On the stationary Einstein–Maxwell equations

C. Hoenselaers

Research Institute for Fundamental Physics, Kyoto University, Kyoto, Japan [a)]
(Received 11 April 1977)

The ansatz $\Phi = \Phi(\mathcal{E}, \bar{\mathcal{E}})$ for a solution of the stationary Einstein–Maxwell equations is analyzed. The possible forms of this function are listed and it is shown that one obtains from every solution of the vacuum Ernst equations an at most two-parameter solution of the Einstein–Maxwell equations.

1. INTRODUCTION

It is well known that the problem of stationary Einstein–Maxwell fields can be formulated in a three-dimensional manifold with two complex functions $\mathcal{E}$ and Φ, the Ernst potential and the electromagnetic potential, in it.[1–4] It has furthermore been shown that the internal symmetry group of the equations emerging from that formalism is isomorphic to SU(2, 1).[2,4] Still it is a formidable task to solve those equations, and one is thus tempted to restrict oneself to solutions not incorporating some of the fields, for instance the pure vacuum fields or the electrostatic fields. Also the ansatz $\Phi =$ linear function of $\mathcal{E}$ has been made and used to obtain the charged version of the Kerr and Tomimatsu–Sato solutions from the uncharged ones.[1,5]

While this ansatz is fairly obvious, we shall derive in Sec. 3 by the method outlined in Sec. 2 the most general form in which an ansatz $\Phi = \Phi(\mathcal{E}, \bar{\mathcal{E}})$ may be made.

Throughout the paper ∇ will denote the covariant derivative operator and we shall suppress coordinate indices.

2. THE METHOD

Suppose we are given a system of N interacting massless fields, denoted by f^a, and the field equations are to be derived from the Lagrangian

$$L = G_{ab}(f^m)\nabla f^a \nabla f^b, \tag{2.1}$$

(cf. Refs. 4,6), where G_{ab} is nonsingular. The field equations are

$$\nabla^2 f^a + \Gamma^a_{bc}\nabla f^b \nabla f^c = 0, \tag{2.2}$$

where the Γ's are the usual Christoffel symbols formed with respect to the G's.

As an ansatz for a solution of (2.2) we take here at most $N-1$ functions of the f^a and assume them to vanish identically, i.e.,

$$F^A(f^m) \equiv 0. \tag{2.3}$$

This implies

$$\nabla F^A = F^A{}_{;a}\nabla f^a \equiv 0,$$

$$\nabla^2 F^A = F^A{}_{;ab}\nabla f^a \nabla f^b \equiv 0.$$

Combining those, we obtain

$$F^A{}_{;ab} h^a_m h^b_n \nabla f^m \nabla f^n \equiv 0, \tag{2.4}$$

where h^a_b is the projection tensor onto the submanifold of the space whose metric is given by G_{ab} defined by (2.3).

One may now use (2.3) to replace some of the field equations (2.2) by (2.4). If we chose the F^A to satisfy

$$F^A{}_{;ab} h^a_m h^b_n = 0, \tag{2.5}$$

then (2.4) is satisfied identically and we have reduced the number of variables in (2.1) and (2.2). Equation (2.5) is, of course, satisfied if

$$F^A{}_{;ab} = 0,$$

but this implies the existence of a covariantly constant vector and is thus only possible if G_{ab} is decomposible.

To summarize: If we want to simplify the Lagrangian (2.1) and (2.2) by reducing the number of independent variables, we have to solve (2.5).

3. STATIONARY EINSTEIN-MAXWELL FIELDS

We now apply the method of the preceeding section to stationary Einstein–Maxwell fields. For the definitions let us recall the following: Let M be a four-dimensional manifold with metric g (signature $-+++$) satisfying the Einstein–Maxwell equations with electromagnetic field tensor $\mathbf{F}$ and admitting a Killing vector ξ. Define λ, ω, ϵ, β, ϕ so that

$$\lambda = -\xi\xi > 0, \quad \omega = \epsilon(\xi\nabla\xi) = \nabla\phi + \epsilon\nabla\beta - \beta\nabla$$

$$\nabla\epsilon = (1/\sqrt{2})\mathbf{F}\xi, \ \nabla\beta = (1/\sqrt{2})\,\boldsymbol{\epsilon}(\xi\mathbf{F}) \quad (\boldsymbol{\epsilon} = \text{Levi-Civita tensor}).$$

The Einstein–Maxwell equations can be formulated in the manifold S of trajectories of ξ in M, endowed with the metric

$$\mathbf{h} = \lambda \mathbf{g} + \xi \circ \xi.$$

(For details of the derivation cf., e.g., Ref. 2.) The Lagrangian is

$$L_{\text{tot}} = 2R^{(3)} + L,$$

$$L = (1/\lambda^2)[\nabla\lambda^2 + (\nabla\phi + \epsilon\nabla\beta - \beta\nabla\epsilon)^2 - 2\lambda(\nabla\epsilon^2 + \nabla\beta^2)]. \tag{3.1}$$

Considering first only one function of ϕ, λ, ϵ, β to vanish, i.e.,

$$\phi = \phi(\lambda, \epsilon, \beta), \tag{3.2}$$

one finds after fairly long calculations that this ansatz does not provide a solution of (2.5).

Rewriting (3.1) in terms of the variables

$$\mathcal{E} = \lambda - \tfrac{1}{2}\Phi\bar{\Phi} + i\phi, \qquad \Phi = \epsilon + i\beta, \tag{3.3}$$

one has

[a)] Present address: Raman Research Institute, Hebbal P.O., Bangalore 560 006, India.

 0022-2488/78/1903-0539$1.00

$$L=(1/\lambda^2)(\nabla\mathcal{E}\nabla\bar{\mathcal{E}}+\Phi\nabla\mathcal{E}\nabla\bar{\Phi}+\bar{\Phi}\nabla\bar{\mathcal{E}}\nabla\Phi-(\mathcal{E}+\bar{\mathcal{E}})\nabla\Phi\nabla\bar{\Phi}), \tag{3.4}$$

and assuming

$$\Phi=\Phi(\mathcal{E},\bar{\mathcal{E}}), \tag{3.5}$$

one finds for (2.5) the equations (a dot and a prime denote differentiation with respect to $\mathcal{E}$ and $\bar{\mathcal{E}}$, respectively; thus under complex conjugation $\mathcal{E}$ goes into $\bar{\mathcal{E}}$, Φ into $\bar{\Phi}$, dot into prime, and vice versa)

$$\begin{aligned}&\ddot{\Phi}=0,\\ &\Phi''=(1/\lambda)\Phi'(\bar{\Phi}\Phi'-\Phi\bar{\Phi}'-1),\\ &2\dot{\Phi}'=(1/\lambda)\Phi'(1+\bar{\Phi}\dot{\Phi}-\Phi\dot{\bar{\Phi}}),\end{aligned} \tag{3.6}$$

and their complex conjugates. The integrability conditions are

$$2\dot{\Phi}'\dot{\bar{\Phi}}+\ddot{\bar{\Phi}}\Phi'=0 \tag{3.7}$$

or

$$2\ln\Phi'+\ln\dot{\bar{\Phi}}=h(\bar{\mathcal{E}}),$$

from which by use of (3.6) follows

$$(\Phi''/\Phi')^{\cdot}=0.$$

With

$$\Phi=f(\bar{\mathcal{E}})\mathcal{E}+g(\bar{\mathcal{E}}),$$

one finds

$$f(\bar{\mathcal{E}})=\alpha g(\bar{\mathcal{E}})+\beta,$$

and then from (3.7)

$$\Phi=\alpha\mathcal{E}+\beta+(\mathcal{E}+\delta)/(\bar{\mathcal{E}}+\delta). \tag{3.8}$$

By inserting this expression into (3.6) one obtains conditions on the constants α, β, γ, δ:

$$\begin{aligned}&1+\beta\bar{\alpha}-\alpha\bar{\gamma}-\alpha\bar{\alpha}\delta=0,\\ &\beta\bar{\beta}-\gamma\bar{\gamma}-\delta-\alpha\bar{\beta}\delta-\alpha\bar{\gamma}\bar{\delta}=0.\end{aligned} \tag{3.9}$$

Before entering into further calculations we observe that we still have the freedom of performing Kinnersley transformations. In particular, the following can be used to simplify (3.8):

$$\begin{aligned}&\Phi\to e^{ia}\Phi, && \mathcal{E}\to\mathcal{E},\\ &\Phi\to a\Phi, && \mathcal{E}\to a^2\mathcal{E},\\ &\Phi\to\Phi, && \mathcal{E}\to\mathcal{E}+ia.\end{aligned}$$

We can transform such that $\delta=d=1,\ -1,\ 0,\ \alpha=a$ = real. In the case $d=0$ a can, moreover, be transformed to 1. Thus one finds from (3.9) that

$$\begin{aligned}&\Phi=a\mathcal{E}+\beta+(\mathcal{E}+d)/(\bar{\mathcal{E}}+d),\\ &-\beta=\gamma+ad=\tfrac{1}{2}(1/a+ib),\\ &d=1,\ -1,\ 0,\quad d=0\to a=1.\end{aligned} \tag{3.10a}$$

For completeness one has to consider the cases for $\Phi''=\Phi^{\cdot\prime}=0$. They are in already simplified form:

$$\Phi=a(\mathcal{E}+d), \tag{3.10b}$$

$$\Phi=\mathcal{E}-\bar{\mathcal{E}}-\tfrac{1}{2}. \tag{3.10c}$$

The question if the relations (3.10) are covariant under Kinnersley's SU(2,1) group can be answered by simply arguing that the infinitesimal Kinnersley transformations are just the Killing vectors of the metric defined by (3.4) and thus (3.10), being the solution of (2.5), has to be covariant under that group. Another way of obtaining this result is using the variables in Kinnersley's space y^a ($\mathcal{E}=y^1/y^2$, $\Phi=y^3/2y^2$), where (3.10) can be written in the form

$$\bar{y}^a A_{ab} y^b=0\quad(|A_{ab}|=0),$$

which is manifestly covariant under Lorentz rotations in that space. This form of Eqs. (3.10a), (3.10c) also serves to answer the question whether they are connected by some Kinnersley transformations. The matrices A are

$$\begin{pmatrix}-a & \beta & \frac{1}{2}\\ -\beta & -d^2a & -\frac{1}{2}\\ 0 & 0 & 0\end{pmatrix}$$

and

$$\begin{pmatrix}0 & 1 & 0\\ -1 & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & 0\end{pmatrix},$$

respectively. Define the interior product of two vectors in Kinnersley's space by

$$(\bar{a},b):=\eta_{ab}\bar{a}^a b^b,\quad \eta=\begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 1 & 1\end{pmatrix}.$$

One finds then that one of the eigenvectors of the second matrix is a null vector, while none of the eigenvectors of the first matrix has this property. Thus the two matrices cannot be transformed into each other by Kinnersley's SU(2,1) transformations.

Equation (3.10b) has been used by Ernst to "charge" the Kerr and Tomimatsu–Sato solutions. Equation (3.10c) describes magnetostatic solutions as one can easily calculate that $\epsilon=-\frac{1}{2}$ and $\beta=2\phi$, and consequently the twist of the Killing vector vanishes.

Finally the Lagrangian (3.4) becomes

$$\begin{aligned}\lambda^2L=&[\Phi-(\mathcal{E}+\bar{\mathcal{E}})\dot{\Phi}]\dot{\bar{\Phi}}\nabla\mathcal{E}^2+[\bar{\Phi}-(\mathcal{E}+\bar{\mathcal{E}})\bar{\Phi}']\Phi'\nabla\bar{\mathcal{E}}^2\\ &+[1+\Phi\bar{\Phi}'+\bar{\Phi}\dot{\Phi}-(\mathcal{E}+\bar{\mathcal{E}})(\dot{\Phi}\bar{\Phi}'+\dot{\bar{\Phi}}\Phi')]\nabla\mathcal{E}\nabla\bar{\mathcal{E}},\end{aligned} \tag{3.11}$$

and one can calculate, using (3.6) via

$$R_{ab}=-KG_{ab},$$

the Gaussian curvature of the hypersurfaces defined by (3.10). One finds in the case of (3.10b) $K=-1$ while $K=-\frac{1}{4}$ in the other cases. In the course of the calculation one encounters for $\Phi'\neq 0$ the relation

$$(\mathcal{E}+\bar{\mathcal{E}})(\dot{\bar{\Phi}}\Phi'-\bar{\Phi}'\dot{\Phi})+\bar{\Phi}\dot{\Phi}+\Phi\bar{\Phi}'+1=0,$$

which is identically satisfied for Φ given by (3.10a), (3.10c) and may be used to write the last term in Eq. (3.11) in the form

$$- 2(\mathcal{E}+\bar{\mathcal{E}})\mathring{\bar{\Phi}}\Phi'.$$

It also may be used to verify that in this case

$$|G_{ab}| = (2/\lambda^3)\dot{\bar{\Phi}}\Phi'.$$

For Φ given by (3.10a) Eq. (3.11) may be cast into the form

$$L = [-16/(1+mn)^2]\,\nabla m \nabla n, \tag{3.12a}$$

while the transformations accomplishing that are given by

$$\mathrm{Re}\,\frac{1}{\mathcal{E}+d} - \frac{d}{2} = \frac{1}{2}\left(1 + \frac{a(\gamma+\bar{\gamma})}{2\gamma\bar{\gamma}d}\right)^{1/2} \frac{m-n}{m+n},$$

$$\mathrm{Im}\,\frac{1}{\mathcal{E}+d} = \frac{ab}{2[2\gamma\bar{\gamma}d(a(\gamma+\bar{\gamma})+2\gamma\bar{\gamma}d)]^{1/2}}\,\frac{m-n}{m+n} + \frac{1}{2[a(\gamma+\bar{\gamma})+2\gamma\bar{\gamma}d]^{1/2}}\,\frac{mn-1}{m+n} \tag{3.13a'}$$

for $d \neq 0$ and

$$\mathrm{Re}\frac{1}{\mathcal{E}} = \frac{1}{8\gamma\bar{\gamma}}\left(\frac{m}{n} - 1\right),\qquad \mathrm{Im}\frac{1}{\mathcal{E}} = \frac{1}{4\gamma\bar{\gamma}}\left[\frac{1}{2}\left(m - \frac{1}{n}\right) - b\right], \tag{3.13a''}$$

for $d=0$ and $a=1$. The remaining cases are fairly trivial; we list them for completeness:

Case (3.10b):

$$L = \frac{4}{(k+\bar{k})^2}\nabla k \nabla \bar{k}, \tag{3.12b}$$

$$k = \frac{\sqrt{d}(c+1)\mathcal{E} + \sqrt{d}(c-1)}{\sqrt{d}(c-1)\mathcal{E} + \sqrt{d}(c+1)}, \quad c = (1+2a^2d)^{1/2} \quad (d \neq 0) \tag{3.13b'}$$

$$k = \tfrac{1}{2} + 1/\mathcal{E}, \quad d=0, \quad a=1. \tag{3.13b''}$$

Case (3.10c):

$$L = \frac{16}{(m+n)^2}\nabla m \nabla n, \tag{3.12c}$$

$$\left.\begin{matrix} m \\ n \end{matrix}\right\} = \sqrt{\lambda} \pm \frac{1}{\sqrt{2}}\phi. \tag{3.13c}$$

The Lagrangians (3.12) are very similar to or the same as the Ernst Lagrangian. Thus for every solution of the Ernst equations one obtains by the transformations (3.13) and the relations (3.10) a one- or two-parameter solution of the Einstein–Maxwell equations. It is, however, straightforward to check that the easiest solution of the field equations belonging to (3.12a) in prolate spheroidal coordinates (cf. Ref. 7) $m=x$, $n=-x$ gives, by (3.13a″), rise to a solution which is not asymptotically Schwarzschild like, as $\lambda \approx 4\gamma\bar{\gamma} + \mathrm{const}\, x^{-2}$. The physical meaning of the parameters a, b, is thus far from obvious.

ACKNOWLEDGMENT

I am grateful to Professor H. Sato and Professor R.M. Misra for helpful discussions.

[1]F. Ernst, Phys. Rev. 168, 1415 (1968).
[2]W. Kinnersley, J. Math. Phys. 14, 651 (1973).
[3]R.O. Hansen, J. Math. Phys. 15, 46 (1974).
[4]G. Neugebauer, D. Kramer, Ann. Physik 24, 62 (1969).
[5]F. Ernst, Phys. Rev. D 7, 2520 (1973).
[6]C. Hoenselaers, J. Math. Phys. 17, 7 (1976).
[7]F. Ernst, J. Math. Phys. 15, 1409 (1974).

Evaluation of simple Feynman graphs

Bernie G. Nickel

Physics Department, University of Guelph, Guelph, Ontario, Canada N1G 2W1
(Received 29 April 1977)

A number of one-loop graphs with arbitrary external momenta and internal masses contributing to the perturbation expansion of a Euclidean ϕ^4 theory are evaluated exactly in three and two dimensions. The final expressions are simple closed forms involving elementary functions only. A method for handling the multidimensional angular integrations that arise in calculations of massless QED or ϕ^4 in four dimensions is also discussed.

I. INTRODUCTION

The calculations described below have been motivated by the recent interest shown in the ϕ^4 theory as a model for second-order phase transitions observed in three and two dimensions. Critical exponents and correlation functions can in principle be calculated in perturbation theory by using either the ϵ-expansion method developed by Wilson[1] or renormalized perturbation methods in the space dimension of interest.[2] However, in either case it has become apparent[3,4] that the series are such that high order terms must be calculated before accurate quantitative information can be obtained. It is therefore essential that one devise efficient methods for handling the multidimensional integrals appearing in the expansion.

Comparison of different integration procedures reveals that the situation in four dimensions is rather different from that in three and two. With each increasing order in the perturbation expansion of the ϕ^4 theory one must evaluate graphs with two additional propagators and one additional loop integral. Since in four dimensions each loop integral is an extra four-dimensional integral in momentum space, it is generally preferable to evaluate the graphs in Feynman parameter space where one parameter is associated with each propagator and only two extra integrals in each order are required. A direct momentum space evaluation seems preferable only when all masses can be set equal to zero and every propagator can be expressed in terms of the difference of two momenta. The propagators then have a simple expansion in four-dimensional hyperspherical harmonics and the angular integrations decouple from the integrations over momentum magnitudes. Furthermore, the magnitude integrals are of the form $\int k^n dk$ and thus trivial: the angular integrals can be reduced to well-known integrals if one exploits the connection between the four-dimensional hyperspherical harmonics and the three-dimensional rotation matrices.[5] A particular example of this procedure, namely the evaluation of the 12-dimensional angular integral described in Rosner[6] as the analog of the Racah coefficient, is described in Sec. V.

In three dimensions, the simple counting argument given above still suggests Feynman parameter methods preferable to direct integration; in two dimensions both methods require roughly the same number of integrations. However, by using the analytical expressions derived in Secs. II and III for single-loop graphs, the dimensionality of the integrals that must be handled by numerical quadrature methods is actually less in the momentum space representation. Furthermore, because the explicit formulas given below for the three- and two-dimensional one-loop graphs involve only elementary functions of the external momenta, these formulas are very convenient for use in any subsequent numerical integration.

The results of the one-loop graph integrations can be expressed in terms of the dimensionless invariants which are conventionally used in discussions of the Landau singularities of these graphs.[7] Let the internal propagators of an n-vertex one-loop graph be $(m_i^2+K_i^2)^{-1}$, $i=1,2,\ldots,n$ and define the external momenta

$$\mathbf{k}_{ij}=\mathbf{K}_i-\mathbf{K}_j\,. \tag{1}$$

The dimensionless parameters

$$y_{ij}=(m_i^2+m_j^2+k_{ij}^2)/(2m_im_j), \tag{2}$$

together with the masses m_i, then completely characterize the graph. From the $n\times n$ matrix y_{ij} construct first the determinant $D^{(n)}$,

$$D^{(n)}=\det|y_{ij}|\,, \tag{3}$$

then the determinants $F_l^{(n)}$ obtained by replacing the elements y_{il} in the lth column by m_l/m_i,

$$F_l^{(n)}=\det|y_{ij}(1-\delta_{jl})+(m_l/m_i)\delta_{jl}| = \frac{\partial}{\partial m_l^2}(m_l^2D^{(n)}), \tag{4}$$

and finally the principle minors $D_l^{(n-1)}$ obtained by eliminating the l^{th} row and column from y_{ij},

$$D_l^{(n-1)}=\det|y_{ij}|\,,\ i,j\neq l. \tag{5}$$

The subscript l will be dropped from $D_l^{(n-1)}$ if no ambiguity can result. The three-dimensional triangle graph is then given by

$$T^{(3d)}=m_1m_2m_3\pi^{-2}\int d^3K_1\prod_{i=1}^{3}(m_i^2+K_i^2)^{-1} =\arctan\left((D^{(3)})^{1/2}/C\right)/(D^{(3)})^{1/2}, \tag{6}$$

where

$$C=1+y_{12}+y_{13}+y_{23}. \tag{7}$$

For physically allowed momenta, $D^{(3)}$ is nonnegative and the arctan lies in the interval $[0,\pi/2]$. Henceforth it will be understood that square root denotes positive square root, and arctan and ln respectively denote the principal branch of Arctan and Ln, whenever the momenta lie in a physically allowed region. The three-dimensional box graph is

0022-2488/78/1903-0542$1.00

$$Q^{(3d)} = m_1 m_2 m_3 m_4 \pi^{-2} \int d^3K_1 \prod_{i=1}^{4} (m_i^2 + K_i^2)^{-1}$$
$$= \tfrac{1}{2}\left(\sum_{l=1}^{4} \frac{1}{m_l} F_l^{(4)} T_l^{(3d)}\right) / D^{(4)}, \qquad (8)$$

where $T_l^{(3d)}$ are the three-dimensional triangles

$$T_l^{(3d)} = \arctan\left((D_l^{(3)})^{1/2}/C_l\right)/(D_l^{(3)})^{1/2}. \qquad (9)$$

The C_l are constructed from the elements of $D_l^{(3)}$ exactly as C is from $D^{(3)}$ [cf. Eq. (7)]. The two-dimensional triangle is

$$T^{(2d)} = m_1 m_2 m_3 \pi^{-1} \int d^2K_1 \prod_{i=1}^{3} (m_i^2 + K_i^2)^{-1}$$
$$= \tfrac{1}{2}\left(\sum_{l=1}^{3} \frac{1}{m_l} F_l^{(3)} B_l^{(2d)}\right) / D^{(3)}, \qquad (10)$$

where the $B_l^{(2d)}$ are the two-dimensional bubble graphs

$$B_3^{(2d)} = m_1 m_2 \pi^{-1} \int d^2K_1 (m_1^2 + K_1^2)^{-1} (m_2^2 + K_2^2)^{-1}$$
$$= \ln(y_{12} + (y_{12}^2 - 1)^{1/2})/(y_{12}^2 - 1)^{1/2} \qquad (11)$$

and $B_1^{(2d)}$ and $B_2^{(2d)}$ are obtained by a cyclic permutation of 1,2,3. The three-dimensional bubble, included for the sake of completeness, is

$$B^{(3d)} = m_1 m_2 \pi^{-2} \int d^3K_1 (m_1^2 + K_1^2)^{-1} (m_2^2 + K_2^2)^{-1}$$
$$= 2 m_1 m_2 \arctan(k_{12}/(m_1 + m_2))/k_{12}. \qquad (12)$$

The evaluation of the bubble graphs is an elementary excercise and is not discussed in this paper. The three-dimensional triangle graph is determined in Sec. II by a direct evaluation of the integral in Eq. (6) in momentum space and again in Sec. III by Feynman parameter methods. The two-dimensional triangle graph has been evaluated as an intermediate step in four-dimensional calculations,[8] but can also be obtained, together with the three-dimensional box graph, by using our analysis in Sec. III. There we show that in d dimensions, any one-loop n-vertex graph for $n \geq d+1$ is related to the $(n-1)$-vertex graphs by connection formulas such as (8) and (10). Another example is the four-dimensional formula

$$P^{(4d)} = \tfrac{1}{2}\left(\sum_{l=1}^{5} \frac{1}{m_l} F_l^{(5)} Q_l^{(4d)}\right) \Big/ D^{(5)} \qquad (13)$$

for the one-loop, five-vertex "production" graph P in terms of the known[8,9] one-loop, four-vertex scattering graphs.

Although the expressions given above for the one-loop graphs are in some sense the simplest possible, they are not always the most suitable if their actual numerical value is needed for, say, use in numerical integration routines. Therefore, in Sec. IV we give a few alternative formulas which are both simpler to evaluate and more stable against roundoff error.

The four-dimensional "Racah" coefficient, derived in Sec. V, is

$$R = \left(\prod_{i=1}^{4} \int \frac{d\Omega_i}{2\pi^2}\right) C_l(\cos\theta_{12})\, C_m(\cos\theta_{13})\, C_n(\cos\theta_{23})$$
$$\times C_p(\cos\theta_{34})\, C_q(\cos\theta_{24})\, C_r(\cos\theta_{14}) \qquad (14)$$
$$= \begin{Bmatrix} p/2 & q/2 & r/2 \\ l/2 & m/2 & n/2 \end{Bmatrix}^2,$$

where the brace expression on the right is the 6-j symbol in three dimensions.[10] The integrals in (14) are over the four-dimensional solid angles Ω_i that respectively specify the directions of vectors $\mathbf{k}_i$. The Chebyshev polynomial $C_l(\cos\theta_{ij}) = \sin((l+1)\theta_{ij})/\sin\theta_{ij}$ is the four-dimensional analog of the Legendre polynomial; θ_{ij} is the angle between the vectors $\mathbf{k}_i$ and $\mathbf{k}_j$.

As with all multidimensional integration, the correct choice and sequencing of integration variables is crucial in determining how hard or easy the eventual calculation turns out to be. Therefore, in the following sections we present mainly those details which display the choice of variables and coordinate systems.

II. THE THREE-DIMENSIONAL TRIANGLE GRAPH

To evaluate the three-dimensional triangle, we rewrite the propagators in Eq. (6) in the partial fraction form

$$(m_2^2 - m_1^2 + K_2^2 - K_1^2)^{-1}(m_3^2 - m_1^2 + K_3^2 - K_1^2)^{-1}$$
$$\times (m_1^2 + K_1^2)^{-1} + \text{cyclic perm.}$$
$$= (m_2^2 - m_1^2 + k_{12}^2 - 2\mathbf{k}_{12}\cdot\mathbf{K}_1)^{-1}(m_3^2 - m_1^2 + k_{13}^2 - 2\mathbf{k}_{13}\cdot\mathbf{K}_1)^{-1}$$
$$\times (m_1^2 + K_1^2)^{-1} + \text{cyclic perm.} \qquad (15)$$

and consider each of the three resulting integrals separately. The singularities introduced into the individual integrands by this procedure are inconsesequential. Since the real part of $(A \pm i\epsilon)^{-1}(B \pm i\epsilon)^{-1}$ is a principal part term plus the term $\pm\pi^2\delta(A)\delta(B)$, the only possible error in (15) is

$$\pm 2 m_1 m_2 m_3 \int d^3K_1 \delta(m_2^2 - m_1^2 + k_{12}^2 - 2\mathbf{k}_{12}\cdot\mathbf{K}_1)\delta(m_3^2 - m_1^2 + k_{13}^2$$
$$- 2\mathbf{k}_{13}\cdot\mathbf{K}_1)(m_1^2 + K_1^2)^{-1} + \text{cyclic perm.} \qquad (16)$$

By integrating (16) we find a possible error of $\pm\frac{1}{2}\pi/(D^{(3)})^{1/2}$ for each of the three terms but this is already implicitly contained in the principal part formula derived below [cf. formula (24)]. The correct additive term is finally chosen by the constraint that in the limit $k_{ij}^2 = 0$, $T^{(3d)} = 2m_1m_2m_3(m_1+m_2)^{-1}(m_1+m_3)^{-1}(m_2+m_3)^{-1}$ as determined by direct integration of Eq. (6).

In the integral over the first term in (15) we choose cylindrical coordinates with the k_z axis perpendicular to $\mathbf{k}_{12}$ and $\mathbf{k}_{13}$, and then write $d^3K_1 = k\,dk\,d\theta\,dk_z$, $\mathbf{k}_{12}\cdot\mathbf{K}_1 = k_{12}k\cos\theta$, and $\mathbf{k}_{13}\cdot\mathbf{K}_1 = k_{13}k\cos(\theta-\phi)$ where ϕ is the angle between $\mathbf{k}_{12}$ and $\mathbf{k}_{13}$. Since the k_z dependence is entirely contained in $(m_1^2+K_1^2)^{-1} = (m_1^2 + k^2 + k_z^2)^{-1}$, elementary integration of this first term in (15) then yields

$$2m_1m_2m_3 \int \frac{k\,dk}{(m_1^2+k^2)^{1/2}} \int \frac{d\theta}{2\pi} (m_2^2 - m_1^2 + k_{12}^2 - 2k_{12}k\cos\theta)^{-1}$$
$$\times [m_3^2 - m_1^2 + k_{13}^2 - 2k_{13}k\cos(\theta-\phi)]^{-1}. \qquad (17)$$

The integral over θ is most easily done by residue methods which yield

$$2m_1m_2m_3 \int \frac{k\,dk}{(m_1^2+k^2)^{1/2}}\, \frac{1}{S_{12}}\, \frac{k_{12}}{\alpha_3 - iS_{12}k_{13}\sin\phi} + (2 \leftrightarrow 3) \qquad (18)$$

where

$$S_{12} = [\sigma_{12} - 4k_{12}^2(m_1^2 + k^2)]^{1/2},$$
$$\sigma_{12} = (m_2^2 - m_1^2 + k_{12}^2)^2 + 4m_1^2k_{12}^2, \qquad (19)$$
$$\alpha_3 = k_{12}(m_3^2 - m_1^2 + k_{13}^2) - k_{13}(m_2^2 - m_1^2 + k_{12}^2)\cos\phi.$$

The constants in (19) are related to the dimensionless invariants defined in the Introduction by

$$\sigma_{12} = 4m_1^2 m_2^2 (y_{12}^2 - 1),$$
$$\alpha_3 k_{12} = 2m_1^2 m_2^2 F_3^{(3)}, \qquad (20)$$
$$\alpha_3^2 + \sigma_{12} k_{13}^2 \sin^2\phi = 4m_1^2 m_2^2 m_3^2 D^{(3)}.$$

The changes from geometrical to algebraic notation in (20) are readily achieved with the use of the scalar product, $2\mathbf{k}_{12}\cdot\mathbf{k}_{13} = 2k_{12}k_{13}\cos\phi = k_{12}^2 + k_{13}^2 - k_{23}^2$, and area, $4k_{12}^2 k_{13}^2 \sin^2\phi = 2k_{12}^2 k_{13}^2 + 2k_{12}^2 k_{23}^2 + 2k_{13}^2 k_{23}^2 - k_{12}^4 - k_{13}^4 - k_{23}^4$, formulas.

