambda(\lambda^2 - \lambda_3^2)\nabla_y^{*^2}\phi_+(\mathbf{y}, t) = \Phi(\mathbf{y}, t\,|\,\mathbf{w}, \lambda), \tag{2.23}$$

where

$$\Phi(\mathbf{y}, t\,|\,\mathbf{w}, \lambda) = \frac{1}{2\pi^2}\iiint_{R_3} \frac{\cos(\boldsymbol{\alpha}\cdot\mathbf{y} - |\boldsymbol{\alpha}|\lambda t)}{|\boldsymbol{\alpha}|(\boldsymbol{\alpha}\cdot\mathbf{w} - |\boldsymbol{\alpha}|\lambda)}\,d\boldsymbol{\alpha}. \tag{2.24}$$

Clearly,

$$4\pi\lambda(\lambda^2 - \lambda_3^2)\nabla_y^{*^2}\phi_+(\mathbf{s}, 0) = \Phi(\mathbf{s}, 0\,|\,\mathbf{w}, \lambda). \tag{2.25}$$

Likewise,

$$4\pi\lambda(\lambda^2 - \lambda_3^2)\nabla_y^{*^2}\phi_-(\mathbf{y}, t) = \frac{-\lambda}{2\pi^2}\iiint_{R_3} \frac{\cos[\boldsymbol{\alpha}\cdot\mathbf{y} - \sigma_-(\boldsymbol{\alpha})t]}{\sigma_-(\boldsymbol{\alpha})[\boldsymbol{\alpha}\cdot\mathbf{w} - \sigma_-(\boldsymbol{\alpha})]}\,d\boldsymbol{\alpha}, \tag{2.26}$$

$$\equiv -\Phi(\bar{\mathbf{y}}, t\,|\,\bar{\mathbf{w}}, \lambda_3) \tag{2.27}$$

which follows from (2.26) by comparison with (2.24) after first noting the form (2.21), substituting the α_3 integration variable appropriately in (2.26), and then introducing new vectors

$$\bar{\mathbf{y}} = (y_1, y_2, \lambda^{-1}\lambda_3 y_3), \bar{\mathbf{w}} = (w_1, w_2, \lambda^{-1}\lambda_3 w_3). \tag{2.28}$$

Moreover, if $\bar{\mathbf{s}} = \bar{\mathbf{y}} - \bar{\mathbf{w}}t = (s_1, s_2, \lambda^{-1}\lambda_3 s_3)$, (2.29)

then, $4\pi\lambda(\lambda^2 - \lambda_3^2)\nabla_y^{*2}\phi_-(\mathbf{s},0) = -\Phi(\bar{\mathbf{s}},0|\bar{\mathbf{w}},\lambda_3)$; (2.30)

whereupon, applying (2.23), (2.25), (2.27), and (2.30) to (2.17) and defining

$$\Delta = 4\pi\lambda(\lambda^2 - \lambda_3^2)\nabla_y^{*2}\phi, \tag{2.31}$$

$$\Psi \equiv \Psi(\mathbf{y},t|\mathbf{w},\lambda) = \Phi(\mathbf{y},t|\mathbf{w},\lambda) - \Phi(\mathbf{s},0|\mathbf{w},\lambda), \tag{2.32}$$

$$\bar{\Psi} \equiv \Psi(\bar{\mathbf{y}},t|\bar{\mathbf{w}},\lambda_3), \tag{2.33}$$

we obtain $\Delta \equiv \Delta(\mathbf{y},t) = \Psi - \bar{\Psi}$, (2.34)

a net scalar field whose convolution with a logarithmic Poisson kernel provides the inverse to (2.31), viz.,

$$\phi = \frac{1}{16\pi^2\lambda(\lambda^2 - \lambda_3^2)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Delta(\zeta_1,\zeta_2,y_3,t) \times \ln[(y_1 - \zeta_1)^2 + (y_2 - \zeta_2)^2] d\zeta_1 d\zeta_2. \tag{2.35}$$

This together with (2.14) serve as a formal representation for the vector field $\mathbf{E}_{tt}$.

3. THE FUNCTIONS Φ and Ψ

To determine the Δ solution explicitly, we need the value of the Φ function represented by (2.24). To evaluate the triple integral involved, it will be expedient to use another reference frame. For this purpose, we first perform a positive rotational transformation on the integration variable $\boldsymbol{\alpha}$ to get

$$\boldsymbol{\kappa} = \boldsymbol{\alpha}\mathbf{Z}. \tag{3.1}$$

Regarding the matrix of rotation, one has $\mathbf{Z}^{-1} = \mathbf{Z}^T$, while the Jacobian equals

$$\frac{\partial\boldsymbol{\alpha}}{\partial\boldsymbol{\kappa}} = \det\mathbf{Z} \equiv 1. \tag{3.2}$$

Note that $|\boldsymbol{\alpha}\mathbf{Z}| = |\boldsymbol{\alpha}|$ and $\boldsymbol{\alpha}\cdot\mathbf{y} = \boldsymbol{\kappa}\cdot(\mathbf{y}\mathbf{Z})$. Thus, (2.24) becomes

$$\Phi(\mathbf{y},t|\mathbf{w},\lambda) = \frac{1}{2\pi^2}\iiint_{R_3} \frac{\cos[\boldsymbol{\kappa}\cdot(\mathbf{y}\mathbf{Z}) - |\boldsymbol{\kappa}|\lambda t]}{|\boldsymbol{\kappa}|[\boldsymbol{\kappa}\cdot(\mathbf{w}\mathbf{Z}) - |\boldsymbol{\kappa}|\lambda]} d\boldsymbol{\kappa}, \tag{3.3}$$

$$= \frac{1}{2\pi}\iint_{\Omega_3} \frac{\delta\{\hat{\boldsymbol{\kappa}}\cdot(\mathbf{y}\mathbf{Z}) - \lambda t\}}{\hat{\boldsymbol{\kappa}}\cdot(\mathbf{w}\mathbf{Z}) - \lambda} d\Omega, \tag{3.4}$$

which ranges with the unit position $\hat{\boldsymbol{\kappa}} = \boldsymbol{\kappa}|\boldsymbol{\kappa}|^{-1}$ over the three dimensional unit spherical surface Ω_3 with element $d\Omega$; (3.4) follows from (3.3) via $d\boldsymbol{\kappa} = \boldsymbol{\kappa}^2 d|\boldsymbol{\kappa}|d\Omega$ and the rule[34]

$$\int_0^\infty \cos(\alpha X)d\alpha = \pi\delta(X). \tag{3.5}$$

Suppose the matrix $\mathbf{Z}$ is chosen to lead to a coordinate frame $\boldsymbol{\kappa} = (\kappa_1,\kappa_2,\kappa_3)$ whose κ_2 axis is coplanar with the (original) vectors $\mathbf{y}$ and $\mathbf{w}$, and whose κ_3 axis is aligned with the vector $\mathbf{y}$. Then, relative to this $\boldsymbol{\kappa}$ frame,

$$\mathbf{y}\mathbf{Z} = (0,0,|\mathbf{y}|), \quad \mathbf{w}\mathbf{Z} = |\mathbf{w}|(0,\sin\beta,\cos\beta), \tag{3.6}$$

β being the angle between $\mathbf{y}\mathbf{Z}$ and $\mathbf{w}\mathbf{Z}$. Furthermore, in terms of a colatitude $\theta\in[0,\pi]$ and an azimuthal angle $\psi\in[0,2\pi)$:

$\hat{\kappa} = (\sin\theta\cos\psi, \sin\theta\sin\psi, \cos\theta)$, $d\Omega = \sin\theta d\theta d\psi$.

Hence, from (3.4),

$$\Phi(\mathbf{y},t|\mathbf{w},\lambda) = \int_0^\pi I(\theta)\delta(|\mathbf{y}|\cos\theta - \lambda t)\sin\theta d\theta, \tag{3.7}$$

where

$$I(\theta) = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\psi}{\hat{\boldsymbol{\kappa}}\cdot(\mathbf{w}\mathbf{Z}) - \lambda} = \frac{1}{2\pi i}\oint_{\mathscr{L}} \frac{dz}{D(z)}, \tag{3.8}$$

with

$$D(z) = z^2(2i)^{-1}|\mathbf{w}|\sin\beta\sin\theta + z(|\mathbf{w}|\cos\beta\cos\theta - \lambda) - (2i)^{-1}|\mathbf{w}|\sin\beta\sin\theta. \tag{3.9}$$

In (3.8), $\oint_{\mathscr{L}}$ denotes an integral over a closed anticlockwise-described unit circle $\mathscr{L}: z = \exp(i\psi)$ $(0 \leqslant \psi < 2\pi)$.

Suppose

$$|\mathbf{w}|\sin\beta\sin\theta \neq 0. \tag{3.10}$$

Then $D(z)$ can be factorized as

$$D(z) = (2i)^{-1}|\mathbf{w}|\sin\beta\sin\theta(z - z_+)(z - z_-), \tag{3.11}$$

where

$$z_\pm = \{|\mathbf{w}|\cos\beta\cos\theta - \lambda \pm [(|\mathbf{w}|\cos\beta\cos\theta - \lambda)^2 - \mathbf{w}^2\sin^2\beta\sin^2\theta]^{1/2}\}/i|\mathbf{w}|\sin\beta\sin\theta \tag{3.12}$$

is a pole of the z integrand $[D(z)]^{-1}$ in (3.8). Observe that

$$|z_+||z_-| = 1. \tag{3.13}$$

Case 1:

$$(|\mathbf{w}|\cos\beta\cos\theta - \lambda)^2 < \mathbf{w}^2\sin^2\beta\sin^2\theta; \tag{3.14}$$

here, $z_+ \neq z_-$ but $|z_+| = |z_-| = 1$, so that $z_+(\neq z_-)\in\mathscr{L}$ and $z_-\in\mathscr{L}$; therefore, in the sense of a Cauchy principal value,

$$I(\theta) = \tfrac{1}{2}\underset{z=z_-}{\text{residue}}[D(z)]^{-1} + \tfrac{1}{2}\underset{z=z_-}{\text{residue}}[D(z)]^{-1} \equiv 0, \tag{3.15}$$

since

$$\underset{z=z_\pm}{\text{residue}}[D(z)]^{-1} = \pm 2i(|\mathbf{w}|\sin\beta\sin\theta)^{-1}(z_+ - z_-)^{-1}. \tag{3.16}$$

Case 2:

$$(|\mathbf{w}|\cos\beta\cos\theta - \lambda)^2 > \mathbf{w}^2\sin^2\beta\sin^2\theta\,(>0); \tag{3.17}$$

here $z_\pm$ is purely imaginary and, moreover,

$$z_-\in\text{int}\mathscr{L}, \text{ but } z_+\in\text{ext}\mathscr{L} \text{ if } |\mathbf{w}|\cos\beta\cos\theta > \lambda, \tag{3.18}$$

$$z_+\in\text{int}\mathscr{L}, \text{ but } z_-\in\text{ext}\mathscr{L} \text{ if } |\mathbf{w}|\cos\beta\cos\theta < \lambda; \tag{3.19}$$

so (3.8) yields, via (3.16),

$$I(\theta) = 2i(|\mathbf{w}|\sin\beta\sin\theta)^{-1}(z_+ - z_-)^{-1} \times \text{sgn}(\lambda - |\mathbf{w}|\cos\beta\cos\theta). \tag{3.20}$$

Accounting for (3.12), (3.20) together with (3.15) imply

$$I(\theta) = \text{sgn}(|\mathbf{w}|\cos\beta\cos\theta - \lambda) \times \frac{H\{(|\mathbf{w}|\cos\beta\cos\theta - \lambda)^2 - \mathbf{w}^2\sin^2\beta\sin^2\theta\}}{\{(|\mathbf{w}|\cos\beta\cos\theta - \lambda)^2 - \mathbf{w}^2\sin^2\beta\sin^2\theta\}^{1/2}}, \tag{3.21}$$

pertaining to either (3.14) or (3.17). The situation

$$|\mathbf{w}|\cos\beta\cos\theta = \lambda \tag{3.22}$$

falls under (3.14), for which (3.15) holds and is clearly covered by the overall result (3.21).

Alternatively, suppose

$$|\mathbf{w}|\sin\beta\sin\theta = 0. \tag{3.23}$$

Then, provided

$$|\mathbf{w}|\cos\beta\cos\theta \neq \lambda, \tag{3.24}$$

it can be seen directly from (3.8) that

$$I(\theta) = (|\mathbf{w}|\cos\beta\cos\theta - \lambda)^{-1}, \tag{3.25}$$

which is actually the limiting form of (3.21) at attainment of (3.23). Therefore, our result (3.21) also covers the situation (3.23) accompanied by (3.24).

We now turn to (3.7). At $\mathbf{y} = \mathbf{0}$,

$$\Phi(\mathbf{0},t\,|\mathbf{w},\lambda) \equiv 0 \quad (t>0). \tag{3.26}$$

Otherwise, assuming $\mathbf{y}\neq\mathbf{0}$ tentatively,

$$\Phi(\mathbf{y},t\,|\mathbf{w},\lambda) = |\mathbf{y}|^{-1}\int_{-\infty}^{\infty}[I(\theta)]_{\cos\theta=\alpha}H(1-\alpha^2) \times\delta(\alpha - |\mathbf{y}|^{-1}\lambda t)d\alpha \tag{3.27}$$

$$= |\mathbf{y}|^{-1}H(\mathbf{y}^2-\lambda^2t^2)[I(\theta)]_{\cos\theta=|\mathbf{y}|^{-1}\lambda t}\,. \tag{3.28}$$

Now, from (3.6),

$$|\mathbf{w}|\cos\beta = (\hat{\mathbf{y}}\mathbf{Z})\cdot(\mathbf{w}\mathbf{Z}) = \hat{\mathbf{y}}\mathbf{Z}\mathbf{Z}^T\mathbf{w}^T = \hat{\mathbf{y}}\cdot\mathbf{w}, \tag{3.29}$$

in terms of the unit vector $\hat{\mathbf{y}} = \mathbf{y}|\mathbf{y}|^{-1}$. Whence

$$[(|\mathbf{w}|\cos\beta\cos\theta - \lambda)^2 - \mathbf{w}^2\sin^2\beta\sin^2\theta]_{\cos\theta=|\mathbf{y}|^{-1}\lambda t} = \chi|\mathbf{y}|^{-2}, \tag{3.30}$$

with

$$\chi \equiv \lambda^2(|\mathbf{y}| - \hat{\mathbf{y}}\cdot\mathbf{w}t)^2 - (\mathbf{y}^2-\lambda^2t^2)[\mathbf{w}^2 - (\hat{\mathbf{y}}\cdot\mathbf{w})^2], \tag{3.31}$$

$$\equiv \lambda^2(\mathbf{y}-\mathbf{w}t)^2 + (\mathbf{w}\cdot\mathbf{y})^2 - \mathbf{w}^2\mathbf{y}^2. \tag{3.32}$$

Consequently, (3.21) and (3.28) lead to

$$\Phi(\mathbf{y},t\,|\mathbf{w},\lambda) = \begin{cases} \chi^{-1/2}H(\chi)\operatorname{sgn}(\hat{\mathbf{y}}\cdot\mathbf{w}t - |\mathbf{y}|) & (\mathbf{y}^2>\lambda^2t^2), \quad (3.33)\\ 0 & (\mathbf{y}^2<\lambda^2t^2), \quad (3.34)\end{cases}$$

valid for

$$\chi \neq 0. \tag{3.35}$$

In view of (3.26), the range $\mathbf{y}^2<\lambda^2t^2$ for (3.34) may be point extended to include $\mathbf{y} = \mathbf{0}$.

The situation

$$|\mathbf{y}| = \hat{\mathbf{y}}\cdot\mathbf{w}t \tag{3.36}$$

is admissible throughout provided, in view of (3.31) and (3.35), that

$$\mathbf{y}^2\neq\lambda^2t^2, \quad \mathbf{w}\neq\mathbf{0} \quad \text{and} \quad \hat{\mathbf{y}}\neq\hat{\mathbf{w}}, \tag{3.37}$$

where $\hat{\mathbf{w}} = \mathbf{w}|\mathbf{w}|^{-1}$; in particular, we have $\chi<0$ when $\mathbf{y}^2>\lambda^2t^2$, in which case the associated result (3.33) vanishes identically corresponding to the effect (3.15) under (3.14) with (3.22); alternatively we have $\chi>0$ when $\mathbf{y}^2<\lambda^2t^2$, in which case the trivial result (3.34) holds. On the other hand, avoiding situation (3.36), one can allow

$$\mathbf{w} = \mathbf{0}(\mathbf{y}\neq\mathbf{0}) \quad \text{or} \quad \hat{\mathbf{y}} = \hat{\mathbf{w}}(\mathbf{y}\neq\mathbf{w}t), \tag{3.38}$$

corresponding to (3.23) under (3.24); nonetheless, it must again be understood that $\mathbf{y}^2\neq\lambda^2t^2$. Regarding (3.38), (2.6) and (2.14) indicate an absolute vanishment of the time derivative $\mathbf{E}_{tt}$ (but not necessarily of $\mathbf{E}$) when $\mathbf{w} = \mathbf{0}$; however $\phi\neq0$ and, according to (2.15), represents the scalar field of an abruptly activated stationary source.

Consider the expression (3.32). From (2.7) we deduce

$$(\mathbf{w}\cdot\mathbf{y})^2 - \mathbf{w}^2\mathbf{y}^2 = (\mathbf{w}\cdot\mathbf{s})^2 - \mathbf{w}^2\mathbf{s}^2. \tag{3.39}$$

So

$$\chi \equiv (\lambda^2 - \mathbf{w}^2)\mathbf{s}^2 + (\mathbf{w}\cdot\mathbf{s})^2, \tag{3.40}$$

whose dependence on $\mathbf{y}$ and t arises solely through their combination $\mathbf{s}$; (3.35) demands that $\mathbf{s}\neq\mathbf{0}$. By (3.33), then,

$$\Phi(\mathbf{s},0|\mathbf{w},\lambda) = -\chi^{-1/2}H(\chi). \tag{3.41}$$

Thus, applying (3.33), (3.34), (3.41) to (2.32), we arrive at

$$\Psi = \begin{cases} 2\chi^{-1/2}H(\chi)H(\hat{\mathbf{y}}\cdot\mathbf{w}t - |\mathbf{y}|) & (\mathbf{y}^2>\lambda^2t^2), \quad (3.42)\\ \chi^{-1/2}H(\chi) & (\mathbf{y}^2<\lambda^2t^2). \quad (3.43)\end{cases}$$

4. GEOMETRICAL INTERPRETATION IN THE ORIGINAL REFERENCE FRAME

All results should be recast in the original $\mathbf{x}$ frame. Only then can actual propagation phenomena be satisfactorily understood.

First we appeal to the following fact[33]: Since the real symmetric matrices $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ are positive definite, their respective eigenvalues ϵ_j, μ_j ($j = 1,2,3$) are all positive; furthermore, by the principal axes theorem, there exist orthogonal matrices $\mathbf{M}$, $\mathbf{N}$ whereby

$$\mathbf{M}^{-1}\boldsymbol{\epsilon}\mathbf{M} = \begin{pmatrix}\epsilon_1 & 0 & 0\\ 0 & \epsilon_2 & 0\\ 0 & 0 & \epsilon_3\end{pmatrix}, \quad \mathbf{N}^{-1}\boldsymbol{\mu}\mathbf{N} = \begin{pmatrix}\mu_1 & 0 & 0\\ 0 & \mu_2 & 0\\ 0 & 0 & \mu_3\end{pmatrix}. \tag{4.1}$$

Note that $\mathbf{M}^{-1} = \mathbf{M}^T$ and $\mathbf{N}^{-1} = \mathbf{N}^T$. The eigenvectors involved in the diagonalizations may be arranged to yield positive determinants viz. $\det\mathbf{M} = \det\mathbf{N}\equiv1$. On introducing the vectors

$$\mathbf{X} = \mathbf{x}\mathbf{N} = (X_1,X_2,X_3), \quad \mathbf{V} = \mathbf{v}\mathbf{N} = (V_1,V_2,V_3), \tag{4.2}$$

it can be verified from (2.5) and (2.6) that

$$\mathbf{y}^2 \equiv \mathbf{x}\boldsymbol{\mu}^{-1}\mathbf{x}^T\det\boldsymbol{\mu} \equiv (\det\boldsymbol{\mu})\sum_{j=1,2,3}\mu_j^{-1}X_j^2, \tag{4.3}$$

$$\mathbf{w}^2 \equiv \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu} \equiv (\det\boldsymbol{\mu})\sum_{j=1,2,3}\mu_j^{-1}V_j^2, \tag{4.4}$$

$$\mathbf{y}\cdot\mathbf{w} \equiv \mathbf{x}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu} \equiv (\det\boldsymbol{\mu})\sum_{j=1,2,3}\mu_j^{-1}X_jV_j. \tag{4.5}$$

Throughout, $\det\boldsymbol{\mu} = \mu_1\mu_2\mu_3 > 0$.