We now note that with two variable changes, the integration of the first term in (18) is elementary. First, write $m_1^2 + k^2 = x^2$ so that $k\,dk/(m_1^2 + k^2)^{1/2} = dx$ and $S_{12} = (\sigma_{12} - 4k_{12}^2 x^2)^{1/2}$. Second, write $2k_{12}x = z + \sigma_{12}/(4z)$ to eliminate the remaining square root $S_{12} = i[z - \sigma_{12}/(4z)]$. The limit on the z integration corresponding to the limit $k=0$ is $z_0 = m_1 k_{12} - i(m_2^2 - m_1^2 + k_{12}^2)/2$. With these changes the first integral in (18) becomes

$$2m_1 m_2 m_3 \int \frac{dz}{i} \frac{2k_{13}\sin\phi}{(2zk_{13}\sin\phi + \alpha_3)^2 - 4m_1^2 m_2^2 m_3^2 D^{(3)}} \qquad (21)$$

which reduces, upon integration, to

$$\frac{i}{2(D^{(3)})^{1/2}} \mathrm{Ln}\left(\frac{2zk_{13}\sin\phi + \alpha_3 - 2m_1 m_2 m_3 (D^{(3)})^{1/2}}{2zk_{13}\sin\phi + \alpha_3 + 2m_1 m_2 m_3 (D^{(3)})^{1/2}}\right)\Bigg|_{z=z_0} \qquad (22)$$

with real part given by

$$\frac{1}{2(D^{(3)})^{1/2}} \mathrm{Arctan}\left(\frac{m_3(m_2^2 - m_1^2 + k_{12}^2)}{2m_1^2 m_2} \frac{(D^{(3)})^{1/2}}{F_3^{(3)}}\right). \qquad (23)$$

The corresponding formula obtained by integration of the second term in (18) can be added to (23) with the aid of the addition formula $\mathrm{Arctan}(x) + \mathrm{Arctan}(y) = \mathrm{Arctan}((x+y)/(1-xy))$. The sum is

$$\frac{1}{2(D^{(3)})^{1/2}} \mathrm{Arctan}\left(\frac{2m_1^2(D^{(3)})^{1/2}}{m_2 m_3} \frac{m_3^2(m_2^2 - m_1^2 + k_{12}^2)F_2^{(3)} + m_2^2(m_3^2 - m_1^2 + k_{13}^2)F_3^{(3)}}{4m_1^4 F_2^{(3)} F_3^{(3)} - (m_2^2 - m_1^2 + k_{12}^2)(m_3^2 - m_1^2 + k_{13}^2)D^{(3)}}\right) = \frac{1}{2(D^{(3)})^{1/2}} \mathrm{Arctan}\left(\frac{(D^{(3)})^{1/2}}{y_{23} - y_{12}y_{13}}\right), \qquad (24)$$

where the result on the right follows upon cancellation of the common factor $4m_1^2(m_2^2 m_3^2 D^{(3)} - k_{12}^2 k_{13}^2 \sin^2\phi)$ in the argument of the Arctan function.

Finally, with the help of $\mathrm{Arctan}(x) + \mathrm{Arctan}(y) + \mathrm{Arctan}(z) = \mathrm{Arctan}((x+y+z-xyz)/(1-xy-xz-yz))$, one can add to (24) the results of the integrations of the remaining two terms in (15). In this case a common factor $(y_{12}-1)(y_{13}-1)(y_{23}-1)$ can be cancelled from the argument of the final Arctan and one obtains

$$T^{(3d)} = \frac{1}{2(D^{(3)})^{1/2}} \mathrm{Arctan}\left(\frac{2(D^{(3)})^{1/2}(1 + y_{12} + y_{13} + y_{23})}{(1 + y_{12} + y_{13} + y_{23})^2 - D^{(3)}}\right). \qquad (25)$$

Use of the double angle formula, $\mathrm{Arctan}(2x/(1-x^2)) = 2\,\mathrm{Arctan}(x)$, and the known value of the integral at $k_{ij}^2 = 0$ then completes the derivation of Eq. (6).

It is by no means obvious from either the present calculation or the alternative derivation given in the following section whether the cancellations in, say, Eq. (24) are accidental or whether there is a profound reason for the very simple structure of the final answer. Compared with the apparent complexity of the four-dimensional triangle and box formulas,[8] the result for the three-dimensional triangle is indeed surprising.

III. FEYNMAN PARAMETER METHODS

The connection formulas (8), (10), and (13) are derived below with the help of the Feynman parameter representations for the corresponding integrals. In this representation a one-loop n-vertex graph in d dimensions is[7]

$$I_d^{(n)} = K_d^{(n)} \int \Big(\prod_i^n d\alpha_i\Big)\, \delta\Big(1 - \sum_i^n \alpha_i\Big) (Y^{(n)})^{d/2-n} \qquad (26)$$

for $n > d/2$. The constant $K_d^{(n)}$ depends on the particular normalization used in the momentum space representation of the graph and $Y^{(n)}$ is the quadratic form

$$Y^{(n)} = \sum_{ij}^n \tilde{y}_{ij}\alpha_i\alpha_j \qquad (27)$$

with

$$\tilde{y}_{ij} = \tfrac{1}{2}(m_i^2 + m_j^2 + k_{ij}^2). \qquad (28)$$

The $\tilde{y}_{ij}$ differ from the y_{ij} in Eq. (2) only by mass factors. The constraint that the α_i are positive, together with $\sum_i^n \alpha_i = 1$, defines the region of integration in (26) as the interior of an $(n-1)$-dimensional hypertetrahedron. Its "volume" is bounded by $\binom{n}{1}$ "faces" or "areas" on which one of the α_i vanishes; the "faces" meet at $\binom{n}{2}$ "edges" on which two α_i vanish. This sequence eventually terminates at the vertices of the hypertetrahedron where all but one α_i is zero.

The expression (26) explicitly displays the hypertetrahedral symmetry of the graph but this symmetry is lost in intermediate stages of a calculation if one proceeds directly by choosing some particular α_i to begin the integration. To maintain explicit symmetry we instead replace the region of integration by a sum of $n!$ regions interior to other hypertetrahedra; the latter are characterized as follows. Each has one vertex at the point where the quadratic form $Y^{(n)}$ is stationary. The second vertex is the stationary point of $Y^{(n)}$ on one of the n "faces" and the third is the stationary point on one of the $n-1$ "edges" of the "face." The nth and last vertex is one of the two vertices of the original hypertetrahedron that is common to the particular "face", "edge", $\cdots$ sequence chosen above. Thus each of the new hypertetrahedra is a cone fanning out from a vertex at the stationary point of $Y^{(n)}$ toward a particular subregion on some "face." It is clear that the total original "volume" is included in the sum of all these cones provided all subregions on the "faces" are covered. But this argument can be repeated by replacing "volume" by "face" and "face" by "edge" and we eventually conclude that the new hypertetrahedra indeed cover correctly the original "volume" defined in (26).

Consider now the particular integration "volume" $\Omega(l, m, \cdots)$ in which $\alpha_l = 0$ defines the "face," $\alpha_l = \alpha_m = 0$ defines the "edge," and so on. Denote the α_i at the stationary point of $Y^{(n)}$ by $f_i^{(n)}$ and define a new integration variable β_l,

$$\alpha_l = f_l^{(n)}(1-\beta_l), \tag{29}$$

which is just α_l shifted and scaled so that the integration interval is $[0,1]$. Next shift and rescale the remaining α_i to eliminate all dependence of the integration limits on β_l. Specifically, replace α_i by $f_i^{(n)}(1-\beta_l) + \alpha_i\beta_l$; note that the new α_i are identical to the old on the "face" $\alpha_l = 0$. Then

$$\int \Big(\prod_i^n d\alpha_i\Big)\,\delta\Big(1-\sum_i^n \alpha_i\Big)(Y^{(n)})^{d/2-n} = f_l^{(n)}\int_0^1 \beta_l^{n-2}d\beta_l \int \Big(\prod_{i\neq l}^n d\alpha_i\Big)\,\delta\Big(1-\sum_{i\neq l}^n \alpha_i\Big)(Y^{(n)})^{d/2-n}. \tag{30}$$

The α_i on the right-hand side of (30) are understood to be restricted to lie not only on the "face" $\alpha_l = 0$ but also within the boundaries of $\Omega(l, m, \cdots)$; that is, they lie within a cone fanning out from a vertex at the stationary point of $Y^{(n)}$ on the "face" to a particular subregion on the "edge" $\alpha_l = \alpha_m = 0$. We deduce the dependence of $Y^{(n)}$ on β_l and on the remaining α_i from the fact that $\alpha_i = \text{const}$, for all $i \neq l$, is a straight line emanating radially from the stationary point of $Y^{(n)}$ so that $Y^{(n)} = A + \beta_l^2 B(\alpha)$ is the only functional form consistent with its quadratic nature. From the known values of this function at the points $\beta_l = 0$ and $\beta_l = 1$ we determine

$$Y^{(n)} = \delta^{(n)} + \beta_l^2(Y^{(n-1)} - \delta^{(n)}), \tag{31}$$

where $\delta^{(n)}$ is defined as the stationary value of $Y^{(n)}$ and

$$Y^{(n-1)} = Y^{(n)}(\alpha_l = 0). \tag{32}$$

The steps leading to (30) for the "volume" integral can clearly be repeated for the "face" integral to yield

$$\int \Big(\prod_{i\neq l}^n d\alpha_i\Big)\delta\Big(1-\sum_{i\neq l}^n \alpha_i\Big)(Y^{(n)})^{d/2-n} = f_m^{(n-1)}\int_0^1 \beta_m^{n-3}d\beta_m \int\Big(\prod_{i\neq l,m}^n d\alpha_i\Big)\delta\Big(1-\sum_{i\neq l,m}^n \alpha_i\Big)(Y^{(n)})^{d/2-n}, \tag{33}$$

where $f_m^{(n-1)}$ is the value of α_m at the stationary point of $Y^{(n-1)}$ on the "face" $\alpha_l = 0$. The quadratic form is now

$$Y^{(n)} = \delta^{(n)} + \beta_l^2[\delta^{(n-1)} - \delta^{(n)} + \beta_m^2(Y^{(n-2)} - \delta^{(n-1)})] \tag{34}$$

with $\delta^{(n-1)}$ the stationary value of $Y^{(n-1)}$ and $Y^{(n-2)} = Y^{(n-1)}(\alpha_m = 0)$. Transformations of the form (29) can be used to rewrite the remaining α_i integrations in (33) but the resulting explicit formulas are not needed in the following discussion.

Explicit expressions for the $f_i^{(p)}$ and $\delta^{(p)}$ are obtained as follows. The α_i at the stationary point of $Y^{(n)}$, subject to the constraint $\sum_i^n \alpha_i = 1$, are given by the solution of $\sum_j^n \tilde{y}_{ij}\alpha_j = \lambda$ and hence

$$f_l^{(n)} = \lambda \tilde{F}_l^{(n)}/\tilde{D}^{(n)} = \tilde{F}_l^{(n)}/\Lambda^{(n)}, \tag{35}$$

where

$$\tilde{D}^{(n)} = \det|\tilde{y}_{ij}|,\qquad \tilde{F}_l^{(n)} = \sum_i^n \text{cofactor }(\tilde{y}_{li}) = \frac{\partial}{\partial m_l^2}\tilde{D}^{(n)},\qquad \Lambda^{(n)} = \sum_i^n \tilde{F}_i^{(n)}. \tag{36}$$

The derivative expression in (36) for $F_l^{(n)}$ follows from the specific form (28) for the matrix elements $\tilde{y}_{ij}$. Note that $\tilde{D}$ and $\tilde{F}$ differ only by mass factors from the dimensionless D and F defined by Eqs. (3) and (4). The stationary value of $Y^{(n)}$ is

$$\delta^{(n)} = \tilde{D}^{(n)}/\Lambda^{(n)}. \tag{37}$$

There are corresponding formulas for the lower order $f_i^{(p)}$ and $\delta^{(p)}$ in terms of the lower order $Y^{(p)}$. The remark following Eq. (5) applies here; the subscripted symbols $\tilde{D}_l^{(n-1)}$, $\Lambda_l^{(n-1)}$, $\tilde{F}_{l,m}^{(n-1)} = (\partial/\partial m_m^2)\tilde{D}_l^{(n-1)}$, $\tilde{D}_{lm}^{(n-2)}, \cdots$ will be used if it is necessary to identity explicitly the absent rows and columns.

Some insight into the above formulas can be obtained if one makes use of certain identities derived by Wu[8] both by explicit manipulations of $Y^{(n)}$ and by use of some theorems on determinants. One of these identities is

$$\Lambda^{(n)} = \sum_i^n \tilde{F}_i^{(n)} = \det|-\mathbf{k}_{1i}\cdot\mathbf{k}_{1j}|,\quad i,j = 2,3,\ldots,n \tag{38}$$

so that to within numerical factors the $\Lambda^{(n)}$ are the squares of the k-space "volumes" of hypertetrahedra constructed by joining n points with the lengths k_{ij}. The first few $\Lambda^{(n)}$ are $\Lambda^{(1)} = 1$, $\Lambda^{(2)} = -k_{12}^2$, $\Lambda^{(3)} = |\mathbf{k}_{12}\times\mathbf{k}_{13}|^2$, $\Lambda^{(4)} = -|\mathbf{k}_{12}\cdot\mathbf{k}_{13}\times\mathbf{k}_{14}|^2$. Another identity is

$$(\tilde{F}_l^{(n)})^2 = \Lambda^{(n)}\tilde{D}^{(n-1)} - \Lambda^{(n-1)}\tilde{D}^{(n)} \tag{39}$$

from which we deduce that

$$\delta^{(n-1)} - \delta^{(n)} = \frac{\Lambda^{(n)}}{\Lambda^{(n-1)}}\,(f_l^{(n)})^2 \tag{40}$$

and, since the right-hand side of (40) is less than or equal to zero for physically allowed momenta, that the stationary point of $Y^{(n)}$ is a maximum. For the triangle graph one can derive what is an off-diagonal variant of (40),

$$\tilde{F}_2^{(3)}\tilde{F}_3^{(3)} - \Lambda^{(3)}\tilde{F}_{2,3}^{(2)}\tilde{F}_{3,2}^{(2)} = \mathbf{k}_{12}\cdot\mathbf{k}_{13}(m_1^2\Lambda^{(3)} - \tilde{D}^{(3)}), \tag{41}$$

and, by direct manipulations of the determinants or otherwise, show that

$$\tilde{F}_2^{(3)}\tilde{F}_{3,2}^{(2)} + \tilde{F}_3^{(3)}\tilde{F}_{2,3}^{(2)} = m_1^2\Lambda^{(3)} - \tilde{D}^{(3)}, \tag{42}$$

$$m_1^2\tilde{F}_2^{(3)}\tilde{F}_3^{(3)} - \tilde{D}^{(3)}\tilde{F}_{2,3}^{(2)}\tilde{F}_{3,2}^{(2)} = (\tilde{y}_{12}\tilde{y}_{13} - m_1^2\tilde{y}_{23})(m_1^2\Lambda^{(3)} - \tilde{D}^{(3)}). \tag{43}$$

We return now to the proof of the connection formulas (8), (10), and (13) by explicitly performing the β_l parameter integration in Eq. (30). The relevant part of that integral is

$$f_l^{(n)}\int_0^1 \beta_l^{n-2}\,d\beta_l (Y^{(n)})^{d/2-n}$$
$$=f_l^{(n)}\int_0^1 \beta_l^{n-2}\,d\beta_l[\delta^{(n)}+\beta_l^2(Y^{(n-1)}-\delta^{(n)})]^{d/2-n} \qquad (44)$$

which, for $n \geq d+1$, can be evaluated recursively beginning with

$$f_l^{(n)}\int_0^1 \beta_l^{n-2}\,d\beta_l (Y^{(n)})^{d/2-n}$$
$$=(2n-d-2)^{-1}(f_l^{(n)}/\delta^{(n)})\{(Y^{(n-1)})^{d/2-n+1} + (n-d-1)\int_0^1 \beta_l^{n-2}\,d\beta_l (Y^{(n)})^{d/2-n+1}\}. \qquad (45)$$

The second term in the brace expression in (45) vanishes; obviously so because of the prefactor of the integral for $n=d+1$, and for the reason discussed below in the general case. If $n>d+1$, the matrix elements $\tilde{y}_{ij}$ are not all independent and the k-space "volumes" $\Lambda^{(n)}$ vanish so that $f_l^{(n)}$ and $\delta^{(n)}$ are infinite. However, by arbitrarily reinterpreting the external momenta $\mathbf{k}_{ij}$ for the graph as vectors in an $(n-1)$-dimensional space, and then displacing them appropriately, we can ensure that all expressions including the integral (45) are well defined. Then, having performed the integration in (45), we may legitimately take the limit in which the momenta $\mathbf{k}_{ij}$ again take on their physical values in a space of d dimensions. The ratio $f_l^{(n)}/\delta^{(n)}=\tilde{F}_l^{(n)}/\tilde{D}^{(n)}$ remains finite whereas the factor $Y^{(n)}$ is infinite for all $\beta_l \neq 1$ and the integral in the brace expression in (45) vanishes.

The remaining factor $(Y^{(n-1)})^{d/2-n+1}$ in the right-hand side of (45), when inserted into the α_i integrals in (30), can be recognized as a contribution to an $(n-1)$-vertex graph $I_d^{(n-1)}(l)$ on the "face" $\alpha_l=0$. The result of performing the summation over all integration "volumes" $\Omega(l,m,\cdots)$ is

$$I_d^{(n)}=(2n-d-2)^{-1}(K_d^{(n)}/K_d^{(n-1)}) \times\sum_l \tilde{F}_l^{(n)} I_d^{(n-1)}(l)/\tilde{D}^{(n)} \qquad (46)$$

if we simply reinsert the constants dropped in going from (26) to (30) and then to (45). Finally, if the graph normalization in momentum space depends only on the dimension, then $(2n-d-2)^{-1}K_d^{(n)}/K_d^{(n-1)}=\frac{1}{2}$ and thus

$$I_d^{(n)}=\tfrac{1}{2}\sum_l \tilde{F}_l^{(n)} I_d^{(n-1)}(l)/\tilde{D}^{(n)} \qquad (47)$$

valid for all $n \geq d+1$. If (47) is expressed in terms of the dimensionless variables $F_l^{(n)}$ and $D^{(n)}$, the connection formulas (8), (10), and (13) are obtained as special cases.

The formalism developed in this section can also be used to check the triangle graph calculation presented in Sec. II. If we write $T^{(3d)}$ as the sum of six terms of the form (30) and (33) we obtain

$$T^{(3d)}=m_1m_2m_3\sum T_{lm}, \qquad (48)$$

where, for example,

$$T_{32}=\tfrac{1}{2}f_3^{(3)}f_2^{(2)}\int_0^1\beta_3\,d\beta_3\int_0^1 d\beta_2 \times\{\delta^{(3)}+\beta_3^2[\delta^{(2)}-\delta^{(3)}+\beta_2^2(\delta^{(1)}-\delta^{(2)})]\}^{-3/2} \qquad (49)$$

which after two elementary integrations becomes

$$T_{32}=\tfrac{1}{2}f_3^{(3)}f_2^{(2)}[\delta^{(3)}(\delta^{(3)}-\delta^{(2)})(\delta^{(2)}-\delta^{(1)})]^{-1/2} \times\left\{\mathrm{Arctan}\left[\left(\frac{\delta^{(2)}-\delta^{(1)}}{\delta^{(3)}-\delta^{(2)}}\cdot\frac{\delta^{(3)}}{\delta^{(1)}}\right)^{1/2}\right] - \mathrm{Arctan}\left[\left(\frac{\delta^{(2)}-\delta^{(1)}}{\delta^{(3)}-\delta^{(2)}}\right)^{1/2}\right]\right\}. \qquad (50)$$

This last formula can be rewritten as

$$T_{32}=\frac{1}{2(\tilde{D}^{(3)})^{1/2}}\left\{\mathrm{Arctan}\left(\frac{-\tilde{F}_{3,2}^{(2)}}{m_1\tilde{F}_3^{(3)}}(\tilde{D}^{(3)})^{1/2}\right) - \mathrm{Arctan}\left(\frac{-\tilde{F}_{3,2}^{(2)}}{\tilde{F}_3^{(3)}}(\Lambda^{(3)})^{1/2}\right)\right\} \qquad (51)$$

if we use the identities (40) and the explicit value $\delta^{(1)}=m_1^2$. The first term in (51) is just the expression (23) and furthermore the sum over the six terms in (48) is exactly the sum that led from (23) to (25). The only new insight we obtain by the present method is that the mysterious cancellation in (24) is seen to be related to the determinant identities (42) and (43). (I do not know whether this might generalize in some way to simplify, say, the four-dimensional box graph calculation). The second term in (51) does not appear in (23) but with the use of the Arctan addition formula, the identities (41) and (42), and the explicit expression $\Lambda^{(3)}=|\mathbf{k}_{12}\times\mathbf{k}_{13}|^2$, we find

$$\mathrm{Arctan}\left(\frac{\tilde{F}_{3,2}^{(2)}}{\tilde{F}_3^{(3)}}(\Lambda^{(3)})^{1/2}\right)+\mathrm{Arctan}\left(\frac{\tilde{F}_{2,3}^{(2)}}{\tilde{F}_2^{(3)}}(\Lambda^{(3)})^{1/2}\right) =\mathrm{Arctan}((\Lambda^{(3)})^{1/2}/\mathbf{k}_{12}\cdot\mathbf{k}_{13})=\phi \qquad (52)$$

which is the angle between $\mathbf{k}_{12}$ and $\mathbf{k}_{13}$. The complete six term sum thus simply leads to an extra contribution $\frac{1}{2}\pi/(\tilde{D}^{(3)})^{1/2}$ that only serves to shift the Riemann sheet on which one evaluates the Arctan function in Eq. (25). We may ignore this contribution just as we ignored the δ function contribution (16) and rely instead on the boundary value $T^{(3d)}(k_{ij}=0)=2m_1m_2m_3(m_1+m_2)^{-1}\times(m_1+m_3)^{-1}(m_2+m_3)^{-1}$ to determine that for physically allowed momenta the correct Riemann sheet is the principal sheet in Eq. (6).

IV. NUMERICAL EVALUATION

The one-loop results derived above are not in the most suitable form for use in numerical integration routines. The external momenta at which formulas such as (6) and (8) must be evaluated can range over many orders of magnitude and a direct evaluation of the determinants can lead to a tremendous loss of significance in the numerical results due to cancellations between large terms of opposite sign. For the case of equal internal masses which we discuss below, the simplest way of eliminating this problem is to rewrite the graph expressions in geometrical notation. For example, with all $m_i=1$, $D^{(3)}$ in Eq. (6) can be written

$$D^{(3)}=\tfrac{1}{4}k_{12}^2k_{13}^2k_{23}^2+4A_{123}^2, \qquad (53)$$

where A_{123} is the area of a triangle with sides k_{12}, k_{13}, and k_{23}. It is relatively easy to set up the numerical integration procedure so that the area, and hence $D^{(3)}$ via Eq. (53), is specified without loss of significance.

To evaluate the three-dimensional box graph, with all $m_i=1$, we require

$$D^{(4)} = -\Gamma^2 - 36\Omega^2, \tag{54}$$

where

$$\begin{aligned}16\Gamma^2 = &(k_{12}k_{34} + k_{13}k_{24} + k_{23}k_{14})(k_{12}k_{34} + k_{13}k_{24} \\ &- k_{23}k_{14})(k_{12}k_{34} - k_{13}k_{24} + k_{23}k_{14}) \\ &\times(-k_{12}k_{34} + k_{13}k_{24} + k_{23}k_{24})\end{aligned} \tag{55}$$

and Ω is the volume of a tetrahedron with sides k_{ij} joining the vertices $i = 1, 2, 3, 4$. Again it is relatively easy to avoid loss of significance in calculating the volume Ω but a direct evaluation of Γ using (55) can still involve substantial cancellation between large terms. However, Eq. (55) does suggest that we interpret Γ as the area of a triangle with sides $k_{12}k_{34}$, $k_{13}k_{24}$, and $k_{23}k_{14}$ and although this area is not of direct geometrical significance in momentum space, we can exploit this information to reduce the cancellations in its evaluation by further manipulations as follows. We first determine the largest of the products $k_{12}k_{34}$, $k_{13}k_{24}$, and $k_{23}k_{14}$; for purposes of illustration below, we assume this pair is $k_{23}k_{14}$. By expressing both k_{23} and k_{14} in terms of the remaining momenta, we can derive the identity

$$\begin{aligned}&k_{12}^2k_{34}^2 + k_{13}^2k_{24}^2 - k_{23}^2k_{14}^2 \\ &\quad = 2k_{12}k_{34}k_{13}k_{24}(\cos\phi_1\cos\phi_4 + \sin\phi_1\sin\phi_4\cos\chi) \\ &\quad \equiv 2k_{12}k_{34}k_{13}k_{24}\cos\psi,\end{aligned} \tag{56}$$

where ϕ_1 is the angle between $\mathbf{k}_{12}$ and $\mathbf{k}_{13}$, ϕ_4 is the angle between $\mathbf{k}_{24}$ and $\mathbf{k}_{34}$, and χ is the angle between the two faces of the tetrahedron containing the vertices 1,2,3 and 2,3,4. Eq. (56) is of the form of a scalar product formula and we can interpret ψ as the angle in the triangle between the two sides $k_{12}k_{34}$ and $k_{13}k_{24}$, and opposite the side $k_{23}k_{14}$. The area expression (55) reduces to

$$\Gamma = \tfrac{1}{2}k_{12}k_{34}k_{13}k_{24}\sin\psi, \tag{57}$$

which, since it expresses Γ in terms of the two smallest products $k_{12}k_{34}$ and $k_{13}k_{24}$, can only entail loss of significance in the case of an "accidental" cancellation whenever $\psi \approx \pi$. Now $\psi = \pi$ implies $\chi = \pi$, $\phi_1 = \pi - \phi_4$ and this is just the condition that the k_{ij} making up the tetrahedron be coplanar and that the points $i = 1, 2, 3, 4$ lie on a circle. Since it is easy to design numerical integration routines that avoid coplanar momenta, the expression (57) together with the definition of ψ given by (56) should be adequate for most purposes.

For completeness we give the formulas for $F_i^{(4)}$ corresponding to (54) for $D^{(4)}$. Again, all $m_i = 1$.

$$\begin{aligned}F_1^{(4)} &= -A_{234}k_{12}k_{34}k_{13}k_{24}(\sin\phi_4\cos\phi_1 \\ &\quad - \sin\phi_1\cos\phi_4\cos\chi), \\ F_4^{(4)} &= -A_{123}k_{12}k_{34}k_{13}k_{24}(\sin\phi_1\cos\phi_4 \\ &\quad - \sin\phi_4\cos\phi_1\cos\chi), \\ F_2^{(4)} &= -(\mathbf{k}_{31}\cdot\mathbf{k}_{32}F_1^{(4)} + \mathbf{k}_{34}\cdot\mathbf{k}_{32}F_4^{(4)})/k_{23}^2 - 18\Omega^2, \\ F_3^{(4)} &= -(\mathbf{k}_{21}\cdot\mathbf{k}_{23}F_1^{(4)} + \mathbf{k}_{24}\cdot\mathbf{k}_{23}F_4^{(4)})/k_{23}^2 - 18\Omega^2\end{aligned} \tag{58}$$

The angles ϕ_1, ϕ_4, and χ are as defined for Eq. (56) and A_{ijk} is the area of a triangle with sides k_{ij}, k_{ik}, and k_{jk}.

V. ANGULAR INTEGRATIONS IN FOUR DIMENSIONS

Four-dimensional angular integrations such as the ones needed to evaluate the "Racah" coefficient in Eq. (14) can in principle be done by expanding the Chebyshev polynomials in terms of conventional four-dimensional polar hyperspherical harmonics.[11] However, an equivalent but technically much simpler solution is possible because there exists a particular coordinate system for which the four-dimensional hyperspherical harmonics are the three-dimensional rotation matrices and all the known three-dimensional results can immediately be applied.

If we label these particular coordinates r, α, β, γ, then the four-dimensional Cartesian coordinates x, y, z, w are given by[12]

$$\begin{aligned}x &= r\sin\frac{\beta}{2}\sin\frac{\alpha-\gamma}{2}, \quad y = r\sin\frac{\beta}{2}\cos\frac{\alpha-\gamma}{2}, \\ z &= r\cos\frac{\beta}{2}\sin\frac{\alpha+\gamma}{2}, \quad w = r\cos\frac{\beta}{2}\cos\frac{\alpha+\gamma}{2}.\end{aligned} \tag{59}$$

The four-dimensional solid angle is completely covered if we take $0 \leq \alpha \leq 2\pi$, $0 \leq \beta \leq \pi$, $0 \leq \gamma \leq 4\pi$. The Jacobian

$$\left|\frac{\partial(xyzw)}{\partial(r\alpha\beta\gamma)}\right| = \tfrac{1}{8}r^3\sin\beta$$

and hence the normalized integral over the four-dimensional solid angle can be written as

$$\int\frac{d\Omega}{2\pi^2} = \frac{1}{16\pi^2}\int_0^{4\pi}d\gamma\int_0^{\pi}d\beta\sin\beta\int_0^{2\pi}d\alpha \tag{60}$$

which is exactly of the form of an integral over Euler angles in three dimensions.

Given the coordinate definitions (59) and the definition of the rotation matrices[10]

$$D_{m'm}^{(j)}(\Omega) = e^{im'\gamma}d_{m'm}^{(j)}(\beta)e^{im\alpha}, \tag{61}$$

together with the specific values

$$\begin{aligned}d_{1/2\,1/2}^{(1/2)}(\beta) &= d_{-1/2\,-1/2}^{(1/2)}(\beta) = \cos\frac{\beta}{2}, \\ d_{1/2\,-1/2}^{(1/2)}(\beta) &= -d_{-1/2\,1/2}^{(1/2)}(\beta) = \sin\frac{\beta}{2},\end{aligned} \tag{62}$$

we deduce that the four-dimensional scalar product $\mathbf{r}_1\cdot\mathbf{r}_2 = x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2$ can be written

$$\mathbf{r}_1\cdot\mathbf{r}_2 = \tfrac{1}{2}r_1r_2\sum_{mm'}D_{m'm}^{(1/2)}(\Omega_1)D_{m'm}^{*(1/2)}(\Omega_2). \tag{63}$$

Furthermore, if we define

$$F_l = F_l(\Omega_1, \Omega_2) = \sum_{mm'}D_{m'm}^{(l/2)}(\Omega_1)D_{m'm}^{*(l/2)}(\Omega_2), \tag{64}$$

then it follows that

$$F_lF_1 = F_{l-1} + F_{l+1} \tag{65}$$

as can be shown by using the product formula[10]

$$D_{m_1'm_1}^{(j_1)}(\Omega)D_{m_2'm_2}^{(j_2)}(\Omega) = \sum_j(2j+1)\begin{pmatrix}j_1 & j_2 & j\\ m_1' & m_2' & m'\end{pmatrix}\begin{pmatrix}j_1 & j_2 & j\\ m_1 & m_2 & m\end{pmatrix}D_{m'm}^{*(j)}(\Omega) \tag{66}$$

and the orthogonality of the 3-j symbols on summation over all m_i' and m_i. With the aid of (65) we can derive the recursion relation

$$\frac{r_1F_l - r_2F_{l-1}}{r_1^2+r_2^2-r_1r_2F_1} = \frac{1}{r_1}F_l + \frac{r_2}{r_1}\frac{r_1F_{l+1}-r_2F_l}{r_1^2+r_2^2-r_1r_2F_1} \tag{67}$$

and hence the expansion

$$\frac{1}{(\mathbf{r}_1-\mathbf{r}_2)^2} = \frac{1}{r_1^2+r_2^2-r_1r_2F_1} = \sum_{l=0}^{\infty}\frac{r_2^l}{r_1^{l+2}}F_l \tag{68}$$

for $r_2 < r_1$. With the conventional choice of angular variables in four dimensions, the expansion coefficients in (68) are the Chebyshev polynomials $C_l(\cos\theta_{12}) = \sin((l+1)\theta_{12})/\sin\theta_{12}$. Therefore, we make the identification

$$C_l(\cos\theta_{12}) = F_l(\Omega_1, \Omega_2), \tag{69}$$

a result which can also be found in Englefield.[12]

By substituting F_l for the C_l in Eq. (14) we can reduce the expression for the "Racah" coefficient to a sum of products of four integrals of the form[10]

$$\begin{aligned}\frac{1}{16\pi^2}\int_0^{4\pi}d\gamma\int_0^{\pi}d\beta\sin\beta\int_0^{2\pi}d\alpha\, D^{(j_1)}_{m_1'm_1}(\Omega)\\ \times D^{(j_2)}_{m_2'm_2}(\Omega)D^{(j_3)}_{m_3'm_3}(\Omega)\\ = \begin{pmatrix} j_1 j_2 j_3 \\ m_1'm_2'm_3'\end{pmatrix}\begin{pmatrix} j_1 j_2 j_3 \\ m_1 m_2 m_3\end{pmatrix}.\end{aligned} \tag{70}$$

Finally, the resulting sum of products of eight 3-j symbols can be written as the square of a sum of products of four 3-j symbols and the latter sum is the invariant combination that yields the 6-j symbol in Eq. (14). The algebra in the derivation of (14) is somewhat complicated by the need to keep track of all phase factors arising from the transformations $D^{*(j)}_{m'm}(\Omega) = (-)^{m'-m}D^{(j)}_{-m'-m}(\Omega)$ but is otherwise straightforward.

VI. CONCLUDING REMARKS

The purpose of the present calculation has been to derive expressions for simple graphs in ϕ^4 theory that can subsequently be used in numerical integration routines to evaluate more complicated graphs. The three-dimensional triangle and box graph expressions derived here have made possible the evaluation of all contributions to coupling-constant and propagator renormalization constants up to the 6-loop level.[4,13] Higher order calculations of the renormalization constants would be worthwhile but may not be feasible unless the analytical calculations presented here are extended to include simple two- and three-loop graphs. Unfortunately, we cannot say whether such an extension is possible. The simplicity of, say, the three-dimensional triangle expression is intriguing but whether this is accidental or whether correspondingly simple expressions for other graphs will result is not known.

ACKNOWLEDGMENTS

It is a pleasure to thank Chris Gray for pointing out the connection between the four-dimensional hyperspherical harmonics and the three-dimensional rotation matrices, and to thank George Leibbrandt for numerous comments and suggestions. The hospitality of George Baker, Jr. and other members of the Los Alamos Scientific Laboratory where parts of this calculation were performed is gratefully acknowledged. I thank the National Research Council of Canada for its financial support through grant A9348.

[1]K.G. Wilson, Phys. Rev. Lett. 28, 548 (1972).
[2]G. Parisi, in Proceedings of the 1973 Cargese summer school, unpublished.
[3]E. Brezin, J.C. LeGuillou, J. Zinn-Justin, and B.G. Nickel, Phys. Lett. A 44, 227 (1973).
[4]G.A. Baker, Jr., B.G. Nickel, M.S. Green, and D.I. Meiron, Phys. Rev. Lett. 36, 1351 (1976).
[5]C.G. Gray, unpublished notes (1966); R.L. Armstrong, S.M. Blumenfeld, and C.G. Gray, Can. J. Phys. 46, 1331 (1968). More details will be found in K.E. Gubbins and C.G. Gray, *Theory of Simple Molecular Liquids* (Oxford U.P., Oxford, to be published).
[6]J.L. Rosner, Ann. Phys. (N.Y.) 44, 11 (1967).
[7]R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).
[8]A.C.T. Wu, Mat. Fys. Medd. Dan. Vid. Selsk. 33, no. 3 (1961).
[9]D.S. Kershaw, Phys. Rev. D 8, 2708 (1973).
[10]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1960).
[11]M. Gourdin, Nuovo Cimento 7, 338 (1958).
[12]The angles α, β, γ are linear combinations of the angles used in the $R_3 \times R_3$ representation of R_4 discussed in L.C. Biedenharn, J. Math. Phys. 2, 433 (1961). See also M.J. Englefield, *Group Theory and the Coulomb Problem* (Wiley, New York, 1972).
[13]B.G. Nickel, D.I. Meiron, and G.A. Baker, Jr., "Compilation of 2-pt. and 4-pt. graphs for continuous spin models," University of Guelph Report, 1977, and "Evaluation of series for critical exponents for a ϕ^4 model in three and two dimensions," to be submitted to Physical Review.

The geometry of the gravitational field at spacelike infinity

Paul Sommers[a), b)]

Department of Physics, University of North Carolina, Chapel Hill, North Carolina 27514
(Received 28 September 1977)

The asymptotic structure of gravitational fields at large spacelike separation from sources is studied. Limits of spacetime fields are discussed in terms of a three-dimensional boundary manifold representing spacelike infinity. The boundary is endowed with the metric of a timelike unit hyperboloid. With sufficiently stringent conditions on the asymptotic spacetime geometry, the total energy–momentum and angular momentum emerge as integrals over any cross section of the hyperboloid at infinity. It is possible to identify physically relevant weaker conditions under which the energy–momentum, but not the angular momentum, is well defined. Under still weaker conditions, the energy–momentum also loses its meaning even though the spacetime admits a Minkowskian asymptote.

1. INTRODUCTION

An isolated system of gravitating sources is modelled, in relativity theory, by a spacetime which is asymptotically flat. The geometry of such a spacetime resembles the Minkowski space of special relativity at large distance from the sources. There is, however, considerable ambiguity in the notion of an isolated system and in the notion of an asymptotically flat spacetime. Relativity theory itself does not require a definition of either of these. The concepts are introduced in order to make certain classes of spacetimes understandable by describing their gross features in terms which are familiar from the physics of special relativity. In particular, definitions are normally chosen so one can ascribe a meaning to the total energy—momentum (or mass) of the system. It may be desirable to employ a definition which is restrictive enough that the angular momentum of the system is also well defined in terms of the asymptotic gravitational field.

A major ambiguity in the notion of asymptotic flatness stems from the different ways one can be far away from the sources. Based on the work of Bondi *et al.*,[1] Penrose[2] adopted conditions which usefully characterize spacetimes in which the geometry becomes Minkowskian asymptotically along outgoing *null* geodesics. For such spacetimes one has a boundary manifold $\mathscr{I}^+$ representing future null infinity. Certain fields induced on $\mathscr{I}^+$ from the physical spacetime have significant physical interpretations. In particular, the mass of the system at a given retarded time is an integral of such quantities over a cross section of $\mathscr{I}^+$. If the integration is performed over an earlier cross section, the mass will have an equal or greater value, the difference being the energy radiated to infinity between the two retarded times. Allowing the cross section to recede indefinitely to the past on $\mathscr{I}^+$, one would expect to get a limiting mass which represents all the mass of the spacetime. In a certain sense the past limit of $\mathscr{I}^+$ is spacelike infinity, and the limit of the mass integral is the (time-independent) mass defined at spacelike infinity. This is certainly true in the trivial case of Minkowski space. If the spacetime mass is nonzero, however, it appears that spacelike infinity cannot be represented by a completely regular point of the Penrose boundary. Considerable care must then be taken in regarding spacelike infinity as a past endpoint of $\mathscr{I}^+$. Ashtekar and Hansen[3] are exploring this matter in detail.

Predating the concept of null infinity was the idea that spacetime should become Minkowskian at large *spacelike* separation from the sources. Various notions of spacelike infinity in terms of the asymptotic properties of the spacetime geometry are as old as relativity itself. Relevant asymptotic conditions were sharpened by ADM,[4] who were able to produce integral expressions for the spacetime energy-momentum and angular momentum in terms of the asymptotic geometry. Their method involves locating any spacelike hypersurface which behaves asymptotically like a hyperplane of Minkowski space, and then integrating, over large 2-spheres, quantities derived from the metric and extrinsic curvature of the hypersurface. The energy—momentum and angular momentum are limits of such integrals as the integration surface expands to infinity.

York,[5] along with O'Murchadha, has refined the ADM formalism, freeing it from its dependence on a background metric in the initial data hypersurface and making it manifestly covariant. The form of the energy—momentum integral adopted by York is an expression which Brill[6] had observed to be valid in the case of an initial data hypersurface with vanishing extrinsic curvature.

Using conformal rescaling techniques similar to those employed in the construction of $\mathscr{I}^+$, Geroch[7] has shown how to represent spatial infinity as a single point which represents the completion of an asymptotically flat initial data hypersurface just as Euclidean 3-space can be conformally mapped onto a 3-sphere by adding a single point at infinity. Limits of fields on the initial data hypersurface are represented by direction-dependent tensors at the point of infinity. The spacetime energy—momentum is an integral of one such quantity over the sphere of directions at spatial infinity.

The discussion of spacelike infinity to be presented here differs somewhat from the formulations mentioned above. Spacelike infinity will be represented by a three-dimensional boundary of the four-dimensional spacetime. The characterization of asymptotic flatness

[a)]Present address: Department of Mathematics, North Carolina State University, Raleigh, NC 27607.
[b)]This work was supported in part by NSF Grant No. MPS74-14191-A01.

therefore applies directly to the spacetime and not to some initial data hypersurface. One consequence is that it does not lead to awkward questions such as whether or not an asymptotically flat initial data hypersurface can be arbitrarily translated and boosted (in some sense) without disrupting its asymptotic flatness.

The existence of the three-dimensional boundary representing spacelike infinity will mean, roughly speaking, that the spacetime metric reduces to a Minkowski metric along outgoing asymptotic spacelike directions. There remains the question of how fast the two metrics should converge. If the spacetime's energy—momentum is to be well defined, it is appropriate to require, in effect, that the curvature tensor fall off like $1/r^3$ without oscillatory behavior near spacelike infinity. Although this restriction is not quite strong enough to guarantee a well-defined angular momentum, it is already restrictive enough to exclude certain spacetimes which one might wish to regard as asymptotically flat. For example, one might think that a model for a bounded source which radiated only a finite amount of energy during its entire history should be an asymptotically flat spacetime. One can imagine, however, a source which emits gravitational waves of fixed amplitude in spurts of variable duration. If, looking to the past, the spurts were to become progressively shorter, such a source could have been emitting for all time with only finite energy at its disposal. Looking off in spacelike directions from such a source, one would see occasional oscillatory behavior even in the $1/r$ part of the curvature beyond any fixed distance, and so the spacetime energy—momentum at spacelike infinity could not be defined. Such a source is, of course, rather contrived and unphysical.

Even restricting attention to sources which radiate smoothly in the past, it is easy to see that the finiteness of radiated energy is not sufficient to ensure a $1/r^3$ dependence of the curvature tensor in spacelike directions. Suppose the time dependence of the radiated power is given by $(-t)^n$ in the remote past. Then, provided $n<-1$, the total energy radiated will be finite. If $n>-2$, however, the curvature will fall off less rapidly than $1/r^3$ in spacelike directions. One way to understand this is to recognize that, at $\mathscr{I}^+$, the energy flux is given by the square of the news function N; i.e., $N\sim(-u)^{n/2}$, where u is the retarded time on $\mathscr{I}^+$. The radiation part of the curvature is given by the time derivative of the news function and therefore has a $(-u)^{n/2-1}$ dependence on $\mathscr{I}^+$. Along any outgoing null geodesic, the radiation part of the curvature decreases like $1/r$, where r can be regarded as affinely parametrizing outgoing null geodesics. Since, at spacelike distance r and $t=0$, the radiation was emitted at $-t\sim r$, the r dependence of the curvature is given, in this rough argument, by $r^{n/2-2}$. If $n>-2$, therefore, the curvature will not fall off as fast as $1/r^3$. Examples of this type also may be judged to be unphysical on the grounds that if the sources are bounded in space, then some unrealistic mechanism is needed to excite the sources so they will radiate. The mechanism may be incoming radiation or an artifical time-dependent equation of state for the sources.

If one wants to examine realistic physical models which are not stationary, it may be necessary to turn to systems which coalesce from unbounded distances. No exact solutions of this type are available for study, and the motions of source particles are not yet rigorously understood.[8] Nevertheless, it is reasonable to assume that at early times such a system would evolve in a Newtonian fashion. From this it follows that the dimension of the system would behave like $(-t)^{2/3}$ in the remote past. The quadrupole moment would be expected to go as $(-t)^{4/3}$, and one may suppose that the power radiated would go as the square of the third time derivative of this moment[9]; i.e., the power radiated would behave like $(-t)^{-10/3}$. By the reasoning of the preceding paragraph, the radiation part of the curvature would fall off like $r^{-11/3}$ at large spacelike distances. Such a system would therefore be expected to have a well-defined energy—momentum at spacelike infinity since the radiation part of the curvature would not mask the $1/r^3$ limit at spacelike infinity.

For many purposes it is desirable to have a definition of a system's angular momentum in terms of its asymptotic gravitational field. A satisfactory definition of angular momentum seems to require that the radiation part of the curvature fall off faster than $1/r^4$. For coalescing systems of the type just described, therefore, the angular momentum may not be defined. That the definition of angular momentum should require stronger asymptotic conditions than that of energy—momentum is not surprising. The same is true for matter fields in Minkowski space, due to the fact that rotational Killing vector fields grow larger with separation from the origin, whereas translational Killing vector fields do not. Worthy of emphasis is that, in general relativity theory, bounded *stationary* systems may be the only realistic physical systems for which the angular momentum is well defined by the spacetime geometry near spacelike infinity.

The precise concept of spacelike infinity adopted here will be modelled on a boundary ρ, of Minkowski space (M,η_{ab}), which represents projective spacelike infinity. Every spacelike ray of (M,η_{ab}) acquires an endpoint on ρ, and each point of ρ can be identified with an equivalence class of parallel spacelike rays. The points of ρ can therefore be identified with the unit spacelike vectors at any point of M. The space of such vectors is a timelike unit hyperboloid, and the boundary ρ also acquires this metric structure. There are noteworthy advantages of representing infinity as a boundary manifold. The limits of fields on spacetime, if the limits exist, become ordinary tensor fields on ρ. For example, a linearized gravitational field, if it has the proper asymptotic behavior, induces on ρ a pair of symmetric trace-free valence-2 tensor fields (E_{ab} and B_{ab}) whose divergences with respect to the hyperboloid metric connection vanish. Integrals of these tensor fields over arbitrary cross sections of ρ yield quantities which may be identified as energy—momentum and angular momentum of the linearized gravitational field.

A curved spacetime is asymptotically flat if, first of all, it admits a boundary manifold ρ with properties (specified in Sec. 3) in common with spacelike infinity of Minkowski space. Conditions on the asymptotic curvature may also be imposed. Roughly speaking, if the curvature is of order $1/r^3$ and nonoscillatory asymptotically, then the energy—momentum is well defined. If the part of the curvature which is magnetic (with respect to a foliation whose limit is $\mathcal{P}$) has a limiting $1/r^4$ behavior, then the spacetime angular momentum integral also exists.

2. SPACELIKE INFINITY FOR MINKOWSKI SPACE

The points of spacelike infinity $\mathcal{P}$ for Minkowski space (M, η_{ab}) represent equivalence classes of parallel rays in (M, η_{ab}). An explicit construction demonstrates that $\mathcal{P}$ can be realized as a boundary manifold of (M, η_{ab}): Let x be any point of M and let M_x be the points of M other than x itself which lie on spacelike lines through x; i.e., M_x is the exterior of the lightcone of x. On M_x is a positive distance function $r := |\eta_{ab} x^a x^b|^{1/2}$, where x^a is the position vector of points in M_x relative to x. The hypersurfaces of constant r are timelike hyperboloids centered on x. The congruence of curves normal to this foliation are the rays emanating from x. Labeling these rays by angles χ, θ, φ, which then serve as coordinates on each hyperboloid, the chart $(r, \chi, \theta, \varphi)$ covers M_x (aside from the points at which the spherical coordinates are singular). Define Σ by $\Sigma := r^{-1}$. Then the boundary $\mathcal{P}$ of M_x is incorporated by extending the range of Σ to $0 \leq \Sigma < \infty$. The hypersurface $\Sigma = 0$ is the boundary $\mathcal{P}$. Every spacelike ray of M intersects M_x and acquires an endpoint on $\mathcal{P}$. The points of $\mathcal{P}$ represent angles in M_x (or M), parallel rays all meeting the same point of $\mathcal{P}$.

Choosing a different point $\hat{x}$, the above construction for $M_{\hat{x}}$ yields the *same* boundary points $\mathcal{P}$ representing angles of spacelike rays in M. Since every spacelike ray of M enters both M_x and $M_{\hat{x}}$, the intersection of the two regions contains a neighborhood of $\mathcal{P}$. The two charts, $(\Sigma, \chi, \theta, \varphi)_x$ and $(\Sigma, \chi, \theta, \varphi)_{\hat{x}}$ are analytically related in the overlap. Allowing x to vary over all of M, the charts for M_x constitute an atlas for M. An atlas for $\bar{M} = M \cup \mathcal{P}$ is obtained by extending any or all of these charts to the boundary $\mathcal{P}$.

The Minkowski metric on M_x is $ds^2 = -\Sigma^{-4} d\Sigma^2 + \Sigma^{-2} d\sigma^2$, where

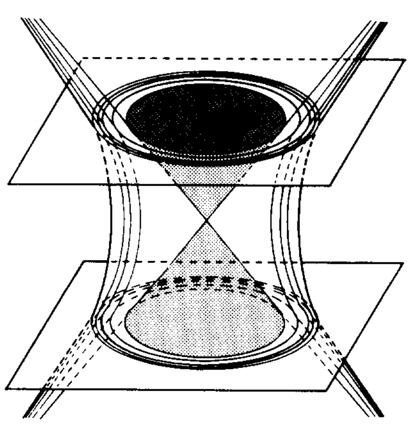


FIG. 1. The figure depicts the region M_x foliated by hyperboloids of constant Σ. The outermost hyperboloid represents the boundary $\mathcal{P}$ where $\Sigma = 0$.

$$d\sigma^2 = d\chi^2 - \cosh^2\chi \, (d\theta^2 + \sin^2\theta \, d\varphi^2)$$

is the metric of a unit timelike hyperboloid. The Minkowski metric is singular at $\Sigma = 0$, and no conformal rescaling can make it regular on the three-dimensional boundary $\mathcal{P}$. On the other hand, there *is* a metric on $\mathcal{P}$ induced from (M, η_{ab}). On each $\Sigma = \text{const}$ leaf is an induced metric $h_{ab} = \Sigma^{-2} p_{ab}$, where p_{ab} is the unit hyperboloid metric. The tensorfield $\Sigma^2 h_{ab}$ has a smooth extension to $\mathcal{P}$, inducing the unit hyperboloid metric on $\mathcal{P}$.

Representing infinity as a boundary manifold of spacetime makes it easier to discuss asymptotic properties of fields on spacetime. A scalar field on spacetime has a limit at spacelike infinity if it is the restriction to M of a continuous field on $\bar{M}$. A scalar field f on M will be said to be of order Σ^n provided $\Sigma^{-n} f$ has a limit which does not vanish everywhere on $\mathcal{P}$.

The notion of the limit of a tensor field, on the other hand, is slightly more subtle than the existence of an extension of the tensor field to $\mathcal{P}$. The reason is that the tangential components tend to behave differently from the orthogonal components. (Here "tangential" and "orthogonal" refer to a decomposition of the tensor field with respect to a Σ-foliation and its unit normal. Although the leaves are here timelike hypersurfaces, the algebra differs in no essential respect from standard "3 plus 1" decompositions, and the notation here will follow that of Ashtekar and Geroch.[10]) To obtain the limit of a tensor field on M, one first represents it as a set of tangential tensor fields. A tangential tensor field has a limit if it can be extended continuously to $\mathcal{P}$.

The "order" of a tensor field can be confusing because the charts near $\mathcal{P}$ use angular coordinates. Thus, for example, although the vector field ∂_θ and the 1-form $d\theta$ are smooth at $\mathcal{P}$, the fields $\Sigma\,\partial_\theta$ and $-\Sigma^{-1} d\theta$ both have norms of order Σ^0 and these two are related via the metric. In other words $\Sigma\,\partial_\theta$ and $\Sigma^{-1} d\theta$ are what one would regard as tensor fields of order Σ^0. More generally, a tangential tensor field $T^{a_1 a_2 \cdots a_p}_{b_1 b_2 \cdots b_q}$ is of order Σ^n if $\Sigma^{-n-p+q} T^{a_1 a_2 \cdots a_p}_{b_1 b_2 \cdots b_q}$ has a continuous extension to $\bar{M}$ which does not vanish everywhere on $\mathcal{P}$.

As an example, consider an electromagnetic field F_{ab} on M. If the field is the retarded field of some bouned source which was nonradiative in the distant past, then the electromagnetic field would be expected to fall off like r^{-2} in spacelike directions. If n^a is the unit normal to Σ-leaves, then $E_a := n^b F_{ab}$ and $B_a := n^b F^*_{ab} \equiv \frac{1}{2} n^b \epsilon_{abcd} F^{cd}$ should be tangential tensor fields of order Σ^2. That is to say, the 1-forms $\Sigma^{-1} E_a$ and $\Sigma^{-1} B_a$ should have extensions to $\mathcal{P}$.

The natural derivative operation on $\mathcal{P}$ is the symmetric covariant derivative D_a which annihilates the unit hyperboloid metric: $D_a p_{bc} = 0$. It is natural in the sense that for tangential tensor fields with smooth limits at $\mathcal{P}$, the derivative of the limit is the limit of the derivative. Denote by D_a also the intrinsic covariant derivative of Σ-leaves (for $\Sigma > 0$). If $\Sigma^m T^{\cdots}_{\cdots}$ is differentiable at $\mathcal{P}$, then $D_\circ(\Sigma^m T^{\cdots}_{\cdots}) = \Sigma^m D_\circ T^{\cdots}_{\cdots}$ is continuous at $\mathcal{P}$. If $T^{\cdots}_{\cdots}$ is of order Σ^n, then $D_\circ T^{\cdots}_{\cdots}$ is of order $\sum^{n+1}$ since it has one more lower index than $T^{\cdots}_{\cdots}$ has.

An electromagnetic field F_{ab} on M satisfies $\nabla^a F_{ab} = J_b$ and $\nabla^a F^*_{ab} = 0$, where J_b is the charge current. If the $\sum$-leaves are hyperboloids centered on a point x, as discussed above, then $\nabla_a n_b = \sum h_{ab}$, where h_{ab} is the intrinsic metric of the hyperboloids. The Maxwell equations include $D^a E_a = n^a J_a$ and $D^a B_a = 0$. For a bounded electromagnetic source, J_a vanishes in some neighborhood of $\mathcal{P}$, so $D^a E_a = 0 = D^a B_a$ near $\mathcal{P}$. On $\mathcal{P}$ as well, then, $D^a E_a = 0 = D^a B_a$. By virtue of Stokes' theorem, therefore, $\oint E_a dS^a$ and $\oint B_a dS^a$ are both independent of choice of cross section for integration on $\mathcal{P}$. (Regarding $\mathcal{P}$ as a fiber bundle over S^2 with fiber $\mathbb{R}$, a cross section is homeomorphic to S^2.) These quantities are numbers associated with the electromagnetic field and determined completely by the asymptotic form of the field. They are, of course, simply the total electric and magnetic charges in M.

Of somewhat more interest are the quantities defined in an analogous way using the limits of a linearized gravitational field. Such a field may be represented on M by a tensor field C_{abcd} having the algebraic symmetries of a conformal curvature tensor and satisfying the field equations $\nabla^a C_{abcd} = 0$. Decomponsing C_{abcd} into tangential tensor fields yields two symmetric tracefree tensor fields:

$$E_{ab} = n^c C_{cabd} n^d$$

and

$$B_{ab} = n^c C^*_{cabd} n^d = \tfrac{1}{2} n^c \epsilon_{ca}{}^{ef} C_{efbd} n^d.$$

Using a hyperboloid foliation of the type already discussed, for which the extrinsic curvature is $K_{ab} = -\nabla_a n_b = -\sum h_{ab}$, the field equation becomes the following four equations:

$$D^a E_{ab} = 0, \quad D^a B_{ab} = 0, \quad \mathcal{L}_n E_{ab} = -\epsilon_b{}^{cd} D_c B_{da} - \sum E_{ab},$$

and

$$\mathcal{L}_n B_{ab} = \epsilon_b{}^{cd} D_c E_{da} - \sum B_{ab}.$$

(The tensor field $\epsilon_{bcd} := n^a \epsilon_{abcd}$ is the normalized three-dimensional alternating tensor field.) Suppose now that the fields satisfy these asymptotic conditions: E_{ab} is of order $\sum^3$ and B_{ab} is of order $\sum^4$.

In order to define the quantities analogous to electric and magnetic charges, it is expedient to use the Killing vector fields and conformal Killing vector fields of $\mathcal{P}$. There are six independent Killing vector fields K^a satisfying $D_{(a}K_{b)} = 0$ and four independent conformal Killing vector fields ξ^a satisfying $D_a \xi_b = \frac{1}{3}(D_c \xi^c) p_{ab}$. These may be visualized by thinking of $\mathcal{P}$ as embedded as a unit hyperboloid centered on a point x of an abstract Minkowski space. Then the six independent Lorentz rotations about x, when restricted to $\mathcal{P}$, are Killing vector fields of $\mathcal{P}$. The four independent conformal Killing vector fields can be obtained by projecting any four constant vector fields of the Minkowski space into $\mathcal{P}$.