Let us define in $\mathbf{x}$ space the surface ξ:

$$\mathbf{x}\boldsymbol{\mu}^{-1}\mathbf{x}^T\det\boldsymbol{\mu} = \lambda^2t^2. \tag{4.6}$$

Now the right-handed $\mathbf{x}$ system is related to a right-handed $\mathbf{X}$ system by a rotation which preserves the shape and dimensions of any surface, e.g., that of ξ. By (4.3), (4.6) is equivalent to

$$(\det \boldsymbol{\mu}) \sum_{j=1,2,3} \mu_j^{-1} X_j^2 = \lambda^2 t^2; \tag{4.7}$$

i.e., ξ is an expanding ellipsoid centered at $\mathbf{x} = \mathbf{0}$. It represents a fundamental wavefront. Its principal (permeability) axes are the X_j axes ($j = 1,2,3$). Clearly,

$$\mathbf{y}^2 \gtreqless \lambda^2 t^2 \text{ if and only if } \mathbf{x} \in \begin{cases} \text{ext}\xi \text{ (exterior to } \xi) \\ \xi \\ \text{int}\,\xi \text{ (interior to } \xi). \end{cases} \tag{4.8}$$

Thus (3.42) holds if $\mathbf{x} \in \text{ext}\xi$, while (3.43) holds if $\mathbf{x} \in \text{int}\xi$. These results involve χ. Now, we can show from (3.32) that

$$\chi t^2 \equiv (\mathbf{y} \cdot \mathbf{w} t - \lambda^2 t^2)^2 - (\mathbf{w}^2 t^2 - \lambda^2 t^2)(\mathbf{y}^2 - \lambda^2 t^2); \tag{4.9}$$

$$\begin{aligned} &\equiv \Big[(\det \boldsymbol{\mu}) \sum_{j=1,2,3} \mu_j^{-1} X_j V_j t - \lambda^2 t^2 \Big]^2 \\ &\quad - \Big[(\det \boldsymbol{\mu}) \sum_{j=1,2,3} \mu_j^{-1} V_j^2 t^2 - \lambda^2 t^2 \Big] \\ &\quad \times \Big[(\det \boldsymbol{\mu}) \sum_{j=1,2,3} \mu_j^{-1} X_j^2 - \lambda^2 t^2 \Big], \end{aligned} \tag{4.10}$$

which follows from (4.9) via (4.3)–(4.5). Whence,

$$\chi = 0, \tag{4.11}$$

with χ expressed in the principal $\mathbf{X}$ frame by (4.10), is now recognized as the Joachimsthal's equation for the surface C which is tangential to the expanding ellipsoid ξ and is point constricted at the traveling current edge at $\mathbf{x} = \mathbf{v}t$, provided $\mathbf{v}t \in \text{ext}\xi$. If $\mathbf{v}t \in \text{int}\xi$, C never exists. Actually C comprises two convertexed cones C_+, C_-, one of which C_-, say, touches ξ. We refer to Fig. 1. The domain interior to C_- but exterior to ξ is separated by ξ into two portions $\mathscr{D}_+$ and $\mathscr{D}_-$, say:

$$\mathscr{D}_+ \cup \mathscr{D}_- = \text{int}C_- \cap \text{ext}\xi. \tag{4.12}$$

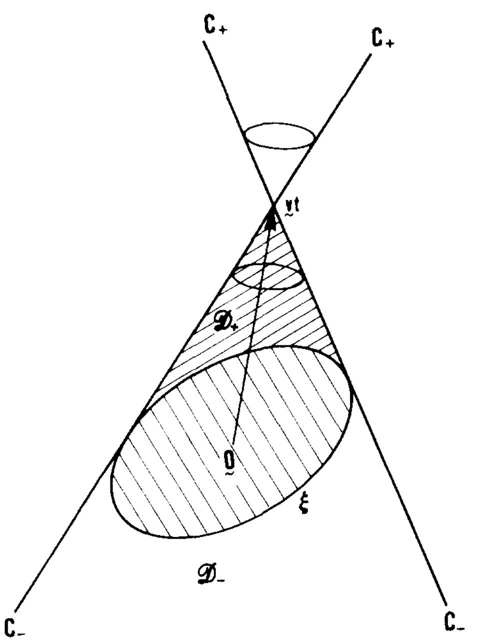


FIG. 1. *Case* $\mathbf{v}t \in \text{ext}\xi$: Generation of the tangent cone C_- and its complement C_+. Nontrivial values of Ψ are obtained within the darkened "ice-cream cone" domain $\mathscr{D}_+ \cup \text{int}\xi$. Elsewhere and off the boundaries ξ, C_+ and C_-: $\Psi \equiv 0$. *N.B.* The path traced by the source current over time t is represented by the vector $\mathbf{v}t$.

In particular, $\mathscr{D}_+$ is a finite conical region vertexed at $\mathbf{x} = \mathbf{v}t$ and adjacent to $\text{int}C_+$. The vector $\mathbf{v}t$ passes into $\mathscr{D}_+$ from within $\text{int}\xi$ and is directed away from the partially infinite complementary domain $\mathscr{D}_-$. In Sec. 5 we shall establish the following explicit results:

(I) If

$$\mathbf{v}t \in \text{ext}\xi, \tag{4.13}$$

$$\Psi = \begin{cases} \chi^{1/2}, & \forall \mathbf{x} \in \text{int}\xi, \quad (4.14) \\ 2\chi^{-1/2}, & \forall \mathbf{x} \in \mathscr{D}_+, \quad (4.15) \\ 0, & \forall \mathbf{x} \notin \mathscr{D}_+ \cup \text{int}\xi \cup \xi \cup C. \quad (4.16) \end{cases}$$

(II) If

$$\mathbf{v}t \in \text{int}\xi, \tag{4.17}$$

$$\Psi = \begin{cases} \chi^{-1/2} & \forall \mathbf{x} \in \text{int}\xi \text{ but } \neq \mathbf{v}t, \quad (4.18) \\ 0 & \forall \mathbf{x} \in \text{ext}\xi. \quad (4.19) \end{cases}$$

Note that the Ψ values expressed by (4.14)–(4.16), in the case (4.13), hold at reception points off the double cone C, i.e., as required by (3.35).

Defining the moving frame

$$\mathbf{r} = (r_1, r_2, r_3) = \mathbf{x} - \mathbf{v}t, \tag{4.20}$$

and applying (2.5)–(2.7) to (3.40), we have

$$\begin{aligned} \chi = &(\lambda^2 - \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T \det \boldsymbol{\mu}) \mathbf{r} \boldsymbol{\mu}^{-1} \mathbf{r}^T \det \boldsymbol{\mu} \\ &+ (\mathbf{r} \boldsymbol{\mu}^{-1} \mathbf{v}^T \det \boldsymbol{\mu})^2. \end{aligned} \tag{4.21}$$

Relative to the moving current edge at $\mathbf{r} = \mathbf{0}$, χ is obviously time independent. Our earlier requirement $\mathbf{s} \neq \mathbf{0}$ is now equivalent to $\mathbf{r} \neq \mathbf{0}$, a necessity for $\mathbf{x} \notin C$; it is also a condition of (4.18). We emphasize that, regarding both (4.14) and (4.18), a reception point at the initial current origin $\mathbf{x} = \mathbf{0}$ ($\in \text{int}\xi$) is acceptable unless $\mathbf{v} = \mathbf{0}$.

By (4.3)–(4.5), (3.36) is equivalently given by

$$(\det \boldsymbol{\mu}) \sum_{j=1,2,3} \mu_j^{-1} (X_j - \tfrac{1}{2} V_j t)^2 = \tfrac{1}{4} \mathbf{w}^2 t^2, \tag{4.22}$$

the equation of another ellipsoid ξ_0. Its center is the traveling point $\mathbf{x} = \frac{1}{2}\mathbf{v}t$ along the current path. Like ξ, ξ_0 expands, but possesses a fixed point at $\mathbf{x} = \mathbf{0}$. Additionally, it passes through the current edge at $\mathbf{x} = \mathbf{v}t$. The three principal axes of ξ_0 are parallel to those of ξ. By virtue of (3.31), (3.36), (4.8), and (4.11), if $\mathbf{v}t \in \text{ext}\xi$, ξ_0 intersects ξ along the latter's contacts with C_-. But if $\mathbf{v}t \in \text{int}\xi$, then $\xi_0 \cup \text{int}\xi_0 \subset \text{int}\xi$. According to (3.36) and (3.37), the reception point $\mathbf{x}$ may be taken along ξ_0 provided $\mathbf{x} \notin \xi \cap \xi_0$ when $\mathbf{v}t \in \text{ext}\xi$, $\mathbf{v} \neq \mathbf{0}$ and that $\mathbf{x}$ avoids the current edge at $\mathbf{v}t$, a consistency with the requirement $\mathbf{r} \neq \mathbf{0}$. Corresponding comments apply to (3.38).

5. ESTABLISHMENT OF THE GEOMETRICALLY EXPLICIT Ψ VALUES

The geometrically explicit expressions (4.13)–(4.19) must be derived from (3.42) and (3.43). The Heaviside functions $H(\chi)$ and $H(\hat{\mathbf{y}} \cdot \mathbf{w}t - |\mathbf{y}|)$ evidently play key roles.

Case I $\mathbf{v}t \in \text{ext}\xi$: Here, by virtue of (4.4) and (4.8),

$$\mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T \det \boldsymbol{\mu} > \lambda^2. \tag{5.1}$$

The geometrical configuration of Fig. 1 can be constructed. Now,

$$\mathbf{y}^2 - \lambda^2 t^2 \equiv \mathbf{s}^2 + 2\mathbf{s} \cdot \mathbf{w}t - (\lambda^2 - \mathbf{w}^2) t^2, \tag{5.2}$$

$$\equiv \mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{r}^T\det\boldsymbol{\mu} + 2\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{v}^T t \det\boldsymbol{\mu} - (\lambda^2 - \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu})t^2. \tag{5.3}$$

Consider, in $\mathbf{r}$ space, the point $\alpha\mathbf{r}$. If $\alpha\mathbf{r}\in\xi$, (4.8) and (5.3) imply that

$$\alpha^2\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{r}^T\det\boldsymbol{\mu} + 2\alpha\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{v}^T t \det\boldsymbol{\mu} - (\lambda^2 - \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu})t^2 = 0\,, \tag{5.4}$$

a quadratic equation in α with leading coefficient

$$\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{r}^T\det\boldsymbol{\mu} \equiv \mathbf{s}^2 > 0, \tag{5.5}$$

it being implicit that the reception point $\mathbf{r}\neq\mathbf{0}$. Both α roots to (5.4) are, on accommodating (4.21),

$$\alpha_\pm = t(-\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu} \pm \chi^{1/2})(\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{r}^T\det\boldsymbol{\mu})^{-1}. \tag{5.6}$$

For our proof, we consult Fig. 2. First, we select any $\mathbf{r}\in\text{ext}C_- \cap \text{ext}C_+$. Then the line passing through the $\mathbf{r}$ origin (at $\mathbf{x} = \mathbf{v}t$) and point $\mathbf{r}$ never intersects ξ, so that $\alpha_\pm$ must be complex. Therefore the real function

$$\chi < 0 \quad \forall\mathbf{x}\in\text{ext}C_- \cap \text{ext}C_+\,. \tag{5.7}$$

Furthermore, $\text{ext}C_- \cap \text{ext}C_+ \subset \text{ext}\xi$. Accounting for (4.8), then, (3.42) yields:

$$\Psi \equiv 0 \quad \forall\mathbf{x}\in\text{ext}C_- \cap \text{ext}C_+\,, \tag{5.8}$$

a partial verification of (4.16). On the other hand, for any $\mathbf{r}\in\text{int}C_- \cup \text{int}C_+$, such a line always intersects ξ at two separate $\mathbf{r}$ points, viz., $\alpha_+\mathbf{r}$ and $\alpha_-\mathbf{r}$. Hence $\alpha_\pm$ must be real and distinct, so that

$$\chi > 0 \quad \forall\mathbf{x}\in\text{int}C_- \cup \text{int}C_+\,. \tag{5.9}$$

Then from (4.21), (5.1) and (5.5),

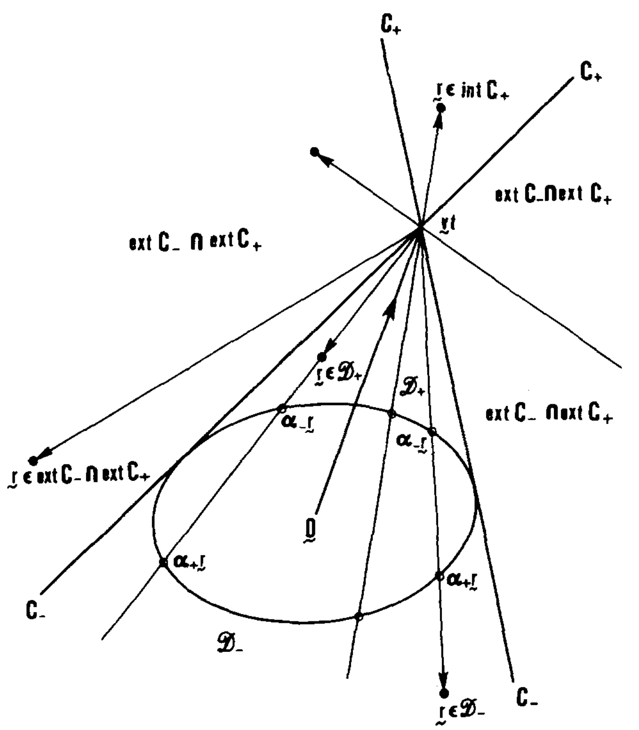


FIG. 2. *Case* $\mathbf{v}t\in\text{ext}\xi$: Geometric consideration for the $\alpha_\pm$ ranges initially indicates the following: (1) $\mathbf{r}\in\text{ext}C_- \cap \text{ext}C_+ : \alpha_\pm$ is complex, (2) $\mathbf{r}\in\mathscr{D}_+ : \alpha_\pm > 1$, (3) $\mathbf{r}\in\mathscr{D}_- : 0 < \alpha_\pm < 1$, (4) $\mathbf{r}\in\text{int}C_+ : \alpha_\pm < 0$.

$$\chi^{1/2} < |\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu}| \quad \forall\mathbf{x}\in\text{int}C_- \cup \text{int}C_+\,. \tag{5.10}$$

We note in passing that if $\mathbf{r}\in C$, then $\alpha_+ = \alpha_-$, confirming satisfaction of (4.11).

Now $\text{int}\xi \subset \text{int}C_-$. Whereupon, applying (5.9) and (4.8) to (3.43), the result (4.14) follows.

It now remains to turn our attention to the domains $\mathscr{D}_+$, $\mathscr{D}_-$ and $\text{int}C_+$, each of which is inside C_- or C_+ but outside ξ, wherein at any reception point $\mathbf{r}$, $\mathbf{y}\neq\mathbf{0}$ so that

$$H(\hat{\mathbf{y}}\cdot\mathbf{w}t - |\mathbf{y}|) \equiv H(-\mathbf{s}\cdot\mathbf{y}) \equiv H(-\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{x}^T\det\boldsymbol{\mu}). \tag{5.11}$$

Consider any $\mathbf{r}\in\mathscr{D}_+$. Fig. 2 reveals that the vectors $\mathbf{r}$ and $\alpha_\pm\mathbf{r}$ are in the same direction, i.e., $\alpha_\pm > 0$. So, necessarily, by (5.5), (5.6), and (5.10),

$$\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu} < 0. \tag{5.12}$$

But, as Fig. 2 further discloses, $|\alpha_\pm\mathbf{r}| > |\mathbf{r}|$. Hence $\alpha_+ > \alpha_- > 1$. In particular, $\alpha_- > 1$ implies

$$\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{x}^T\det\boldsymbol{\mu} < -\chi^{1/2}t < 0. \tag{5.13}$$

Thus, via (4.8), (5.9), (5.11), and (5.13), (3.42) simplifies into the form (4.15).

At any $\mathbf{r}\in\mathscr{D}_-$, inequality (5.12) holds for the same reason. However, Fig. 2 indicates that

$$1 > \alpha_+ > \alpha_- \; (>0). \tag{5.14}$$

Using $1 > \alpha_+$, we deduce from (5.6):

$$\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{x}^T\det\boldsymbol{\mu} > \chi^{1/2}t > 0 \quad \forall\mathbf{x}\in\mathscr{D}_-\,. \tag{5.15}$$

Alternatively, for any $\mathbf{r}\in\text{int}C_+$, the two vectors $\mathbf{r}$ and $\alpha_\pm\mathbf{r}$ are oppositely directed, i.e., $\alpha_\pm < 0$, so that by (5.6) and (5.10), we must have

$$\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{x}^T\det\boldsymbol{\mu} > 0, \tag{5.16}$$

which, combined with (5.5) implies

$$\mathbf{r}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu} > 0 \quad \forall\mathbf{x}\in\text{int}C_+\,. \tag{5.17}$$

Consequently, incorporating (4.8), (5.11), (5.15), and (5.17), (3.42) reduces to

$$\Psi \equiv 0 \quad \forall\mathbf{x}\in\mathscr{D}_- \cup \text{int}C_+\,, \tag{5.18}$$

which together with (5.8) implies (4.16).

Case II $\mathbf{v}t\in\text{int}\xi$: In view of (4.4) and (4.8),

$$\mathbf{w}^2 \equiv \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu} < \lambda^2. \tag{5.19}$$

So by (4.21) and (5.5),

$$\chi > 0 \quad \forall\mathbf{x}(\neq\mathbf{v}t). \tag{5.20}$$

Thus, via (4.8), the result (4.18) follows from (3.43). Finally, we choose any $\mathbf{r}\in\text{ext}\xi$; then

$$\hat{\mathbf{y}}\cdot\mathbf{w}t \leqslant |\mathbf{w}|t < \lambda t < |\mathbf{y}| \tag{5.21}$$

via (4.8) and (5.19). Whereupon (4.19) follows from (3.42). Our verification is now complete.

6. RELATIVE EVOLUTION

Let us consider the propagation of Ψ relative to the advancing current edge. Although χ is time independent, nonetheless (4.13)–(4.19) constitute an unsteady representation, primarily because of the transitions at finite instances across the expanding ellipsoidal wavefront ξ. The latter's direct involvement is a switch-on effect of the current source.

If, instead, the current is switched on during some infinite negative time, then at any finite positive time, no direct participation of ξ is physically detectable, at least within finite ranges, because ξ now exists at infinity. The study of Čerenkov radiation from a particle moving in an isotropic or crystalline medium is normally conducted under such a steady state environment. However, in his investigation into the moving isotropic medium problem, Compton[10] (see also Ref. 26) did discuss unsteady possibilities arising from a suddenly activated point source. From our present analysis, the steady state behavior can be extracted as a corollary from the evolution process.

The vanishing of expression (5.3) provides an equation for ξ within the $\mathbf{r}$ frame. Let $\hat{\mathbf{r}} = \mathbf{r}|\mathbf{r}|^{-1}$. We deduce via Fig. 2 that if $\mathbf{v}t \in \operatorname{ext}\xi$, then for any $\hat{\mathbf{r}}$ directed into $\operatorname{int}C_-$, ξ's equation must possess two real $|\mathbf{r}|$ roots representable by

$$|\mathbf{r}| = c_+(\hat{\mathbf{r}})t, \quad c_-(\hat{\mathbf{r}})t, \tag{6.1}$$

where $c_\pm(\hat{\mathbf{r}})$ are $(|\mathbf{r}|,t)$-independent positive c roots to

$$c^2\hat{\mathbf{r}}\boldsymbol{\mu}^{-1}\hat{\mathbf{r}}^T\det\boldsymbol{\mu} + 2c\hat{\mathbf{r}}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu} - (\lambda^2 - \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu}) = 0. \tag{6.2}$$

Likewise, it can be demonstrated that if $\mathbf{v}t \in \operatorname{int}\xi$, then for each $\hat{\mathbf{r}}$, there is only one real $|\mathbf{r}|$ root, say,

$$|\mathbf{r}| = c(\hat{\mathbf{r}})t, \tag{6.3}$$

$c(\hat{\mathbf{r}})$ being a positive c root to (6.2), to which $c(-\hat{\mathbf{r}})$ is, incidentally, also a positive c root.