Since $D^a E_{ab} = 0 = D^a B_{ab}$ on M, these equations are satisfied also by the limit fields on $\mathcal{P}$. Let L^a be any of the vector fields K^a or ξ^a. Then the integrals $\oint E_{ab} L^a dS^b$ and $\oint B_{ab} L^a dS^b$ are insensitive to which cross section of $\mathcal{P}$ is used for the integration, since $D^b(E_{ab}L^a) = 0 = D^b(B_{ab}L^a)$. At first sight, then, there would appear to be twenty quantities. Actually, however, ten of these vanish. Since B_{ab} is of higher order in $\sum$ than E_{ab}, the fourth field equation above yields $D_{[a}E_{b]c} = 0$. From this one can show[11] that

$$\oint E_{ab} K^a dS^b = -2\oint D_a (E^{[a}_m D^{b]} K^m) dS_b,$$

and, since the integration surface is compact, this vanishes by virtue of Stokes' theorem. Now examine once again the fourth field equation above. Because B_{ab} is of order $\sum^4$, it is $\sum^{-2} B_{ab}$ which has a non-zero limit at $\mathcal{P}$. Since $n^a \partial_a = -\sum^2 \partial_\Sigma$, the term $\mathcal{L}_n B_{ab}$ is asymptotically equal to $-2\sum B_{ab}$, and hence

$$B_{ab} = -\sum^{-1} \epsilon_b{}^{cd} D_c E_{da}.$$

Although the order $\sum^3$ part of E_{ab} is curl-free, the order $\sum^4$ part serves as a symmetric tensor potential for B_{ab}. (It is assumed that the fields are sufficiently smooth at $\mathcal{P}$ that the $\sum^4$ parts are differentiable.) Denoting the symmetric potential by κ_{ab}, Stokes' theorem can be invoked to show that $\oint B_{ab} \xi^a dS^b$ vanishes:

$$\oint B_{ab} \xi^a dS^b = \oint (\epsilon_b{}^{cd} D_c \kappa_{da}) \xi^a dS^b = \oint D_c (\epsilon^{bcd} \xi^a \kappa_{da}) dS_b.$$

There remain, therefore, the integrals

$$\oint E_{ab} \xi^a dS^b \quad \text{and} \quad \oint B_{ab} K^a dS^b.$$

The first integral is a linear mapping from constant vector fields of the embedding Minkowski space to the real numbers, and, as such, is a candidate for the definition of energy—momentum. If $\tilde{\xi}^a$ is the constant vector field whose projection is ξ^a, then the energy—momentum P_a acts on $\tilde{\xi}^a$ according to

$$P_a \tilde{\xi}^a := \oint E_{ab} \xi^a dS^b.$$

The second integral maps Lorentz rotational Killing vector fields linearly to the reals, as does an angular momentum tensor in Minkowski space. Since any Killing vector field K^a of the hyperboloid can be expressed in terms of a pair of constant vector fields of the embedding space by $K^a = \epsilon^{abc} \tilde{\xi}_b \tilde{\eta}_c$, the angular momentum can be represented as a skew tensor M_{ab}:

$$M_{ab} \tilde{\xi}^a \tilde{\eta}^b := \oint B_{ab} \epsilon^{acd} \xi_c \eta_d dS^b.$$

The origin dependence of this angular momentum integral is contained in its dependence on the particular hyperboloid foliation chosen. A foliation centered on a point $\hat{x}$ will produce a different angular momentum tensor than a foliation centered on x. Under a change of center point, the unit normal is affected asymptotically according to $\hat{n}^a \approx n^a + \sum \psi^a$. Here ψ^a is a tangential vector field which can be described this way: The change of center point in M defines a vector $\tilde{\psi}^a = \hat{x}^a - x^a$ which can be regarded as a constant vector field on M. Its projection ψ^a into the unit hyperboloid identifies a vector field on $\mathcal{P}$, and, by Lie dragging along the original normal congruence, a vector field in a neighborhood of $\mathcal{P}$. The change in foliation and unit normal causes a change in the magnetic curvature:

$$\hat{B}_{ab} = \hat{n}^c C^*{}_{cabd} \hat{n}^d \approx B_{ab} - \sum E^m_{(a} \epsilon_{b)mn} \psi^n.$$

Then

$$\hat{M}_{ab}\tilde{\xi}^a\tilde{\eta}^b = \oint \hat{B}_{ab}\epsilon^{acd}\xi_c\eta_d dS^b$$
$$= M_{ab}\tilde{\xi}^a\tilde{\eta}^b + \oint E^m_{(a}\epsilon_{b)mn}\psi^n\epsilon^{acd}\xi_c\eta_d dS^b.$$

Now examine this final integral. Suppose $\tilde{\psi}^a$ is timelike and the integration is carried out over the unit sphere section of $\mathcal{P}$ orthogonal to ψ^a. Then $dS^b \propto \psi^b$ and so only the integral

$$\tfrac{1}{2}\oint E^m_b\epsilon_{amn}\epsilon^{acd}\psi^n\xi_c\eta_d dS^b = \tfrac{1}{2}\oint(E^c_b\psi^d - E^d_b\psi^c)\xi_c\eta_d dS^b$$

survives. This is equal to $\frac{1}{2}(P_a\tilde{\xi}^a\tilde{\psi}_b\tilde{\eta}^b - P_a\tilde{\eta}^a\tilde{\psi}_b\tilde{\xi}^b)$. The angular momentum dependence on foliation is therefore of the familar form:

$$\hat{M}_{ab} = M_{ab} + P_{[a}\tilde{\psi}_{b]}.$$

(No generality was lost by assuming $\tilde{\psi}^a$ to be timelike since one can choose a timelike basis for constant vector fields in Minkowski space.)

The asymptotic linearized gravitational field in this way gives rise to quantities which have the properties of energy—momentum and angular momentum. Calculating these quantities for simple examples yields the expected answers. The energy—momentum of the Kerr metric (linearized) is a future-oriented timelike vector P_a of norm m, and the angular momentum tensor is orthogonal to a plane containing P_a and has magnitude ma. Here m and a are the familiar parameters appearing in the Kerr metric.

In summary, then, a linearized gravitational field of a bounded system induces on $\mathcal{P}$ a tensor field E_{ab} from which one calculates the energy—momentum of the field. Because the field B_{ab} is required to fall off faster than E_{ab}, its limit at $\mathcal{P}$ is foliation-dependent. Restricting consideration to hyperboloid foliations centered on points of M, one obtains for each foliation an angular momentum tensor. Changing the foliation center point causes the angular momentum to transform as an angular momentum tensor should under change of origin.

3. ASYMPTOTICALLY FLAT SPACETIMES

In order to qualify as being asymptotically flat, a spacetime $(\mathcal{M}, g_{ab})$ should have some properties in common with Minkowski space (M, η_{ab}) asymptotically. In particular, there should exist a manifold with boundary, $\bar{\mathcal{M}} = \mathcal{M} \cup \mathcal{P}$, with $\mathcal{P} \approx \mathbb{R}\times S^2$, and a scalar field Σ on some neighborhood $\bar{\mathcal{M}}_*$ of $\mathcal{P}$, with $\Sigma = 0$ on $\mathcal{P}$ and $\Sigma > 0$ on $\mathcal{M}_* = \bar{\mathcal{M}}_* - \mathcal{P}$. The Σ-foliation should resemble the hyperboloid foliations in Minkowski space in the following sense. On $\mathcal{M}_*$ there should be a continuous tensor-field which agrees with $\Sigma^2 h_{ab}$ on $\mathcal{M}_*$ and is the unit hyperboloid metric p_{ab} on $\mathcal{P}$. The extrinsic curvature of the foliation should be such that ΣK_{ab} has the limit $-p_{ab}$ at $\mathcal{P}$. It is assumed, moreover, that the curves orthogonal to the Σ-foliation are the restrictions of a congruence on $\bar{\mathcal{M}}_*$ which meets $\mathcal{P}$ transversely.

If these conditions are satisfied, then one can display a flat metric η_{ab} on $\mathcal{M}_*$ which is an asymptote of the spacetime metric. Let p_{ab} be the tangential tensor field on $\bar{\mathcal{M}}_*$ obtained by Lie dragging the unit hyperboloid metric from $\mathcal{P}$ along the orthogonal congruence. Then the flat metric is

$$\eta_{ab} = -\Sigma^{-4}(\partial_a\Sigma)\partial_b\Sigma + \Sigma^{-2}p_{ab}.$$

The spacetime metric g_{ab} is asymptotic to η_{ab} in the sense that the two tensor fields, when decomposed into tangential fields, have the same limits at $\mathcal{P}$. By construction, both metrics give the same congruence orthogonal to the Σ-foliation. The tangential part of either metric, when multiplied by Σ^2, has the unit hyperboloid metric p_{ab} as limit at $\mathcal{P}$. To see that the normal parts of the metrics also agree at $\mathcal{P}$, one notes that, for the flat metric, the unit normal satisfies $\Sigma^{-2}n^a\partial_a\Sigma = -1$. From the assumptions that $\Sigma^2 h_{ab} \to p_{ab}$ and $\Sigma K_{ab} \to -p_{ab}$, it is not difficult to verify that $\Sigma^{-2}n^a\partial_a \times\Sigma \to -1$ for the unit normal of the nonflat metric as well.

This flat "background" metric facilitates comparison of conventional ADM formalism with the $\mathcal{P}$-boundary description of infinity. Also, if one knows a suitable flat "background" metric for a spacetime, the construction of the $\mathcal{P}$-boundary is straightforward: the Σ-foliation may be taken to be one of the hyperboloid foliations of the flat metric.

It is important that the boundary $\mathcal{P}$, when it exists, be unambiguous. Of course, a spacetime may have more than one asymptotic region. (This occurs for example in the Schwarzschild—Kruskal spacetime which has two asymptotic regions joined by the Einstein—Rosen bridge.) It is presumably the case, however, that, for any single asymptotic spacetime region $\mathcal{M}_*$, there is at most one C^1 manifold with boundary $\bar{\mathcal{M}}_*$ which satisfies the conditions above. This conjecture seems very likely to be true.

In order for the energy—momentum of the spacetime to be well defined in terms of the asymptotic geometry, it is necessary that the scalar field Σ can be chosen so that E_{ab} is of order Σ^3, where $E_{ab} = n^c C_{cabd} n^d$ and n^a is the unit vector field normal to the Σ-foliation. The expression for the energy—momentum is formally identical to the one in linearized theory.

The spacetime must satisfy slightly stronger asymptotic conditions if angular momentum is to be well defined. It must be possible to choose Σ so that E_{ab} is of order Σ^3, B_{ab} is of order Σ^4, and ΣK_{ab} is differentiable at $\mathcal{P}$. The angular momentum integral is then formally the same as in the linearized theory. In contrast to the linearized case, however, the Σ-foliation cannot be chosen to be exactly a family of hyperboloids. If Σ is a scalar field satisfying the asymptotic conditions, then so is $\hat{\Sigma}$ if $\hat{\Sigma} \approx \Sigma + f\Sigma^2$, where f is constant along the congruence orthogonal to Σ-leaves and $\approx$ indicates that higher order terms may also be present. Rather than a four-parameter family of allowed foliations as in Minkowski space, the allowed refoliations now depend on an arbitrary function on $\mathcal{P}$. The limit of B_{ab}, and hence also the angular momentum tensor M_{ab}, varies with the function f. Is there a way to restrict further the asymptotic foliations so that one obtains a suitable four-parameter family of angular momentum tensors? Ashtekar[12] has suggested such a method. One notices that, since $B_{ab} = 0$ to order Σ^3, $D_{[a}K_{b]c} = 0$ to order Σ^3. This implies that the trace-free part of K_{ab} (which is of order Σ^2 or higher) is derived from a scalar field[13]:

$$K_{ab} \approx D_a D_b f + \text{trace terms}.$$

Under a refoliation $\Sigma \mapsto \hat{\Sigma} = \Sigma + f\Sigma^2$, the trace-free part of K_{ab} changes by[14] $D_a D_b f$, so one can always choose a foliation to make the trace-free part of K_{ab} vanish to order Σ^2. The allowed refoliations preserving this condition are then of the form $\hat{\Sigma} = \Sigma + \alpha\Sigma^2$, where α is one of a four-parameter family of scalar fields on $\mathcal{P}$ obtained as the divergence of a conformal Killing vector field on $\mathcal{P}$. To the relevant order, these restricted refoliations are the same at $\mathcal{P}$ as the refoliations by change of center point in Minkowski space. One obtains, then, a four-parameter family of angular momentum tensors with the familiar transformation behavior. The property that K_{ab} be pure trace to order Σ^2 mimics the hyperboloid foliations of Minkowski space for which K_{ab} is exactly proportional to the 3-metric. It is for this reason that the Ashtekar method is natural.

In linearized theory, by virtue of E_{ab} being curl-free and B_{ab} being the curl of a potential, ten conserved integrals vanished. Do those same formal quantities vanish for asymptotically flat spacetimes as well? The integrals $\oint E_{ab} K^a dS^b$ vanish as before provided $D_{[a}E_{b]c} = 0$, and this will hold if B_{ab} is of order Σ^4 and K_{ab} is pure trace to order Σ^2. In other words, when angular momentum is well defined, then the electric curvature limit gives only energy—momentum. The argument that $\oint B_{ab}\xi^a dS^b$ vanishes in linearized theory used the fact that B_{ab} was the curl of a symmetric tensor. Using the relation between magnetic curvature and extrinsic curvature, $B_{ab} = \epsilon_a{}^{cd} D_c K_{db}$, it is seen that the Σ^3 part of K_{ab} is a potential for the tensor B_{ab} which is of order Σ^4. Provided ΣK_{ab} is twice differentiable at $\mathcal{P}$, therefore, the potential exists and those four conserved quantities vanish.

In summary, there are several distinct degrees of asymptotic flatness which may be usefully identified. A spacetime may admit the boundary $\mathcal{P}$ and asymptotic metric η_{ab} but fail to have well-defined energy—momentum and angular momentum. Under more restrictive conditions the energy—momentum exists but not necessarily the angular momentum. Still stronger conditions are needed to have both energy—momentum and angular momentum exist.

ACKNOWLEDGMENTS

Thanks are due to a large number of persons for helpful discussions about spacelike infinity. It is a pleasure to acknowledge especially J.W. York and A. Ashtekar.

[1]H. Bondi, M.G.J. van der Burg, and A.W.K. Metzner, Proc. Roy. Soc. Lond. A **269**, 21 (1962).
[2]R. Penrose, Proc. Roy. Soc. Lond. A **284**, 159 (1965).
[3]A. Ashtekar and R.O. Hansen, "Unified Null and Spatial Infinity," in preparation.
[4]R. Arnowitt, S. Deser, and C.W. Misner, "The Dynamics of General Relativity," in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[5]N. O'Murchadha and J.W. York, Phys. Rev. D **10**, 428 (1974).
[6]D. Brill, Ann. Phys. (N.Y.) **7**, 466 (1959).
[7]R. Geroch, J. Math. Phys. **13**, 956 (1972). Also the Geroch article in *Asymptotic Structure of Spacetime*, edited by P. Esposito and L. Witten (Plenum, New York, 1977).
[8]J. Ehlers, A. Rosenblum, J.N. Goldberg, and P. Havas, Astrophys. J. **208**, L77 (1976).
[9]L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields*, (Pergamon, New York, 1962), rev. 2nd ed., Eq. 104.12.
[10]A. Ashtekar and R. Geroch, Rep. Prog. Phys. **37**, 1211 (1974).
[11]The integrands are seen to be equal by noting that $D_a E^a_m = 0$, $D^a K^b$ is skew-symmetric, $D_{[a}E^b_{m]} = 0$, and $D_a D^b K^m = p^b_a K^m - K^b p^m_a$. This last equation is the general relation for Killing vectors, $\nabla_a \nabla_b K_c = R_{cbad} K^d$, specialized to the hyperboloid with metric p_{ab} and curvature tensor $R_{abcd} = p_{ad} p_{bc} - p_{ac} p_{bd}$.
[12]Privately communicated aspects of the Ashtekar and Hansen "Spi Story."
[13]The trace-free part of a symmetric tensor S_{ab} with vanishing curl on the hyperboloid is necessarily of the form $D_a D_b f - \frac{1}{3} p_{ab} D_c D^c f$. Note first that if ξ^a satisfies $D_a \xi_b = \xi p_{ab}$, then $D_{[a}(S_{b]c}\xi^c) = 0$, i.e., $S_{bc}\xi^c = D_b g$ for some g. By removing the gauge freedom of additive constants in g, one sees S_{ab} as a linear map from conformal Killing vector fields to functions. At any point, a conformal Killing vector field (which can be identified with a constant vector field $\tilde{\xi}^a$ on an embedding Minkowski space) can be represented by a vector ξ^a and scalar $\xi = \tilde{\xi}^a \eta_a$ at the point. Let σ_a and σ be fields such that their action on conformal Killing vectors agrees with that of S_{ab}:

$$\sigma_a \xi^a + \sigma\xi = g.$$

Taking the gradient of both sides and using $S_{ab}\xi^b = D_a g$ and $D_a \xi = \xi_a$, one sees that this can be valid for all conformal Killing fields only if $\sigma_a = -D_a\sigma$ and $S_{ab} = D_a D_b \sigma + \sigma p_{ab}$. This method of proof was suggested by Ashtekar.
[14]The unit normal vector fields are related by $\hat{n}_a \approx n_a + D_a f$. Using the definitions $\hat{h}_{ab} := g_{ab} + \hat{n}_a \hat{n}_b$ and $\hat{K}_{ab} := -\hat{h}^m_a \hat{h}^n_b \nabla_{(m}\hat{n}_{n)}$ together with the fact that $\Sigma K_{ab} \approx -p_{ab} + \Sigma q_{ab}$, one finds that $\hat{\Sigma}\hat{K}_{ab} \approx -\hat{p}_{ab} + \hat{\Sigma}(q_{ab} + D_a D_b f)$, where $\hat{p}_{ab} = p_{ab} + 2n_{(a}D_{b)}f$.

On the Palatini method of variation

Michael Tsamparlis

Mathematics Department, [a] *Imperial College, London SW7, United Kingdom*
(Received 2 May 1977)

The Palatini method of variation is compared with the Hilbert method for symmetric metrics and affine connections. It is found that the two methods are in general inequivalent. The Hilbert method is recommended as being more general.

1. INTRODUCTION

The use and the range of the Palatini method of variation and its equivalence to the Hilbert action principle has been discussed in the literature.[1–4] However, recently published work[4] shows that a reexamination of this method within the context of general relativity is necessary.

The Palatini method of variation[5] requires the action integral

$$I=\int L_{\rm geom}d^4x \tag{1}$$

to be stationary under arbitrary independent variations of the (symmetric) metric g_{ij} and the (symmetric) connection Γ^i_{jk}. The $L_{\rm geom}$ in Eq. (1) is defined by

$$L_{\rm geom}\equiv\sqrt{-g}B, \tag{2}$$

where B is a curvature scalar of the symmetric connection $\Gamma^i{}_{jk}$. Two sets of field equations result. The first set reduces to the geometric identity (metricity condition)

$$\nabla g_{ij}=0 \tag{3}$$

which forces $\Gamma^i{}_{jk}=\{^{\,i}_{jk}\}$. The second set is the Einstein vacuum field equations when (3) is taken into account.

The Hilbert action principle is another method for deriving the Einstein vacuum field equations. In this method one assumes the connection to be the Levi-Civita one, i.e., $\Gamma^i{}_{jk}=\{^{\,i}_{jk}\}$, and takes the unique $L_{\rm geom}\equiv\sqrt{-g}R$, where R is the curvature scalar of the Levi-Civita connection. Variation of this action yields one set of field equations which is the Einstein vacuum field equations. Equivalently, instead of considering $\Gamma^i{}_{jk}=\{^{\,i}_{jk}\}$ one introduces Lagrange multipliers and considers the *a priori* assumption $\Gamma^i{}_{jk}=\{^{\,i}_{jk}\}$ as a constraint[3], i.e., one redefines the Lagrangian to be

$$L'_{\rm geom}(g,\Gamma,\Lambda)\equiv\sqrt{-g}\,B+\Lambda_i{}^{jk}(\Gamma^i{}_{jk}-\{^{\,i}_{jk}\}), \tag{4}$$

where B is the *a* curvature scalar of a general connection.

It has been pointed out[5,6] that the results of these two methods of variation are equivalent for the Lagrangian

$$L_{\rm matter}\equiv\sqrt{-g}\,B+L_\phi, \tag{5}$$

where $L_\phi(g,\phi,\partial\phi)$ is the Lagrangian of some tensor fields ϕ (indices suppressed) which does not contain derivatives of the metric. B is the curvature scalar of a symmetric affine connection.

However, these two methods are fundamentally different. The main difference lies in the way they treat the *a priori* assumption(s)

$$\Gamma^i{}_{jk}=\Gamma^i{}_{kj}, \tag{6a}$$

$$g_{ij}|_k=0. \tag{6b}$$

The Palatini method assumes the first as a constraint (i.e., $\delta\Gamma^i_{jk}=\delta\Gamma^i_{kj}$) and shows that the second is trivially satisfied, so to speak, an unnecessary selection rule. However, in generalizations of the Palatini method, e.g. in the derivation of the field equations in the Einstein–Cartan theory,[7] the role of (6b) as a selection rule is clearly not trivial. The Hilbert action principle treats both (6a) and (6b) as constraints either by assuming $\Gamma^i_{jk}=\{^{\,i}_{jk}\}$ or by introducing Lagrange multipliers and adding the term

$$\Lambda^i{}_{jk}(\Gamma^i{}_{jk}-\{^{\,i}_{jk}\})$$

to the Lagrangian. We will use the term Hilbert for the *a priori* assumptions and the term Hilbert–Lagrange multiplier for the latter. They are obviously equivalent.

In the following we show that:

(I) The Palatinin method and the Hilbert action principle are equivalent for empty space no matter which curvature scalar of Γ is taken.

(II) When matter is present, the two methods are not in general equivalent.[6] More precisely, for the Palatini method to produce the Einstein field equations, we require that L_ϕ does not contain covariant derivatives. The Hilbert action principle allows the presence of covariant derivatives. Also, the Palatini method does not yield field equations consistent with the metricity condition (6b) when L_ϕ contains derivatives with respect to the symmetric affine connection Γ^i_{jk} whereas the Hilbert method does.

2. THE LAGRANGIAN $L_{\rm geom}$

If we exclude the possibility of a nonsymmetric metric tensor, $g_{\mu\nu}$, one can easily show that the components of a general linear symmetric connection Γ^i_{jk} are given by

$$\Gamma^i_{jk}=\{^{\,i}_{jk}\}-\tfrac12 g^{il}\Delta_{jlk}, \tag{7}$$

where

$$\Delta_{jlk}=g_{jl|k}+g_{lk|j}-g_{kj|l} \tag{8}$$

with "|" denoting covariant derivatives. The curvature

[a]Correspondence address: 16, Admitou Street, Agios Ierotheos, Peristeri, Athens, Greece.

 0022-2488/78/1903-0555$1.00

tensor of Γ^i_{jk} is defined as usual by

$$B^i{}_{jkl}=2\partial_{[k}\Gamma_{l]}{}^i{}_j+2\Gamma^i{}_{s[k}\Gamma^s_{|j|l]}. \tag{9}$$

One proves easily that

$$B^r{}_{rjk}=2\delta^r{}_{[j}\delta^s{}_{k]}[(ln\sqrt{-g})_{|r}]_{;s} \tag{10}$$

(not sum over r). Thus

$$B_{ijkl}=-B_{jikl} \quad \text{if and only if} \quad g_{ij|k=0}. \tag{11}$$

So, in general one has three Ricci tensors

$$B_{jk}\equiv B^r{}_{jrk},\quad B'_{jk}\equiv B^r{}_{rjk},\quad B''_{jk}\equiv B_j{}^r{}_{rk} \tag{12}$$

and one curvature scalar

$$B=g^{jk}B_{jk},\quad B_1=g^{jk}B''_{jk}=-B. \tag{13}$$

The third possibility was not considered by Schrödinger.[8]

3. THE PALATINI METHOD

Let us assume the Lagrangian of the geometry to be

$$L_{\text{geom}}=\sqrt{-g}B. \tag{14}$$

Varying this by the Palatini method, we find the follow ing sets of field equations:

$$B_{(ij)}-\tfrac{1}{2}g_{ij}B=0, \tag{15}$$

$$2\delta^{(c}_{[d}{}_{a]}g^{b)d}+\tfrac{1}{2}g^{r(b}\delta^{c)}_a g^{kl}\Delta_{rlk}-\tfrac{1}{2}g^{r(b}\delta^{c)}_r g^{sl}\Delta_{als}=0, \tag{16}$$

with ( ) denoting symmetrization and [ ] antisymmetrization. Equation (16) can be rewritten as

$$[-\delta^b_j\delta^c_l\delta^k_a+\tfrac{1}{2}\delta^k_a\delta^b_i\delta^k_j+\tfrac{1}{2}\delta^b_a\delta^c_j\delta^k_i+\tfrac{1}{2}\delta_a{}^{(c}g^{b)k}g_{ij}$$
$$-\tfrac{1}{2}g^{bc}g_{ij}\delta^k_a]g^{ij}{}_{|k}=0. \tag{17}$$

This equation appears to be different than the one usually found in textbooks [see Ref. 5, Eq. (21.25)]. In Appendix A we show that Eq. (17) which is a system of 40 linear simultaneous equations in the 40 unknowns $g^{ab}{}_{|c}=0$, that is, (6b). Thus Eqs. (15) reduce to the standard Einstein vacuum field equations.

Let us assume now that there are matter fields present and let the Lagrangian be

$$L_{\text{matter}}=L_{\text{geom}}+L_\Phi \tag{18}$$

where $L_\Phi(g,\partial g;\Gamma,\partial\Gamma;\phi,\partial\phi)$ is the Lagrangian of the matter fields, ϕ, and L_{geom} is defined in Eq. (14). If we define

$$-\sqrt{-g}k,\ T_{ij}\equiv\frac{\partial L_\Phi}{\partial g_{ij}} \tag{19}$$

$$\sqrt{-g}k_2M_i{}^{jk}\equiv\frac{\partial L_\Phi}{\partial\Gamma^i{}_{jk}} \tag{20}$$

and vary the L_{matter} by the Palatini method, we obtain the following sets of field equations:

$$B_{(ij)}-\tfrac{1}{2}g_{ij}B=K_1T_{ij}, \tag{21}$$

$$[g^k{}_cg_{ji}\delta^b_a-g^{bc}g_{ji}\delta^k_a+\delta^k_j\delta^b_a\delta^c_i+2\delta^b_i\delta^c_j\delta^k_a-2\delta^k_i\delta^b_a\delta^c_j$$
$$-\delta^c_a\delta^b_i\delta^k_j]g^{ij}{}_{|k}=K_2M_a{}^{bc}, \tag{22}$$

$$\frac{\partial L_\Phi}{\delta\phi}=0. \tag{23}$$

Equation (22) has a unique solution (because the corresponding homogeneous system has a unique solution) which is not zero unless, and only unless,

$$M_a{}^{bc}=0, \tag{24}$$

i.e., when L_Φ does not contain any covariant derivatives. Therefore, for nonempty space—times the Palatini method does not yield field equations which are consistent with the metricity condition (6b). In conclusion, a necessary and sufficient condition that the Palatini method of variation applied to L_{matter} give the Einstein field equations is that $M_a{}^{bc}=0$.

4. PALATINI-HILBERT

We look now at the Hilbert action principle and compare it with the Palatini method. The Hilbert Principle gives identical results to a Lagrange multiplier approach.[3] We introduce the Lagrange multipliers $\Lambda_a{}^{bc}$ and redefine the Lagrangian L_{matter} as follows:

$$L'_{\text{matter}}(g,\partial g;\Lambda;\Gamma;\phi,\partial\phi)\equiv L_{\text{matter}}+\Lambda_a{}^{bc}(\Gamma^a{}_{bc}-\{^a_{bc}\}), \tag{25}$$

where $\Lambda_a{}^{bc}=\Lambda_a{}^{cb}$. Independent variations of the fields Λ,ϕ,g,Γ give respectively

$$\Gamma^i{}_{jk}=\{^i_{jk}\}, \tag{26}$$

$$\frac{\delta L_\Phi}{\delta\phi}=0, \tag{27}$$

$$R_{ab}-\tfrac{1}{2}g_{ab}R=K_1T_{ab}+(1/\sqrt{|g|})\Lambda_r{}^{kl}g_{mb}\delta_a{}^{(r}\{^{m)}_{kl}\}$$
$$-(1/2\sqrt{|g|})[\Lambda_r{}^{kl}(\delta^r{}_{(a}g_{b)k}\delta^c_l$$
$$+\delta^r{}_{(a}g_{b)l}\delta^c_k-g_{k(a}g_{b)l}g^{rc})]_{,c}, \tag{28}$$

$$\Lambda_a{}^{bc}=K_2M_a{}^{bc}, \tag{29}$$

where T_{ab}, $M_a{}^{bc}$ are defined in (19) and (20), respectively. We look for the necessary and sufficient conditions that these equations reduce to the Einstein field equations, but not necessarily the vacuum. Choose coordinate conditions $g_{ij,k}=0$. Then Eq. (28) reduces to

$$G_{ab}\equiv T_{ab}-\tfrac{1}{2}g_{ab}R$$
$$=K_1T_{ab}-(1/2\sqrt{|g|})[\delta^r{}_{(a}g_{b)k}\delta^c_l+\delta^r{}_{(a}g_{b)l}\delta^c_k$$
$$-g_{k(a}g_{b)l}g^{rc}]\Lambda^{kl}{}_{r,c}. \tag{30}$$

Taking into account Eq. (29) this becomes

$$G_{ab}=K_1T_{ab}-(K_2/2\sqrt{|g|})[M_{ab}{}^c+M_{ba}{}^c-M^c{}_{ab}]_{;c}.$$

Hence, we conclude that:

A necessary and sufficient condition that the Hilbert action principle applied to L'_{matter} give the Einstein field equations is that

$$(M_{ab}{}^c+M_{ba}{}^c-M^c{}_{ab})_{;c}=0.$$

We see that this condition is weaker than the corresponding condition of the Palatini method. We conclude that, in general, the two methods are not equivalent but the Hilbert action principle is more general than the Palatini method. More precisely they are equivalent only when the Lagrange multipliers $\Lambda_c{}^{ab}=K_2M_c{}^{ab}$ vanish,[6,3] a well-known result. Besides the fact that the Hilbert action principle is more general, it is also more useful because it produces field equations for any kind of fields, which are consistent with the *a priori* assumed Riemannian (or even more general) structure(s) of space–time.

Considering these results and also the results of the previous work done on the Palatini method, we suggest that, perhaps, this method should be abandoned in favor of the Hilbert action principle (with or without Lagrange multipliers).

ACKNOWLEDGMENTS

I am indebted to Dr. P. Dolan, Professor J. L. Safko, and Dr. F. Elston for a critical reading of the manuscript and for constructive suggestions.

APPENDIX

In this appendix we show that (17) has the unique solution $g^{ab}{}_{|c}=0$. Choose a coordinate system in which the metric tensor g_{ab} reduces to its diagonal form, i.e., η_{ab}, where η_{ab} is the Minkowski metric. Then (17) reads

$$[-\delta^b_t\delta^c_i\delta^m_a+\tfrac{1}{2}\delta^c_a\delta^b_t\delta^m_i+\tfrac{1}{2}\delta^b_a\delta^c_i\delta^m_t + \tfrac{1}{2}\delta^{(c}_a\eta^{b)m}\eta_{ti}-\eta^{bc}\eta_{ti}\delta^m_a]A^{ti}{}_m=0, \tag{A1}$$

where $A^{ab}{}_c=g^{ab}{}_{|c}=A^{ba}{}_c$. We consider the following cases.