Not only does ξ expand about the fixed origin $\mathbf{x} = \mathbf{0}$, but it also suffers, according to (6.1) and (6.3), a relative retreat from the current edge at $\mathbf{x} = \mathbf{v}t$, assuming $\mathbf{v} \neq \mathbf{0}$. In fact, along ξ any particular point which maintains a fixed $\hat{\mathbf{r}}$ direction retreats from $\mathbf{x} = \mathbf{v}t$ with a uniform relative velocity. Now both ends of the current, viz., the initial and subsequent positions of its edge at $\mathbf{x} = \mathbf{0}$ and $\mathbf{v}t$, serve as locations of two principal energy sources. The energy released is trapped partially ($\mathbf{v}t \in \operatorname{ext}\xi$) or fully ($\mathbf{v}t \in \operatorname{int}\xi$) inside ξ. The retreat of ξ from both principal energy sources confirms the Sommerfeld radiation principle that energy propagates away from any source. Our present analysis never incorporates this principle. Its satisfaction may be interpreted as an indirect outcome of our applied initial condition (2.2) or

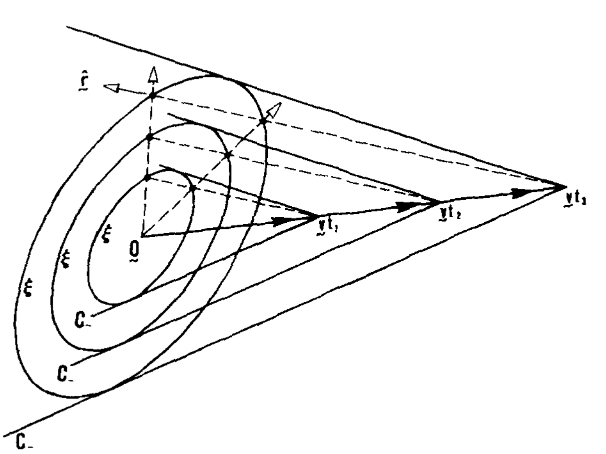


FIG. 3. *Case* $\mathbf{v}t \in \operatorname{ext}\xi$: Relative evolution over three consecutive times.

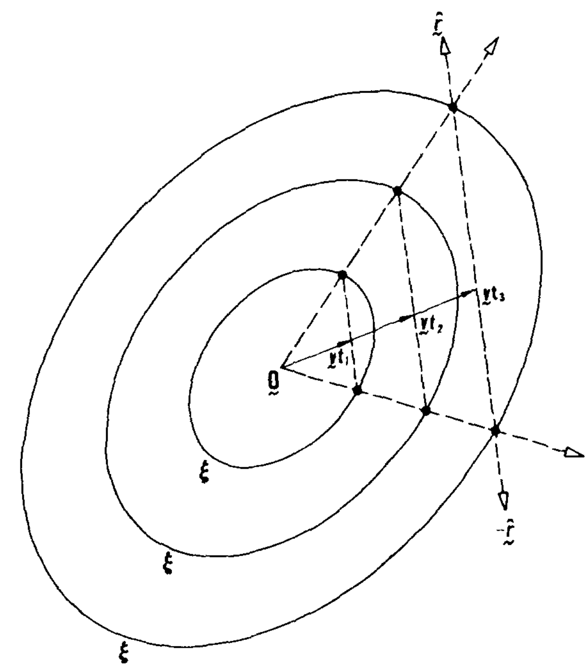


FIG. 4. Relative evolution when $\mathbf{v}t \in \operatorname{int}\xi$.

(2.16). Now, nontrivial Ψ values are proportional to $\chi^{-1/2}$ and hence to $|\mathbf{x} - \mathbf{v}t|^{-1}$ by virtue of (4.21). Thus at any fixed $\mathbf{x}$ point of reception, there is an attenuation with increasing time (unless $\mathbf{v} = \mathbf{0}$). But at any fixed $\mathbf{r}$ point for reception following the current edge, there is no attenuation. A possible explanation for the attenuation is that, although the current keeps flowing, the principal source at $\mathbf{x} = \mathbf{0}$ acts impulsively with current activation; it then immediately ceases transmitting so that its initial energy flux is never sustained. On the other hand the principal source at $\mathbf{x} = \mathbf{v}t$ does maintain a continuous energy supply to the extent of preserving permanently the nontrivial Ψ field at any fixed $\mathbf{r}$ position. According to (4.21), $\chi = 0$ $(\mathbf{r} = \mathbf{0})$, $\lambda^2 t^2 \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu}$ $(\mathbf{x} = \mathbf{0})$. Consequently, Ψ is singular at $\mathbf{r} = \mathbf{0}$ but not at $\mathbf{x} = \mathbf{0}$ (for $\mathbf{v} \neq \mathbf{0}$). This reinforces the proposition that the principal source at $\mathbf{x} = \mathbf{v}t$, but not that at $\mathbf{x} = \mathbf{0}$, stays active throughout $t > 0$.

Figure 3 portrays the evolution scheme for three consecutive periods $t = t_1, t_2, t_3$ over which $\mathbf{v}t \in \operatorname{ext}\xi$. The paths relative to $\mathbf{x} = \mathbf{0}$ and $\mathbf{r} = \mathbf{0}$ for two typical positions of ξ with $|\mathbf{r}|$-distances given by (6.1) are clearly traced (broken lines). Both these positions diverge as ξ expands. Meanwhile the gaps between them and the current edge increase. The cones C_+ and C_- are, in a sense, translated along with their common vertex at $\mathbf{x} = \mathbf{v}t$. Simultaneously, they expand with ξ but maintain a fixed solid angle at the vertex, relative to which, therefore, they do not appear to vary. This process develops uninterruptedly. Ultimately, when $t = \infty$, ξ is at infinity in relation to the two infinitely separated origins $\mathbf{x} = \mathbf{0}$ and $\mathbf{r} = \mathbf{0}$; so is the domain $\mathscr{D}_-$; however, the domain $\mathscr{D}_+$ virtually occupies the interior of C_-. Thereupon, a steady state prevails with solution

$$\Psi = \begin{cases} 2\chi^{-1/2} & \text{at any finite } \mathbf{r} \in \operatorname{int}C_-, \quad (6.4) \\ 0 & \text{at any } \mathbf{r} \notin C \cup \operatorname{int}C_-. \quad (6.5) \end{cases}$$

In the steady state, (4.21) again represents χ which is involved in (6.4) and Eq. (4.11) for $C = C_+ \cup C_-$. Results (6.4) and (6.5) correspond to Čerenkov radiation from the moving current edge. Evidently, on approaching the steady state,

$$\Psi \to 0 \quad \text{at every finite } \mathbf{x} \text{ point.} \tag{6.6}$$

The case: $\mathbf{v}t \in \mathrm{int}\xi$ is depicted in Fig. 4, again over three consecutive times t_1, t_2, t_3. Broken lines indicate the paths, relative to $\mathbf{x} = \mathbf{0}$ and $\mathbf{r} = \mathbf{0}$, described by two $\mathbf{r}$ points of the type

$$\hat{\mathbf{r}}c(\hat{\mathbf{r}})t, \quad -\hat{\mathbf{r}}c(-\hat{\mathbf{r}})t, \tag{6.7}$$

with reference to (6.3). When $t = \infty$, the evolution also attains a steady state wherein the points $\mathbf{x} = \mathbf{0}$, $\mathbf{r} = \mathbf{0}$ and the ellipsoid ξ are infinitely apart from each other; moreover

$$\Psi = \chi^{-1/2} \quad \text{at every finite } \mathbf{r}(\neq \mathbf{0}). \tag{6.8}$$

At such a point, the infinite ellipsoid ξ is undetectable. The same remark applies with regard to (6.4) and (6.5).

7. THE COMPLEMENT $\bar{\Psi}$ AND THE Δ SOLUTION

The resultant expressed by (2.34) involves the complement $\bar{\Psi}$. From (2.32), (2.33), (3.42), and (3.43), we deduce

$$\bar{\Psi} = \begin{cases} 2\bar{\chi}^{-1/2} H(\bar{\chi}) H(\hat{\bar{\mathbf{y}}}\cdot\bar{\mathbf{w}}t - |\bar{\mathbf{y}}|) & (\bar{\mathbf{y}}^2 > \lambda_3^2 t^2), \quad (7.1) \\ \bar{\chi}^{-1/2} H(\bar{\chi}) & (\bar{\mathbf{y}}^2 < \lambda_3^2 t^2), \quad (7.2) \end{cases}$$

where $\hat{\bar{\mathbf{y}}} = \bar{\mathbf{y}}|\bar{\mathbf{y}}|^{-1}$ and, via (3.40) and (4.9),

$$\bar{\chi} \equiv (\lambda_3^2 - \bar{\mathbf{w}}^2)\bar{\mathbf{s}}^2 + (\bar{\mathbf{w}}\cdot\bar{\mathbf{s}})^2, \tag{7.3}$$

$$\equiv t^{-2}(\bar{\mathbf{y}}\cdot\bar{\mathbf{w}}t - \lambda_3^2 t^2)^2 - (\bar{\mathbf{w}}^2 - \lambda_3^2)(\bar{\mathbf{y}}^2 - \lambda_3^2 t^2). \tag{7.4}$$

The overall validity reception criterion is, in view of (3.35),

$$\bar{\chi} \neq 0. \tag{7.5}$$

Again an effective interpretation would be a geometrical one given in the original $\mathbf{x}$ or $\mathbf{r}$ frame. This should, in particular, provide a clear contrast with our geometrically based Ψ values. For this purpose, we first write the symmetric matrix $\boldsymbol{\mu}^{-1/2}$ in the form

$$\boldsymbol{\mu}^{-1/2} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix} = (\mathbf{a}_1^T, \mathbf{a}_2^T, \mathbf{a}_3^T), \tag{7.6}$$

$\mathbf{a}_j$ being the jth row (vector) of $\boldsymbol{\mu}^{-1/2}$. By (2.5), then,

$$\mathbf{y} = (\det\boldsymbol{\mu})^{1/2}(\mathbf{x}\mathbf{a}_1^T, \mathbf{x}\mathbf{a}_2^T, \mathbf{x}\mathbf{a}_3^T). \tag{7.7}$$

In view of (2.4), we can write $\boldsymbol{\tau} = \boldsymbol{\tau}^{1/2}\boldsymbol{\tau}^{1/2}$ with

$$\boldsymbol{\tau}^{1/2} = \begin{pmatrix} \lambda_1^{-1} & 0 & 0 \\ 0 & \lambda_2^{-1} & 0 \\ 0 & 0 & \lambda_3^{-1} \end{pmatrix} \quad (\lambda_1 = \lambda_2 = \lambda). \tag{7.8}$$

Thus,

$$\boldsymbol{\mu}^{-1/2}\boldsymbol{\tau}^{-1/2} = (\lambda\mathbf{a}_1^T, \lambda\mathbf{a}_2^T, \lambda_3\mathbf{a}_3^T). \tag{7.9}$$

Whence, via (2.28) and (7.7),

$$\bar{\mathbf{y}} = (\det\boldsymbol{\mu})^{1/2}(\mathbf{x}\mathbf{a}_1^T, \mathbf{x}\mathbf{a}_2^T, \lambda^{-1}\lambda_3\mathbf{x}\mathbf{a}_3^T) \tag{7.10}$$

$$= \lambda^{-1}(\det\boldsymbol{\mu})^{1/2}\mathbf{x}\boldsymbol{\mu}^{-1/2}\boldsymbol{\tau}^{-1/2}. \tag{7.11}$$

Likewise, starting from (2.6), we can show that

$$\bar{\mathbf{w}} = \lambda^{-1}(\det\boldsymbol{\mu})^{1/2}\mathbf{v}\boldsymbol{\mu}^{-1/2}\boldsymbol{\tau}^{-1/2}, \tag{7.12}$$

so that by (2.29),

$$\bar{\mathbf{s}} = \lambda^{-1}(\det\boldsymbol{\mu})^{1/2}\mathbf{r}\boldsymbol{\mu}^{-1/2}\boldsymbol{\tau}^{-1/2}. \tag{7.13}$$

From (2.3),

$$\boldsymbol{\epsilon}^{-1} = \boldsymbol{\mu}^{-1/2}\boldsymbol{\tau}^{-1}\boldsymbol{\mu}^{-1/2}. \tag{7.14}$$

We can then proceed to derive from (7.11), (7.12), (7.14), and (2.4):

$$\bar{\mathbf{y}}^2 \equiv \lambda_3^2\lambda^2\mathbf{x}\boldsymbol{\epsilon}^{-1}\mathbf{x}^T\det\boldsymbol{\epsilon} \equiv \lambda_3^2\lambda^2\det\boldsymbol{\epsilon}\sum_{j=1,2,3}\epsilon_j^{-1}\bar{X}_j^2, \tag{7.15}$$

$$\bar{\mathbf{w}}^2 \equiv \lambda_3^2\lambda^2\mathbf{v}\boldsymbol{\epsilon}^{-1}\mathbf{v}^T\det\boldsymbol{\epsilon} \equiv \lambda_3^2\lambda^2\det\boldsymbol{\epsilon}\sum_{j=1,2,3}\epsilon_j^{-1}\bar{V}_j^2, \tag{7.16}$$

$$\bar{\mathbf{y}}\cdot\bar{\mathbf{w}} \equiv \lambda_3^2\lambda^2\mathbf{x}\boldsymbol{\epsilon}^{-1}\mathbf{v}^T\det\boldsymbol{\epsilon} \equiv \lambda_3^2\lambda^2\det\boldsymbol{\epsilon}\sum_{j=1,2,3}\epsilon_j^{-1}\bar{X}_j\bar{V}_j, \tag{7.17}$$

where, accounting for (4.1),

$$\bar{\mathbf{X}} = \mathbf{x}\mathbf{M} = (\bar{X}_1, \bar{X}_2, \bar{X}_3), \quad \bar{\mathbf{V}} = \mathbf{v}\mathbf{M} = (\bar{V}_1, \bar{V}_2, \bar{V}_3). \tag{7.18}$$

Note that $\det\boldsymbol{\epsilon} = \epsilon_1\epsilon_2\epsilon_3 > 0$. The right-handed $\bar{\mathbf{X}}$ system is related to the $\mathbf{x}$ system by a rotation with orthogonal matrix $\mathbf{M}$, and therefore to the $\mathbf{X}$ system of (4.2) by a combined rotation with orthogonal matrix $\mathbf{N}^{-1}\mathbf{M}$.

Through (7.15), we see that

$$\mathbf{x}\boldsymbol{\epsilon}^{-1}\mathbf{x}^T\det\boldsymbol{\epsilon} \equiv (\det\boldsymbol{\epsilon})\sum_{j=1,2,3}\epsilon_j^{-1}\bar{X}_j^2 = \lambda^{-2}t^2 \tag{7.19}$$

is the equation for another expanding ellipsoid $\bar{\xi}$; its principal (permittivity) axes $\bar{X}_1, \bar{X}_2, \bar{X}_3$ are generally unaligned with the principal axes of ξ, which is cocentered with $\bar{\xi}$ at $\mathbf{x} = \mathbf{0}$. More precisely,

$$\bar{\mathbf{y}}^2 \gtreqless \lambda_3^2 t^2 \text{ if and only if } \mathbf{x} \in \begin{cases} \mathrm{ext}\bar{\xi} \\ \bar{\xi} \\ \mathrm{int}\bar{\xi} \end{cases}. \tag{7.20}$$

Furthermore, from (7.4) and (7.15)–(7.17), we have

$$\frac{\bar{\chi}t^2}{\lambda_3^4\lambda^4} \equiv \left[(\det\boldsymbol{\epsilon})\sum_{j=1,2,3}\epsilon_j^{-1}\bar{X}_j\bar{V}_j t - \lambda^{-2}t^2\right]^2 - \left[(\det\boldsymbol{\epsilon})\sum_{j=1,2,3}\epsilon_j^{-1}\bar{V}_j^2 t^2 - \lambda^{-2}t^2\right] \times \left[(\det\boldsymbol{\epsilon})\sum_{j=1,2,3}\epsilon_j^{-1}\bar{X}_j^2 - \lambda^{-2}t^2\right]. \tag{7.21}$$

So, if $\mathbf{v}t \in \mathrm{ext}\bar{\xi}$, then $\bar{\chi} = 0$ is another Joachimsthal's equation for the tangent surface $\bar{C}$ (to $\bar{\xi}$) comprising two cones $\bar{C}_+$ and $\bar{C}_-$ which are convertexed at $\mathbf{x} = \mathbf{v}t$, with $\bar{C}_-$, say, touching $\bar{\xi}$. Criterion (7.5) requires that each reception point $\mathbf{x} \notin \bar{C}$. Comparison of (7.1) and (7.2) with (3.42) and (3.43) reveals that a radiation pattern analogous to that for Ψ can be easily deduced via the following substitutions:

$$\xi \Rightarrow \bar{\xi}, \quad C \Rightarrow \bar{C}, \quad C_\pm \Rightarrow \bar{C}_\pm, \quad \chi \Rightarrow \bar{\chi}. \tag{7.22}$$

Suppose the finite conical domain $\bar{\mathscr{D}}_+$ and the partially infinite domain $\bar{\mathscr{D}}_-$ are defined via the same analogy: $\mathscr{D}_\pm \Rightarrow \bar{\mathscr{D}}_\pm$. If $\mathbf{v}t \in \mathrm{int}\bar{\xi}$, then analogous to the situation where $\mathbf{v}t \in \mathrm{int}\xi$, $\bar{C}$ and hence $\bar{C}_\pm$ and $\bar{\mathscr{D}}_\pm$ never exist. We can now assert:

$$\Psi \Rightarrow \bar{\Psi}. \tag{7.23}$$

Therefore, $\bar{\Psi}$ is analogously determined from (4.14)–(4.16) and with reference to Fig. 1 when $\mathbf{v}t\in \operatorname{ext}\bar{\xi}$, as well as from (4.18) and (4.19) when $\mathbf{v}t\in \operatorname{int}\bar{\xi}$. It depends on

$$\bar{\chi} \equiv \lambda_3^4\lambda^4[(\lambda^{-2} - \mathbf{v}\epsilon^{-1}\mathbf{v}^T\det\epsilon)\mathbf{r}\epsilon^{-1}\mathbf{r}^T\det\epsilon + (\mathbf{r}\epsilon^{-1}\mathbf{v}^T\det\epsilon)^2], \tag{7.24}$$

obtained from (7.3), (7.12)–(7.14) and (7.16). Hence, within the $\mathbf{r}$ frame, $\bar{\chi}$ is time independent like χ.

The combination (2.34) for the Δ solution is naturally influenced by the geometric disposition between ξ and $\bar{\xi}$. Some idea of the geometry is therefore crucial. Consider

$$\mathbf{n}(\xi) = \frac{\nabla(\mathbf{y}^2 - \lambda^2 t^2)}{|\nabla(\mathbf{y}^2 - \lambda^2 t^2)|} = \frac{\nabla\mathbf{y}^2}{|\nabla\mathbf{y}^2|},$$

$$\mathbf{n}(\bar{\xi}) = \frac{\nabla(\bar{\mathbf{y}}^2 - \lambda_3^2 t^2)}{|\nabla(\bar{\mathbf{y}}^2 - \lambda_3^2 t^2)|} = \frac{\nabla\bar{\mathbf{y}}^2}{|\nabla\bar{\mathbf{y}}^2|}, \tag{7.25}$$

the instantaneous unit normal vectors to ξ and $\bar{\xi}$ respectively, ∇ being the gradient operator in $\mathbf{x}$ space. But

$$\nabla y_j^2 = 2(\det\boldsymbol{\mu})^{1/2} y_j \nabla(\mathbf{x}\mathbf{a}_j^T) = 2y_j\mathbf{a}_j(\det\boldsymbol{\mu})^{1/2}. \tag{7.26}$$

So

$$\mathbf{n}(\xi) = \sum_{j=1,2,3} y_j\,\mathbf{a}_j \left|\sum_{j=1,2,3} y_j\,\mathbf{a}_j\right|^{-1}, \tag{7.27}$$

$$\mathbf{n}(\bar{\xi}) = \frac{y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \lambda^{-2}\lambda_3^2\,y_3\mathbf{a}_3}{|y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \lambda^{-2}\lambda_3^2\,y_3\mathbf{a}_3|}. \tag{7.28}$$

Next, we note that ξ meets $\bar{\xi}$ if and only if

$$\mathbf{y} = \pm(0,0,\lambda t). \tag{7.29}$$

Let $\mathbf{b}_j$ denote the jth row of

$$\boldsymbol{\mu}^{1/2} = \begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\end{pmatrix}. \tag{7.30}$$

Then, via (2.5), the encounter $\xi\cap\bar{\xi}$ occurs at two moving points $\mathbf{x} = \mathbf{x}_+, \mathbf{x}_-$ given by

$$\mathbf{x}_\pm = \pm\mathbf{b}_3\lambda t(\det\boldsymbol{\mu})^{-1/2}. \tag{7.31}$$

Moreover, from (7.27)–(7.29) we have

$$\mathbf{n}(\xi)|_{\mathbf{x}=\mathbf{x}_\pm} = \pm\mathbf{a}_3|\mathbf{a}_3|^{-1} = \mathbf{n}(\bar{\xi})|_{\mathbf{x}=\mathbf{x}_\pm}, \tag{7.32}$$

confirming that the ellipsoids ξ and $\bar{\xi}$ meet tangentially at the points $\mathbf{x}_+$ and $\mathbf{x}_-$. Evidently, one ellipsoid is enclosed by the other. To be more precise, we consider any $\mathbf{x}\in\operatorname{int}\bar{\xi}$. Then by (7.20),

$$\lambda^2 t^2 > \lambda_3^{-2}\lambda^2(y_1^2 + y_2^2) + y_3^2 \geqslant \mathbf{y}^2,$$

so that $\mathbf{x}\in\operatorname{int}\xi$. As $\operatorname{int}\xi \neq \operatorname{int}\bar{\xi}$, therefore

$$\operatorname{int}\bar{\xi}\subset\operatorname{int}\xi. \tag{7.33}$$

Observe that the tangential contacts between ξ and $\bar{\xi}$ occur along the uniaxis (Sec. 2) which is obviously parallel to $\mathbf{b}_3$.