Case I: $a\neq b\neq c$

The system (A1) reduces to

$$A^{ab}{}_c=0. \tag{A2}$$

Case II: $a=b\neq c$

The system (A1) reduces to

$$-A^{bc}{}_b+\tfrac{1}{2}A^{mc}{}_m+\tfrac{1}{4}\eta^{bm}\eta_{ti}A^{ti}{}_m=0. \tag{A3}$$

Case III: $a=c\neq b$

The system reduces to Case II because of the symmetry $A^{ab}{}_c=A^{ba}{}_c$.

Case IV: $a\neq b=c$

The system (A1) reduces to

$$-A^{bb}{}_a-\tfrac{1}{2}\eta^{bb}\eta_{ti}A^{ti}{}_a=0. \tag{A4}$$

Case V: $a=b=c$

The system (A1) reduces to

$$-A^{bb}{}_b+A^{bm}{}_m+\tfrac{1}{2}\eta^{bm}\eta_{ti}A^{ti}{}_m-\tfrac{1}{2}\eta^{bb}\eta_{ti}A^{ti}{}_b=0. \tag{A5}$$

In (A4) we have the following subcases:

If $b=0$, then

$$A^{00}{}_c-\eta_{ti}A^{ti}{}_c=0. \tag{A6}$$

If $b=k$ ($k=1,2,3$), then

$$A^{kk}{}_c+\eta_{ti}A^{ti}{}_c=0. \tag{A7}$$

Equation (A7) implies

$$A^{kk}{}_c=0 \quad\text{and}\quad \eta_{ti}A^{ti}{}_c=0.$$

Substituting in (A6), we find $A^{00}{}_c=0$. These imply

$$A^{bb}{}_c=0 \tag{A8}$$

and

$$\eta_{ti}A^{ti}{}_c=0. \tag{A9}$$

Using Eq. (A9), Eq. (A3) reduces to $-A^{bc}{}_b+A^{mc}{}_m=0$ ($b\neq c$), which implies

$$A^{bc}{}_b=0 \tag{A10}$$

and

$$A^{mc}{}_m=0. \tag{A11}$$

Finally, using Eqs. (A9) and (A11), Eq. (A5) gives

$$A^{bb}{}_b=0. \tag{A12}$$

Equations (A2), (A8), (A10), and (A12) prove that the solutions of the system (A1) are

$$A^{ab}{}_c=0. \qquad \text{QED}$$

[1]H. Buchdahl, Proc. Camb. Phil. Soc. 56, 398 (1960).
[2]A. Palatini, Circ. Mat. Palermo 43, 203 (1919).
[3]J.L. Safko and F. Elston, J. Math. Phys. 17, 1531 (1976).
[4]J.L. Safko, M. Tsamparlis, and F. Elston, Phys. Lett. A 60, 1 (1977).
[5]C. Misner, K. Thorne and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[6]J. Ray. Nuovo Cimento B 25, 706 (1975).
[7]A. Trautman, Symposia Mathematica 12, 139 (1973).
[8]E. Schrödinger, *Space-Time Structure* (Cambridge U.P., Cambridge, 1963).

Global thermodynamical stability and correlation inequalities

M. Fannes[a] and A. Verbeure

Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Heverlee, Belgium
(Received 10 August 1977)

Using spatially homogeneous dissipative perturbations, we derive a correlation inequality for states satisfying the variational principle for infinitely extended quantum lattice systems.

1. INTRODUCTION

For finite quantum systems it is well known that the Gibbs states can be characterized by the KMS property. In Ref. 1 it is proved that this property persists in the thermodynamic limit for a large class of systems. It is therefore generally believed that this property is characteristic for the equilibrium states of infinite systems. To analyze this idea, different notions of stability have been analyzed. In particular two essentially different types of stability have been studied. On the one hand, it has been proved[2] that dynamical stability, supplemented with cluster properties leads to KMS states. On the other hand, we have the notion of thermodynamic stability, as expressed by different variational principles for the free energy. We distinguish the global thermodynamical stability (GTS) and the local thermodynamical stability (LTS), which both have been proved to be equivalent with KMS for quantum lattice systems.[3,4]

The implication LTS implies KMS in Ref. 4 relies on the derivation of correlation inequalities from LTS. One shows then as in Ref. 5, that these inequalities imply KMS. Also these inequalities follow from KMS[6,7,8] and therefore they are equilibrium conditions.

The main result of this paper (Sec. 3) is to derive a correlation inequality from GTS. The derivation in Ref. 4 from LTS was based on the use of local dissipative perturbations with bounded generators of the Lindblad type.[9] These local perturbations however are irrelevant for considerations on intensive observables like energy and entropy density. Therefore, we have to introduce spatially homogeneous dissipative perturbations. These are studied in Sec. 2, where we construct strongly-continuous spatially homogeneous semigroups of completely positive maps as volume limits of local ones. The physical idea behind these dissipative perturbations, in contradistinction with automorphic perturbations, is that the system can be looked upon as a subsystem in interaction with a heat bath. The appropriate heat bath is found by considering the dilation of the corresponding dissipative evolutions.

Finally we introduce the following notations, borrowed from Ref. 10. We consider the lattice set $\mathbb{Z}^\nu$ ($\nu\in\mathbb{N}^+$). For each finite subset $\Lambda\subset\mathbb{Z}^\nu$, $N(\Lambda)$ denotes the number of points in Λ. Denote by $\mathcal{A}$ the quasilocal algebra of a spin lattice system; $\mathcal{A}$ is the closure of the union of all local algebras $\mathcal{A}_\Lambda$. Let $\mathcal{A}_L=\cup_\Lambda\mathcal{A}_\Lambda$ be the algebra of strictly local elements.

Furthermore let $\{\tau_k, k\in\mathbb{Z}^\nu\}$ be the group of space translation automorphisms on $\mathcal{A}$. For simplicity the volume limits which we consider are taken in the sense of increasing cubes.

2. TRANSLATION-INVARIANT SEMI-GROUPS

In this section we construct spatially homogeneous semigroups by taking the Lindblad[9] form of the generator for local elements and by translating it all over the lattice. For any element $v\in\mathcal{A}$, denote by $\mathcal{L}_v$ the generator. Then we have

$$\mathcal{L}_v:\ a\in\mathcal{A}\to\mathcal{L}_v(a)=v^*av-\tfrac{1}{2}(v^*va+av^*v).$$

Theorem 2.1[15]: Let $v\in\mathcal{A}_\Lambda$, then for any $x\in\mathcal{A}_L$ the map γ from $\mathcal{A}_L$ into $\mathcal{A}_L$,

$$\gamma(x)=\sum_{k\in\mathbb{Z}^\nu}\mathcal{L}_{\tau_k(v)}(x), \tag{1}$$

is well defined and generates a spatially homogeneous strongly continuous semigroup $(e^{\gamma t})_{t\geq 0}$ of completely positive (CP) unity preserving transformations of $\mathcal{A}$.

Proof: Because of the local commutativity property of the algebra, i.e., if $\Lambda\cap\Lambda'=\emptyset$, then $[\mathcal{A}_\Lambda,\mathcal{A}_{\Lambda'}]=0$, the series in (1) contains only a finite number of terms; therefore γ is well defined.

Furthermore, for $x\in\mathcal{A}_{\Lambda'}$,

$$\|\gamma(x)\|\leq 2N(\Lambda)N(\Lambda')\|v\|^2\|x\|.$$

Indeed $\gamma(x)$ is a sum of at most $N(\Lambda)N(\Lambda')$ elements belonging to algebras $\mathcal{A}_{\Lambda'\cup\Lambda+k}$, $k\in\mathbb{Z}^\nu$.

By induction,

$$\begin{aligned}\|\gamma^n(x)\|&\leq 2^nN(\Lambda)^nN(\Lambda')[N(\Lambda')+N(\Lambda)]\cdots[N(\Lambda')+(n-1)N(\Lambda)]\|v\|^{2n}\|x\|\\&\leq n!\,2^nN(\Lambda)^2\frac{(N(\Lambda')/N(\Lambda)+(n-1))^n}{n!}\|v\|^{2n}\|x\|\\&\leq n!\,[2eN(\Lambda)^2\|v\|^2]^n\exp\left(\frac{N(\Lambda')}{N(\Lambda)}\right)\|x\|.\end{aligned}$$

Hence for $0\leq t<t_0=[2eN(\Lambda)^2\|v\|^2]^{-1}$ the series

$$\sum_{n=0}^{\infty}t^n\frac{\|\gamma^n(x)\|}{n!}<\infty.$$

converges and defines $(e^{\gamma t})_{0\leq t<t_0}$. It is also clear that

$$e^{\gamma t}=\operatorname*{s-lim}_{M\to\mathbb{Z}^\nu}\exp(\gamma_Mt),\quad 0\leq t<t_0$$

where

$$\gamma_M(x)=\sum_{k\in M\subset\mathbb{Z}^\nu}\mathcal{L}_{\tau_k(v)}(x).$$

[a]Aangesteld Navorser NFWO, Belgium.

 0022-2488/78/1903-0558$1.00

By Ref. 9 $\exp(\gamma_M t)$ is a CP unity preserving transformation, hence $e^{\gamma t}$ shares this property on $\mathcal{A}_L$. Further as $\|\exp(\gamma_M t)\| = 1$, $e^{\gamma t}$ extends to $\mathcal{A}$ by continuity for $0 \le t < t_0$.

Removal of the condition on t goes as follows: for all t_1, t_2 such that $0 \le t_1, t_2, t_1 + t_2 < t_0$ from the series expansion, $\exp[\gamma(t_1+t_2)] = \exp(\gamma t_1)\exp(\gamma t_2)$. For arbitrary $t \ge 0$, choose $N \in \mathbb{N}$ such that $t/N < t_0$ and define $e^{\gamma t} = (e^{\gamma t/N})^N$. The strong continuity $t \to e^{\gamma t}$ follows immediately.

3. CORRELATION INEQUALITY FROM GTS

As usual consider the lattice sites occupied by particles interacting through many-body potentials $\phi(X) \in \mathcal{A}_X$, $X \subset \mathbb{Z}^\nu$. We consider the interactions ϕ satisfying:

(i) $\phi(X)$ is Hermitian;

(ii) $(\tau_k\phi)(X) = \phi(X+k)$, translation invariance;

(iii) $\|\phi\| = \sum_{0 \in X} \|\phi(X)\| < \infty$.

Let $\mathcal{B}$ be the set of these potentials. For any $\phi \in \mathcal{B}$ the local Hamiltonians are given by: For $\Lambda \subset \mathbb{Z}^\nu$,

$$H_\phi(\Lambda) = \sum_{X \subset \Lambda} \phi(X).$$

Definition 3.1: A spatially homogeneous state ω on $\mathcal{A}$ satisfies the variational principle or is globally thermodynamically stable (GTS) if it minimizes the expression

$$f(\omega) = \omega(A_\phi) - \frac{1}{\beta} s(\omega)$$

where $\beta = 1/kT$ and

$$A_\phi = \sum_{0 \in X} \frac{\phi(X)}{N(X)},$$

which by (iii) is an element of $\mathcal{A}$ representing the energy per lattice site; $s(\omega)$ is the entropy density (see Ref. 11)

Lemma 3.2: With the notation used above, for any state ω on $\mathcal{A}$ satisfying Definition 3.1 $v \in \mathcal{A}_L$ and γ as in (1), we have

$$\lim_{t \to 0^+} \omega\left(\frac{(e^{\gamma t}-1)}{t} A_\phi\right) = \lim_{\Gamma \to Z^\nu} \omega(v^*[H_\phi(\Gamma), v]).$$

Proof: As $v \in \mathcal{A}_L$, there exists a $\Lambda \subset \mathbb{Z}^\nu$, such that $v \in \mathcal{A}_\Lambda$. From the proof of Theorem 2.1,

$$\|\gamma(\phi(X))\| \le 2N(\Lambda)N(X)\|v\|^2 \|\phi(X)\|.$$

Therefore,

$$\|\gamma \sum_{0 \in X \subset M \subset \mathbb{Z}} \phi(X)/N(X)\| \le \|\phi\| 2N(\Lambda)\|v\|^2.$$

We can define

$$\gamma(A_\phi) = \lim_{M \to \mathbb{Z}^\nu} \gamma\left[\sum_{0 \in X \subset M} \phi(X)/N(X)\right].$$

Because of the strong continuity of the semigroup $t \to e^{\gamma t}$ (Theorem 2.1),

$$\lim_{t \to 0^+} \frac{(e^{\gamma t}-1)}{t} A_\phi = \gamma(A_\phi).$$

Applying the state ω,

$$\lim_{t \to 0^+} \omega\left(\frac{(e^{\gamma t}-1)}{t} A_\phi\right) = \omega(\gamma(A_\phi)). \tag{2}$$

Using the spatial homogeneity of the state,

$$\omega(\gamma(A_\phi)) = \sum_k \omega(v^*[\tau_k A_\phi, v]) + \tfrac{1}{2}\omega([v^*v, \tau_k A_\phi]). \tag{3}$$

We note that for any local observable $b = b^* \in \mathcal{A}_L$,

$$\lim_{\Gamma \to \mathbb{Z}^\nu} [H_\phi(\Gamma), b] = \sum_{k \in \mathbb{Z}^\nu} [\tau_k A_\phi, b] \tag{4}$$

and that, if $b_\Gamma = \sum_{k \in \Gamma} \tau_k b$, the following limit,

$$\lim_{\Gamma \to \mathbb{Z}^\nu} \exp(i\lambda b_\Gamma) c \exp(-i\lambda b_\Gamma) \quad \text{for all } c \in \mathcal{A} \text{ and } \lambda \in \mathbb{R},$$

exists and defines a strongly continuous group of *automorphisms α_λ of $\mathcal{A}$:[12] By Ref. 13, Theorem 5, we have $s(\omega \circ \alpha_\lambda) = s(\omega)$ for all $\lambda \in \mathbb{R}$.

As ω is GTS, for all λ,

$$f(\omega \circ \alpha_\lambda) - f(\omega) = (\omega \circ \alpha_\lambda - \omega)(A_\phi) \ge 0$$

and

$$\lim_{\lambda \to 0} \frac{f(\omega \circ \alpha_\lambda) - f(\omega)}{\lambda} = \omega\left(\sum_k [\tau_k b, A_\phi]\right) \ge 0.$$

As this remains true with b replaced by $-b$ we obtain

$$\omega\left(\sum_k [\tau_k b, A_\phi]\right) = 0.$$

Using the space homogeneity of ω and the above remark, we get for all $b \in \mathcal{A}_L$

$$\lim_{\Gamma \to \mathbb{Z}^\nu} \omega([b, H_\Lambda]) = 0. \tag{5}$$

This relation expresses the time invariance of the state. Combining Eqs. (2)–(5) we get the lemma. ∎

Lemma 3.3: Let ρ be any density matrix on $(\mathbb{C}^n)$, ω the corresponding state $\omega(\cdot) = \mathrm{Tr}\rho$ and $v_i \in \mathcal{B}(\mathbb{C}^n)$, $i = 1, \ldots, k$, For $\gamma = \sum_{i=1}^k \mathcal{L}_{v_i}$,

$$-\omega(\gamma \ln \rho)) \ge \omega\left(\sum_i v_i^* v_i\right) \ln \frac{\omega(\sum_i v_i^* v_i)}{\omega(\sum_i v_i v_i^*)}.$$

Proof: As this lemma is a slight generalization of Ref. 4, Lemma 6, the proof goes along the same lines as in Ref. 4. ∎

Now we have the main theorem:

Theorem 3.4: For any interaction $\phi \in \mathcal{B}$ and state ω satisfying GTS (Definition 3.1 and for all v_i, $i = 1, \ldots, n$, in $\mathcal{A}_L$, the following inequality holds:

$$\beta \lim_{\Gamma \to \mathbb{Z}^\nu} \omega\left(\sum_{i=1}^n v_i^* [H_\phi(\Gamma), v_i]\right) \ge \omega\left(\sum_{i=1}^n v_i^* v_i\right) \ln \frac{\omega(\sum_{i=1}^n v_i^* v_i)}{\omega(\sum_{i=1}^n v_i v_i^*)}.$$

Proof: For notational convenience we restrict ourself to a single local observable $v(n=1)$. The generalization to arbitrary n is immediate. The idea of the proof consists in the claculation of the following limit,

$$\delta f \equiv \lim_{t \to 0^+} \frac{f(\omega \circ e^{\gamma t}) - f(\omega)}{t},$$

γ as in (1), which is nonnegative by the assumption on ω.

We note that

$$\delta f = \lim_{t=0^+} \omega\left(\frac{(e^{\gamma t}-1)}{t} A_\Phi\right) - \delta s(\omega) \geq 0,$$

where

$$\delta s(\omega) = \lim_{t\to 0^+} \frac{s(\omega \circ e^{\gamma t}) - s(\omega)}{t}.$$

By Lemma 3.3

$$\delta f = \lim_{\Gamma \to \mathbb{Z}^\nu} \omega(v^*[H_\Phi(\Gamma), v]) - \delta s(\omega) \geq 0. \qquad (6)$$

Now we compute $\delta s(\omega)$. Consider $(\Gamma(n))_n$ an increasing absorbing set of cubes for $\mathbb{Z}^\nu$. For n large enough, consider the set.

$$I_n = \{k \in \mathbb{Z}^\nu \mid \tau_k v \in \mathcal{A}_{\Gamma(n)}\} \text{ and } \gamma_n = \sum_{k\in I_n} L_{\tau_k v}.$$

Let ω_n be the restriction of ω to $\mathcal{A}_{\Gamma(n)}$ with density matrix ρ_n, and $\tilde{\omega}_n^s$, $s \in \mathbb{R}^+$, the periodic extension on $\mathcal{A}$ of $\omega_n \circ \exp(\gamma_n s)$. The density matrix of the state $\omega_n \circ \exp(\gamma_n s)$ is then given by $\exp(\gamma_n^* s)\rho_n$ where

$$\gamma_n^*(\cdot) = \sum_{k\in I_n} \tau_k v \cdot \tau_k v^* - \tfrac{1}{2}[\tau_k v^* v, \cdot]_+.$$

One checks that

$$w^* - \lim_{n\to\infty} \tilde{\omega}_n^s = \omega \cdot e^{\gamma s}.$$

Using the upper semicontinuity of the entropy density on periodic states,[14]

$$\delta s(\omega) \geq \lim_{t\to 0^+} \lim_{n\to\infty} \frac{s(\tilde{\omega}_n^t) - s(\omega)}{t}$$

$$= \lim_{t=0^+} \lim_{n\to\infty} \frac{s(\omega_n \circ \exp(\gamma_n t)) - S(\omega_n)}{t N(\Gamma(n))},$$

where $S(\omega_n) = -\mathrm{Tr}\rho_n \log \rho_n$.

By Ref. 4, formula 2.13,

$$\delta s(\omega) \geq \lim_{t\to 0^+} \lim_{n=\infty} \frac{-1}{tN(\Gamma(n))} \times \int_0^t ds\, \omega_n \circ \exp(\gamma_n s)\{\gamma_n \ln[\exp(\gamma_n^* s)\rho_n]\}.$$

By Lemma 3.3.,

$$\delta s(\omega) \geq \lim_{t\to 0^+} \lim_{n\to\infty} \frac{1}{t} \int_0^t ds\, \frac{\omega_n \circ \exp(\gamma_n s)(\sum_{k\in I_n} \tau_k(v^*v))}{N(\Gamma(n))} \times \ln \frac{\omega_n \circ \exp(\gamma_n s)[\sum_{k\in I_n} \tau_k(v^*v)]}{\omega_n \circ \exp(\gamma_n s)[\sum_{k\in I_n} \tau_k(vv^*)]}.$$

We note that

$$\lim_{n\to\infty} \frac{\omega_n(\gamma_n[\sum_{k\in I_n} \tau_k(v^*v)])}{N(\Gamma(n))}$$

$$= \lim_{n\to\infty} \frac{1}{N(\Gamma(n))} \omega(\textstyle\sum_{k\in I_n} \tau_k \sum_{l\in I_n - I_n} L_{\tau_l v}(v^*v))$$

$$= \lim_{n\to\infty} \frac{N(I_n)}{N(\Gamma(n))} \omega(\textstyle\sum_{l\in I_n - I_n} L_{\tau_l(v)}(v^*v))$$

$$= \omega(\gamma(v^*v)).$$

Analogously this can be done for any power of γ. Because v^*v is strictly local, we can use the series expansion as in Theorem 2.1., to derive

$$\lim_{n\to\infty} \frac{\omega_n \circ \exp(\gamma_n s)[\sum_{k\in I_n} \tau_k(v^*v)]}{N(\Gamma(n))} = \omega(e^{\gamma s}(v^*v)),$$

uniformly in $s \in [0, \epsilon]$ for ϵ small enough.

Therefore,

$$\delta s(\omega) \geq \omega(v^*v) \ln \frac{\omega(v^*v)}{\omega(vv^*)}. \qquad (7)$$

Combining (6) and (7) we get the proof. ∎

We note that the inequality of Theorem 3.4 is a generalization to n observables v_i, $i = 1, \ldots, n$, of the corresponding inequality which can be found in Refs. 4, 5, and 8. The same kind of generalization of the Roepstorff inequality, Ref. 7, Theorem III.1, can be given and looks as follows,

$$\sum_i (v_i, v_i)_\sim \leq \omega(\sum_i [v_i^*, v_i]) \ \ln \frac{\omega(\sum_i v_i^* v_i)}{\omega(\sum_i v_i v_i^*)},$$

where $(v_i, v_i)_\sim$ is the Duhamel two-point function, given by

$$(v_i, v_i)_\sim = \omega\left(v_i^* \frac{e^H - 1}{H} v_i\right),$$

where

$$Hv_i = \lim_{\Gamma\to\mathbb{Z}^\nu} [H(\Lambda), v_i].$$

[1] R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).
[2] R. Haag, D. Kastler, and E. B. Trych-Pohlmeyer, Commun. Math. Phys. **38**, 173 (1974).
[3] H. Araki, Commun. Math. Phys. **38**, 1 (1974).
[4] G. L. Sewell, Commun. Math. Phys. **55**, 53 (1977).
[5] M. Fannes and A. Verbeure, Commun. Math. Phys. **55**, 125 (1977); **57**, 165 (1977).
[6] G. Roepstorff, Commun. Math. Phys. **46**, 253 (1976).
[7] J. Naudts and A. Verbeure, J. Math. Phys. **17**, 419 (1976).
[8] A. Verbeure, "Characterizations of Equilibrium States," preprint KUL-TF-77/018 to appear in *Algebras of Operators and their Applications to Mathematical Physics*, proceedings on the Marseille Conference 1977, edited by D. Kastler and D. W. Robinson (to be published).
[9] G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).
[10] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[11] D. W. Robinson, Commun. Math. Phys. **6**, 151 (1967).
[12] D. W. Robinson, Commun. Math. Phys. **7**, 337 (1968).
[13] O. E. Lanford, III and D. W. Robinson, Commun. Math. Phys. **9**, 327 (1968).
[14] D. W. Robinson, *The Thermodynamic Pressure in Quantum Statistical Mechanics* (Springer-Verlag, Heidelberg, 1971).
[15] An analogous result can be found in: E. B. Davies, Commun. Math. Phys. **55**, 231 (1977) (Theorem 2.1).

Tachyonic scalar waves in the Schwarzschild space–time

S. V. Dhurandhar

Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India
(Received 24 March 1977)

The scalar wave equation of a tachyon is investigated in the background of Schwarzschild geometry. The scalar field is split up into partial waves of all integral momentum states and the space development of each partial wave is studied as it approaches the singularity. The problem is mainly considered in the light of the assumption that the tachyon mass-parameter is comparable to the mass of a atomic particle while the black hole mass is comparable to that of an average star. The reflection and transmission properties of these partial waves at the effective potential barrier, arising partly from their angular momentum and partly from the curvature of space–time are discussed. It is found that in the radial case ($l = 0$) the criteria for the bounce are different from the purely classical behavior of spacelike geodetic trajectories.

1. INTRODUCTION

In the recent years several authors have considered the interaction of tachyons with gravitation. Narlikar and Sudarshan[1] have showed that a primordial tachyon in a big bang universe is eventually reflected at a time barrier. Dhurandhar[2] has considered the propagation of tachyons inside a white hole. Honig *et al.*[3] and Narlikar and Dhurandhar[4] have discussed the classical geodetic trajectories of these particles in Schwarzschild space—time. They have shown that, the curvature of space—time gives rise to a potential barrier from which the low energy tachyons, i. e., those with energy per unit mass parameter[5] less than the maximum of the potential barrier, are reflected. Such tachyons re-emerge out of the event horizon at $r=2m$ of the extended Schwarzschild space—time, while the high energy ones fall into the singularity. It is quite intriguing to observe that, these faster than light particles rob the $r=2m$ surface of its property as a one way membrane of information propagation. Also the gravitational field of the central mass tends to oppose the infall of the tachyon rather than abet it, while the angular momentum tends to lower this potential barrier. These effects are opposite to those expected for tardiyons.

The above papers deal with the problem classically, that is, the tachyons are classical particles having definite position and momentum at a given point of space—time. In the present paper we propose to analyze the problem within a semiclassical frame work, in the sense that, the tachyons are treated as quantum wavepackets, but the gravitational field of the black hole is unquantized. This type of approach has already yielded interesting properties of quantum effects on ordinary matter near black holes as shown by the pioneering work of Hawking.[6] It would be interesting to find out whether the quantum tachyons encounter a potential barrier and if so, whether they tunnel through it. To study the tunnelling effect we shall extensively make use of the WKB method.

2. KLEIN-GORDON EQUATION FOR TACHYONS

One of the basic assumptions we make here is that the Schwarzschild geometry is not perturbed significantly by the tachyon. The assumption is not unjustified, if the total tachyon mass parameter is negligible compared to the mass of the black hole, responsible for the background Schwarzschild geometry. Although tachyons still remain to be detected in a "typical" case we will assume the mass parameter of the tachyon comparable to the mass of an electron. The mass of the black hole will be taken to be order of the mass of the sun $M_\odot$.

It is possible to determine the spin of a tachyon from the considerations of the unitary irreducible representations of the inhomogeneous Lorentz group. These representations are either spinless or infinite dimensional. Following Narlikar and Sudarshan,[1] we treat here the simplest case of the spinless tachyon. The tachyon wavefunction in empty space then satisfies the Klein—Gordon equation,

$$(\Box^2 - M_0^2)\,\psi = 0, \tag{1}$$

in which, we have chosen units with $c=1$, $\hbar=1$. $\psi(R,\theta,\varphi,t)$ is the scalar wavefunction of the tachyon, M_0 is the mass parameter of the tachyon, the operator

$$\Box^2 = \frac{1}{\sqrt{-g}}\,\frac{\partial}{\partial x^j}\left(\sqrt{-g}\,g^{ij}\,\frac{\partial}{\partial x^i}\right),$$

where g_{ij} is the metric tensor of Schwarzschild space—time with the following nonzero components:

$$g_{00} = \left(1 - \frac{2MG}{R}\right), \quad g_{11} = -\left(1 - \frac{2MG}{R}\right)^{-1},$$

$$g_{22} = -R^2, \quad g_{33} = -R^2\sin^2\theta.$$

M is the mass of the black hole and G is the gravitational constant.

The negative sign in Eq. (1) arises because the tachyon has imaginary rest mass iM_0 and this can be immediately seen to be consistent with the fact that the wavepackets of the solutions to this equation have group velocities greater than that of light. The plane wave solutions to (1) are of the form $\exp(i\bar{k}\cdot\bar{r} - i\omega t)$ with $\bar{k}$ and ω satisfying the relation $k^2 - \omega^2 = M_0^2$. Therefore, the group velocity $d\omega/dk = k/\omega$ is greater than unity.

The curvature of space—time is incorporated in the $\Box^2$ operator and this gives rise to an effective potential which the tachyonic waves encounter.

One may remark here that, when the sign of M_0^2 in (1) is positive, that is, in the case of ordinary particles, there do not exist physically acceptable static solutions

 0022-2488/78/1903-0561$1.00

of the Klein–Gordon equation in Schwarzschild space–time, but in our case such solutions do exist (Price[7]). They correspond to zero energy tachyons.

Expansion of (1) leads to the equation

$$\frac{\partial^2\psi}{\partial t^2} - \frac{1}{R^2}\left(1-\frac{2MG}{R}\right)\frac{\partial}{\partial R}\left[R^2\left(1-\frac{2MG}{R}\right)\frac{\partial\psi}{\partial R}\right] - \frac{1}{R^2}\left(1-\frac{2MG}{R}\right)\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) - \left(1-\frac{2MG}{R}\right)M_0^2\psi = 0, \tag{2}$$

where due to azimuthal symmetry of the problem about $\theta=0$, the wavefunction is independent of φ.

ψ can be expanded in terms of a complete set of eigenfunctions of the angular momentum operator, that is, solutions of the form $[\psi_l(R)/R]P_l(\cos\theta)$. Setting $\psi=\sum_{l=0}^{\infty}[\psi_l(R)/R]P_l(\cos\theta)$ and substituting in (2), we get a partial differential equation for $\psi_l(R,t)$, namely,

$$\frac{\partial^2\psi}{\partial t^2} - \frac{1}{R}\left(1-\frac{2MG}{R}\right)\frac{\partial}{\partial R}\left[R^2\left(1-\frac{2MG}{R}\right)\frac{\partial}{\partial R}\left(\frac{\psi_l}{R}\right)\right] + \left(1-\frac{2MG}{R}\right)\frac{l(l+1)}{R^2}\psi_l - \left(1-\frac{2MG}{R}\right)M_0^2\psi_l = 0. \tag{3}$$

To separate out the t dependence of $\psi_l(R,t)$ and since the calculations are going to be linear throughout, we can proceed by Fourier analysis.

Writing

$$\psi_l(R,t)=\int_{-\infty}^{\infty}A(\Omega)\,\psi_l^{\Omega}(R)\exp(-i\Omega t)\,d\Omega$$

and substituting in (3) gives an ordinary differential equation for a typical Fourier component,

$$\frac{d^2\psi_l^{\Omega}}{dR^2} + \frac{2MG}{R(R-2MG)}\frac{d\psi_l^{\Omega}}{dR} + \frac{R^2}{(R-2MG)^2}\left[\Omega^2 + \left(1-\frac{2MG}{R}\right)M_0^2 - \frac{2MG}{R^3}\left(1-\frac{2MG}{R}\right) - \frac{l(l+1)}{R^2}\left(1-\frac{2MG}{R}\right)\right]\psi_l^{\Omega}=0. \tag{4}$$

In Eq. (4) the coefficient of ψ_l^{Ω} consists of two parts:

(i) the term involving Ω^2 corresponds to the energy or frequency of the partial wave;

(ii) the other three terms constitute the effective potential barrier, which the Fourier component ψ_l^{Ω} has to surmount; the first two of which arise due to the curvature of space–time, while the last term represents the "centrifugal barrier."