We are now ready to formulate explicitly the Δ function of (2.34) with the aid of (4.13)–(4.19) accompanied by Fig. 1, plus (7.22), (7.23), and (7.33).

The simplest case is that for a *slow current*: $\mathbf{v}t\in\operatorname{int}\bar{\xi}$; here

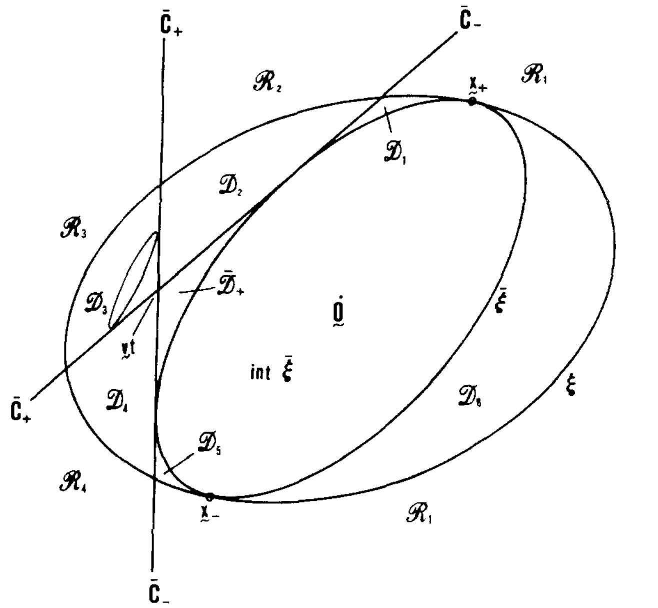


FIG. 5. Intermediate current–velocity regime: $\mathbf{v}t\in \operatorname{int}\xi\cap\operatorname{ext}\bar{\xi}$. The various subdomains of $\operatorname{int}\xi$ normally support nontrivial representations of Δ. Trivial values occur over the four subdomains $\mathscr{R}_1, \mathscr{R}_2, \mathscr{R}_3, \mathscr{R}_4$ of $\operatorname{ext}\xi$.

$$\Delta = \begin{cases} \chi^{-1/2} - \bar{\chi}^{-1/2} & \text{inside } \bar{\xi}\text{: } \mathbf{x}\neq\mathbf{v}t, \quad (7.34)\\ \chi^{-1/2} & \text{between } \xi \text{ and } \bar{\xi}, \quad (7.35)\\ 0 & \text{outside } \xi. \quad (7.36)\end{cases}$$

Next, suppose the current flows within an *intermediate velocity regime*: $\mathbf{v}t$ lies between ξ and $\bar{\xi}$. Then a single double-conical tangent surface, viz., $\bar{C} = \bar{C}_+\cup\bar{C}_-$, to $\bar{\xi}$ is point constricted at $\mathbf{x} = \mathbf{v}t$. Let us refer to Fig. 5. The domain $\operatorname{int}\xi$ is normally partitioned by $\bar{\xi}$ and $\bar{C}$ into eight subdomains, viz. $\operatorname{int}\bar{\xi}$ and $\bar{\mathscr{D}}_+$ (both already defined) together with $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_6$. The domain $\operatorname{ext}\xi$ is partitioned by $\bar{C}_+$ and $\bar{C}_-$ into four partially infinite portions $\mathscr{R}_1, \mathscr{R}_2, \mathscr{R}_3, \mathscr{R}_4$. Generally, then,

$$\Delta = \begin{cases} \chi^{-1/2} - 2\bar{\chi}^{-1/2} \text{ over } \bar{\mathscr{D}}_+, & (7.37)\\ \chi^{-1/2} - \bar{\chi}^{-1/2} \text{ inside } \bar{\xi}, & (7.38)\\ \chi^{-1/2} \quad \text{over } \mathscr{D}_1, \mathscr{D}_2, \ldots \mathscr{D}_6, & (7.39)\\ 0 \quad \text{over } \mathscr{R}_1, \mathscr{R}_2, \mathscr{R}_3, \mathscr{R}_4. & (7.40)\end{cases}$$

Finally, we examine the case for a *fast current*: $\mathbf{v}t\in\operatorname{ext}\xi$. Here, two double-conical tangent surfaces, viz., $C = C_+\cup C_-$ to ξ and $\bar{C} = \bar{C}_+\cup\bar{C}_-$ to $\bar{\xi}$, are coincidentally point constricted at $\mathbf{x} = \mathbf{v}t$ (see Fig. 6). We ignore degenerate situations wherein adjacent $C_-\cap\xi$ and $\bar{C}_-\cap\bar{\xi}$ contacts overlap the $\xi\cap\bar{\xi}$ contact at $\mathbf{x}_+$ or $\mathbf{x}_-$. Then $\bar{C}_-$ partitions the finite conical domain $\mathscr{D}_+$ (represented in Fig. 1) into three segments $\mathscr{D}^+, \mathscr{D}^0, \mathscr{D}^-$. Also, $\bar{\xi}$ and $\bar{C}_-$ divide $\operatorname{int}\xi$ into $\operatorname{int}\bar{\xi}$, subdomains $\mathscr{D}_1, \mathscr{D}_2, \ldots\mathscr{D}_6$, while the remaining space outside ξ is divided by C and $\bar{C}$ into $\mathscr{R}_1, \mathscr{R}_2, \ldots\mathscr{R}_8$. Whereupon,

$$\Delta = \begin{cases} \chi^{-1/2} - \bar{\chi}^{-1/2} & \text{inside } \bar{\xi}, \quad (7.41)\\ 2\chi^{-1/2} - 2\bar{\chi}^{-1/2} & \text{over } \mathscr{D}^0 \quad (7.42)\\ 2\chi^{-1/2} & \text{over } \mathscr{D}^+, \mathscr{D}^-, \quad (7.43)\\ \chi^{-1/2} - 2\bar{\chi}^{-1/2} & \text{over } \mathscr{D}_1, \mathscr{D}_2, \quad (7.44)\\ \chi^{-1/2} & \text{over } \mathscr{D}_3, \mathscr{D}_4, \mathscr{D}_5, \mathscr{D}_6, \quad (7.45)\\ 0 & \text{over } \mathscr{R}_1, \mathscr{R}_2, \ldots, \mathscr{R}_8. \quad (7.46)\end{cases}$$

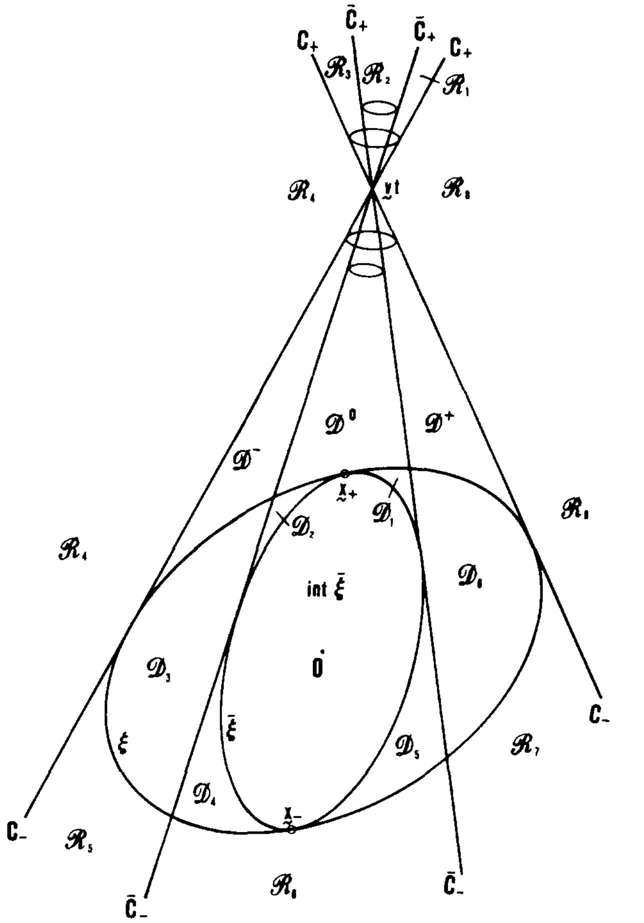


FIG. 6. Fast current: $\mathbf{v}t\in \text{ext}\xi$. Nontrivial representations of Δ are found within the various subdomains of $\text{int}\xi\cup\mathscr{D}_+$, viz., the darkened region (Fig. 1) of nontrivial Ψ. Elsewhere, precisely over the subdivisions $\mathscr{R}_1, \mathscr{R}_2,\dots,\mathscr{R}_8$: $\Delta\equiv 0$.

As $t\rightarrow\infty$, ξ and $\bar{\xi}$ progress towards infinity relative to both $\mathbf{x}=\mathbf{0}$ and $\mathbf{r}=\mathbf{0}$. Amongst the results (7.34)–(7.36) for the slow current, only (7.34) remains appropriate. Thus, during the ultimate steady state,

$$\Delta=\chi^{-1/2}-\bar{\chi}^{-1/2}\quad\text{at any finite }\mathbf{r}(\neq\mathbf{0}),\tag{7.47}$$

whenever

$$\mathbf{v}\boldsymbol{\epsilon}^{-1}\mathbf{v}^T\det\boldsymbol{\epsilon}<\lambda^{-2},\tag{7.48}$$

an explicit t independent expression of slowness. For a current velocity within the intermediate regime, equivalently,

$$\mathbf{v}\boldsymbol{\epsilon}^{-1}\mathbf{v}^T\det\boldsymbol{\epsilon}>\lambda^{-2},\quad\text{but}\quad \mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu}<\lambda^{2},\tag{7.49}$$

we note in particular that the originally bounded conical domain $\overline{\mathscr{D}}_+$ (Fig. 5) grows into $\text{int}\bar{C}_-$ itself; in the limit (7.37)–(7.40) reduce to

$$\Delta=\begin{cases}\chi^{-1/2}-2\bar{\chi}^{-1/2} & \text{at any finite }\mathbf{r}\in\text{int}\bar{C}_-, \quad (7.50)\\ \chi^{-1/2} & \text{at any finite }\mathbf{r}\notin\bar{C}\cup\text{int}\bar{C}_-, \quad (7.51)\end{cases}$$

the complete steady state solution under (7.49). Finally, in the case of the fast current, i.e.,

$$\mathbf{v}\boldsymbol{\mu}^{-1}\mathbf{v}^T\det\boldsymbol{\mu}>\lambda^{2},\tag{7.52}$$

it is evident from Fig. 6 that

$$\mathscr{D}^0\rightarrow\text{int}\bar{C}_-\ \text{and}\ \mathscr{D}^+\cup\mathscr{D}^-\rightarrow\text{int}C_-\cap\text{ext}\bar{C}_-,\tag{7.53}$$

in the course of which $\mathscr{R}_5,\mathscr{R}_6,\mathscr{R}_7$ are repelled towards infinity by ξ, while $\mathscr{R}_1,\dots,\mathscr{R}_4$ and $\mathscr{R}_8$ remain intact; from (7.41)–(7.46), we deduce that in the steady state,

$$\Delta=\begin{cases}2\chi^{-1/2}-2\bar{\chi}^{-1/2} & \text{at any finite }\mathbf{r}\in\text{int}\bar{C}_-, \quad (7.54)\\ 2\chi^{-1/2} & \text{at any finite }\mathbf{r}\in\text{int}C_-\cap\text{ext}\bar{C}_-, \quad (7.55)\\ 0 & \text{at any }\mathbf{r}\in\mathscr{R}_1,\dots,\mathscr{R}_4,\mathscr{R}_8. \quad (7.56)\end{cases}$$

Now, in $\mathbf{y}$ space, ξ is a sphere while $\bar{\xi}$ is a prolate spheroid. Consider the fundamental propagation with $\mathbf{v}=\mathbf{0}$. [*NB* Although, by (2.14), $\mathbf{E}_{tt}=\mathbf{0}$ when $\mathbf{v}=\mathbf{0}$, the associated fundamental problem as posed by (2.15) with $\mathbf{w}=\mathbf{0}$ and (2.16) is, nonetheless, never a trivial one. In particular, its ϕ solution as well as the inducing Δ quantity, both related by (2.35), remain nontrivial.] The function $\chi\equiv\lambda^2\mathbf{y}^2$ is spherically symmetric about $\mathbf{y}=0$, while $\bar{\chi}\equiv\lambda_3^2\bar{\mathbf{y}}^2$, like the surface $\bar{\xi}$, is axisymmetric about the y_3 axis, i.e., the uniaxis. The net effect measured by the corresponding Δ solution of (7.34)–(7.36) is consequently axisymmetric about this uniaxis. It is within this context, and with reference to the $\mathbf{y}$ frame, that our terminology of compounded uniaxiality applies. Strictly speaking, unless the current velocity is oriented along the uniaxis, it disrupts such a fundamental uniaxiality, causing χ and hence Ψ to be axisymmetric about the $\mathbf{w}$ direction instead in $\mathbf{y}$ space. It additionally causes $\bar{\chi}$ and hence $\bar{\Psi}$ to be axisymmetric about the $\bar{\mathbf{w}}$ direction in $\bar{\mathbf{y}}$ space. However ξ and $\bar{\xi}$, being independent of current velocity, remain unchanged.

We shall next verify that the fundamental uniaxiality can be recovered by aligning the current flow along the uniaxis, viz.,

$$\mathbf{v}=\pm|\mathbf{v}|\,|\mathbf{b}_3|^{-1}\mathbf{b}_3=\pm|\mathbf{v}|\,|\mathbf{b}_3|^{-1}(0,0,1)\boldsymbol{\mu}^{1/2},\tag{7.57}$$

i.e., by (2.6) and (2.28),

$$\mathbf{w}=(0,0,w_3),\quad \bar{\mathbf{w}}=\lambda^{-1}\lambda_3(0,0,w_3),$$

with

$$w_3=\pm|\mathbf{v}|\,|\mathbf{b}_3|^{-1}(\det\boldsymbol{\mu})^{1/2}.\tag{7.58}$$

Hence from (3.40) and (7.3),

$$\begin{aligned}\chi&\equiv(\lambda^2-\mathbf{w}^2)(s_1^2+s_2^2)+\lambda^2 s_3^2\\ &\equiv(\lambda^2-\mathbf{w}^2)(y_1^2+y_2^2)+\lambda^2(y_3-w_3t)^2,\end{aligned}\tag{7.59}$$

$$\begin{aligned}\bar{\chi}&\equiv(\lambda_3^2-\bar{\mathbf{w}}^2)(s_1^2+s_2^2)+\lambda^{-2}\lambda_3^4 s_3^2\\ &\equiv(\lambda_3^2-\bar{\mathbf{w}}^2)(y_1^2+y_2^2)+\lambda^{-2}\lambda_3^4(y_3-w_3t)^2,\end{aligned}\tag{7.60}$$

which are obviously axisymmetric in $\mathbf{y}$ space about the y_3 axis. Thus we recover the fundamental uniaxiality. Note that χ and $\bar{\chi}$ are also axisymmetric about the s_3 axis in $\mathbf{s}$ space. As the current conductor obviously passes through a contact point between ξ and $\bar{\xi}$, only two current velocity regimes are admissible, viz., the slow and fast regimes, corresponding to which the respective solutions are given by (7.34)–(7.36) and (7.41)–(7.46).

We now turn temporarily to the situation wherein $\mathbf{M}=\mathbf{N}=\mathbf{I}$ in (4.1), i.e., $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ are diagonal (such a restriction being abandoned elsewhere throughout this paper). In particular, one can presently take

$$\mu^{1/2} = \begin{pmatrix} \mu_1^{1/2} & 0 & 0 \\ 0 & \mu_2^{1/2} & 0 \\ 0 & 0 & \mu_3^{1/2} \end{pmatrix}. \tag{7.61}$$

Whence, (2.3) and (2.4) yield $\lambda_\nu^{-2} = \epsilon_\nu \mu_\nu^{-1} (\nu = 1,2,3)$, so that the compounded uniaxiality criterion of (2.19) simplifies to

$$\epsilon_1 \epsilon_2^{-1} = \mu_1 \mu_2^{-1}, \tag{7.62}$$

which represents the criterion exploited by Besieris,[1] Majumdar and Pal,[16] Lewandowski.[21] From (4.2) and (7.18), $\mathbf{X} = \bar{\mathbf{X}} = \mathbf{x}$, $\mathbf{V} = \bar{\mathbf{V}} = \mathbf{v}$. According to (4.7) and (7.19),

$$(\det\mu) \sum_{j=1,2,3} \mu_j^{-1} x_j^2 = \lambda^2 t^2,$$

$$(\det\epsilon) \sum_{j=1,2,3} \epsilon_j^{-1} x_j^2 = \lambda^{-2} t^2 \tag{7.63}$$

are the new equations for ξ and $\bar{\xi}$; these now share common permittivity and permeability principal axes x_1, x_2 and x_3. The new uniaxis is the principal x_3 axis through the $\xi \cap \bar{\xi}$ contacts of (7.31) which now occur at $\mathbf{x}_\pm = \pm \mu_1^{-1/2} \times \epsilon_2^{-1/2} t\,(0,0,1)$. As in the general case, axisymmetry of the fundamental structure normally exists only in the $\mathbf{y}$ space. However, should (7.62) be satisfied by $\epsilon_1 = \epsilon_2$ and $\mu_1 = \mu_2$, this axisymmetry extends into the $\mathbf{x}$ space.

8. CONSTANT χ SURFACES

As we already know, when $\mathbf{v}t \in \mathrm{ext}\xi$, the function χ takes the constant zero along the tangent surface $C_+ \cup C_-$ and nowhere else. In the case: $\mathbf{v}t \in \mathrm{int}\xi$, however, χ never vanishes. A point of curiosity therefore arises. This concerns the possible existence of other surfaces of constant χ values, particularly, positive χ values. Such surfaces would then be symmetry surfaces of Ψ within the domains wherein $\Psi \not\equiv 0$ and is nonsingular. They need not, however, be surfaces of constant Ψ, e.g., due to the latter's discontinuity across ξ when $\mathbf{v}t \in \mathrm{ext}\xi$.