It is fruitful at this juncture to use dimensionless units, by setting $r=RM_0$, $m=GMM_0$, $\omega=\Omega/M_0$. Equation (4) then takes the form

$$\frac{d^2\psi_l^{\Omega}}{dr^2} + \frac{2m}{r(r-2m)}\frac{d\psi_l^{\Omega}}{dr} + \frac{r^2}{(r-2m)^2}\left(\omega^2 + \frac{r-2m}{r} - \frac{2m(r-2m)}{r^4} - \frac{l(l+1)}{r^3}(r-2m)\right)\psi_l^{\Omega}=0. \tag{5}$$

Equation (5) cannot be solved in terms of any known functions, but its application to the typical situation cited before furnishes a simpler equation. For effecting the approximation, it is necessary to get an idea of the magnitude of the various terms appearing in Eq. (5). If the mass of the black hole $M\sim M_{\odot}=2\times10^{33}$ gm where M is the mass of the sun, and if the mass parameter of the tachyon $M_0\sim m_e=9\times10^{-28}$ gm is the mass of the electron, then

$$m=\frac{GMM_0}{c\hbar}\ 3.84\times10^{15},$$

$$r=\frac{RM_0c}{\hbar}=(R\text{ in cm})\ (2.56\times10^{10}\text{cm}^{-1}).$$

One immediately sees that the coefficient of $d\psi_l^{\Omega}/dr$ is very small everywhere, except inside the extremely thin shell $|r-2m|<\epsilon$ where ϵ is of the order of unity in dimensionless units. Also the term $2m(r-2m)/r^4$ is negligible everywhere outside a small sphere centered at the singularity $r=0$. We can neglect these terms provided we do not enter these exceptional regions. Then Eq. (5) simplifies to the form

$$\frac{d^2\psi_l^{\Omega}}{dr^2} + \frac{r^2}{(r-2m)^2}\left(\omega^2+1-\frac{2m}{r}-\frac{l(l+1)(r-2m)}{r^3}\right)\psi_l^{\Omega}=0. \tag{6}$$

3. THE RADIALLY INFALLING TACHYON

This case does not just deal with the $l=0$ partial wave, but applies also to eigenfunctions of higher angular momentum states, provided the centrifugal term $l(l+1)(r-2m)/r^3$ for those eigenfunctions is small compared to the other terms in the parenthesis. To get an idea of the value of l for which this term is negligible, we look at it in more detail. This term is of the order of unity, when l is roughly of the order of m, that is, $l\sim10^{15}$, say. One need not be surprised by this large value of l, because the tachyon is at an astronomical distance from $r=0$ and consequently its angular momentum is several magnitudes higher than the angular momenta encountered in the atomic encounters. Here $l\lesssim10^{9}$ is sufficient to ensure the smallness of the term in the region, where (6) is valid. One neglects this term and obtains the differential equation

$$\frac{d^2\phi}{dr^2} + \frac{r^2}{(r-2m)^2}\left(\omega^2+1-\frac{2m}{r}\right)\phi=0, \tag{7}$$

where ϕ is one of the eigenfunctions which fall into this category.

At large distances from the black hole, i.e., for $r\gg2m$, Eq. (7) becomes

$$\frac{d^2\phi}{dr^2}+k^2\phi=0, \tag{8}$$

where $k^2=1+\omega^2$ and k is the wavenumber of the wave at infinity. The solutions of (8) are of the form $\exp(\pm ikr)$.

A. A crude approximation

In the vicinity of the black hole, however, the curvature effects come into picture. One can then obtain a solution of (7) in the neighborhood of a given point and determine the form of the solution, that is, whether the solution is oscillatory or damped.

Setting $r = r_0 + \epsilon = 2m\rho + \epsilon$, say, where $\rho = O(1)$, one obtains

$$\frac{d^2\phi}{d\epsilon^2} + \frac{\rho^2}{(\rho-1)^2}\left(k^2 - \frac{1}{\rho}\right)\phi = 0. \tag{9}$$

From the definition of k^2, $k^2 > 1$ and so in the region $r > 2m$, corresponding to $\rho > 1$, the solution is always oscillatory. Inside the Schwarzschild radius the coefficient of ϕ is positive or negative according as $\rho > 1/k^2$ or $\rho < 1/k^2$. Accordingly the solution is oscillatory if $r > r_1$ and damped if $r < r_1$ where $r_1 = 2m/k^2 = 2m\rho_1$ (say).

One may remark here that a similar result was first obtained by Raychaudhuri[8] by purely classical considerations. $2m/k^2$ was the point of reflection of the classical tachyon, the k in this result represented the momentum per unit mass parameter of the tachyon. Whereas quantum mechanically there are limitations to this result. For, if k becomes arbitrarily large, r_1 decreases without bound and falls inside the region, where Eq. (7) is invalid, as the term $2m(r - 2m/r^4$ can no longer be neglected from Eq. (5) to obtain (6) or (7). In order that r_1 lies outside the small sphere (the radius of which is about 10^7 in dimensionless units) centered at the origin, k should be less than 10^4. Then the result of the reflection of the tachyon at $2m/k^2$ is also valid in the quantum mechanical case.

B. Application of the WKB method

The situation in the vicinity of $r = r_1$ therefore deserves a more accurate analysis than given above. A solution accurate enough for the present purpose can be obtained by the Wentzel—Kramers—Brillouin (WKB) method described below. Setting

$$p^2(r) = \frac{r}{(2m-r)^2}\left(k^2 - \frac{2m}{r}\right), \tag{10}$$

we have

$$\frac{d^2\phi}{dr^2} + p^2(r)\,\phi = 0. \tag{11}$$

The solution of (11) can be approximated by

$$\phi(r) = \alpha(r)\exp(\pm i \textstyle\int p\,dr), \tag{12}$$

under the assumption that the effective potential is linear near r_1 and $p'(r)/p^2(r) \ll 1$, where the prime denotes differentiation with respect to r. We shall first apply the method to (7) to get the required solution and then discuss the constraints on its validity.

A direct substitution of $\phi(r)$ in (11) leads to the determination of $\alpha(r)$ to be

$$\alpha(r) = Ap^{-1/2}(r), \tag{13}$$

where A is an arbitrary constant. Setting $r = 2m\rho_1 + \xi$ in (11) and retaining only the first order terms in ξ we get

$$\frac{d^2\phi}{d\xi^2} + \frac{\xi}{2m(1-\rho_1)^2}\phi = 0. \tag{14}$$

Putting $\zeta = -\xi/[2m(1-\rho_1)^2]^{1/3}$ in (14), we get

$$\frac{d^2\phi}{d\zeta^2} - \zeta\phi = 0. \tag{15}$$

This is the well-known Airy equation and has the solutions $Ai(\zeta)$ and $Bi(\zeta)$. For $\zeta \gg 1$,

$$Ai(\zeta) \sim \frac{1}{2\sqrt{\pi}}\zeta^{-1/4}\exp(-2\zeta^{3/2}/3),$$

$$Bi(\zeta) \sim \frac{1}{\sqrt{\pi}}\zeta^{-1/4}\exp(2\zeta^{3/2}/3).$$

In the limit $\zeta \ll -1$,

$$Ai(\zeta) \sim \frac{1}{\sqrt{\pi}}(-\zeta)^{-1/4}\sin(\tfrac{2}{3}(-\zeta)^{3/2} + \pi/4),$$

$$Bi(\zeta) \sim \frac{1}{\sqrt{\pi}}(-\zeta)^{-1/4}\cos(\tfrac{2}{3}(-\zeta)^{3/2} + \pi/4).$$

In the solution by the use of Airy functions two scales of distances are involved, namely, the astronomical scale [$O(10^{15})$ dimensionless units] and the atomic scale [$O(1)$ dimensionless units]. The effective potential varies over distances comparable to the astronomical scale, while the Compton wavelength of the tachyon measures with atomic length scale. A unit length of parameter ζ, used in Eq. (15), can incorporate several oscillations of the tachyon wavefunction, but is small enough for the change in potential to occur approximately linearly. Equation (14) is obtained by an approximation carried out in the astronomical scale, while the asymptotic formulas to its solutions are obtained pertaining to the atomic scale.

For the application of asymptotic formulas the lower bound for r_1 should be about 10^9, greater than the one required for (6) to be valid. This further restricts the permitted range of k to less than 10^3.

The extent of the region $r < r_1$ is large enough to let the solution $Bi(\zeta)$ grow extremely large, hence it is physically unacceptable. Therefore, we choose $\phi = Ai(\zeta)$.

The WKB method is applied to the given problem by writing (14) in the form

$$\frac{d^2\phi}{dr^2} + p^2(r)\,\phi = 0,$$

where

$$p^2(r) = \frac{r - r_1}{2m(1-\rho_1)^2} = \alpha^3(r - r_1) \quad \text{(say)}.$$

We have $\zeta = -\alpha(r - r_1)$, which yields

$$\int_{r_1}^{r} p(r)\,dr = \tfrac{2}{3}(-\zeta)^{3/2} \quad \text{in the region } r > r_1.$$

In view of the asymptotic form of $Ai(\zeta)$ for $\zeta \ll -1$, the solution in the region $r > r_1$ can be written as

$$\phi_1 = Ap^{-1/2}\sin[\textstyle\int_{r_1}^{r} p(r)\,dr + \pi/4]. \tag{16}$$

The Airy function provides the continuation of ϕ_1, into the region $r < r_1$ corresponding to $\zeta > 0$,

$$\phi_2 = B|p|^{-1/2}\exp(-\textstyle\int_{r}^{r_1}|p(r)|\,dr, \tag{17}$$

where B is a constant.

We immediately see from the asymptotic forms of the Airy functions, for $\zeta \gg 1$ and $\zeta \ll -1$, that $A = 2B$, and conclude that a total reflection of the wave occurs at $r=r_1$. The wave which penetrates the potential barrier is heavily damped.

C. Test of validity of the WKB method

It is required to test whether $p'(r)/p^2(r) \ll 1$ in the vicinity of $r=r_1$. We have

$$\frac{p'(r)}{p^2(r)}=\frac{1}{2m}\,\frac{p'(\rho)}{p^2(\rho)}. \qquad (18)$$

A straightforward calculation shows that

$$\frac{p'(\rho)}{p^2(\rho)}=\frac{1-\rho}{\rho^2(1/\rho-k^2)^{3/2}}\left\{\frac{1}{1-\rho}\left(\frac{1}{\rho}-k^2\right)-\frac{1}{2\rho}\right\}. \qquad (19)$$

If k is not very large—say of the order of unity—$p'(\rho)/p^2(\rho)$ is of the same order, so that $p'(r)/p^2(r)$ is very small. Therefore, WKB is valid in this case.

For $k\gg 1$, $\rho_1=1/k^2\ll 1$ and in the limit the simplified expression $p'(\rho)/p^2(\rho)$ becomes

$$\frac{p'(\rho)}{p^2(\rho)}=\frac{1}{\rho^2(1/\rho-k^2)^{3/2}}\left(\frac{1}{2\rho}-k^2\right). \qquad (20)$$

Setting $\rho=a/k^2$ where $a=O(1)$, we get $p'(\rho)/p^2(\rho)=bk^3$ where $b=O(1)$ which yields,

$$\frac{p'(r)}{p^2(r)}\sim\frac{k^3}{m}. \qquad (21)$$

With regard to the permitted range of k discussed before, the ratio $p'(r)/p^2(r)$ is very small and WKB method is valid.

D. Exact solution

It is possible to solve Eq. (7) exactly in terms of the confluent hypergeometric function. We consider the equation for $r<2m$; the case for $r>2m$ is on similar lines.

Putting $x=2m-r$ in (7) gives

$$\frac{d^2\phi}{dx^2}+\left(k^2-\frac{2m\alpha}{x}-\frac{n(n+1)}{x^2}\right)\phi=0, \qquad (22)$$

where $\alpha=2k^2-1$ and $n(n+1)=-4m^2\omega^2$. Let $\phi(x)=x^{n+1}\exp(ikx)f(x)$, and putting $y=-2ikx$ we get an equation for f,

$$y\frac{d^2f}{dy^2}+[2(n+1)-y]\frac{df}{dy}-(n+1+i\alpha m/k)f=0. \qquad (23)$$

The solutions of (23) are $F(n+1+i\alpha m/k,\ 2n+2,y)$ and $F(-n+i\alpha m/k,\ -2n,y)y^{-2n-1}$. The solutions of (22) are correspondingly multiplied by $x^{n+1}\exp(ikx)$. The appropriate combinations of these solutions describe the incoming and the outgoing waves.

Though the above solution is exact, unfortunately it provides little physical insight into the problem. It seems difficult to extract relevant asymptotic information from it.

E. High energy tachyonic waves

Very high energy partial waves ($\omega\sim 10^5$) penetrate to the region very near singularity $r\lesssim 10^7$ (say), so the term $2m(r-2m)/r^4$ can no longer be neglected from Eq. (5). But another approximation in (5) is possible for distances very close to $r=0$, that is, r can be neglected compared to $2m$. With this approximation, (5) takes the form

$$\frac{d^2\phi}{dr^2}-\frac{1}{r}\frac{d\phi}{dr}+\frac{r^2}{4m^2}\left(k^2-\frac{2m}{r}+\frac{4m^2}{r^4}\right)\phi=0, \qquad (24)$$

where ϕ is one of the "radial" eigenfunctions. Setting $g=r^{-1/2}\phi(r)$, (24) becomes

$$\frac{d^2g}{dr^2}+\frac{r^2}{4m^2}(k^2-V^2(r))g=0, \qquad (25)$$

where

$$V^2(r)=\frac{2m}{r}-\frac{m^2}{r^4}.$$

We have

$$V^2_{\max}=\frac{3}{2}\left(\frac{m^2}{2}\right)^{1/3}. \qquad (26)$$

Therefore, if $k>\sqrt{3}/2^{2/3}m^{1/3}$, then the partial wave with such high energy surmounts the effective potential barrier and goes into the singularity. This result is markedly different from the classical one, where reflection takes place for all energies of the radially infalling tachyon.

4. TACHYON WITH ORBITAL ANGULAR MOMENTUM

In this case we require Eq. (6) in full. The value of l considered here is about 10^{15}, so that the centrifugal term $l(l+1)(r-2m)/r^3$ in (6) is comparable to the other terms in the parentheses.

We shall now discuss the general behavior of the wavefunction as the wave approaches the black hole from infinity. For very large distances from the singularity, the very large values of l come into the picture. If $r\sim 100m$ and $l\sim 10^{17}$, Eq. (6) reduces to

$$\frac{d^2\psi_l^\Omega}{dr^2}+\left(\omega^2+1-\frac{l(l+1)}{r^2}\right)\psi_l^\Omega=0. \qquad (27)$$

One may write (27) as

$$\frac{d^2\psi_l^\Omega}{dr^2}+p^2(r)\psi_l^\Omega=0 \qquad (28)$$

with

$$p^2(r)=\omega^2+1-\frac{l(l+1)}{r^2}.$$

For a fixed ω and l, and at a large distance from $r=0$, when the value of r is sufficiently large, $p^2(r)>0$ and the Fourier component ψ_l^Ω is oscillatory. As the wave approaches $r=0$, $p^2(r)$ decreases, so that the oscillations slow down until finally $p^2(r)$ becomes negative, and the partial wave is damped out. It is also easy to see that the partial waves with larger angular momenta die out at a larger distance from the black hole. So if one looks at the totality of all the partial waves of an incident wave with a fixed ω, the partial waves with larger angular momenta are damped out first, while the rest of the partial waves proceed towards the black hole.

For comparatively smaller distances one can include comparatively smaller values of l; to be more precise, if $r\sim 20m$ and $l\sim 2m$ it is possible to treat a special case for low frequency waves. With this approximation equation (6) becomes

$$\frac{d^2\psi_l^\Omega}{dr^2} + \left(\omega^2 + 1 - \frac{2m}{r} - \frac{l(l+1)}{r^2}\right)\psi_l^\Omega = 0. \tag{29}$$

In the low energy limit, we see that the solution is oscillatory if $r > r_l$, where $r_l = m + [m^2 + l(l+1)]^{1/2}$. A parallel result was obtained by Narlikar and Dhurandhar[4] for the classical tachyon.

When r is of the order of the Schwarzschild radius Eq. (6) has to be treated in its full form and can be written as

$$\frac{d^2\psi_l^\Omega}{dr^2} + \frac{r^2}{(r-2m)^2}(\omega^2 - V_l^2(r))\psi_l^\Omega = 0, \tag{30}$$

where

$$V_l^2(r) = \left(1 - \frac{2m}{r}\right)\left(\frac{l(l+1)}{r^2} - 1\right)$$

is the effective potential.

This potential is impenetrable for low frequency waves, but is transparent to those of high frequency.

It is convenient, at this juncture, to use a length variable of the astronomical scale. Putting $r = 2m\rho$ in (30) but retaining the derivative in terms of r, we have

$$\frac{d^2\psi_l^\Omega}{dr^2} + \frac{\rho^2}{(1-\rho)^2}(\omega^2 - V_l^2(\rho))\psi_l^\Omega = 0, \tag{31}$$

with

$$V_l^2(\rho) = \left(\frac{1}{\rho} - 1\right)\left(1 - \frac{L^2}{\rho^2}\right),$$

where

$$L^2 = \frac{l(l+1)}{4m^2}.$$

The potential $V_l^2(\rho)$ is of the same form, as obtained in the classical case by Honig *et al.*,[3] who discussed its properties in full detail. Hence we shall be brief.

A. Comparison with classical case

The roots of $V_l^2(\rho) = 0$ are at $\rho = \pm L$ and 1, and a maximum of the potential occurs at $\rho = \rho_m$ lying between $\rho = 1$ and $\rho = L$. There is no minimum of potential. We have $\rho_m = (L^4 + 3L^2)^{1/2} - L^2$ and $\lim_{L\to\infty}\rho_m = \frac{3}{2}$. For $L = 1$, $V_l^2(\rho) \le 0$ for all ρ. Hence the partial waves with $L \cong 1$ do not see any potential barrier and proceed into the singularity undamped.

There is, however, one difference which occurs in our treatment as contrasted with the purely classical treatment. In the classical case L can vary continuously from 0 to ∞: while this case L is a discrete variable, it takes only those values for which l is a nonnegative integer. Therefore, one gets a discrete series of potential curves for the various values of l.

If $\omega^2 > V_l^2(\rho_m)$, the partial wave does not encounter any potential barrier, but when $\omega^2 < V_l^2(\rho_m)$ an interesting situation arises. The equation $\omega^2 - V_l^2(\rho) = 0$ has two positive real roots, say ρ_{l_1} and ρ_{l_2} with $\rho_{l_1} > \rho_{l_2}$. Hence one can anticipate qualitatively the general behavior of the solution as follows:

(i) For $\rho > \rho_{l_1}, \rho_{l_2}$, the solution is going to be oscillatory.

(ii) For $\rho_{l_1} < \rho < \rho_{l_2}$, the solution is going to be damped.

(iii) For $\rho < \rho_{l_1}, \rho_{l_2}$, the solution is again going to be oscillatory.

B. Application of the WKB method

We shall again apply the WKB method and later discuss its validity.

The outline of the method will be as follows:

(a) We divide the ρ axis into three regions:

(1) $\rho > \rho_{l_1}$,

(2) $\rho_{l_2} < \rho < \rho_{l_1}$,

(3) $\rho < \rho_{l_2}$.

(b) We obtain approximate forms of the wavefunction in the regions (1), (2), and (3) and match them at the barriers with the aid of Airy functions.

Our next step is to find the form of the solution near the potential barriers at ρ_{l_1} and ρ_{l_2}. To this end, we set $r = 2m\rho_l + \xi$, where ρ_l represents either ρ_{l_1}, or ρ_{l_2}. Then a Taylor expansion of the potential near $\rho = \rho_l$ leads to the differential equation

$$\frac{d^2\psi_l^\Omega}{d\xi^2} + \frac{\xi}{2m(1-\rho_l)^2}\left(1 + \frac{2L^2}{\rho_l} - \frac{3L^2}{\rho_l^2}\right)\psi_l^\Omega = 0. \tag{32}$$

Putting

$$\beta^3 = \frac{1}{2m(1-\rho_l)^2}\left(1 + \frac{2L^2}{\rho_l} - \frac{3L^2}{\rho_l^2}\right) \quad \text{and} \quad \zeta = \beta\xi,$$

we get the Airy equation,

$$\frac{d^2\psi_l^\Omega}{d\zeta^2} - \zeta\psi_l^\Omega = 0. \tag{33}$$

Since there seems to be a possibility of the tachyon tunneling across the potential barrier, one has to take into consideration both the solutions $Ai(\zeta)$ and $Bi(\zeta)$ of the Airy equation. As was the situation in the radial case, the parameter ζ has a length scale lying between the astronomical and the atomic scale. So in the vicinity of the potential barrier the Airy equation, together with the asymptotic forms of its solutions, is valid.

What we are going to do henceforth has been schematically represented in Fig. 1. We assume the wavefunction in region (3) to be

$$\phi_3 = Ap^{-1/2}\exp\left(-i\int_r^{r_2} p(r)\,dr - i\pi/4\right), \quad r < r_2, \tag{34}$$

where $p^2(r)$ is the coefficient of ψ_l^Ω in (30) and we have set $r_1 = 2m\rho_{l_1}$ and $r_2 = 2m\rho_{l_2}$. Since we are considering a single angular momentum state at a time, no ambiguity arises if we drop the suffix l from the variable r. An extra phase factor of $\exp(-i\pi/4)$ has been added to facilitate the matching of solutions across the barriers. (34) represents a wave moving radially away from the singularity $r = 0$.

In the vicinity of r_2, ϕ_3 has the approximation

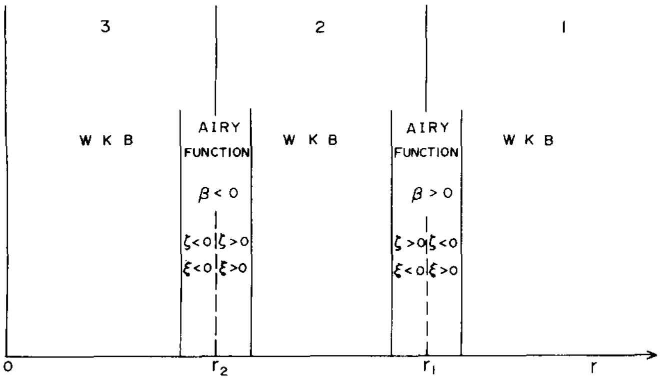


FIG. 1. Figure showing the region of validity of the WKB approximation and the Airy function approximation of the wave function $\psi_l^\Omega(r)$. The WKB approximation is valid everywhere except in the neighborhoods of r_1 and r_2, the width of the neighborhoods being given by $|\zeta| \lesssim O(10)$.

$$\phi_3 \sim A(-\zeta)^{-1/4} \exp[-i(\tfrac{2}{3}(-\zeta)^{3/2} + \pi/4)]. \tag{35}$$

The Airy functions provide the required connections across the barrier. (35) shows that ϕ_3 represents a combination of the asymptotic forms of the Airy functions (except for the constant factor $\sqrt{\pi}$) for $\zeta \ll 1$. Since the Airy functions are well behaved at $\zeta = 0$, we can continue the solutions into region (2) and obtain its asymptotic form for $\zeta \gg 1$,

$$\phi_2 \sim A(\zeta)^{1/4}[\exp(2\zeta^{3/2}/3) - \tfrac{1}{2}i \exp(-2\zeta^{3/2}/3)]. \tag{36}$$

In view of this, in region (2) the wavefunction can be written as

$$\phi_2(r) = A\,|p|^{-1/2}[\exp(\textstyle\int_{r_2}^{r} |p|\,dr) - \tfrac{1}{2}i[\exp(-\textstyle\int_{r_2}^{r} |p|\,dr)] \tag{37}$$

for $r > r_2$. Define $T = \exp(-\int_{r_2}^{r_1} |p|\,dr)$. Then (37) takes the form

$$\phi_2(r) = A\,|p|^{-1/2}\left(\frac{1}{T}\exp\left(-\int_r^{r_1} |p|\,dr\right) - \frac{iT}{2}\exp\left(\int_r^{r_1} |p|\,dr\right)\right). \tag{38}$$

By going through a similar procedure as followed above, it is possible to obtain the wavefunction in region (1) as

$$\phi_1(r) = Ap^{-1/2}\left[\frac{2}{T}\sin\left(\int_{r_1}^{r} p\,dr + \frac{\pi}{4}\right) - \frac{iT}{2}\cos\left(\int_{r_1}^{r} p\,dr + \frac{\pi}{4}\right)\right]. \tag{39}$$

The solution is more revealing in the form

$$\phi_1(r) = \frac{Ap^{-1/2}}{i}\left\{\left(\frac{1}{T} + \frac{T}{4}\right)\exp\left[i\left(\int_{r_1}^{r} p\,dr + \frac{\pi}{4}\right)\right] - \left(\frac{1}{T} - \frac{T}{4}\right)\exp\left[-i\left(\int_{r_1}^{r} p\,dr + \frac{\pi}{4}\right)\right]\right\}. \tag{40}$$

The first term of (40) represents the reflected wave and the second one represents the incident wave. The transmitted wave is given by $\phi_3(r)$ in Eq. (34).

C. Transmission and reflection coefficients

The transmission coefficient is computed by the formula

transmission coefficient

$$= \frac{|\text{amplitude of transmitted wave}|}{|\text{amplitude of incident wave}|} = \frac{T}{1 - T^2/4} \cong T \text{ if } T \ll 1.$$

In the same approximation the reflection coefficient is unity.

The computation of T requires the evaluation of the integral in its definition,

$$\int_{r_2}^{r_1} |p|\,dr = 2m \int_{\rho_{l_2}}^{\rho_{l_1}} \frac{\rho}{1-\rho}\,(V_l^2(\rho) - \omega^2)^{1/2}\,d\rho \tag{41}$$

unless $\rho_{l_1} \cong \rho_{l_2}$, the integral is large in the typical situation, because of the multiplying factor m. Therefore, T is practically zero implying that total reflection occurs at the potential barrier.

The case when $\rho_{l_1} \cong \rho_{l_2}$ does not carry much physical significance, as it would be a rare coincidence for the energy of the tachyonic wave to lie within such a small range, but for the sake of completeness, it has been treated in the Appendix.

D. Test of validity of the WKB approximation

The validity of the above procedure can be seen by computing $p'(r)/p^2(r)$. We have

$$\frac{p'(r)}{p^2(r)} = \frac{1}{2m}\,\frac{p'(\rho)}{p^2(\rho)} = \frac{1-\rho}{2m\rho^2\,|V_l^2(\rho) - \omega^2|^{1/2}}\left(\frac{1}{1-\rho} - \frac{\rho^2 + 2L^2\rho - 3L^2}{2\rho^3(V_l^2(\rho) - \omega^2)}\right). \tag{42}$$

The factor m occurring in the denominator of (42) implies $p'(r)/p^2(r) \ll 1$, except when $V_l^2(\rho) - \omega^2$ is small. This is so if $r \cong r_1$ or r_2 or if two roots of the equation $V_l^2(\rho) - \omega^2 = 0$ lie close together.

5. CONCLUDING REMARKS

The above quantum mechanical analysis shows up somewhat different results compared to the earlier discussions of classical tachyons.[3,4] In the classical case, the identification of the regions I and III in the Kruskal diagram led to the result that a radially infalling tachyon is bounced by the black hole and emerges from inside the event horizon (Narlikar and Dhurandhar[4]). Here the quantum tachyon shows a similar behavior at low energies. At high energies, however, it tunnels through the potential barrier and hits the singularity.

This curious behavior has some resemblance to the quantum effects on ordinary matter (tardiyons) near black holes. As shown by Hawking[5] the quantum mechanical tunnel effect allows the outward movement of ordinary matter near black holes. The results presented here show a similar effect operating on tachyons moving inwards.

ACKNOWLEDGMENT

I am grateful for the guidance of Professor J. V. Narlikar.

APPENDIX

We treat here the case when the two roots of $V_l^2(\rho) - \omega^2 = 0$ lie close together. Near the maximum of the potential, the potential curve can be approximated by a small parabolic arc. Hence one can write the two roots of $V_l^2(\rho) - \omega^2 = 0$ as $\rho_m + \epsilon$ and $\rho_m - \epsilon$. Then a Taylor expansion of the potential near the point $\rho = \rho_m$ yields

$$V_l^2(\rho_m) - V_l^2(\rho_m \pm \epsilon) = \epsilon^2 A^2, \tag{A1}$$

where

$$A^2 = -\frac{1}{2} \left. \frac{d^2 V_l^2}{d\rho^2} \right|_{\rho=\rho_m} = \frac{L^2}{\rho_m^5}(3 - \rho_m). \tag{A2}$$

We note that, as $\rho_m < \frac{3}{2}$ for all L, $A^2 > 0$.

Our interest is to investigate whether there is a finite probability for the tachyon to tunnel through the potential barrier and also to simultaneously determine how far in the limit as $\epsilon \to 0$ can we apply the WKB method to achieve our end. The validity of WKB is determined from the computation of $p'(r)/p^2(r)$ at an intermediate point $\rho = \rho_m + \delta$ with $|\delta| < \epsilon$. Taking δ as the same order as ϵ and setting $\delta = a\epsilon$ with $a \sim O(1)$ and $|a| < 1$, yields

$$\frac{p'(\rho_m + \delta)}{p^2(\rho_m + \delta)} \cong \frac{1}{A\rho^2(1 - a^2)^{1/2}} \times \frac{1}{\epsilon} - \frac{a}{(1 - a^2)^{3/2}} \frac{1 - \rho_m}{A\rho_m} \times \frac{1}{\epsilon^2}. \tag{A3}$$

Therefore, the WKB method can be applied when $m\epsilon^2 \gg 1$ or, i.e., when $\epsilon \gg m^{-1/2}$.