We need a certain canonical representation for χ. First we perform on the $\mathbf{s}$ frame of (2.7) a rotation with orthogonal matrix $\mathbf{K}$: $\mathbf{K}^T = \mathbf{K}^{-1}$ and $\det\mathbf{K} = 1$, until we arrive at the $\mathbf{s}^*$-frame:

$$\mathbf{s}^* = (s_1^*, s_2^*, s_3^*) = \mathbf{sK}. \tag{8.1}$$

The third column of $\mathbf{K}$ is chosen to be the vector $\mathbf{w}^T |\mathbf{w}|^{-1}$ with unit magnitude compatible with orthogonality. Then $s_3^* = \mathbf{s}\cdot\mathbf{w}|\mathbf{w}|^{-1}$; also $\mathbf{s}^{*2} = \mathbf{s}^2$. Whereupon, defining the diagonal matrix

$$\mathbf{H} = \begin{pmatrix} \lambda^2 - \mathbf{w}^2 & 0 & 0 \\ 0 & \lambda^2 - \mathbf{w}^2 & 0 \\ 0 & 0 & \lambda^2 \end{pmatrix}, \tag{8.2}$$

(3.40) leads to

$$\chi \equiv \mathbf{s}^* \mathbf{H} \mathbf{s}^{*T} \equiv \mathbf{r} \mathbf{G} \mathbf{r}^T \det\mu, \tag{8.3}$$

where

$$\mathbf{G} \equiv \mu^{-1/2} \mathbf{K} \mathbf{H} \mathbf{K}^T \mu^{-1/2}. \tag{8.4}$$

Clearly,

$$\mathbf{G}^T \equiv \mu^{-1/2} \mathbf{K} (\mu^{-1/2} \mathbf{KH})^T \equiv \mu^{-1/2} \mathbf{K} (\mathbf{KH})^T \mu^{-1/2} \equiv \mathbf{G}, \tag{8.5}$$

which is therefore symmetric and so possesses real eigenvalues G_1, G_2, G_3, say. Furthermore, according to the principal axes theorem,

$$\mathbf{L}^{-1} \mathbf{GL} = \begin{pmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & G_3 \end{pmatrix}, \tag{8.6}$$

for some orthogonal matrix $\mathbf{L}$: $\mathbf{L}^T = \mathbf{L}^{-1}$ and $\det\mathbf{L} = 1$, say. By (8.2), (8.4), and (8.6), then,

$$G_1 G_2 G_3 \equiv \det\mathbf{G} \equiv (\det\mathbf{H})(\det\mu)^{-1}$$
$$\equiv \lambda^2 (\lambda^2 - \mathbf{w}^2)^2 (\det\mu)^{-1}, \tag{8.7}$$

a positive value. Hence either

$$G_1 > 0, \quad G_2 > 0 \quad \text{and} \quad G_3 > 0, \tag{8.8}$$

or two of the G_j's are negative and one is positive, say,

$$G_1 < 0, \quad G_2 < 0 \quad \text{and} \quad G_3 > 0. \tag{8.9}$$

We next introduce a new coordinate $\mathbf{r}^*$ frame derived from the $\mathbf{r}$ frame by a rotation with the orthogonal matrix $\mathbf{L}$:

$$\mathbf{r}^* = (r_1^*, r_2^*, r_3^*) = \mathbf{rL}. \tag{8.10}$$

Thus, (8.3) becomes

$$\chi \equiv (\det\mu) \sum_{j=1,2,3} G_j r_j^{*2}, \tag{8.11}$$

the desired canonical form. For any real constant χ_ν,

$$\chi = \chi_\nu \tag{8.12}$$

is now recognized as the equation of a quadric surface Q_ν centered at the current edge $\mathbf{r} = \mathbf{0}$ and whose principal axes are the r_j^* axes. By letting χ_ν range over suitable values, we generate a family $\{Q_\nu\}$ of quadric surfaces of constant χ values. Within this context, we say that the function Ψ possesses *quadrical symmetry*. The analogy associated with (7.22) and (7.23) implies that the complement $\bar{\Psi}$ is, likewise, quadrically symmetric.

According to (8.3) or (8.11), Q_ν is time invariant relative to the translated $\mathbf{r}$ or $\mathbf{r}^*$ frame, i.e., it is transported, without change, with the current edge. Now as we already know, the ellipsoidal wavefront ξ not only expands in relation to the $\mathbf{x}$ frame, but also retreats from $\mathbf{r} = \mathbf{0}$. Inevitably, at some stage, it makes contact with Q_ν. We seek more information on the modes of contact, the subsequent intersections between Q_ν and ξ as well as the time durations of such intersections which will, among other things, indicate whether they persist indefinitely or terminate eventually. Some of these questions will be answered in Secs. 9 and 10. Significant physical situations arise only under $\chi_\nu > 0$, which we tentatively assume.

The fundamental case where $\mathbf{v} = \mathbf{0}$ is special; since $\mathbf{r} = \mathbf{x}$, (4.21) reduces to

$$\chi = \lambda^2 \mathbf{x} \mu^{-1} \mathbf{x}^T \det\mu, \tag{8.13}$$

which identifies Q_ν as an ellipsoid similar to and concentric with ξ. However, unlike ξ, Q_ν is a time-invariant ellipsoid. It is crossed, with complete instantaneous coincidence, by the expanding ξ when $t = \chi_\nu^{1/2} \lambda^{-2}$.

Hereafter, we assume that $\mathbf{v}\neq\mathbf{0}$, i.e., $\mathbf{w}\neq\mathbf{0}$. From (4.8) and (4.9), we deduce the following whenever ξ meets Q_ν. Regarding $\mathbf{s}^*$ of (8.1), its third component

$$s_3^* = s_3^{*\pm}, \quad (8.14)$$

where

$$|\mathbf{w}|s_3^{*\pm} \equiv (\lambda^2 - \mathbf{w}^2)t \pm \chi_\nu^{1/2}; \quad (8.15)$$

furthermore,

$$s_1^{*2} + s_2^{*2} + (s_3^* + |\mathbf{w}|t)^2 = \lambda^2 t^2. \quad (8.16)$$

We then go on to conclude that ξ may meet Q_ν along, at most, two possible sets of $\mathbf{s}^*$ points, viz.,

$$\{\mathbf{s}^{*+}\} \text{ and } \{\mathbf{s}^{*-}\} : \mathbf{s}^{*\pm} = (s_1^{*\pm}, s_2^{*\pm}, s_3^{*\pm}), \quad (8.17)$$

with $s_3^{*\pm}$ given by (8.15) while

$$\mathbf{w}^2[(s_1^{*+})^2 + (s_2^{*+})^2] \equiv [\lambda t(|\mathbf{w}| - \lambda) - \chi_\nu^{1/2}] \times [\lambda t(|\mathbf{w}| + \lambda) + \chi_\nu^{1/2}], \quad (8.18)$$

$$\mathbf{w}^2[(s_1^{*-})^2 + (s_2^{*-})^2] \equiv [\lambda t(|\mathbf{w}| - \lambda) + \chi_\nu^{1/2}] \times [\lambda t(|\mathbf{w}| + \lambda) - \chi_\nu^{1/2}]. \quad (8.19)$$

The meeting along $\{\mathbf{s}^{*\pm}\}$ occurs if and only if

$$(s_1^{*\pm})^2 + (s_2^{*\pm})^2 \geqslant 0. \quad (8.20)$$

Equation (8.14) is equivalently

$$\mathbf{r}\cdot(\mathbf{v}\boldsymbol{\mu}^{-1})\det\boldsymbol{\mu} = |\mathbf{w}|s_3^{*\pm}; \quad (8.21)$$

i.e., any single meeting between ξ and Q_ν under (8.20) occurs on a plane $\Pi_\nu^\pm$ having an invariant orientation with its permanent unit normal

$$\mathbf{n}(\Pi_\nu^\pm) = \mathbf{v}\boldsymbol{\mu}^{-1}|\mathbf{v}\boldsymbol{\mu}^{-1}|^{-1}. \quad (8.22)$$

However, according to (8.15) and (8.21), $\Pi_\nu^\pm$ is a traveling plane with instantaneous distance

$$(\det\boldsymbol{\mu})^{-1}|\mathbf{v}\boldsymbol{\mu}^{-1}|^{-1}|\mathbf{w}|\,|s_3^{*\pm}| \quad (8.23)$$

measured from $\mathbf{r}=\mathbf{0}$. More precisely, we deduce from (8.17)–(8.19) that if the strict inequality in (8.20) holds, then the meeting on $\Pi_\nu^\pm$ occurs along a time-dependent closed circuit with equation

$$[\mathbf{r}\cdot(\mathbf{k}_1\boldsymbol{\mu}^{-1/2})]^2 + [\mathbf{r}\cdot(\mathbf{k}_2\boldsymbol{\mu}^{-1/2})]^2 = (\det\boldsymbol{\mu})^{-1}[(s_1^{*\pm})^2 + (s_2^{*\pm})^2], \quad (8.24)$$

where $\mathbf{k}_1$ and $\mathbf{k}_2$ denote, respectively, the first and second rows of the matrix $\mathbf{K}^T$.

At this stage, we suspect that when the equality in (8.20) holds, ξ touches Q_ν tangentially. To test this, we first verify via (2.5) and (7.6) that

$$\nabla(\mathbf{y}\cdot\mathbf{w} - \lambda^2 t)^2 = 2(\det\boldsymbol{\mu})^{1/2}(\mathbf{y}\cdot\mathbf{w} - \lambda^2 t)\sum_{j=1,2,3} w_j\,\mathbf{a}_j, \quad (8.25)$$

which is then applied together with (7.26) to (4.9) to get

$$\nabla(\chi - \chi_\nu) = 2\lambda^2(\det\boldsymbol{\mu})^{1/2}\sum_{j=1,2,3}\mathbf{a}_j(y_j - w_j t) + 2(\det\boldsymbol{\mu})^{1/2}\sum_{j=1,2,3}(w_j\mathbf{y} - \mathbf{w}y_j)\mathbf{w}^T\mathbf{a}_j. \quad (8.26)$$

Now, suppose the equality in (8.20) holds. Then $s_1^{*\pm} = 0 = s_2^{*\pm}$, so that by (8.17), ξ meets Q_ν whenever

$$\mathbf{s}^* = (0,0,s_3^{*\pm}), \quad (8.27)$$

i.e., via (8.1),

$$\mathbf{s} = (0,0,s_3^{*\pm})\mathbf{K}^T = \mathbf{w}|\mathbf{w}|^{-1}s_3^{*\pm}, \quad (8.28)$$

or

$$\mathbf{y} = \mathbf{w}|\mathbf{w}|^{-1}(s_3^{*\pm} + |\mathbf{w}|t). \quad (8.29)$$

Thus,

$$w_j\mathbf{y} = \mathbf{w}y_j \quad (j = 1,2,3). \quad (8.30)$$

Since $t>0$ and it is implicit that the $\mathbf{x}$ point $\mathbf{v}t\notin\xi$, then (8.16) discloses that $s_3^{*\pm}\neq 0$ and $s_3^{*\pm} + |\mathbf{w}|t\neq 0$. Whereupon, applying (8.29) and (8.30) to (8.26) and comparing with (7.27), we deduce that if

$$\mathbf{n}(Q_\nu) = \frac{\nabla(\chi-\chi_\nu)}{|\nabla(\chi-\chi_\nu)|}, \quad (8.31)$$

the instantaneous unit normal to Q_ν, then

$$\mathbf{n}(Q_\nu)\operatorname{sgn}(s_3^{*\pm} + |\mathbf{w}|t) = \mathbf{n}(\xi)\operatorname{sgn}s_3^{*\pm} \quad (8.32)$$

at the encounter $\xi\cap Q_\nu$. Hence this encounter is indeed tangential. Now, from (2.6) and (7.6),

$$\mathbf{w} = (\det\boldsymbol{\mu})^{1/2}(\mathbf{v}\mathbf{a}_1^T, \mathbf{v}\mathbf{a}_2^T, \mathbf{v}\mathbf{a}_3^T); \quad (8.33)$$

also,

$$\sum_{j=1,2,3}\mathbf{a}_j^T\mathbf{a}_j = \boldsymbol{\mu}^{-1}. \quad (8.34)$$

Then using (8.29), we have

$$\sum_{j=1,2,3} y_j\,\mathbf{a}_j = \mathbf{v}\boldsymbol{\mu}^{-1}(\det\boldsymbol{\mu})^{1/2}|\mathbf{w}|^{-1}(s_3^{*\pm} + |\mathbf{w}|t). \quad (8.35)$$

Consequently, by (7.27), (8.32) and (8.22):

$$\mathbf{n}(Q_\nu) = \mathbf{v}\boldsymbol{\mu}^{-1}|\mathbf{v}\boldsymbol{\mu}^{-1}|^{-1}\operatorname{sgn}s_3^{*\pm} = \mathbf{n}(\Pi_\nu^\pm)\operatorname{sgn}s_3^{*\pm}, \quad (8.36)$$

i.e., as should be expected, $\Pi_\nu^\pm$ serves as a common tangent plane at the contact between ξ and Q_ν. Finally, we establish from (8.28), (2.5)–(2.7) and (4.20) that this contact occurs at

$$\mathbf{r} = \mathbf{v}|\mathbf{w}|^{-1}s_3^{*\pm}, \quad (8.37)$$

which lies on the current axis. It corresponds to the single element of $\{\mathbf{s}^{*\pm}\}$.

9. HYPERBOLOIDAL SYMMETRY

In this section, we shall extract specific details on the quadric surface Q_ν of constant $\chi(=\chi_\nu)$ for the case where the current edge $\mathbf{v}t\in\operatorname{ext}\xi$. Here, the function χ vanishes everywhere on the tangent surface $C = C_+\cup C_-$ and, according to (5.7) and (5.9), changes sign across C. This phenomenon together with the canonical form of (8.11) rule out possibility (8.8). Thus only (8.9) is admissible. As $\Psi\equiv 0$ throughout $\operatorname{ext}C_-\cap\operatorname{ext}C_+$, we shall restrict all interests to $\operatorname{int}C_-\cup\operatorname{int}C_+$ wherein $\chi>0$. In particular, the symmetry quadric surface Q_ν lies within $\operatorname{int}C_-\cup\operatorname{int}C_+$ if and only if $\chi_\nu>0$. Its equation is, therefore, by (8.11),

$$|G_3|r_3^{*2} - |G_2|r_2^{*2} - |G_1|r_1^{*2} = |\chi_\nu|(\det\boldsymbol{\mu})^{-1}, \quad (9.1)$$

i.e., Q_ν is a hyperboloid of two sheets confined within $\operatorname{int}C_-\cup\operatorname{int}C_+$ to two subdomains:

$$|r_3^*| \geqslant |\chi_\nu|^{1/2}\,|G_3|^{-1/2}\,(\det\mu)^{-1/2}. \qquad (9.2)$$

One sheet Q_ν^-, say, lies inside the tangent cone C_- which serves as the asymptotic cone to Q_ν^-. The other sheet Q_ν^+ lies inside the complementary cone C_+, the asymptotic cone to Q_ν^+.

Evidently, Q_ν^+ never meets ξ. However, according to (8.18)–(8.20), a meeting along $\{\mathbf{s}^{*-}\}$ between ξ and Q_ν^- occurs if and only if

$$t \geqslant t_-, \quad \text{where} \quad t_- = \chi_\nu^{1/2}\lambda^{-1}(|\mathbf{w}| + \lambda)^{-1}; \qquad (9.3)$$

moreover, another meeting along $\{\mathbf{s}^{*+}\}$ arises if and only if

$$t \geqslant t_+, \quad \text{where} \quad t_+ = \chi_\nu^{1/2}\lambda^{-1}(|\mathbf{w}| - \lambda)^{-1}. \qquad (9.4)$$

Clearly, $t_\pm$ represents an instant where an encounter between ξ and Q_ν^- originates tangentially. Since $t_+ > t_-$, the tangential contact at $t = t_-$ precedes that at $t = t_+$. Once each encounter is achieved, it is never broken. From (9.3), (9.4), (8.15), and (8.37), it can be deduced that both tangential contacts occur at the same $\mathbf{r}$ location, viz.,

$$(\mathbf{r})_{t=t_-} = (\mathbf{r})_{t=t_+} = -\mathbf{v}\chi_\nu^{1/2}\lambda^{-1}|\mathbf{w}|^{-1}. \qquad (9.5)$$

This repetition should of course be anticipated from our knowledge that both such contacts must occur along the current axis which intersects Q_ν^- exactly once, plus the fact that Q_ν^- (as well as Q_ν^+) is invariant relative to the $\mathbf{r}$ frame. Now

$$|\mathbf{v}t_-| = |\mathbf{v}|\chi_\nu^{1/2}\lambda^{-1}(|\mathbf{w}| + \lambda)^{-1} < |\mathbf{r}|_{t=t_-}, \qquad (9.6)$$

i.e., the preceding tangential contact occurs directly on the left extension of the current path away from the latter's initial origin at $\mathbf{x} = \mathbf{0}$. However

$$|\mathbf{v}t_+| = |\mathbf{v}|\chi_\nu^{1/2}\lambda^{-1}(|\mathbf{w}| - \lambda)^{-1} > |\mathbf{r}|_{t=t_+}, \qquad (9.7)$$

i.e., the succeeding tangential contact occurs directly along the "lighted" portion of the current conductor.

A time sequence relative to the $\mathbf{r}$ frame can be conveniently constructed by first translating each of the observed phenomena depicted, over different instances, in Fig. 3 until the various C_- cones with parallel generators coincide.

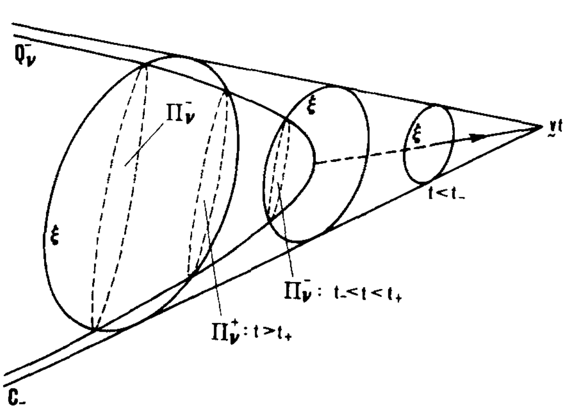


FIG. 7. *Case* $\mathbf{v}t \in \text{ext}\,\xi$: Three stages of development of ξ during its approach towards and subsequent interaction with Q_ν^-, viz., $t < t_-$, $t_- < t < t_+$, $t > t_+$. The current axis intersects Q_ν^- at the common $\mathbf{r}$ location of the two tangential contacts expressed by (9.5). The complementary hyperboloidal sheet Q_ν^+ and its asymptotic cone C^+ are not displayed. (*NB* The vector $\mathbf{v}t$ is shown explicitly for $t < t_-$.)

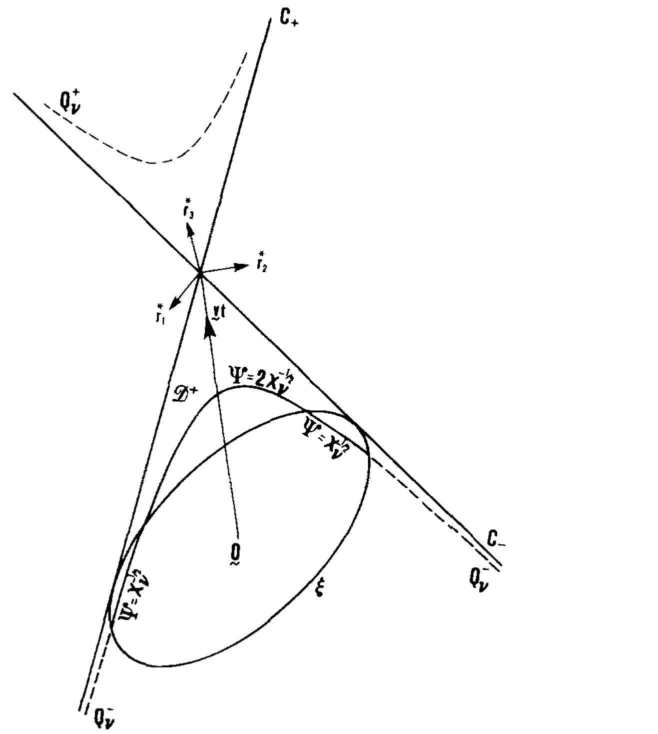


FIG. 8. Hyperboloidal symmetry throughout the time range $t > t_+$.

Meanwhile the centers of the corresponding ξ ellipsoids must become detached. The resultant effect is the same as that of a medium moving with velocity $-\mathbf{v}$ past a stationary point source at $\mathbf{r} = \mathbf{0}$. The various stages are portrayed in Fig. 7. Originally, during $t < t_-$, ξ is apart from Q_ν^- but is approaching Q_ν^- and is expanding simultaneously. It first touches Q_ν^- at instant t_-, at the $\mathbf{r}$ point given by (9.5), and thereafter, over the period (t_-, t_+), maintains a continual intersection with Q_ν^- along an expanding closed circuit on the moving plane Π_ν^- with the invariant unit normal of (8.22). However, the left propagation of the expanding ξ brings it eventually into tangential contact again with Q_ν^- at instant t_+ and at the same $\mathbf{r}$ point of (9.5). As with the first tangential contact, this second contact immediately develops into a closed circuit. The latter expands, throughout $t > t_+$, along another moving plane Π_ν^+ parallel to and trailing Π_ν^-. The plane $\Pi_\nu^\pm$ and its particular circuit of intersection with ξ are described respectively by (8.21) and (8.24), together with (8.15), (8.18) and (8.19).