One next computes the integral appearing in the definition of T,

$$\int_{r_2}^{r_1} |p| dr = 2m \int_{\rho_m - \epsilon}^{\rho_m + \epsilon} \frac{\rho}{1 - \rho} (V_l^2(\rho) - \omega^2)^{1/2} d\rho. \tag{A4}$$

From (A1) the value of the integral turns out to be $2A\rho_m/(1 - \rho_m) \times m\epsilon^2$. Hence the criterion required for the validity of the WKB method, namely $m\epsilon^2 \gg 1$, is responsible in making T negligible. So when the roots of $V_l^2(\rho) - \omega^2 = 0$ are close, but not so close that the relation $\epsilon \gg m^{-1/2}$ fails, total reflection of the partial wave occurs.

In the event when ϵ is less than the critical value for the WKB approximation to hold, the differential equation (30) itself may be approximated. At the point $x = r - r_m$, where $r_m = 2m\rho_m$ and $r \ll r_m$, Eq. (30) takes the form

$$\frac{d^2\psi_l^\Omega}{dx^2} - \frac{r_m^2 A^2}{(2m - r_m)^2} \left(\epsilon^2 - \frac{x^2}{4m^2}\right) \psi_l^\Omega = 0, \tag{A5}$$

where $|x| < 2m\epsilon$ and A^2 is given by (A2). Effecting the transformation,

$$y = x\left(\frac{Ar_m}{2m(2m - r_m)}\right)^{1/2} \text{ and } \phi = \psi_l^\Omega \times \frac{Ar_m}{2m(2m - r_m)}.$$

(A5) becomes

$$\frac{d^2\phi}{dy^2} + y^2\phi = B^2\phi, \tag{A6}$$

where

$$B^2 = \frac{2mAr_m\epsilon^2}{|2m - r_m|}.$$

The solution of (A6) can be obtained by the contour integral method of the generalized Laplace transform type.

Set

$$\phi(y) = \int_C \exp(ty^2) f(t)\, dt, \tag{A7}$$

where after substitution into (48) and solving, $f(t)$ turns out to be

$$f(t) = \text{const} \times \frac{\exp(-\frac{1}{2}B^2 \tan^{-1} 2t)}{(4t^2 + 1)^{3/4}} \tag{A8}$$

and the contour C to be chosen so that

$$\Delta_C\{\exp(ty^2) f(t)(4t^2 + 1)\} = 0. \tag{A9}$$

[1] J. V. Narlikar and E. C. G. Sudarshan, Mon. Not. R. Astron. Soc. 175, 105–16 (1976).
[2] S. V. Dhurandhar, Pramana 8, 133 (1977).
[3] E. Honig, K. Lake, and R. C. Roeder, Phys. Rev. D 10, 3155 (1974).
[4] J. V. Narlikar and S. V. Dhurandhar, Pramana 6, 388 (1976).
[5] The mass parameter M_0 of the tachyon corresponds to its imaginary rest mass iM_0.
[6] S. W. Hawking, Commun. Math. Phys. 43, 199 (1975).
[7] R. H. Price, Phys. Rev. D 5, 2419 (1972).
[8] A. K. Raychaudhuri, J. Math. Phys. 15, 856 (1974).

Slavnov–'t Hooft identities in Mandelstam's formalism

Seichi Naito

Department of Physics, Osaka City University, Sumiyoshiku, Osaka, Japan
(Received 28 January 1977)

In Mandelstam's gauge-independent quantization formalism, Slavnov identities are shown to originate from the fact that auxiliary Green's functions in various gauges (Feynman gauge, Landau gauge, etc.) satisfy the same fundamental equation. Furthermore, we have the practically important result that the infinite number of Slavnov identities are obtained in concise and concrete form. With the help of this form, we are able to easily derive various 't Hooft identities, which assure us of the gauge independence and the unitarity of the S matrix.

I. INTRODUCTION AND PRELIMINARIES

In 1968 Mandelstam[1] quantized the Yang–Mills field in his gauge-independent path-dependent formalism. Starting from the equal time commutation relations among gauge-independent path-dependent quantities, he investigated the auxiliary Green's functions and obtained the same Feynman rules for the Yang–Mills field as prescribed by[2] Feynman, DeWitt, Faddeev, and Popov. However his commutation relations are unfortunately wrong, and there exist consequent systematic errors in his treatment of the covariant Green's functions. In a previous paper,[3] we succeeded in clarifying Mandelstam's quantization scheme by deriving correct commutation relations with the help of the Peierls method.

On the other hand, many people have recently used the Feynman path integral method[4] as the quantization of the Yang–Mills field. Various important results have been obtained by this method. Although this method is very powerful and heuristic, it is not well based on the field theory. Therefore, it might be interesting to develop Mandelstam's formalism by showing how various results are obtained in Mandelstam's quantization scheme. In this paper we discuss the generalized Ward–Takahashi identities.

In the Feynman path integral method, Slavnov has clarified[5] the fact that the generalized Ward–Takahashi identities (hereafter referred to as Slavnov identities) originate from the local gauge invariance of the Yang–Mills field. This fact is not clear in the canonical quantization scheme, since the canonical quantization starts from imposing the gauge condition. On the other hand, Mandelstam's quantization is carried out in gauge-independent manner, so that we can expect that the origin as well as the derivation of the Slavnov identities is clearly understood in the field theoretic Mandelstam's quantization. (Throughout this paper, we leave aside all questions of infrared divergences and we assume that all expressions are dimensionally regulated,[6] so that formal manipulations of Feynman amplitudes are justified.)

In Sec. II, we easily find the infinite number of Slavnov identities among off-shell Green's functions. Moreover we have the practically important result that these identities are obtained in concise and concrete form. This result enables us to easily perform various calculations in Sec. III.

With the help of Feynman rules and the special Ward–Takahashi identities, 't Hooft[7] has perturbationally derived generalized Ward–Takahashi identities by the combinatoric method. These identities (hereafter referred to as 't Hooft identities) have been used in proving the gauge independence and the unitarity of the S matrix. In Sec. III, we rederive 't Hooft identities with the help of Slavnov identities derived in Sec. II.

In Sec. IV, we discuss our results.

We shall consider the self-interacting massless Yang–Mills field, and use the same notations as those in Mandelstam's paper.[1] At this time we give some of Mandelstam's results[1] which are necessary in the following sections. In order to treat the auxiliary Green's functions, Mandelstam has devised operators $\tilde{A}^{\alpha}_{\mu}(x)$ and $\eta^{\beta}_{\nu}(y)$ acting on a linear space of the covariant Green's functions. The auxiliary Green's functions are then given by

$$\Bigl(H\,\Big|\prod_{i=1}^{n}\tilde{A}^{\alpha(i)}_{\mu(i)}(x_i)\Big|\,G\Bigr)\quad(n=1\sim\infty),\tag{1.1}$$

and the equations for them are obtained in the following concise form:

$$[\partial^x_\mu\{\partial^x_\mu\tilde{A}^{\alpha}_{\nu}(x)-\partial^x_\nu\tilde{A}^{\alpha}_{\mu}(x)\}+gj^{\alpha}_{\nu}(x)-i\theta^{\alpha}_{\nu}(x)]\,|G)$$
$$\equiv[(\alpha\,|D^x_\mu\,|\gamma)\tilde{f}^{\gamma}_{\mu\nu}(x)-i\theta^{\alpha}_{\nu}(x)]\,|G)=0,\tag{1.2}$$

where

$$(\alpha\,|D^x_\mu\,|\gamma)\equiv\delta_{\alpha\gamma}\,\partial^x_\mu+g\epsilon_{\alpha\beta\gamma}\tilde{A}^{\beta}_{\mu}(x),$$
$$\tilde{f}^{\alpha}_{\mu\nu}(x)\equiv\partial^x_\mu\tilde{A}^{\alpha}_{\nu}(x)-\partial^x_\nu\tilde{A}^{\alpha}_{\mu}(x)+g\epsilon_{\alpha\beta\gamma}\tilde{A}^{\beta}_{\mu}(x)\tilde{A}^{\gamma}_{\nu}(x),$$

and

$$\theta^{\alpha}_{\nu}(x)\equiv\eta^{\alpha}_{\nu}(x)-\int d^4y[(\gamma\,|D^y_\mu\,|\delta)\,\eta^{\delta}_{\mu}(y)]\,\partial^x_\nu O_{\alpha\gamma}(x,y)$$
$$+g\epsilon_{\alpha\beta\gamma}\partial^x_\nu O_{\beta\gamma}(x,y)\,|_{x=y},\tag{1.3}$$

with

$$[\tilde{A}^{\alpha}_{\mu}(x),\ \eta^{\beta}_{\nu}(y)]=\delta_{\alpha\beta}\delta_{\mu\nu}\delta^4(x-y),\tag{1.4a}$$

$$(H\,|\,\eta^{\beta}_{\nu}(y)=0,\tag{1.4b}$$

and

$$(\alpha\,|D^x_\mu\,|\gamma)\,\partial^x_\mu O_{\gamma\delta}(x,y)=\delta_{\alpha\delta}\delta^4(x-y).\tag{1.5}$$

Partial differential equation (1.5) is solved as

$$O_{\beta\gamma}(x,y)=\frac{-i}{2}\Delta_F(x-y)\,\delta_{\beta\gamma}$$

 0022-2488/78/1903-0568$1.00

$$-\sum_{n=1}^{\infty}\int d^4x_1\cdots d^4x_n\tfrac{1}{2}\Delta_F(x-x_1)$$
$$\times[ig\epsilon_{\beta\delta\epsilon}\tilde{A}^{\delta}_{\lambda}(x_1)\partial^{x_1}_{\lambda}\tfrac{1}{2}\Delta_F(x_1-x_2)]$$
$$\times[ig\epsilon_{\epsilon\zeta\eta}\tilde{A}^{\zeta}_{\mu}(x_2)\partial^{x_2}_{\mu}\tfrac{1}{2}\Delta_F(x_2-x_3)]\times\cdots$$
$$\times[ig\epsilon_{\theta\iota\gamma}\tilde{A}^{\iota}_{\sigma}(x_n)\partial^{x_n}_{\sigma}\tfrac{1}{2}\Delta_F(x_n-y)], \tag{1.6}$$

where

$$\frac{-i}{2}\Delta_F(x)\equiv\int\frac{d^4k}{(2\pi)^4}\,\frac{\exp(-ikx)}{-k^2+i\epsilon}. \tag{1.7}$$

II. SLAVNOV IDENTITIES

The partial differential equation (1.2) cannot be solved uniquely because of the factor

$$\partial^x_\mu\{\partial^x_\mu\tilde{A}^\alpha_\nu(x)-\partial^x_\nu\tilde{A}^\alpha_\mu(x)\}, \tag{2.1}$$

and (2.1) originates from the fact that the Yang—Mills field is a gauge field. Taking account of the ambiguities resulting from (2.1), we try to solve (1.2) in the form

$$[\Box^x\tilde{A}^\alpha_\nu(x)+gj^\alpha_\nu(x)-i\eta^\alpha_\nu(x)-ig\epsilon_{\alpha\beta\gamma}\partial^x_\nu O_{\beta\gamma}(x,y)\,|_{x=y}$$
$$+\partial^x_\nu\Lambda^\alpha(x)]\,|G)=0, \tag{2.2}$$

where $\Lambda^\alpha(x)$ is arbitrary so far as (2.2) can be solved uniquely. However auxiliary Green's functions which we want to obtain are those satisfying (1.2). Therefore, we must investigate whether Green's functions determined by (2.2) *identically* satisfy (1.2). This problem is reduced to proving

$$[-\partial^x_\mu\tilde{A}^\alpha_\mu(x)+i\int d^4y\{(\gamma\,|D^y_\mu\,|\,\delta)\,\eta^\delta_\mu(y)\}O_{\alpha\gamma}(x,y)-\Lambda^\alpha(x)]\,|G)$$
$$\equiv 0, \tag{2.3}$$

with the help of (2.2). In order to prove (2.3), we first move $\eta^\delta_\mu(y)$ in (2.3) to the right of $O_{\alpha\gamma}(x,y)$. This can be carried out in the following way. Both $\theta^{(1)\alpha}_{\ \ \nu}(x)$ defined by

$$\theta^{(1)\alpha}_{\ \ \nu}(x)\equiv\eta^\alpha_\nu(x)-\int d^4y\,\partial^x_\nu O_{\alpha\gamma}(x,y)(\gamma\,|D^y_\mu\,|\,\delta)\,\eta^\delta_\mu(y), \tag{2.4}$$

and $\theta^\alpha_\nu(x)$ given by (1.3) satisfy[1] the same equation

$$(\alpha\,|D^x_\nu|\,\gamma)\,\theta^\gamma_\nu(x)=(\alpha\,|D^x_\nu|\,\gamma)\,\theta^{(1)\gamma}_{\ \ \nu}(x)=0. \tag{2.5}$$

Therefore, we find

$$(\alpha\,|D^x_\mu\,|\,\gamma)[\theta^{(1)\gamma}_{\ \ \nu}(x)-\theta^\gamma_\nu(x)]=0. \tag{2.6}$$

On the other hand, (1.3) and (2.4) show that the difference $\theta^{(1)\alpha}_{\ \ \nu}(x)-\theta^\alpha_\nu(x)$ is in the form

$$\theta^{(1)\alpha}_{\ \ \nu}(x)-\theta^\alpha_\nu(x)=g\epsilon_{\alpha\beta\gamma}\,\partial^y_\nu O_{\beta\gamma}(x,y)\,|_{x=y}+\partial^x_\nu T^\alpha(x). \tag{2.7}$$

Substituting (2.7) into (2.6), we find that $T^\alpha(x)$ in (2.7) is given by

$$T^\alpha(x)=-\int d^4y\,O_{\alpha\gamma}(x,y)(\gamma\,|D^y_\mu\,|\,\delta)g\epsilon_{\delta\zeta\eta}\,\{\partial^z_\mu O_{\zeta\eta}(y,z)\,|_{y=z}\}. \tag{2.8}$$

Thus (1.3), (2.4), (2.7), and (2.8) lead to

$$\int d^4y\,\{(\gamma\,|D^y_\mu\,|\,\delta)\,\eta^\delta_\mu(y)\}O_{\alpha\gamma}(x,y)$$
$$=\int d^4y\,O_{\alpha\gamma}(x,y)(\gamma\,|D^y_\mu\,|\,\delta)\,\eta^\delta_\mu(y)+g\epsilon_{\alpha\beta\gamma}\,O_{\beta\gamma}(x,x)$$
$$-\int d^4y\,O_{\alpha\gamma}(x,y)(\gamma\,|D^y_\mu\,|\,\delta)g\epsilon_{\delta\zeta\eta}\{\partial^z_\mu\,O_{\zeta\eta}(y,z)\,|_{y=z}\}. \tag{2.9}$$

With the help of identity (2.9), the left-hand side of (2.3) can be rewritten into the new form where the factor $\eta^\delta_\mu(y)$ always operates on $|G)$ directly. Next we replace this $\eta^\delta_\mu(y)\,|G)$ with the one given by (2.2). Then (2.3) are found to hold identically by noticing

$$\Box^y O_{\alpha\epsilon}(x,y)-g\epsilon_{\gamma\delta\epsilon}\partial^y_\mu\{\tilde{A}^\delta_\mu(y)\,O_{\alpha\gamma}(x,y)\}$$
$$=\delta_{\alpha\epsilon}\delta^4(x-y), \tag{2.10}$$

which we obtain from (1.6). Identities (2.3) are the infinite number of generalized Ward—Takahashi identities among off-shell Green's functions.

In the following, we investigate the special case where the gauge fixing term $\Lambda^\alpha(x)$ in (2.2) is given by

$$\Lambda^\alpha(x)=\frac{c}{1-c}\,\partial_\mu\tilde{A}^\alpha_\mu(x). \tag{2.11}$$

In this case, (2.2) can be integrated into

$$[\tilde{A}^\alpha_\nu(x)-\int d^4x'\,D^{\nu\rho}(x-x')[igj^\alpha_\rho(x')+\eta^\alpha_\rho(x')$$
$$+g\epsilon_{\alpha\beta\gamma}\partial^{x'}_\rho O_{\beta\gamma}(x',y)\,|_{x'=y}]\,|G)=0, \tag{2.12}$$

with

$$D^{\nu\rho}(x-x')\equiv\left(\delta_{\nu\rho}-c\,\partial^x_\nu\partial^x_\rho\frac{1}{\Box^x}\right)\tfrac{1}{2}\Delta_F(x-x'). \tag{2.13}$$

In the Appendix, we shall prove that (2.12) leads to the same Feynman rules as those prescribed by[2] Feynman, DeWitt, Faddeev, and Popov. [$c=0$ ($c=1$) is the Feynman (Landau) gauge.] Substituting (2.11) into (2.3), we obtain the Slavnov identities

$$\partial^x_\mu\tilde{A}^\alpha_\mu(x)\,|G)=i(1-c)\int d^4y\{(\gamma\,|D^y_\mu\,|\,\delta)\,\eta^\delta_\mu(y)\}O_{\alpha\gamma}(x,y)\,|G), \tag{2.14}$$

which are originally derived[5] by the Feynman path integral method.

III. 'T HOOFT IDENTITIES

With the help of Slavnov identities (2.14), we shall derive 't Hooft identities[7] of three types A, B, and C.

(Type A) We investigate the following Green's functions

$$W\left\{\prod_{i=1}^{m}(\beta_i,p_i),\ \prod_{j=1}^{n}(\alpha_j,x_j)\right\}\equiv\left[\prod_{i=1}^{m}T_{\nu_i}(p_i;y_i)\right]\left[\prod_{j=1}^{n}\partial^{x_j}_{\mu_j}\right]$$
$$\times G\left\{\prod_{i=1}^{m}(\beta_i,\nu_i,y_i);\ \prod_{j=1}^{n}(\alpha_j,\mu_j,x_j)\right\},$$
$$(\text{for }\ m\geq 0\text{ and }n\geq 1), \tag{3.1}$$

where

$$G\left\{\prod_{i=1}^{m}(\beta_i,\nu_i,y_i);\ \prod_{j=1}^{n}(\alpha_j,\mu_j,x_j)\right\}$$
$$\equiv\left(H\,|\,\prod_{i=1}^{m}\tilde{A}^{\beta_i}_{\nu_i}(y_i)\prod_{j=1}^{n}\tilde{A}^{\alpha_j}_{\mu_j}(x_j)\,|\,G\right), \tag{3.2}$$

$$T_\nu(p;y)\equiv\lim_{p^2\to 0}\epsilon_\nu(p)\int d^4y\,\exp(ipy)\,\Box^y\ (\equiv\epsilon_\nu(p)U(p;y)), \tag{3.3}$$

and $\epsilon_\nu(p)$ is the transversal polarization vector

$$p_\nu\epsilon_\nu(p)=0. \tag{3.4}$$

Slavnov identities (2.14) show that (3.1) is equal to

$$\left[\prod_{j=1}^{n} \partial_{\mu_j}^{x_j}\right] G\left\{\prod_{i=1}^{m} (\beta_i, \nu_i, y_i);\ \prod_{j=1}^{n} (\alpha_j, \mu_j, x_j)\right\}$$
$$= i(1-c)\Big(H \Big| \prod_{i=1}^{m} \tilde{A}_{\nu_i}^{\beta_i}(y_i) \prod_{j=1}^{n-1} \partial_{\mu_j}^{x_j} \tilde{A}_{\mu_j}^{\alpha_j}(x_j)$$
$$\times \int d^4z \, \{(\epsilon | D_\phi^\xi | \zeta) \eta_\phi^\xi(z)\} O_{\alpha_n\epsilon}(x_n, z) \Big| G\Big). \qquad (3.5)$$

With the help of (1.4a), we can move $\eta_\phi^\xi(z)$ in (3.5) to the left of all operators. Then, (1.4b), (2.10), (3.1), (3.3), and (3.4) lead to

$$W\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{j=1}^{n} (\alpha_j, x_j)\right\}$$
$$\equiv -i(1-c) \sum_{k=1}^{n-1} \delta_{\alpha_n \alpha_k} \delta^4(x_n - x_k)\, W\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{\substack{j=1 \\ (j\neq k)}}^{n-1} (\alpha_j, x_j)\right\} \quad (\text{for } n \geq 2 \text{ and } m \geq 0), \qquad (3.6)$$

and

$$W\left\{\prod_{i=1}^{m} (\beta_i, p_i); (\alpha, x)\right\} \equiv 0 \quad (\text{for } m \geq 0). \qquad (3.7)$$

In the following, we shall prove that (3.6) and (3.7) lead to 't Hooft identities (3.6) in 't Hooft's paper[7]: In special cases, (3.6) and (3.7) give respectively

$$W\{;(\alpha_1, x_1)(\alpha_2, x_2)\} \equiv -i(1-c)\delta_{\alpha_2\alpha_1}\delta^4(x_2 - x_1), \qquad (3.8)$$

and

$$W\{;(\alpha, x)\} \equiv 0. \qquad (3.9)$$

By using (3.8), (3.6) can be rewritten into

$$W\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{j=1}^{n} (\alpha_j, x_j)\right\}$$
$$\equiv \sum_{k=1}^{n-1} W\{;(\alpha_j, x_j)(\alpha_n, x_n)\}\, W\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{\substack{j=1 \\ (j\neq k)}}^{n-1} (\alpha_j, x_j)\right\}. \qquad (3.10)$$

Then we find from (3.7) and (3.10)

$$W\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{j=1}^{n} (\alpha_j, x_j)\right\} \equiv 0 \quad (\text{for odd } n \text{ and } m \geq 0) \qquad (3.11)$$

and

$$W\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{j=1}^{n} (\alpha_j, x_j)\right\}$$
$$\equiv W\left\{\prod_{i=1}^{m} (\beta_i, p_i);\right\} \sum{}' W\{;(\alpha_j, x_j)(\alpha_k, x_k)\} \cdots$$
$$\times W\{;(\alpha_l, x_l)(\alpha_m, x_m)\} \quad (\text{for even } n (\geq 2) \text{ and } m \geq 0), \qquad (3.12)$$

where $\sum'$ is the summation over all $n/2$ pairs $(j,k), \ldots, (l,m)$ taken from $(1, 2, \ldots, n)$. For the following discussion, we introduce new Green's functions $\tilde{G}$ by

$$\tilde{G}\left\{\prod_{i=1}^{m} (\beta_i, \nu_i, y_i);\ \prod_{j=1}^{n} (\alpha_j, \mu_j, x_j)\right\}$$
$$\equiv G\left\{\prod_{i=1}^{m} (\beta_i, \nu_i, y_i);\ \prod_{j=1}^{n} (\alpha_j, \mu_j, x_j)\right\}$$
$$- G\left\{\prod_{i=1}^{m} (\beta_i, \nu_i, y_i);\right\} \sum{}' G\{;(\alpha_j, \mu_j, x_j)(\alpha_k, \mu_k, x_k)\}$$
$$\times \cdots \times G\{;(\alpha_l, \mu_l, x_l)(\alpha_m, \mu_m, x_m)\}. \qquad (3.13)$$

Then (3.12) can be expressed by

$$\tilde{W}\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{j=1}^{n} (\alpha_j, x_j)\right\} \equiv 0 \quad (\text{for even } n \text{ and } m \geq 0), \qquad (3.14)$$

where $\tilde{W}$ is obtained from (3.1) by replacing G with $\tilde{G}$. In the following, we shall reduce identities (3.11) and (3.14) to identities with respect to W^c, where W^c is obtained from(3.1) by replacing G with the corresponding connected part G^c: We can easily prove by (3.9), (3.11), and (3.14)

$$W^c\left\{;\ \prod_{j=1}^{n} (\alpha_j, x_j)\right\} \equiv 0 \quad (\text{for } n \geq 3). \qquad (3.15)$$

Furthermore (3.7) and (3.9) give

$$W^c\{(\beta, p); (\alpha, x)\} \equiv 0. \qquad (3.16)$$

With the help of (3.9), (3.15), and (3.16), we can find from (3.11) and (3.14)

$$W^c\left\{\prod_{i=1}^{m} (\beta_i, p_i);\ \prod_{j=1}^{n} (\alpha_j, x_j)\right\} \equiv 0 \quad (\text{for } m \geq 1 \text{ and } n \geq 1) \qquad (3.17)$$

by the following mathematical induction: First, (3.16) means that (3.17) is valid for $m = n = 1$. Next, we assume that (3.17) is valid for any m and n provided $m + n \leq L$. Under this assumption, we shall prove (3.17) for any $m + n = L + 1$. For this purpose, we investigate any one of the identities (3.11) and (3.14) for $m + n = L + 1$ (where $m \geq 1$ and $n \geq 1$). Since $\tilde{G}$ (G) in (3.14) [(3.11)] has contributions from various disconnected Feynman diagrams, we consider any one of them. The contribution in consideration can be expressed by the product of some W^c's. In the case when there exists at least one W^c of type $m \geq 1$ and $n \geq 1$, this W^c (for which $m + n \leq L$) identically vanishes by the assumption. In the other case, there necessarily exist W^c's of type $m = 0$. Among these W^c's, there exists at least one W^c of type $n =$ odd [even n (≥ 4)] for G $(\tilde{G})$, so that this W^c vanishes identically by (3.9) and (3.15). In conclusion, contribution from any disconnected Feynman diagram to $\tilde{W}$ (W) in (3.14) [(3.11)] vanishes identically for $m + n = L + 1$. Thus we have proved (3.17) for any $m + n = L + 1$. Q.E.D.

From (A8) in the Appendix, we have

$$G^c\{;(\alpha_1, \mu_1, x_1)(\alpha_2, \mu_2, x_2)\} = \delta_{\alpha_1\alpha_2} D^{\mu_1\mu_2}(x_1 - x_2)$$
$$+ \Delta G\{;(\alpha_1, \mu_1, x_1)(\alpha_2, \mu_2, x_2)\}, \qquad (3.18)$$

where ΔG vanishes in the limit $g = 0$. With the help of (2.13) and (1.7), we find

$$\partial_{\mu_1}^{x_1} \partial_{\mu_2}^{x_2} D^{\mu_1\mu_2}(x_1 - x_2) = -i(1-c)\delta^4(x_1 - x_2). \qquad (3.19)$$

Then (3.8), (3.18), and (3.19) lead to

$$\partial_{\mu_1}^{x_1} \partial_{\mu_2}^{x_2} \Delta G^c\{;(\alpha_1, \mu_1, x_1)(\alpha_2, \mu_2, x_2)\} \equiv 0, \qquad (3.20)$$

so that we obtain

$$\partial^{x_1}_{\mu_1}\partial^{x_2}_{\mu_2}\pi^{\alpha_1\alpha_2}_{\mu_1\mu_2}(x_1,x_2)\equiv 0, \tag{3.21}$$

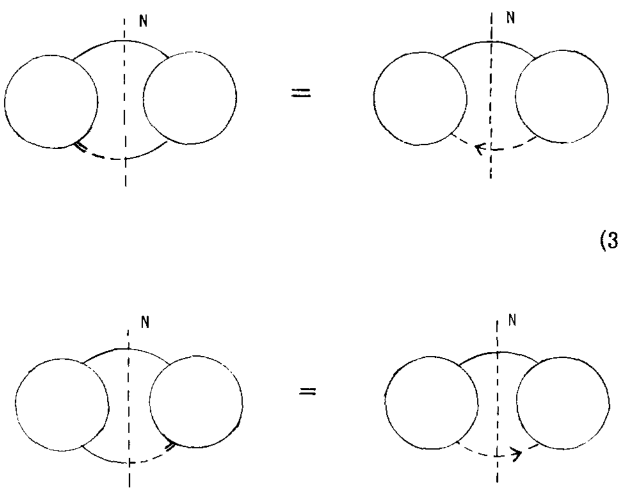
where π is the proper self energy correction. Thus, we find that 't Hooft identities (3.9), (3.21), (3.15), and (3.17) hold for any value of c.

(Type B) By using (1.4a), (1.4b), and the identities (3.5), we find the 't Hooft identities (6.12) in 't Hooft's paper[7]:

$$\begin{aligned}&\left[\prod_{i=1}^{m}T_{\nu_i}(p_i;y_i)\right]\left[\prod_{j=1}^{n}U(q_j,z_j)\right]\Big(H\Big|\prod_{i=1}^{m}\tilde{A}^{\beta_i}_{\nu_i}(y_i)\\&\quad\times\prod_{j=1}^{n}\tilde{A}^{\gamma_j}_{f_j}(z_j)\,\partial^x_\mu\tilde{A}^\alpha_\mu(x)\Big|G\Big)\\&=-i(1-c)\left[\prod_{i=1}^{m}T_{\nu_i}(p_i;y_i)\right]\left[\prod_{j=1}^{n}U(q_j,z_j)\right]\\&\quad\times\left[\sum_{k=1}^{n}\Big(H\Big|\prod_{i=1}^{m}\tilde{A}^{\beta_i}_{\nu_i}(y_i)\prod_{\substack{j=1\\(j\neq k)}}^{n}\tilde{A}^{\gamma_j}_{\rho_j}(z_j)\,\partial^{z_k}_{\rho_k}O_{\alpha\gamma_k}(x,z_k)\Big|G\Big)\right],\end{aligned}\tag{3.22}$$

where the operator U has been defined by (3.3).

(Type C) In Feynman gauge ($c=0$), 't Hooft proved the unitarity of the S matrix, by using the identities (6.19) in his paper.[7] He proved his identities (6.19) with the help of the identities which are graphically expressed by Fig. 1.