Figure 8 provides a detailed instantaneous representation over the partially infinite time range $t > t_+$. Along those finite portions of Q_ν^- within $\mathscr{D}^+$ and intξ, Ψ possesses a *hyperboloidal symmetry* with respective constant values $2\chi_\nu^{-1/2}$, $\chi_\nu^{-1/2}$. But along those indefinite continuations beyond ξ, viz., broken line portions of Q_ν^-, $\Psi \equiv 0$. Likewise, $\Psi \equiv 0$ along the other sheet Q_ν^+ (broken line surface) of the hyperboloid inside the asymptotic cone C_+. The r_3^* axis is the axis of the hyperboloid; it is generally inclined to the current axis, the actual inclination being, in principle, determinable via (8.2), (8.4), (8.6), and (8.10). The inclination depicted in Fig. 8 is acute. This is incidental. It may well turn out to be obtuse; nevertheless, our main interpretation remains effective.

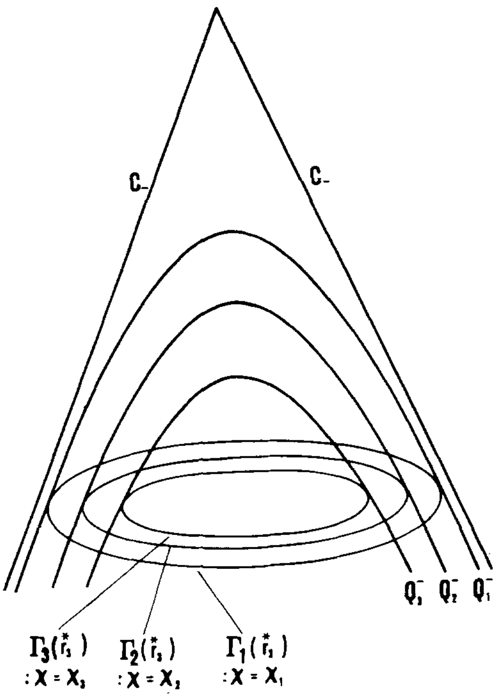


FIG. 9. Variation of χ across colevel elliptical cross sections of Q_ν^- sheets.

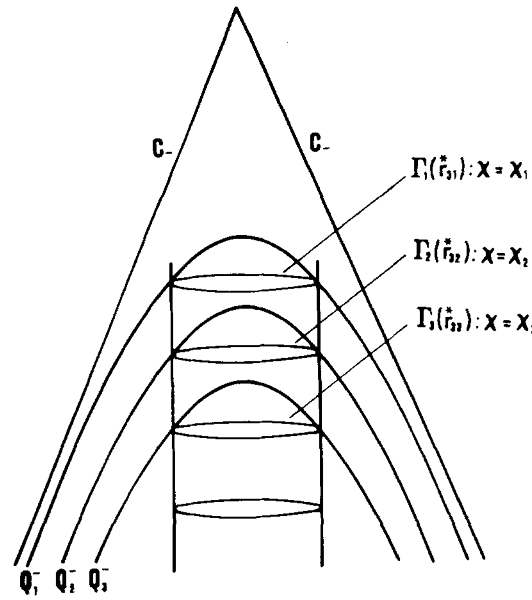


FIG. 10. Elliptical axisymmetry: Single parametric dependence of χ on r_3^* along an elliptic cylindrical surface.

Consider the subfamily $\{Q_\nu^-\}$ of hyperboloidal sheets generated inside C_- by letting χ_ν run over suitable positive values. Suppose $Q_1^-, Q_2^-, Q_3^-, \cdots$ represent a sequence of sheets encountered via a retreat from C_- towards the r_3^* axis (see Fig. 9). Now, for any r_3^* satisfying the strict inequality in (9.2), χ takes the constant value χ_ν along an ellipse $\Gamma_\nu(r_3^*)$, the intersection of Q_ν^- with the plane at distance $|r_3^*|$ from the point $\mathbf{r} = \mathbf{0}$. Furthermore, if the same strict inequality holds over the various ν's, this plane intersects $\{Q_\nu^-\}$ to form a family of colevel concentric ellipses $\{\Gamma_\nu(r_3^*)\}$. We now deduce from (9.1) that

$$0 < \chi_1 < \chi_2 < \chi_3 < \cdots < |G_3| r_3^{*2} \det \boldsymbol{\mu}. \tag{9.8}$$

Evidently, at any r_3^*-level, χ increases through the values $\chi_1, \chi_2, \cdots$ as point $\mathbf{r}$ traverses the ellipses $\Gamma_1(r_3^*), \Gamma_2(r_3^*), \cdots$ to reach, ultimately, the r_3^* axis, thereby attaining a maximum value $|G_3| r_3^{*2} \det \boldsymbol{\mu}$. However, χ is unbounded inside C_-. Observe, on the other hand, that Ψ stays bounded inside C_-.

Consider, next, an elliptic cylinder of uniform cross section, a typical cross section at level $r_3^* = r_{3\nu}^*$, say, being circumscribed by the ellipse $\Gamma_\nu(r_{3\nu}^*)$ where the cylinder intersects the hyperboloidal sheet Q_ν^- (see Fig. 10). Along $\Gamma_\nu(r_{3\nu}^*)$, χ therefore takes the constant value χ_ν. Along the surface of such a cylinder, χ depends solely on the axial coordinate r_3^*. In particular χ, again increases through $\chi_1, \chi_2, \chi_3, \ldots$, this time, indefinitely and as the axial coordinate r_3^* runs over $r_{31}^*, r_{32}^*, r_{33}^*, \cdots$. In these respects, we say that χ exhibits an *elliptical axisymmetry*.

10. ELLIPSOIDAL SYMMETRY

We shall next examine Q_ν in the case where $\mathbf{v}t \in \operatorname{int}\xi$. Here, (5.20) holds for the canonical form (8.11) so that (8.9) becomes incompatible. Rule (8.8) now applies instead. There is another way of viewing this: the real symmetric matrix $\mathbf{H}$ of (8.2) is positive definite as its eigenvalues

$$\lambda^2 > 0, \quad \lambda^2 - \mathbf{w}^2(\text{repeated}) > 0$$

in view of (5.19); also, the real matrix $\boldsymbol{\mu}^{-1/2}\mathbf{K}$ is nonsingular; consequently, the matrix $\mathbf{G}$ of (8.4) is positive definite so that its eigenvalues G_1, G_2, and G_3 do satisfy (8.8). Thus, assuming $\chi_\nu > 0$, the central quadric Q_ν governed by (8.12) is presently an ellipsoid.

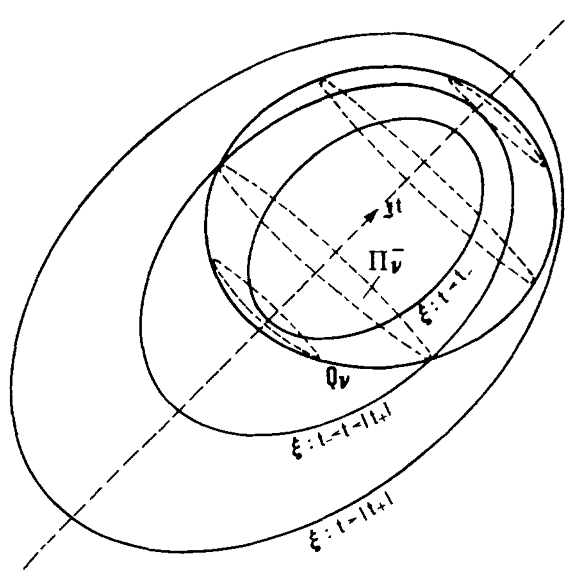


FIG. 11. *Case* $\mathbf{v}t \in \operatorname{int}\xi$: Three stages of development relative to the $\mathbf{r}$ frame, viz., $t < t_-$, $t_- < t < |t_+|$, $t > |t_+|$. The current axis intersects Q_ν at the two symmetric $\mathbf{r}$ points $(\mathbf{r})_{t_-}$ and $(\mathbf{r})_{|t_+|}$. The closed circuit (of intersection) $\xi \cap Q_\nu$, which originates and terminates at these two points and propagates with the plane Π_ν^-, is represented over four different instances during the course of its expansion and subsequent contraction within the period $t_- < t < |t_+|$.

Again, Q_ν plays a significant role only if it supports nontrivial values of Ψ, which we know exist strictly inside the ellipsoid ξ. The latter's variable geometric relationship with Q_ν is therefore an important factor again. Let us consider developments relative to the **r** frame. Now ξ travels and grows from an origin initially coincident with the center, at $\mathbf{r} = \mathbf{0}$, of the invariant ellipsoid Q_ν. Hence, it first evolves inside Q_ν which meanwhile plays an insignificant role. However ξ must eventually cross Q_ν which then acquires significance as a surface of constant Ψ ($\not\equiv 0$). To discuss the matter on firmer grounds, we appeal to (8.18)–(8.20) which disclose: no meeting can be associated with $\{\mathbf{s}^+\}$, while a meeting along $\{\mathbf{s}^-\}$ occurs on the plane Π_ν^- if and only if

$$t_- \leqslant t \leqslant |t_+|, \tag{10.1}$$

with $t_\pm$ formally displayed in (9.3) and (9.4). The equality signs correspond to two tangential contacts. According to (8.15) and (8.37), and in direct contrast to the situation posed by (9.5), both these contacts occur at two symmetric **r** locations, viz.,

$$(\mathbf{r})_{t_-} = -\mathbf{v}\chi_\nu^{1/2}\lambda^{-1}|\mathbf{w}|^{-1}, \quad (\mathbf{r})_{|t_+|} = \mathbf{v}\chi_\nu^{1/2}\lambda^{-1}|\mathbf{w}|^{-1}. \tag{10.2}$$

Developments are schematically presented in Fig. 11. In a primary stage, the evolution of ξ inside Q_ν progresses until the instant t_- when tangential contact between both ellipsoids is first established at $\mathbf{r} = (\mathbf{r})_{t_-}$, directly on the left extension of the current path. Thereafter, in the course of its translation and growth, ξ intersects Q_ν along a closed circuit on the plane Π_ν^- with its normal constantly parallel to $\mathbf{v}\boldsymbol{\mu}^{-1}$. As Π_ν^- propagates in the general direction of **v**, the closed circuit of intersection expands, initially from its origin at $\mathbf{r} = (\mathbf{r})_{t_-}$, and then contracts again. Ultimately, it degenerates at instant $|t_+|$ onto the diametrically opposite point $\mathbf{r} = (\mathbf{r})_{|t_+|}$ where the encounter between ξ and Q_ν turns tangential again, this time along the yet "unlighted" part of the current conductor. Immediately, then, ξ breaks contact permanently with Q_ν. From (8.15), (9.3) and (9.4), we see that

$$|\mathbf{w}|s_3^{*-} = 0 \quad \text{when} \quad t = \tfrac{1}{2}(t_- + |t_+|), \tag{10.3}$$

i.e., in view of (8.23) and (10.2), Π_ν^- crosses the current edge at a mean instant and mean location in relation to contact making and contact breaking. Throughout $t > |t_+|$, $\operatorname{int}\xi \supset Q_\nu$, along which the corresponding Ψ solution of (4.18) takes the constant value $\chi_\nu^{-1/2}$. However, during $t_- < t < |t_+|$, $\Psi \equiv \chi_\nu^{-1/2}$ over that portion of Q_ν inside ξ, but vanishes identically over the complementary portion outside ξ. Thus Ψ is *ellipsoidally symmetric*. As in the previous situation with hyperboloidal symmetry, one can go on to demonstrate an elliptical axisymmetry of χ.

ACKNOWLEDGMENTS

A skeleton for this work was initiated during the brief tenure of a visiting appointment at the Institut für Theoretische Physik der Universität Düsseldorf. It was significantly influenced by discussions with Professor Kurt Suchy and Dr. Helmut Hebenstreit. I owe them a debt of gratitude for this and for the generous help they gave in various indirect forms. I also wish to thank Frau Dr. Suchy, Jörg Thiel, Katja Havranek and Theresia Bernstein, as well as members of that Institute for the very warm hospitality enjoyed.

[1] I.M. Besieris, J. Math. Phys. **10**, 1156 (1969).
[2] I.M. Besieris, J. Math. Phys. **8**, 409 (1967).
[3] I.M. Besieris and R.T. Compton, Jr., J. Math. Phys. **8**, 2445 (1967).
[4] K.S.H. Lee and C.H. Papas, J. Math. Phys. **5**, 1668 (1964).
[5] K.S.H. Lee and C.H. Papas, IEEE Trans. Antennas Propagation **13**, 799 (1965).
[6] R.T. Compton, Jr. and C.T. Tai, IEEE Trans. Antennas Propagation **13**, 574 (1965).
[7] C.T. Tai, IEEE Trans. Antennas Propagation **13**, 322 (1965).
[8] C.T. Tai, Appl. Opt. **4**, 1347 (1965).
[9] C.T. Tai, J. Math. Phys. **8**, 646 (1967).
[10] R.T. Compton, Jr., J. Math. Phys. **7**, 2145 (1966).
[11] K.C. Chen and J.L. Yen, J. Math. Phys. **9**, 2081 (1968).
[12] R.M. Kalafus, Radio Sci. **4**, 977 (1969).
[13] G. Johannsen, J. Math. Phys. **11**, 3251 (1970).
[14] G. Johannsen, J. Math. Phys. **12**, 1349 (1971).
[15] S. Solimeno, J. Math. Phys. **16**, 218 (1975).
[16] S.D. Majumdar and R. Pal, Proc. Roy. Soc. Lond. Ser. A **316**, 525 (1970).
[17] S.D. Majumdar and R. Pal, Ann. Phys. NY **76**, 419 (1973).
[18] S.D. Majumdar, Ann. Phys. NY **76**, 428 (1973).
[19] G.P. Sastry, J. Phys. A: Math., Nucl. Gen. **7**, 600 (1974).
[20] G.P. Sastry and. S. Bhattacharya, J. Phys. A: Math. Gen. **10**, 595 (1977).
[21] S.J. Lewandowski, J. Phys. A: Gen. Phys. **4**, 197 (1971).
[22] S.W. Lee and Y.T. Lo, J. Appl. Phys. **38**, 870 (1967).
[23] J.A. Kong, J. Appl. Phys. **41**, 554 (1970).
[24] J.A. Kong, J. Opt. Soc. Am. **64**, 1304 (1974).
[25] G. Johannsen, J. Math. Phys. **13**, 78 (1972).
[26] R. Handelsman and R.M. Lewis, J. Math. Phys. **7**, 1982 (1966).
[27] P.C. Clemmow, Proc. IEE **110**, 101 (1963).
[28] L.B. Felsen, Proc. IEE **111**, 445 (1964).
[29] L.B. Felsen, Proc. IEE **111**, 454 (1964).
[30] K.K. Mei, IEEE Trans. Antennas Propagation **19**, 665 (1971).
[31] H.S. Lu and K.K. Mei, Alta Freq. N. **4**, 290 (1972) Vol. XLI.
[32] L. Chee-Seng, J. Math. Phys. **21**, 189 (1980).
[33] M. Kline and I.W. Kay, *Electromagnetic Theory and Geometrical Optics* (Interscience, New York, 1965).
[34] M.J. Lighthill, *Fourier Analysis and Generalized Functions* (Cambridge U.P., Cambridge, 1958).

Stability of streets of vortices on surfaces of revolution with a reflection symmetry

David Hally
Department of Physics, University of British Columbia, Vancouver, B. C., Canada V6T 1W5

(Received 16 May 1979; accepted for publication 6 August 1979)

Helmholtz's theory of ideal vortex motion in two dimensions is generalized to flows on curved surfaces. The existence of a generalized vortex stream function is proved and used to generate conservation laws. In particular, the angular moment of circulation is related to invariance under scale transformations. The theory is used to derive criteria for stability of vortex streets on surfaces of revolution having symmetry under reflection in a plane whose normal is the axis of revolution. For the special case of the sphere it is found that only those vortex streets having six or fewer vortices per ring can be stable and that, in contradistinction to the results of von Karman, both symmetric and staggered vortex streets can be stable.

1. INTRODUCTION

The theory of vorticity is a field theory and, as such, is governed by partial differential equations. Helmholtz,[1] however, pointed out that under the restrictive assumptions of perfect, incompressible, two-dimensional flow in which the vorticity is concentrated at N isolated points, the problem of determining the fluid motion is reduced to the solution of a system of first-order ordinary differential equations. Lin[2] proved that this system can be put into Hamilton's form provided that all boundaries are rigid. The Hamiltonian is called the vortex streamfunction and is closely related to the kinetic energy of the fluid.[3]

Lamb, in his well-known text,[4] briefly outlines a method for determining the motion of vortices on a curved surface under the assumption that the depth of the fluid is small in comparison with the principal radii of curvature of the surface. It is the purpose of this paper to examine such vortex systems in more detail. After defining the velocity field (Sec. 2) and the velocity (Sec. 3) of a vortex the equations of motion are derived and it is shown that for a large class of surfaces an analog to the vortex streamfunction exists casting the equations of motion into symplectic form (Sec. 4). The constants of the motion associated with simple symmetries of the surface of flow are compared to with those of a vortex system in the plane (Sec. 5). The connection of the angular moment of circulation with invariance under scale transformations is also discussed. The theory is then applied to vortex streets on surfaces of revolution having symmetry under reflection in a plane whose normal is the axis of revolution (Sec. 6). Criteria for their stability are derived and it is found that in the particular case of the sphere there are both symmetric and staggered stable configurations, in contradistinction to the results of von Karman for infinite vortex streets in the plane.[5]

Ideal vortices have been used for many years to model atmospheric cyclones but almost exclusively in the tangent plane approximation (i.e., the surface of the earth in the immediate vicinity of the cyclone is assumed flat).[6-8] The atmosphere is assumed of constant density and very close to hydrostatic equilibrium so that its thickness is nearly uniform. The cores of the vortices are assumed sufficiently large that the velocities induced by bending of their cores[9] is negligible. It is tempting to try to extend such theories beyond the tangent plane approximation by means of the methods introduced in this paper. Unfortunately, none of the models is then acceptable because the variation of the Coriolis parameter with latitude cannot be included satisfactorily. In particular, the Coriolis terms create a source of vorticity so that the vorticity is no longer advected by the velocity field. It is also observed that at mid-latitudes the wind is nearly geostrophic (pressure gradients balance Coriolis forces), whereas geostrophic flow and potential flow are incompatible unless the Coriolis parameter is constant.[10] One must conclude that the vortices treated in this paper would not provide a satisfactory model for terrestrial cyclones. However, they might still be used as a first approximation for cyclones in an atmosphere in which the Coriolis force is not predominant.

2. IDEAL VORTICES ON CURVED SURFACES

We wish to examine flows on a surface characterized by the Riemannian metric: $g_{ij}(x^1,x^2)$, $i = 1,2$, $j = 1,2$, where x^1 and x^2 are some coordinates. In the limit that the depth of the fluid is very small in comparison with the principal radii of curvature of the surface, one may suppose that the fluid velocity is everywhere tangent to the surface and does not vary with depth. The fluid velocity can then be represented by a covariant vector field V_i.

It will prove convenient to choose coordinates x,y such that

$$g_{ij}(x, y) = \delta_{ij} h^2(x, y)$$

The line element for the surface is then[11]

$$ds^2 = h^2(x, y)(dx^2 + dy^2)\,.$$

For simplicity is is assumed that the whole region of flow is parametrized unambiguously by these "harmonic" coordinates. This amounts to a restriction to those surfaces topologically equivalent to arbitrarily connected sub-domains of the complex plane.

Provided that the depth of the fluid is uniform the re-

 0022-2488/80/010211-07$01.00

quirement for incompressible flow is

$$V^i_{;i} = h^{-2}(x,y)\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}\right) = 0\,, \tag{2.1}$$

the general solution for which is

$$V_x = \frac{\partial \Psi}{\partial y} \quad V_y = \frac{-\partial \Psi}{\partial y} \tag{2.2}$$

for some real-valued function $\Psi(x,y)$ having continuous mixed second-order derivatives in the region of flow.

The velocity field of a vortex at $(x^{1'},x^{2'})$ is defined to be the incompressible velocity field having zero vorticity everywhere throughout the region of flow (henceforth denoted D) except at $(x^{1'},x^{2'})$, i.e., satisfying (2.1) and

$$\epsilon^{ij}V_{i;j} = 2\pi\gamma\delta(x^1 - x^{1'})\delta(x^2 - x^{2'}) \tag{2.3}$$

where ϵ^{ij} is the antisymmetric tensor density with $\epsilon^{12} = 1.\gamma$ is a constant known as the vortex strength. In harmonic coordinates (2.3) becomes

$$\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} = 2\,\pi\gamma\delta(x - x')\delta(y - y') \tag{2.4}$$

Substituting (2.2) into (2.4) one obtains

$$\nabla^2\Psi = -2\pi\delta(x - x')\delta(y - y')\,, \tag{2.5}$$

where $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ and we consider only vortices of unit strength.

If $\Lambda(x,y;x',y')$ is regular throughout D and satisfies

$$\nabla^2\Lambda = 0\,, \tag{2.6}$$

then

$$\Psi(x,y;x',y') = \Lambda(x,y;x',y') - \tfrac{1}{2}\ln[(x-x')^2 + (y-y')^2] \tag{2.7}$$

is a solution of (2.5) provided that the point at infinity lies outside D. By suitable choice of coordinates this is always possible unless there are no boundaries and the surface is closed. Koebe[12] has proved the existence of functions Ψ and Λ satisfying (2.6) and (2.7) for arbitrary rigid boundary conditions, while Lin[2] has proved their uniqueness and reciprocity [i.e., that $\Psi(x,y;x',y') = \Psi(x',y';x,y)$]. Thus the velocity field of a vortex is well-defined on all surfaces topologically equivalent to multiply connected regions of the plane with rigid boundaries.

For surfaces topologically similar to the sphere the above definition is insufficient as it is impossible to have a single point of isolated vorticity. Rather, there is a constraint that the sum of all vortex strengths must vanish. (Choose a closed contour. It divides the surface into two regions. The path integral of the velocity around the contour is equal to the integral of the vorticity in one region and the negative of the integral of the vorticity in the other. This yields $2\pi\,\Sigma\,\gamma_n = 0$.) The function Ψ of Eq. (2.7) must therefore be interpreted as the streamfunction of a two vortex system: one with strength 1 at (x',y') and one with strength -1 at infinity. Notice that a system of vortices at (x_n, y_n) with respective strengths γ_n, $\Sigma\gamma_n = 0$, is still given by: $\Sigma\gamma_n\Psi(x_n, y_n)$ as the vortex at infinity disappears due to the constraints on the γ's.

Since Λ is a real harmonic function in D, if D is simply connected there is a conjugate harmonic function $\Sigma(x,y;x',y')$, unique up to an additive constant such that $(\Sigma + i\Lambda)(x,y;x',y') \equiv g(z;z')$, $z = x + iy$, is an analytic function in D.[13] Moreover, $g(z;z') - i\ln(z - z') \equiv \phi(z;z')$ is analytic everywhere in D except at z', and $\mathrm{Im}\phi = \Psi$. Thus, ϕ is a complex potential for the flow:

$$V_x - iV_y = \frac{d\phi(z)}{dz}\,. \tag{2.8}$$

It is unique up to an additive constant. The complex notation will prove useful for the calculations in Sec. 6.

3. THE VELOCITY OF A VORTEX

If $(x'(t), y'(t))$ is the position of a vortex at time t, then its (physical) velocity is $u_x = h(x',y')\dot{x}'$; $u_y = h(x',y')\dot{y}'$ where the dot denotes time derivative. One determines the velocity of a vortex as follows.

The physical velocity field of a vortex is related to its covariant velocity field by $v_x = h^{-1}(x,y)V_x$; $v_y = h^{-1}(x,y)V_y$. Expanding the physical velocity field in terms of $(z - z')$ and $(\bar{z} - \bar{z}')$ (it is most convenient to use the complex notation) one finds:

$$\begin{aligned} v_x - iv_y &= h^{-1}(z,\bar{z})\frac{d\phi(z;z')}{dz} \\ &= \frac{-ih^{-1}(z',\bar{z}')}{(z - z')} - \frac{i\partial h^{-1}(z',\bar{z}')}{\partial z}\frac{(\bar{z} - \bar{z}')}{(z - z')} \\ &\quad + h^{-1}(z',\bar{z}')\frac{dg(z';z')}{dz} - \frac{i\partial h^{-1}(z',\bar{z}')}{\partial z} \\ &\quad + O(|z - z'|)\,, \end{aligned} \tag{3.1}$$

where for convenience we have used the same symbol, h, to denote the metric function despite the change in arguments from (x,y) to $(z,\bar{z})$. The first term in (3.1) yields a velocity field whose streamlines are concentric circles, the second, one in which the fluid flows radially. Neither of these terms prefers any direction and therefore cannot contribute to the motion of the vortex. The third and fourth terms are uniform fields in which the vorticity concentrated at z' must be convected according to the Helmholtz vorticity theorem.[1] The higher-order terms all vanish as $z \to z'$ so that they, too, cannot contribute to the vortex motion. The velocity of a vortex of strength γ at z' is therefore

$$u_x - iu_y = \gamma h^{-1}(z',\bar{z}')\frac{\partial}{\partial z}\left(g(z;z') + i\ln[h(z,\bar{z})] + \frac{\phi^*(z)}{\gamma}\right)_{z=z'}\,, \tag{3.2}$$

where $\phi^*(z)$ is the complex potential for the flow due to external influences (e.g., other vortices). The first term gives the motion induced by the presence of boundaries. For example, for a vortex in the plane ($h = 1$) bounded by $r = R$, ($r = \{x^2 + y^2\}^{1/2}$), one finds $\phi(z;z') = i\gamma\ln(z - z') + i\gamma\ln(R^2 - z\bar{z}')$ so that $g(z;z') = i\gamma\ln(R^2 - z\bar{z}')$ and there is an induced velocity due to the boundary: $u_x - iu_y = -i\gamma\bar{z}'/(R^2 - z'\bar{z}')$. The second term gives the velocity induced by the curvature of the surface of flow. In the case of a closed surface with no boundaries these also include the effects of the vortex at infinity.

For the more general case when a complex potential cannot be defined similar arguments yield for the velocity of the vortex:

$$u_x = \gamma h^{-1}(x',y')\frac{\partial}{\partial y}\left(\Lambda(x,y;x',y') + \tfrac{1}{2}\ln[h(x,y)] + \frac{\Psi^*(x,y)}{\gamma}\right)_{\substack{x=x'\\ y=y'}}$$

$$u_y = -\gamma h^{-1}(x',y')\frac{\partial}{\partial x}\left(\Lambda(x,y;x',y') + \tfrac{1}{2}\ln[h(x,y)] + \frac{\Psi^*(x,y)}{\gamma}\right)_{\substack{x=x'\\ y=y'}} \qquad (3.3)$$

It might seem paradoxical that a vortex velocity field should contain radial terms such as the second in (3.1). This arises since $\mathbf{v}$ does not satisfy the divergence-free equation $\nabla\cdot\mathbf{v}=0$, but rather $\nabla\cdot\mathbf{v} = -\mathbf{v}\cdot\nabla\ln h$ [from (2.1)]. By projecting onto the plane "fictitious" source terms are introduced giving rise to the radial term.

4. THE VORTEX STREAMFUNCTION

Using (3.3), the equations of motion of a system of vortices with respective strengths and positions γ_n and (x_n, y_n), $n = 1,\ldots,N$, are

$$\dot{x}_n = h^{-2}(x_n,y_n)\frac{\partial}{\partial y}\left(\sum_k{}'\gamma_k\Psi(x,y;x_k,y_k) + \gamma_n\Lambda(x,y;x_n,y_n) + \tfrac{1}{2}\gamma_n\ln[h(x,y)]\right)_{\substack{x=x_n\\ y=y_n}}$$

$$\dot{y}_n = -h^{-2}(x_n,y_n)\frac{\partial}{\partial y}\left(\sum_k{}'\gamma_k\Psi(x,y;x_k,y_k) + \gamma_n\Lambda(x,y;x_n,y_n) + \tfrac{1}{2}\gamma_n\ln[h(x,y)]\right)_{\substack{x=x_n\\ y=y_n}} \qquad (4.1)$$

(the prime denotes a sum over all $k\neq n$).

Making use of the reciprocity of Ψ and Λ, (4.1) can be rewritten

$$\dot{x}_n = (\gamma_n h^2(x_n,y_n))^{-1}\frac{\partial\Omega}{\partial y_n}(x_1,y_1,\ldots,x_N,y_N)$$

$$\dot{y}_n = -(\gamma_n h^2(x_n,y_n))^{-1}\frac{\partial\Omega}{\partial x_n}(x_1,y_1,\ldots,x_N,y_N), \qquad (4.2)$$

where

$$\Omega = \tfrac{1}{2}\sum_{n=1}^{N}\sum_k{}'\gamma_n\gamma_k\Psi(x_n,y_n;x_k,y_k) + \tfrac{1}{2}\sum_{n=1}^{N}\gamma_n^2\{\Lambda(x_n,y_n;x_n,y_n) + \ln[h(x_n,y_n)]\} \qquad (4.3)$$

Ω is a generalization of the vortex streamfunction given by Lin.[2]

As shown in Appendix A, (4.2) can be put in the symplectic form: $\dot{\mathbf{x}} = \sigma^{-1}\nabla\Omega$, where $\mathbf{x}$ is a $2N$-dimensional vector, σ is a symplectic 2-form and ∇ is the exterior derivative.

In the absence of boundaries one has (again making use of the complex notation),

$$\phi(z;z') = -i\ln(z-z'), \quad g(z;z') = 0 \qquad (4.4)$$

and the equations of motion are

$$\dot{\bar{z}}_n = h^{-2}(z_n,\bar{z}_n)\left(\sum_k{}'\frac{-i\gamma_k}{z_n-z_k} + i\gamma_n\frac{\partial\ln[h(z_n,\bar{z}_n)]}{\partial z_n}\right) \qquad (4.5)$$

Routh[14] first drew attention to the properties of the vortex streamfunction under conformal transformations. His results were generalized by Lin[2] who showed that under the conformal transformation $z\to\tilde{z}$, the vortex streamfunction transforms as

$$\tilde{\Omega} = \Omega + \tfrac{1}{2}\sum_n\gamma_n\ln\left|\frac{d\tilde{z}}{dz}\right|, \qquad (4.6)$$

where tildes denote transformed quantities (Lin's κ is our $2\pi\gamma$). Using (4.2) rather than Lin's (5.1) and duplicating his analysis, it is easily shown that on curved surfaces the vortex streamfunction transforms according to (4.6) provided $h(z,\bar{z}) = \tilde{h}(\tilde{z},\bar{\tilde{z}})$. This ensures that the surface of flow remains invariant under the transformation.

5. CONSTANTS OF THE MOTION

Since the equations of motion can be put in symplectic form, there are conservation laws associated with infinitesimal transformations which leave the vortex streamfunction invariant. In particular, if $\boldsymbol{\xi}$ is a Killing vector for the surface of flow which is respected by all boundaries, then symmetry and the uniqueness of Ω imply that Ω is invariant under translations along $\boldsymbol{\xi}$.

If, for example, $h\equiv h(x)$ then $\xi_x = 0$; $\xi_y = 1$ is a Killing vector with which is associated the conservation law

$$\sum_{n=1}^{N}\gamma_n\int h^2(y_n)\,dy_n = \text{const.} \qquad (5.1)$$

If $h\equiv h(r)$ with $r=(x^2+y^2)^{1/2}$, then $\xi_x = -y$; $\xi_y = x$ is a Killing vector with which is associated the conservation law

$$\sum_{n=1}^{N}\gamma_n\int h^2(r_n)r_n\,dr_n = \text{const.} \qquad (5.2)$$

For flow in the plane, (5.1) and (5.2) yield the conservation of center of circulation and moment of circulation, respectively.

The vortex streamfunction itself is conserved as a consequence of the symplectic form of the equations of motion.

For the vortex systems in the plane there is another conserved quantity known as the angular moment of circulation

$$\sum_{n=1}^{N}\gamma_n(\dot{x}_n y_n - x_n\dot{y}_n) = \text{const.} \qquad (5.3)$$

A generalization to curved surfaces is obtained as follows.

Suppose h is a homogeneous function of order λ, i.e., $h(ax,ay) = a^\lambda(x,y)$. Physically this means that the surface is invariant under scale transformations $(x,y)\to(ax,ay)$. If the boundaries are also invariant then the uniqueness of Ω implies that

$$\Omega(ax_1,ay_1,\ldots,ax_N,ay_N) = a^\nu\Omega(x_1,y_1\ldots,x_N,y_N) + b(a) \qquad (5.4)$$

for some constants ν and $b(a)$. Differentiating with respect to a and putting $a=1$ gives:

$$\sum_{n=1}^{N}\left(x_n \frac{\partial \Omega}{\partial x_n} + y_n \frac{\partial \Omega}{\partial y_n}\right) = \nu\Omega + \text{const.} \tag{5.5}$$

Using (4.2) and the fact that Ω is a constant of the motion one finds,

$$\sum_{n=1}^{N} \gamma_n h^2(x_n, y_n)(\dot{x}_n y_n - x_n \dot{y}_n) = \text{const}, \tag{5.6}$$

which reduces to (5.3) when $h = 1$. Thus, the angular moment of circulation is a conserved quantity associated with invariance under scale transformations.

6. STABILITY OF VORTEX STREETS ON SURFACES OF REVOLUTION

The theory of the preceding sections is now applied to a problem similar to the classic questions of Thomson[15] and von Karman[5]: Are double rings of vortices arranged symmetrically around the axis of a surface of revolution stable?

A surface of revolution is characterized by: $h \equiv h(r)$, $r = (x^2 + y^2)^{1/2}$. Introducing the angular coordinate $\phi = \arctan(y/x)$ the line element becomes $ds^2 = h^2(r)(dr^2 + r^2 d\phi^2)$. If in addition it is required that the surface be symmetric under reflection in a plane whose normal is the axis of revolution and which contains the line $r = 1$, the line element must be invariant under the transformation $r \rightarrow r^{-1}$. This imposes the further restriction that

$$r h(r) = r^{-1} h(r^{-1}). \tag{6.1}$$

The symmetries of such a surface suggest that the motion of a ring of vortices each with strength γ, uniformly distributed around the line $r = r_0$, and a similar ring with vortices of strength $-\gamma$ on the line $r = r_0^{-1}$, should be especially simple. Such a configuration is called a vortex street. There are two possible cases: staggered and symmetric.

A. Staggered vortex streets

The vortices of a staggered vortex street are situated initially at

$$r_n = r_0, \quad \phi = 2\pi i n/N, \quad n = 1,\ldots,N, \quad \text{strength } \gamma,$$
$$r_m = r_0^{-1}, \quad \phi = (2m+1)\pi i/N, \quad m = 1,\ldots N, \quad \text{strength } -\gamma \tag{6.2}$$

or, in complex notation,

$$z_n = r_0 \exp[2\pi i n/N], \quad n = 1,\ldots N, \quad \text{strength } \gamma,$$
$$z_m = r_0^{-1} \exp[(2m+1)\pi i/N], \quad m = 1,\ldots,N, \quad \text{strength } -\gamma. \tag{6.3}$$

Trying a solution of the form

$$z_n = r(t) \exp[(2\pi i n/N) + i\omega t],$$
$$z_m = r^{-1}(t) \exp\{[(2m+1)\pi i/N] + i\omega t\} \tag{6.4}$$

in the equations of motion (4.2) one finds:

$$r(t) = r_0, \quad \omega = \frac{N}{1 + r_0^{2N}} - \frac{N}{2} - \frac{p(r_0)}{2}, \tag{6.5}$$

where $p(r) = 1 + [r h'(r)/h(r)]$, and the unit of time has been taken to be $r_0^2 h^2(r_0)/\gamma$. (6.5) also involves the evaluation of the sums,

$$\sum_{n=1}^{N} [1 - \exp(2\pi i n/N)]^{-1} = \tfrac{1}{2}(N-1), \tag{6.6}$$

$$\sum_{n=1}^{N} \{1 - x \exp[(2n+1)\pi i/N]\}^{-1} = N/(1 + x^N) \tag{6.7}$$

(see Appendix B). The vortex street thus rotates rigidly about the axis of revolution with angular velocity given by (6.5).

To examine the stability of this configuration consider small deviations from the motion:

$$z_n(t) = [r_0 \exp(2\pi i n/N) + \epsilon_n(t)]e^{i\omega t}, \quad n = 1,\ldots,N$$
$$z_m(t) = \{r_0^{-1} \exp[(2m+1)\pi i/N] + \delta_m(t)\}e^{i\omega t}, \quad m = 1,\ldots,N \tag{6.8}$$

Substituting into the equations of motion and expanding to first order in the ϵ's and δ's yields:

$$\begin{aligned}\dot{\epsilon}_n = \Bigg(&\sum_k{}' \frac{(\epsilon_n - \epsilon_k)}{\{1 - \exp[2\pi i(k-n)/N]\}^2} \\ &- \sum_{m=1}^{N} \frac{(\epsilon_n - \delta_m)}{\{1 - r_0^{-2} \exp[(2(m-n)+1)\pi i/N]\}^2} \\ &+ P(r_0)\epsilon_n + Q(r_0)\bar{\epsilon}_n \exp(4\pi i n/N)\Bigg) \\ &\times i \exp(-4\pi i n/N),\end{aligned} \tag{6.9}$$

$$\begin{aligned}\dot{\delta}_m = \Bigg(&- \sum_k{}' \frac{(\delta_m - \delta_k)}{\{1 - \exp[2\pi i(k-n)/N]\}^2} \\ &+ \sum_{n=1}^{N} \frac{(\delta_m - \epsilon_n)}{(1 - r_0^2 \exp\{[2(n-m)-1]\pi i/N\})^2} \\ &- P(r_0^{-1})\delta_m - Q(r_0^{-1})\bar{\delta}_m \exp[2(2m+1)\pi i/N]\Bigg) \\ &\times i \exp[-2(2m+1)\pi i/N],\end{aligned} \tag{6.10}$$

where

$$P(r) = \tfrac{1}{4} r p'(r) + (p(r) - 1)(\omega - \tfrac{1}{2}), \tag{6.11}$$

$$Q(r) = \tfrac{1}{4} r p'(r) + p(r)\omega. \tag{6.12}$$

The solutions are of the form

$$\begin{aligned}\epsilon_n &= a \exp[2\pi i(1+M)n/N + i\lambda t] \\ &\quad + b \exp[2\pi i(1-M)n/N - i\lambda t] \\ \delta_m &= c \exp[(2m+1)(1+M)\pi i/N + i\lambda t] \\ &\quad + d \exp[(2m+1)(1-M)\pi i/N - i\lambda t].\end{aligned} \tag{6.13}$$

Substituting into (6.9) and (6.10) gives:

$$\begin{aligned}&[\lambda + Q(r_0)]a + Ab + r_0^4 T_{1-M}(r_0^2)d = 0, \\ &Aa + [-\lambda + Q(r_0)]b + r_0^4 T_{1+M}(r_0^2)c = 0, \\ &T_{1+M}(r_0^2)b + [-\lambda + Q(r_0)]c + Ad = 0, \\ &T_{1-M}(r_0^2)a + Ac + [\lambda + Q(r_0)]d = 0,\end{aligned} \tag{6.14}$$

where

$$\begin{aligned}A &= S_{1+M} + P(r_0) + T_N(r_0^{-2}) \\ &= Q(r_0) - \tfrac{1}{2}M(N-M) + N^2/(r_0^N + r_0^{-N})^2,\end{aligned} \tag{6.15}$$

$$S_L = \sum_{k=1}^{N}{}' \frac{1 - \exp(-2\pi i L k/N)}{[1 - \exp(-2\pi i k/N)]^2} = \tfrac{1}{2}(N-L)(2-L), \tag{6.16}$$

$$T_L(x) = -x^{-2}T_{N-L+2}(x^{-1}) = \sum_{k=1}^{N} \frac{\exp[(2k+1)L\pi i/N]}{\{1 - x\exp[(2k+1)\pi i/N]\}^2} = \frac{Nx^{N-L}[(L-1)x^N - (N-L+1)]}{(1+x^N)^2}, \quad L = 1,\dots,N, \tag{6.17}$$

and where (6.1) has been used to evaluate $P(r_0^{-1})$ and $Q(r_0^{-1})$. There are nontrivial solutions of (6.14) only if $a = \pm r_0^2 d$ and $b = \pm r_0^2 c$. Then

$$[\lambda + Q(r_0) \pm r_0^2 T_{1-M}(r_0^2)]a + Ab = 0,$$
$$Aa + [-\lambda + Q(r_0) \pm r_0^2 T_{1-M}(r_0^2)]b = 0, \tag{6.18}$$

whence for nontrivial solutions:

$$\lambda_M^2 \mp r_0^2[T_{1+M}(r_0^2) - T_{1-M}(r_0^2)]\lambda_M + A^2 - [Q(r_0) \pm r_0^2 T_{1-M}(r_0^2)][Q(r_0) \pm r_0^2 T_{1+M}(r_0^2)] = 0, \tag{6.19}$$

λ is real and the Mth modes are stable if:

$$r_0^2[T_{1+M}(r_0^2) - T_{1-M}(r_0^2)]^2 > 4\{A^2 - [Q(r_0) \pm r_0^2 T_{1-M}(r_0^2)] \times [Q(r_0) \pm r_0^2 T_{1+M}(r_0^2)]\} \tag{6.20}$$

Making use of (6.15)–(6.17), one finds that for the stability of the staggered vortex street:

$$(2C \pm D)(4q - 2C \pm D) > 0, \quad M = 1,\dots,N, \tag{6.21}$$

where

$$C = (Q - A)/N^2 = \tfrac{1}{2}x(1-x) - \tfrac{1}{4}\operatorname{sech}^2(\tfrac{1}{2}y), \tag{6.22}$$

$$D = r_0^2[T_{1+M}(r_0^2) + T_{1-M}(r_0^2)]/N^2 = \frac{x\cosh[(1-x)y] - (1-x)\cosh(xy)}{2\cosh^2(\frac{1}{2}y)}, \tag{6.23}$$

$$q = Q/N^2, \quad x = M/N, \quad y = N\ln r_0^2. \tag{6.24}$$

Since the stability criterion is invariant under $r_0 \to r_0^{-1}$ and under $M \to N - M$, it is sufficient to suppose that $\frac{1}{2} \leqslant x \leqslant 1$ and $y \geqslant 0$.

When $x = 1$ $(M = N)$ then $D = \frac{1}{2}\operatorname{sech}^2(\frac{1}{2}y) = -2C$ whence the lower signs in (6.21) require that for stability

$$q < 0. \tag{6.25}$$

The upper signs give a left side of zero. These modes correspond to a small rotation of the system about the axis of revolution, and to a small enlargement of the separation of the rings of vortices. Both are stable modes. [Note that for each M there are four modes since (6.19) is quadratic in λ_M.]

B. Symmetric vortex streets

The vortices of a symmetric vortex street are situated initially at

$$r_n = r_0, \quad \phi = 2\pi i n/N, \quad n = 1,\dots,N, \quad \text{strength } \gamma,$$
$$r_m = r_0^{-1}, \quad \phi = 2\pi i m/N, \quad m = 1,\dots,N, \quad \text{strength } -\gamma. \tag{6.26}$$

Proceeding exactly as for the staggered case one finds that the configuration rotates rigidly about the axis of revolution with angular velocity

$$\omega^* = \frac{N}{1 - r_0^{2N}} - \frac{N}{2} - \frac{p(r_0)}{2} \tag{6.27}$$

and that it is stable if

$$(2E \pm F)(4q^* - 2E \pm F) > 0, \quad M = 1,\dots,N, \tag{6.28}$$

where

$$E = \tfrac{1}{2}x(1-x) + \tfrac{1}{4}\operatorname{csch}^2(\tfrac{1}{2}y)$$

$$F = \frac{(1-x)\cosh(xy) + x\cosh[(1-x)y]}{2\sinh^2(\frac{1}{2}y)} \tag{6.29}$$

$$q^* = [\tfrac{1}{4}r_0 p'(r_0) + p(r_0)\omega^*]/N^2. \tag{6.30}$$

Again we may suppose that $\frac{1}{2} \leqslant x \leqslant 1$ and $y \geqslant 0$.

When $x = 1$, $E = 2F = \frac{1}{2}\operatorname{csch}^2(\frac{1}{2}y)$. The upper signs in (6.28) then require that $q^* > 0$ for stability. The lower signs are interpreted as in the staggered case.

As shown in Appendix C, $2E \pm F > 0$, so that the criterion for stability is reduced to

$$4q^* - 2E - F > 0, \quad M = 1,\dots,N-1, \quad q^* > 0. \tag{6.31}$$

7. A SPECIAL CASE: THE SPHERE

Suppose the surface of flow is a sphere. Its line element in spherical polar coordinates is $ds^2 = R^2(d\theta^2 + \sin^2\theta d\phi^2)$. Introducing the coordinate $r = \tan(\frac{1}{2}\theta)$ one can rewrite the line element in harmonic form: $ds^2 = 4R^2(1 + r^2)^{-2}(dr^2 + r^2 d\phi^2)$, whence

$$h(r) = 2R(1 + r^2)^{-1}, \quad p(r) = (1 - r^2)/(1 + r^2) \tag{7.1}$$

A. Staggered vortex streets

From (6.5), (6.12), (6.24) and (7.1):

$$Q = -\tfrac{1}{4} - \tfrac{1}{4}\tanh^2(y/2N) + \tfrac{1}{2}N\tanh(y/2N)\tanh(\tfrac{1}{2}y). \tag{7.2}$$

Treating y and N as independent variables and differentiating, it is easily seen that:

$$\frac{\partial Q}{\partial y} \geqslant 0 \text{ if } y \geqslant 0 \text{ and } N \geqslant 1. \tag{7.3}$$

Thus, for each N there is exactly one y_N such that $Q(y_N, N) = 0$. Now, let $\alpha = y/2N$. Then one can write

$$Q = -\frac{\tanh\alpha}{4\alpha}\left(\frac{(1 + \tanh^2\alpha)\alpha}{\tanh\alpha} - y\tanh\tfrac{1}{2}y\right)$$

But

$$1 < \frac{\alpha(1 + \tanh^2\alpha)}{\tanh\alpha} < 1 + 2\alpha \quad \text{if } \alpha > 0$$

since

$$1 < \frac{\alpha}{\tanh\alpha} < 1 + \alpha \quad \text{if } \alpha > 0.$$

Therefore,

$Q < 0$ if $y\tanh\frac{1}{2}y < 1$, i.e., if $y < 1.55$

$Q > 0$ if $y\tanh\frac{1}{2}y > 1 + y/2N$.

Let $y^*_N \tanh y^*_N = 1 + y^*_N/2N$. It is easily shown that $dy^*_N/dN \leqslant 0$. Moreover, $y^*_N < 1.6$ if $N > 12$. Checking $N = 1,\ldots,12$ numerically one finds that $y_N < 1.6$ if $N > 3$ so that $Q > 0$ and the vortex street is unstable, if $y > 1.6$ and $N > 3$. If $y < 1.6$, then

$$\begin{aligned} &4q - 2C + D \\ &\quad > 4Q/N^2 - 2C > -N^{-2} - \tfrac{1}{4} + \tfrac{1}{4}N^{-2} + \tfrac{1}{2}\operatorname{sech}^2(\tfrac{1}{2}y) \\ &\quad > -3/(4N^2) + 0.0297 > 0 \end{aligned} \tag{7.4}$$

if $N > 5$. Hence if $y < 1.6$ and $N > 5$ there is instability if $2C + D < 0$. Numerical analysis shows that if $y < 1.6$ and $x = 0.58$ then

$$2C + D < \tfrac{1}{4} - \tfrac{1}{2}\operatorname{sech}^2(\tfrac{1}{2}y) + D < 0 \tag{7.5}$$

Since $\partial D/\partial x > 0$ (Appendix C), (7.4) holds for $\frac{1}{2} \leqslant x < 0.58$. If $N > 6$ there is always a mode with $x < 0.58$ [either $M = \frac{1}{2}N$ or $M = \frac{1}{2}(N + 1)$] so that there is always an unstable mode. Therefore all staggered vortex streets with $N > 6$ are unstable. Examining all other cases numerically one finds:

$N = 2$, stable if $0 < y < 1.10$, i.e., if $90° > \theta > 74.5°$,

$N = 3$, stable if $1.55 < y < 1.63$, i.e., if $74.6 < \theta < 75.4°$,

$N > 3$, unstable.

B. Symmetric vortex streets

On the sphere

$$\begin{aligned} Q^* &= -\tfrac{1}{4} - \tfrac{1}{4}\tanh^2(y/2N) + \tfrac{1}{2}N\tanh(y/2N)\coth(\tfrac{1}{2}y) \\ &\leqslant -\tfrac{1}{4} + \tfrac{1}{2}N, \text{ since } \tanh(ax)/\tanh x < 1, \\ &\qquad\qquad \text{if } a < 1. \end{aligned} \tag{7.6}$$

Therefore

$$4q^* - 2E + F < (2N - 1)/N^2 - x(1 - x). \tag{7.7}$$

If N is even one can put $x = \frac{1}{2}$, whence

$$4q^* - 2E + F < -(N^2 - 8N + 4)/N^2 < 0, \quad \text{if } N > 7. \tag{7.8}$$

If N is odd one can put $x = \frac{1}{2}(1 + 1/N)$, whence

$$4q^* - 2E + F < -(N^2 - 8N + 3)/N^2 < 0 \text{ if } N > 7. \tag{7.9}$$

Therefore all symmetric vortex streets with $N > 7$ are unstable. Upon examining (6.31) numerically one finds:

$N = 2$, stable if $y > 4.245$, i.e., $\theta < 38.2°$.

$N = 3$, stable if $y > 5.302$, i.e., $\theta < 44.9°$.

$N = 4$, stable if $y > 7.596$, i.e., $\theta < 42.3°$.

$N = 5$, stable if $y > 10.430$, i.e., $\theta < 38.8°$.

$N = 6$, stable if $y > 17.76$, i.e., $\theta < 25.7°$.

$N = 7$, unstable

One sees, then, that the curvature of the surface of flow produces qualitatively different results than those of von Karman for flows in the plane. In particular, there are both staggered and symmetric stable vortex streets.

ACKNOWLEDGMENTS

The author is indebted to Professor F.A. Kaempffer for initiating this work and for his advice and criticism. The author has benefited from discussions with D.M.F. Chapman and G. Mertz. The financial assistance of the University of British Columbia in the form of a U.B.C. Graduate Fellowship is gratefully acknowledged.

APPENDIX A: MATHEMATICAL FORMALISM[16]

Let M be a two-dimensional manifold with metric tensor g. One may define a two-form σ^* by: $\sigma^* = \epsilon(\det g)^{\frac{1}{2}}$, where ϵ is the antisymmetric tensor density with $\epsilon^{12} = 1$. There is a natural extension of σ^* to a two-form on the $2N$-dimensional manifold M^N, namely, the unique two-form σ satisfying: $\sigma(dx^1 \times \ldots \times dx^N, dy^1 \times \ldots \times dy^N) = \Sigma_{n=1}^N \gamma_n \times \sigma^*(dx^n, dy^n)$, where γ_n, $n = 1,\ldots,N$, are constants. Note that $\ker\sigma = 0$ and σ is differentiable everywhere. σ thus induces a symplectic structure on M^N (N.B. $\nabla\sigma^* = 0$ since nontrivial three-forms cannot exist on a two-dimensional manifold).

Suppose now that Ω is some scalar function on M^N. A natural flow is induced, the equations of motion of which are

$$\frac{d\mathbf{x}}{dt} = \sigma^{-1}\nabla\Omega \equiv \operatorname{grad}\Omega, \tag{A1}$$

where $\mathbf{x}$ denotes position on M^N and ∇ is the exterior derivative. Notice that $d\Omega/dt = \nabla\Omega d\mathbf{x}/dt = \sigma(\Omega,\Omega) = 0$ so that Ω is conserved. In harmonic coordinates $\sigma^* = h^2(x, y)\epsilon$ and (A1) becomes (4.2).

In harmonic coordinates, the requirement that G generate an infinitesimal coordinate transformation, $X = \operatorname{grad}G$, becomes:

$$\begin{aligned} X^n &= (\gamma_n h^2(x_n, y_n))^{-1}\frac{\partial G}{\partial y_n}, \quad n = 1,\ldots,N, \\ X^n &= -(\gamma_n h^2(x_n, y_n))^{-1}\frac{\partial G}{\partial x_n}, \quad n = 1,\ldots,N. \end{aligned} \tag{A2}$$

These may be used to generate the constants of the motion in Sec. 5.

APPENDIX B

All the special sums necessary for the calculations of Sec. 6 can be evaluated easily once:

$$R_M(z) = \sum_{n=1}^{N} \frac{\exp(2\pi iMn/N)}{[1 - z\exp(2\pi in/N)]^2}, \qquad z \text{ complex}, \ M = 1,\ldots,N \tag{B1}$$

is known. Suppose first that $|z| < 1$. Then

$$R_M(z) = \sum_{n=1}^{N} \exp(2\pi iMn/N) \times \sum_{k=1}^{\infty} kz^{k-1}\exp[2\pi in(k-1)/N] \tag{B2}$$

The infinite series is absolutely convergent allowing the reordering of the sums:

$$R_M(z) = \sum_{k=1}^{\infty} kz^{k-1}\sum_{n=1}^{N}\exp[2\pi i(M + k - 1)n/N] \tag{B3}$$

The second sum vanishes unless $M + k - 1 = rN$, r an integer.

$$R_M(z) = \sum_{r=1}^{\infty} N(rN - M + 1)z^{rN-M}$$

$$= N d/dz[z^{N-M+1}/(1-z^N)]$$
$$= \frac{N[(N-M+1)z^{N-M}+(M-1)z^{2N-M}]}{(1-z^{2N})^2}. \tag{B4}$$

Both the right side of (B4) and the right side of (B1) are analytic in all regions of the complex plane excluding the $2N$th roots of 1; hence, by analytic continuation, (B4) is valid for all z.

To evaluate $T_M(x)$ put $z = xe^{\pi i/N}$.

$$S_M = \lim_{z\to 1}[R_N(z) - R_M(z)],$$

which is evaluated straightforwardly using l' Hopital's Rule.

APPENDIX C: PROPERTIES OF D AND $2E \pm F$

a) Let $f(x,y) = \cosh[y(1-x)] + (1-1/x)\cosh(yx)$. Then

$$\frac{\partial f}{\partial x} = x^{-2}\cosh(yx)[1 - xy\tanh(yx)]$$
$$+ y(\sinh(yx) - \sinh(y(1-x)))$$
$$> 0 \ \text{if}\ \tfrac{1}{2} \leqslant x \leqslant 1 \ \text{and}\ y < 1.2/x.$$

One can always choose $\frac{1}{2} \leqslant x \leqslant \frac{3}{4}$ whence if $y < 1.6$, $\partial f/\partial x > 0$ and $f(x,y) \geqslant f(\frac{1}{2}, y) = 0$. Therefore, $D = \frac{1}{2}x f(x,y) \times \mathrm{csch}^2(\frac{1}{2}y) > 0$ and $\partial D/\partial x = \frac{1}{2}\mathrm{csch}^2(\frac{1}{2}y)\partial/\partial x(xf) \gg 0$ if $y < 1.6$.

b) Let $g(x,y) = (1-x)\cosh(xy) + x\cosh[(1-x) \times y] - \cosh(\frac{1}{2}y)$. Then

$$\frac{\partial g}{\partial y} = x(1-x)\sinh(xy) - \sinh[(1-x)y] - \tfrac{1}{2}\sinh\tfrac{1}{2}y$$
$$\leqslant \tfrac{1}{4}(2\sinh\tfrac{1}{2}y) - \tfrac{1}{2}\sinh\tfrac{1}{2}y = 0 \quad \text{if}\ \tfrac{1}{2} \leqslant x \leqslant 1.$$

Therefore, $g(x,y) \geqslant g(x,0) = 0$ whence $0 \leqslant F \leqslant \frac{1}{2}\cosh\frac{1}{2}y \times \mathrm{csch}^2\frac{1}{2}y$. Thus $2E \pm F \geqslant 2E - F \geqslant \frac{1}{4} + \frac{1}{2}\mathrm{csch}^2\frac{1}{2}y - \frac{1}{2}\cosh\frac{1}{2}y \times \mathrm{csch}^2\frac{1}{2}y = \frac{1}{4}(1 - \mathrm{sech}\frac{1}{4}y) \geqslant 0$

[1]H. von Helmholtz, Crelle's Journ. (1858), Trans. Philos. Mag. Ser. 4, No. 226 Supp. 33, 485 (1867).
[2]C.C. Lin, *On the Motion of Vortices in Two Dimensions* (Univ. Toronto, Toronto, 1943).
[3]G.K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge U.P., Cambridge, England, 1967), p. 530.
[4]H. Lamb, *Hydrodynamics*, (Cambridge, U.P., Cambridge, England, 1916), 4th ed., Art. 80, Ch. IV, p. 101.
[5]T. von Karman, Phys. Zeitschr. **13**, 49 (1912).
[6]T.C. Yeh, J. Meteor. **7**, 108 (1950).
[7]H. Kuo, J. Meteor. **7**, 247 (1950).
[8]F.A. Berson, Tellus **3**, 172 (1951).
[9]G.K. Batchelor, op. cit., p. 510.
[10]Other objections are voiced by R.W. James, J. Meteor. **9**, 447 (1952).
[11]See for example L.P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J., 1926), Sec. 28.
[12]P. Koebe, Acta Math. **41**, 306 (1918).
[13]See for example J.B. Conway, *Functions of One Complex Variable* (Springer-Verlag, New York, 1973), p. 254.
[14]E.J. Routh, Proc. L.M.S. **12**, 83 (1881).
[15]J.J. Thomson, *Treatise on the Motion of Vortex Rings* (Adams Prize Essay 1882-MacMillan, London, 1883), p. 95.
[16]For further reference see J.-M. Souriau, *Structure des Systemes Dynamique* (Dunod, Paris, 1969).

ERRATA

Erratum: Exact solutions of some multiplicative stochastic processes [J. Math. Phys. 20, 45 (1979)]

K. Wódkiewicz
Institute of Theoretical Physics, Warsaw University, Warsaw 00-681, Poland and Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

Because of a printing error Eq. (2.4) should read:

$$Z[\mathscr{J},\mathscr{J}^*] = \exp\Big[-\tfrac{1}{2}\int d\tau_1\, d\tau_2 (\mathscr{J}^*(\tau_1),\mathscr{J}(\tau_1)) \times \Delta(\tau_1,\tau_2)\begin{pmatrix}\mathscr{J}^*(\tau_2)\\ \mathscr{J}(\tau_2)\end{pmatrix}\Big].$$

Erratum: Note on the stability of the Schwarzschild metric [J. Math. Phys. 20, 1056 (1979)]

Robert M. Wald
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637

The first sentence below Eq. (10) should state dom$A \subset$ dom$A^{1/2}$. Equation (14) should be corrected to read

$$\int |f(t,r_*)|^2\, dr_* \leqslant 2\|f_0\|^2 + 2\|A^{-1/2}\dot{f}_0\|^2. \qquad (14)$$

[That $\dot{f}_0$ lies in dom$A^{-1/2}$—and thus that $\|A^{-1/2}\dot{f}_0\|$ is finite—may be verified as follows. Except for the $l=0$ scalar case (where stability can easily be proven directly), for the radiative modes of scalar, electromagnetic, or gravitational perturbations, analysis of the static solutions shows that for any C^∞ compact support $\dot{f}_0$ one can find a vector $\psi \in$ dom$A$ such that $A\psi = \dot{f}_0$. Thus $\dot{f}_0 \subset$ dom$A^{-1} \subset$ dom$A^{-1/2}$.]

Equation (19) should be corrected to read,

$$|f(t,r_*)|^2 \leqslant \int |f_0|^2 dr_* + \|A^{-1/2}\dot{f}_0\|^2 + \tfrac{1}{2}\int \bar{f}_0 A f_0\, dr_* + \tfrac{1}{2}\int |\dot{f}_0|^2 dr_*. \qquad (19)$$

The stability conclusions are, of course, unaffected.

Erratum: Characteristic surfaces and characteristics initial data for the generalized Einstein–Maxwell field equations [J. Math. Phys. 20, 1745 (1979)]

Gregory Walter Horndeski
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 4 September 1979; accepted for publication 3 October 1979)

The following remark should be added immediately before the last paragraph of Sec. 3.

Remark: Since the data presented in sets I–VIII is characteristic, Eq. (2.12) implies that this data must satisfy a further constraint, viz., that the 9×1 vector appearing on the rhs of Eq. (2.12) must lie in the image of the 9×9 characteristic matrix. This condition is met by the data presented in sets II, V, and VII. The data given in sets I and IV will satisfy this constraint provided $\Lambda = 0$, in which case the resulting data is a special case of the data presented in sets II and V respectively. The data given in set VIII will satisfy the required condition when $B^\alpha B_\alpha \neq -1/k$. At present it does not appear to be possible to modify the data given in sets III and VI to be compatible with the additional constraint.

 0022-2488/80/010218-01$00.00