(3.23)

We refer the graphical meaning of (3.23) to 't Hooft's paper.[7] In the following, we shall prove (3.23): First, we continue to transform the left-hand side of (3.23), as far as we can apply (3.22). The final result can be grouped into factors of the type

$$\begin{aligned}&\left[\prod_{i=1}^{m}T_{\nu_i}(p_i;y_i)\right]\left[\prod_{j=1}^{n}U(q_j;z_j)\right]\left[\prod_{k=1}^{l}U(s_k;v_k)\,U(t_k;w_k)\right]U(r;x)\\&\quad\times Q\Big(H\Big|\prod_{i=1}^{m}\tilde{A}^{\beta_i}_{\nu_i}(y_i)\prod_{j=1}^{n}\tilde{A}^{\gamma_j}_{\rho_j}(z_j)\prod_{k=1}^{l}O_{\delta_k\zeta_k}(v_k,w_k)\,\partial^x_\mu\tilde{A}^\alpha_\mu(x)\Big|G\Big),\end{aligned}\tag{3.24}$$

where Q is the total antisymmetrization operation among (α,x), $(\delta_1,v_2),\ldots,(\delta_l,v_l)$, which corresponds to the fact that the Faddeev and Popov ghosts obey Fermi statistics.[1,2] In order to show the way we calculate (3.24), we consider for simplicity

$$\begin{aligned}&\left[\prod_{i=1}^{m}T_{\nu_i}(p_i;y_i)\right]\left[\prod_{j=1}^{n}U(q_j;z_j)\right]U(s;v)\,U(t;w)\,U(r;x)\\&\quad\times\Big(H\Big|\prod_{i=1}^{m}\tilde{A}^{\beta_i}_{\nu_i}(y_i)\prod_{j=1}^{n}\tilde{A}^{\gamma_j}_{\rho_j}(z_j)[O_{\delta\zeta}(v,w)\,\partial^x_\mu\tilde{A}^\alpha_\mu(x)\\&\quad-O_{\alpha\zeta}(x,w)\,\partial^v_\mu\tilde{A}^\delta_\mu(v)]\Big|G\Big).\end{aligned}\tag{3.25}$$

With the help of the Slavnov identities (2.14), (3.25) in the Feynman gauge $c=0$ is found to be equal to

$$\begin{aligned}&\left[\prod_{i=1}^{m}T_{\nu_i}(p_i;y_i)\right]\left[\prod_{j=1}^{n}U(q_j;z_j)\right]U(s;v)U(t;w)U(r;x)\\&\quad\times\Big[i\Big(H\Big|\prod_{i=1}^{m}\tilde{A}^{\beta_i}_{\nu_i}(y_i)\prod_{j=1}^{n}\tilde{A}^{\gamma_j}_{\rho_j}(z_j)\,O_{\delta\zeta}(v,w)\\&\quad\times\int d^4u\{(\tau|D^u_\mu|\theta)\,\eta^\theta_\mu(u)\}O_{\alpha\tau}(x,u)\\&\quad-\{(\alpha,x)\to(\delta,v)\}\Big|G\Big)\Big].\end{aligned}\tag{3.26}$$

Next, we move $\eta^\theta_\mu(u)$ in (3.26) to the left of all factors, by using (1.4a), (1.4b), and

$$[O_{\delta\zeta}(v,w),\eta^\theta_\mu(u)]=-gO_{\delta\eta}(v,u)\,\epsilon_{\eta\theta\varphi}\,\partial^u_\mu O_{\varphi\zeta}(u,w),\tag{3.27}$$

where (3.27) can be derived from (1.6). Then, (3.26) is found to be

$$\begin{aligned}&-i\left[\prod_{i=1}^{m}T_{\nu_i}(p_i;y_i)\right]\left[\prod_{j=1}^{n}U(q_j;z_j)\right]U(s;v)\,U(t;w)\,U(r;x)\\&\quad\times\left[\sum_{k=1}^{n}\Big(H\Big|\prod_{i=1}^{m}\tilde{A}^{\beta_i}_{\nu_i}(y_i)\prod_{\substack{j=1\\(j\neq k)}}^{n}\tilde{A}^{\gamma_j}_{\rho_j}(z_j)\,\partial^{z_k}_{\rho_k}O_{\alpha\gamma_k}(x,z_k)\right.\\&\quad\left.\times\,O_{\delta\zeta}(v,w)\Big|G\Big)-\{(\alpha,x)\to(\delta,v)\}\right]\end{aligned}\tag{3.28}$$

in the following way: In order to derive (3.28), we have used (3.22) and the following formula:

$$\begin{aligned}&\epsilon_{\eta\theta\varphi}\int d^4u\,O_{\delta\eta}(v,u)\,\partial^u_\mu O_{\varphi\zeta}(u,w)\\&\quad\times\{-\delta_{\tau\theta}\partial^u_\mu+g\epsilon_{\tau\epsilon\theta}\tilde{A}^\epsilon_\mu(u)\}O_{\alpha\tau}(x,u)\\&\quad-\epsilon_{\eta\theta\varphi}\int d^4u\,O_{\alpha\eta}(x,u)\,\partial^u_\mu O_{\varphi\zeta}(u,w)\\&\quad\times\{-\delta_{\tau\theta}\partial^u_\mu+g\epsilon_{\tau\epsilon\theta}\tilde{A}^\epsilon_\mu(u)\}O_{\delta\tau}(v,u)\end{aligned}\tag{3.29}$$

$$\begin{aligned}&=-\epsilon_{\eta\tau\varphi}\int d^4u\,\partial^u_\mu O_{\varphi\zeta}(u,w)\,\partial^u_\mu\{O_{\alpha\tau}(x,u)\,O_{\delta\eta}(v,u)\}\\&\quad+g(\epsilon_{\eta\theta\varphi}\epsilon_{\tau\epsilon\theta}-\epsilon_{\tau\theta\varphi}\epsilon_{\eta\epsilon\theta})\int d^4u\,O_{\alpha\tau}(x,u)\\&\quad\times O_{\delta\eta}(v,u)\tilde{A}^\epsilon_\mu(u)\,\partial^u_\mu O_{\varphi\zeta}(u,w)\end{aligned}\tag{3.30}$$

$$\begin{aligned}&=\epsilon_{\eta\tau\varphi}\int d^4u\,O_{\alpha\tau}(x,u)\,O_{\delta\eta}(v,u)\\&\quad\times\{\Box^u O_{\varphi\zeta}(u,w)+g\epsilon_{\varphi\epsilon\theta}\tilde{A}^\epsilon_\mu(u)\,\partial^u_\mu O_{\theta\zeta}(u,w)\}\end{aligned}\tag{3.31}$$

$$=\epsilon_{\eta\tau\zeta}\,O_{\alpha\tau}(x,w)\,O_{\delta\eta}(v,w).\tag{3.32}$$

In obtaining (3.31) and (3.32), we have used respectively

$$\epsilon_{\theta\eta\varphi}\epsilon_{\theta\tau\epsilon}+\epsilon_{\theta\eta\epsilon}\epsilon_{\theta\varphi\tau}+\epsilon_{\theta\eta\tau}\epsilon_{\theta\epsilon\varphi}=0,\tag{3.33}$$

and (1.5). Then, the factor $U(t;w)$ leads to the fact that (3.32) does not contribute to the right-hand side of (3.25). Q.E.D.

By using (3.22) and identities obtained similarly to (3.28), the left-hand side of (3.23) is generally reduced to the right-hand side of (3.23).

IV. CONCLUSION AND DISCUSSION

We have investigated Slavnov—'t Hooft identities in Mandelstam's formalism. Slavnov identities originate from the fact that auxiliary Green's functions in various gauges satisfy the same fundamental equation derived by Mandelstam. Since the infinite number of Slavnov identities have been obtained in concise and concrete form, we have been able to explicitly prove 't Hooft identities of type C [i.e., (3.23)] which assure us of the unitarity of the S matrix. In the following, the identities (2.14) are shown to lead to the fact that the S matrix is independent of the value c: First, we have from (A8) in the Appendix

$$\frac{\partial}{\partial c}W[J]=-\tfrac{1}{2}\int d^4x\,d^4x'\,\frac{\partial^x_\phi\partial^x_\omega}{\Box^x}\,\tfrac{1}{2}\Delta_F(x-x')\times\exp[i\int d^4y\{\mathcal{L}_{\rm I}(y)+\mathcal{L}_{\rm II}(y)\}]\times J^\beta_\phi(x)J^\beta_\omega(x')\exp[\tfrac{1}{2}\int d^4z\,d^4z'\,J^\alpha_\mu(z)D^{\mu\nu}(z-z')J^\alpha_\nu(z')].\tag{4.1}$$

On the other hand, we have from (A8)

$$\frac{\delta}{\delta J^\beta_\tau(x)}\frac{\delta}{\delta J^\gamma_\theta(x')}W[J]=\delta_{\beta\gamma}\Big(\delta_{\tau\theta}-c\frac{\partial^x_\tau\partial^x_\theta}{\Box^x}\Big)\tfrac{1}{2}\Delta_F(x-x')W[J]+\int d^4v\,d^4w\Big(\delta_{\tau\rho}-c\frac{\partial^x_\tau\partial^x_\rho}{\Box^x}\Big)\tfrac{1}{2}\Delta_F(x-v)\times\Big(\delta_{\theta\sigma}-c\frac{\partial^{x'}_\theta\partial^{x'}_\sigma}{\Box^{x'}}\Big)\tfrac{1}{2}\Delta_F(x'-w)\times\exp\Big[i\int d^4y\{\mathcal{L}_{\rm I}(y)+\mathcal{L}_{\rm II}(y)\}\Big]J^\beta_\rho(v)J^\gamma_\sigma(w)\times\exp\Big[\tfrac{1}{2}\int d^4z\,d^4z'\,J^\alpha_\mu(z)D^{\mu\nu}(z-z')J^\alpha_\nu(z')\Big].\tag{4.2}$$

Then, (4.2), (2.13), and (1.7) lead to

$$\Big(\Box^x\delta_{\phi\tau}+\frac{c}{1-c}\partial^x_\phi\partial^x_\tau\Big)\Big(\Box^{x'}\delta_{\omega\theta}+\frac{c}{1-c}\partial^{x'}_\omega\partial^{x'}_\theta\Big)\frac{\delta}{\delta J^\beta_\tau(x)}\frac{\delta}{\delta J^\gamma_\theta(x')}W[J]=\delta_{\beta\gamma}\Big(\Box^x\delta_{\phi\omega}+\frac{c}{1-c}\partial^x_\phi\partial^x_\omega\Big)i\delta^4(x-x')W[J]-\exp\Big[i\int d^4y\{\mathcal{L}_{\rm I}(y)+\mathcal{L}_{\rm II}(y)\}\Big]J^\beta_\phi(x)J^\gamma_\omega(x')\times\exp\Big[\tfrac{1}{2}\int d^4z\,d^4z'\,J^\alpha_\mu(z)D^{\mu\nu}(z-z')J^\alpha_\nu(z')\Big].\tag{4.3}$$

Substituting (4.3) into (4.1), we find from (1.7)

$$\frac{\partial}{\partial c}W[J]=-\frac{i}{2}\int d^4x\,d^4x'\,\delta^4(x-x')\times\Big\{i\frac{3}{1-c}\delta^4(x-x')W[J]+\frac{1}{1-c}\partial^x_\tau\frac{\delta}{\delta J^\beta_\tau(x)}\frac{1}{1-c}\partial^{x'}_\theta\frac{\delta}{\delta J^\beta_\theta(x')}W[J]\Big\}.\tag{4.4}$$

The last term in the right-hand side of (4.4) can be calculated by (A2) and

$$\partial^x_\tau\tilde{A}^\beta_\tau(x)\,\partial^x_\theta\tilde{A}^\gamma_\theta(x)\,|G)=(1-c)[-i\delta_{\beta\gamma}\delta^4(0)+i\int d^4y\{\delta\,|D^y_\mu|\epsilon)\,\eta^\epsilon_\mu(y)\}\times O_{\beta\delta}(x,y)\,\partial^x_\nu\tilde{A}^\gamma_\nu(x)]\,|G),\tag{4.5}$$

which is derived from (2.14). With the help of (4.4) and (4.5), we can prove that the S *matrix* is independent of c, in much the same way as 't Hooft and Veltman did.[8]

APPENDIX: GENERATING FUNCTIONAL OF GREEN'S FUNCTIONS

In this Appendix, we ascertain Mandelstam's statement that (2.12) leads to the same Feynman rules as those prescribed by Feynman, DeWitt, Faddeev, and Popov. For this purpose, we investigate the generating functional $W[J]$ defined by

$$W[J]\equiv1+\sum_{n=1}^\infty\frac{1}{n!}\Big(\prod_{i=1}^n\int d^4x_i\,J^{\alpha_i}_{\mu_i}(x_i)\Big)\times\Big(H\,|\prod_{i=1}^n\tilde{A}^{\alpha_i}_{\mu_i}(x_i)\,|G\Big).\tag{A1}$$

First, the equations (2.12) determining Green's functions are rewritten into the equation determining $W[J]$, by using

$$(H\,|\tilde{A}^\alpha_\mu(x)\tilde{A}^\beta_\nu(y)\cdots\tilde{A}^\gamma_\rho(z)\,|G)=\frac{\delta}{\delta J^\alpha_\mu(x)}\frac{\delta}{\delta J^\beta_\nu(y)}\cdots\frac{\delta}{\delta J^\gamma_\rho(z)}W[J]\Big|_{J=0}.\tag{A2}$$

The result is

$$\Big[\frac{\delta}{\delta J^\alpha_\nu(x)}-ig\int d^4x'\,D^{\nu\rho}(x-x')\Big\{\epsilon_{\alpha\beta\gamma}\Big(-\frac{\delta}{\delta J^\beta_\rho(x')}\partial^{x'}_\mu\frac{\delta}{\delta J^\gamma_\mu(x')}+2\frac{\delta}{\delta J^\beta_\mu(x')}\partial^{x'}_\mu\frac{\delta}{\delta J^\gamma_\rho(x')}-\frac{\delta}{\delta J^\beta_\mu(x')}\partial^{x'}_\rho\frac{\delta}{\delta J^\gamma_\mu(x')}\Big)+g\epsilon_{\alpha\beta\gamma}\epsilon_{\gamma\delta\epsilon}\frac{\delta}{\delta J^\beta_\mu(x')}\frac{\delta}{\delta J^\delta_\mu(x')}\frac{\delta}{\delta J^\epsilon_\rho(x')}\Big\}-\int d^4x'\,D^{\nu\rho}(x-x')J^\alpha_\rho(x')-g\epsilon_{\alpha\beta\gamma}\int d^4x'\,D^{\nu\rho}(x-x')\,\partial^{x'}_\rho\tilde{O}_{\beta\gamma}(x',y)\,|_{x'=y}\Big]W[J]=0.\tag{A3}$$

In (A3), $\tilde{O}_{\beta\gamma}(x,y)$ is the operator obtained from $O_{\beta\gamma}(x,y)$ (1.6) by replacing $\tilde{A}^\alpha_\mu(x)$ with $\delta/\delta J^\alpha_\mu(x)$. Equation (A3) can be solved as follows: We introduce $W^0[J]$ by

$$W^0[J] \equiv \exp[-i \int d^4x \{\mathcal{L}_{\rm I}(x) + \mathcal{L}_{\rm II}(x)\}] \; W[J], \tag{A4}$$

where

$$\mathcal{L}_{\rm I}(x) \equiv -\tfrac{1}{2} g \epsilon_{\alpha\beta\gamma} \frac{\delta}{\delta J^\beta_\mu(x)} \frac{\delta}{\delta J^\gamma_\nu(x)} \left(\partial^x_\mu \frac{\delta}{\delta J^\alpha_\nu(x)} - \partial^x_\nu \frac{\delta}{\delta J^\alpha_\mu(x)} \right)$$
$$- \tfrac{1}{4} g^2 \epsilon_{\alpha\beta\gamma} \epsilon_{\alpha\delta\epsilon} \frac{\delta}{\delta J^\beta_\mu(x)} \frac{\delta}{\delta J^\gamma_\nu(x)} \frac{\delta}{\delta J^\delta_\mu(x)} \frac{\delta}{\delta J^\epsilon_\nu(x)}$$

and

$$\mathcal{L}_{\rm II}(x) \equiv i \sum_{n=0}^{\infty} \frac{1}{n+1} \int d^4x_1 \cdots d^4x_n \left[ig\epsilon_{\gamma\alpha\beta} \frac{\delta}{\delta J^\alpha_\nu(x)} \partial^x_\nu \tfrac{1}{2} \Delta_F(x - x_1) \right]$$
$$\times \left[ig\epsilon_{\beta\delta\epsilon} \frac{\delta}{\delta J^\delta_\lambda(x_1)} \partial^{x_1}_\lambda \tfrac{1}{2} \Delta_F(x_1 - x_2) \right] \times \cdots$$
$$\times \left[ig\epsilon_{\theta\iota\gamma} \frac{\delta}{\delta J^\iota_\sigma(x_n)} \partial^{x_n}_\sigma \tfrac{1}{2} \Delta_F(x_n - x) \right]. \tag{A5}$$

Then (A3), (A4), and (A5) give

$$\int d^4x' D^{\nu\rho}(x - x') J^\alpha_\rho(x') W^0[J]$$
$$= \int d^4x' D^{\nu\rho}(x - x') \exp\left[-i \int d^4y \{\mathcal{L}_{\rm I}(y) + \mathcal{L}_{\rm II}(y)\} \right]$$
$$\times \{ -i \int d^4z [J^\alpha_\rho(x'), \; \mathcal{L}_{\rm I}(z) + \mathcal{L}_{\rm II}(z)] + J^\alpha_\rho(x') \} W[J]$$
$$= \exp[-i \int d^4y \{\mathcal{L}_{\rm I}(y) + \mathcal{L}_{\rm II}(y)\}] \frac{\delta}{\delta J^\alpha_\nu(x)} W[J]$$
$$= \frac{\delta}{\delta J^\alpha_\nu(x)} W^0[J], \tag{A6}$$

so that we finally find

$$W^0[J] = \exp[\tfrac{1}{2} \int d^4x \, d^4x' \, J^\alpha_\mu(x) D^{\mu\nu}(x - x') J^\alpha_\nu(x')]. \tag{A7}$$

In conclusion, we obtain from (A4) and (A7)

$$W[J] = \exp[i \int d^4x \{\mathcal{L}_{\rm I}(x) + \mathcal{L}_{\rm II}(x)\}]$$
$$\times \exp[\tfrac{1}{2} \int d^4x \, d^4x' \, J^\alpha_\mu(x) D^{\mu\nu}(x - x') J^\alpha_\nu(x')], \tag{A8}$$

which is the same as what is obtained by the Feynman path integral method.[4]

[1]S. Mandelstam, Phys. Rev. **175**, 1580 (1968).

[2]R.P. Feynman, Acta Phys. Polon. **24**, 697 (1963); B.S. DeWitt, Phys. Rev. **162**, 1195, 1239 (1967); L.D. Faddeev and V.N. Popov, Phys. Lett. B **25**, 29 (1967).

[3]S. Naito, Phys. Rev. D **14**, 3584 (1976).

[4]As one of the review articles on the Feynman path integral method, we give E.S. Abers and B.W. Lee, Phys. Rep. C **9**, 1 (1973).

[5]A.A. Slavnov, Theor. Math. Phys. **10**, 99 (1972).

[6]G. 't Hooft and M.T. Veltman, Nucl. Phys. B **44**, 189 (1972); C.G. Bollini, J.J. Giambiagi, and A. Gonzales-Diminguez, Nuovo Cimento **31**, 550 (1964); G. Cicuta and E. Montaldi, Lett. Nuovo Cimento **4**, 329 (1972).

[7]G. 't Hooft, Nucl. Phys. B **33**, 173 (1971).

[8]G. 't Hooft and M. Veltman, Nucl. Phys. B **50**, 318 (1972).

On independent sets of basis functions for irreducible representations of finite groups

Vojtěch Kopský

Institute of Physics, Czechoslovak Academy of Sciences, POB 24, Na Slovance 8, 180 40 Praha 8, Czechoslovakia

(Received 29 March 1977)

Two theorems on bases of irreducible representations of finite groups are compared. It is stressed that their validity depends upon the functional sets for which they are formulated. The first theorem, which states that there are as many linearly independent (modulo the identity representation) sets of basis functions as is the dimension of the representation, is shown to hold only if the considered functional set constitutes a field. Otherwise, more such sets are necessary as shows the second theorem (extended Noether's theorem), which is limited to polynomial algebra. The second theorem seems to be more apt for explicit construction of functional bases.

INTRODUCTION

The first of the two theorems we want to discuss here has been formulated by Hopfield[1]; it is also given together with its proof as theorem 3.8.1 in the book by Lax.[2] The second theorem has been proven by us for finite abelian groups[3]; it holds, however, for any finite group as will be shown in a pending publication.[4]

It is suitable to discuss the matter in the language of covariants which we have borrowed from Weyl's book.[5] Let $A(G)=\Gamma(G;\mathbf{e}_i)$ be a group of linear operators on a linear space L_n (n finite) which, in a basis $\{\mathbf{e}_i\}$ of L_n, are expressed by matrices $D(g)$ of the corresponding matrix group $\Gamma(G)$; $g\in G$ are elements of an abstract finite group G of which $A(G)$ and $\Gamma(G)$ are faithful operator and matrix representations. Further, let $\psi_i^{(\alpha)}(\mathbf{x})$, $i=1,2,\ldots,\sigma_\alpha(\sigma_\alpha=\dim\Gamma_\alpha)$ be a set of functions on L_n having the property

$$\mathcal{A}(g)\psi_i^{(\alpha)}(\mathbf{x})=\tilde{D}_{ji}^{(\alpha)}(g)\psi_j^{(\alpha)}(\mathbf{x}), \tag{1}$$

where $\mathcal{A}(g)\psi(\mathbf{x})=\psi(A(g^{-1})\mathbf{x})$ defines the action of group elements on functions of $\mathbf{x}\in L_n$, $\tilde{D}_{ji}^{(\alpha)}(g)$ are the matrix elements of matrices of an irreducible representation $\tilde{\Gamma}_{0\alpha}(G)$ adjoint (reciprocal and transposed) to a certain irreducible matrix representation $\Gamma_{0\alpha}(G)$ of the equivalency class Γ_α. We denote the set by $\psi^{(\alpha)}(\mathbf{x})=(\psi_1^{(\alpha)}(\mathbf{x}),\psi_2^{(\alpha)}(\mathbf{x}),\ldots,\psi_{\sigma_\alpha}^{(\alpha)}(\mathbf{x}))$ and call it the $\Gamma_{0\alpha}$ covariant.

Notice that, in a basis $\{\mathbf{e}_{\alpha a,i}\}$ of L_n in which

$$A(G)=\bigoplus_{\alpha=1}^{\kappa}\bigoplus_{a=1}^{n_\alpha}\Gamma_{0\alpha}(G;\mathbf{e}_{\alpha a,i}), \tag{2}$$

the vector components $x_{\alpha a,i}$ of $\mathbf{x}=x_{\alpha a,i}\mathbf{e}_{\alpha a,i}$ constitute the $\Gamma_{0\alpha}$ covariants $\mathbf{x}_a^{(\alpha)}=(x_{\alpha a,1},x_{\alpha a,2},\ldots,x_{\alpha a,\sigma_\alpha})$. The total number n_α of these $\Gamma_{0\alpha}$ covariants equals the number of times the irreducible representation of equivalency class Γ_α is contained in $A(G)$, $\alpha=1,2,\ldots,\kappa$ runs the equivalency classes.

INDEPENDENT COVARIANTS

Theorem 1: There are precisely σ_α of $\Gamma_{0\alpha}$ covariants independent modulo the identity representation or, in other words, any $\Gamma_{0\alpha}$ covariant $\psi^{(\alpha)}(\mathbf{x})$ is expressible as

$$\psi^{(\alpha)}(\mathbf{x})=\sum_{i=1}^{\sigma_\alpha}F_i^{(1)}(\mathbf{x})\psi_i^{(\alpha)}(\mathbf{x}), \tag{3}$$

where $\psi_i^{(\alpha)}(\mathbf{x})$ are the σ_α linearly independent $\Gamma_{0\alpha}$ covariants and $F_i^{(1)}(\mathbf{x})$ are invariant under $\mathcal{A}(G)$.

This theorem is an analog of theorem 3.8.1 by Lax[2] and its proof can be carried out in the same way if, instead of about spherical harmonics, we speak about polynomials in components of x. We reproduce here only the part of the proof which is essential for further discussion.

Proof: The $\Gamma_{0\alpha}$ covariants can be interpreted as vectors of σ_α-dimensional vector space. Then for each x, for which the $\psi_i^{(\alpha)}(\mathbf{x})$ are linearly independent, we can determine the values of $F_i^{(1)}(\mathbf{x})$ as a solution of the σ_α linear equations (3). We can certainly find σ_α polynomial $\Gamma_{0\alpha}$ covariants $\psi_i^{(\alpha)}(\mathbf{x})$, which will be linearly independent.[6] Then the determinant $\Delta(\mathbf{x})=\mathrm{Det}|\psi_i^{(\alpha)}(\mathbf{x})|$ does not vanish almost everywhere and the solution $F_i^{(1)}(\mathbf{x})$ accordingly exists almost everywhere. As the last step it remains to be proved that these solutions are invariants for which we refer to Lax.[2]

Theorem 2 (extended Noether's theorem): There exists a finite number of polynomial $\Gamma_{0\alpha}$ covariants $\mathbf{p}_i^{(\alpha)}(\mathbf{x})$ such, that any other polynomial $\Gamma_{0\alpha}$ covariant $\mathbf{p}^{(\alpha)}(\mathbf{x})$ is expressible as

$$\mathbf{p}^{(\alpha)}(\mathbf{x})=\sum_i P_i^{(1)}(\mathbf{x})\mathbf{p}_i^{(\alpha)}(\mathbf{x}), \tag{4}$$

where $P_i^{(1)}(\mathbf{x})$ are polynomial invariants.

For the proof of this theorem we refer to our work on the extended integrity bases of finite groups.[3,4] The set of independent (modulo the identity representation) $\mathbf{p}_i^{(\alpha)}(\mathbf{x})$ is called the linear integrity basis of $\Gamma_{0\alpha}$ covariants and all such sets together with the integrity basis of invariants are called the "extended integrity basis" associated with the typical[7] representation $\Gamma_0(G)=\oplus_{\alpha=1}^{\kappa}\Gamma_{0\alpha}(G)$.

It is known,[8] that the polynomial invariants (Γ_1 covariants) constitute a polynomial algebra $\mathcal{P}_1$ with a finite integrity basis. Theorem 2 could also be equivalently formulated as follows: The space of polynomial $\Gamma_{0\alpha}$ covariants is of finite dimension over $\mathcal{P}_1$ or, there exists a finite number of linearly independent polynomial $\Gamma_{0\alpha}$ covariants modulo the indentity representation.

Let us consider an arbitrary operator group $A(G)$.

 0022-2488/78/1903-0574$1.00

The covariants $\mathbf{x}_a^{(\alpha)}$ are linearly independent in the sense of Theorem 2. At the same time the number n_α of these covariants is not limited; indeed, it is only the matter of the definition of $A(G)$ and of L_n. Additionally, there are many others $\Gamma_{0\alpha}$ covariants of higher orders which are also independent in the sense of Theorem 2. However, in all cases there should be only σ_α of independent $\Gamma_{0\alpha}$ covariants in the sense of Theorem 1.

There seems to be a contradiction between the two theorems; Actually they are addressing different problems.

To make it clear, we have to realize that at different stages of the previous discussion we speak about different kinds of linear independence. A $\Gamma_{0\alpha}$ covariant in a certain point $\mathbf{x}$ is a σ_α-dimensional vector and there could be, of course, maximally σ_α linearly independent vectors in a σ_α-dimensional space. Here we have, however, the vectors and the linear independence over the field of complex numbers in mind. On the other hand, formula (4) is connected with linear independence of covariants over the algebra of invariants. Finally, in Eq. (3) and in Theorem 1 we have also to specify what kind of independence we have in mind; this fails to be done in the proof by Lax.[2] As a result the following weak point appears in the proof of Theorem 1: Taking $\psi_i^{(\alpha)}(\mathbf{x})$ $(i=1,2,\ldots,\sigma_\alpha)$ as linearly independent (in the usual meaning of vector independence) polynomial $\Gamma_{0\alpha}$ covariants, we are able to express any polynomial $\Gamma_{0\alpha}$ covariant $\psi^{(\alpha)}(\mathbf{x})$ as a linear combination (3) but, to do this, we must generally admit rational invariants.

GENERAL FORM OF A COVARIANT

Let us consider a set $\mathcal{J}$ of function on L_n, which is a linear space closed with respect to $G: \mathcal{A}(G)\mathcal{J}=\mathcal{J}$. It is known that such space $\mathcal{J}$ has a basis whose elements are components of $\Gamma_{0\alpha}$ covariants.[5] The $\Gamma_{0\alpha}$ covariants themselves constitute linear spaces $\mathcal{J}^{(\alpha)}$; for any linear combination $\sum_i a_i\psi_i^{(\alpha)}(\mathbf{x})$ of $\Gamma_{0\alpha}$ covariants $\psi_i^{(\alpha)}(\mathbf{x})$ is again a $\Gamma_{0\alpha}$ covariant. Here a_i are the complex numbers. The number of basic $\Gamma_{0\alpha}$ covariants which allow to express any other $\Gamma_{0\alpha}$ covariant as a linear combination depends on the definition of $\mathcal{J}$. It is, for example, denumerable if $\mathcal{J}$ is the algebra of all polynomials, but the index i may also take values from a continuous set etc.

On the other hand, the "linear combination" $\sum_i F_i^{(1)}(\mathbf{x})\psi_i^{(\alpha)}(\mathbf{x})$ of $\Gamma_{0\alpha}$ covariants, where $F_i^{(1)}(\mathbf{x})$ are from the subspace $\mathcal{J}^{(1)}$ of invariants, is also a $\Gamma_{0\alpha}$ covariant. One is naturally tempted to consider $\mathcal{J}^{(\alpha)}$ as a linear space, for which the coefficients of linear combinations are taken from $\mathcal{J}^{(1)}$. However, the axioms of linear space require that the coefficients are elements of a field.[5] This is important for otherwise we can have a set of $(n+1)$ vectors related by a nontrivial linear equation, yet no one of these vectors may be expressible as a linear combination of the others.

Let, for example, $\mathbf{p}_i^{(\alpha)}(\mathbf{x})$, $i=1,2,\ldots,\sigma_\alpha$ be a set of linearly independent (almost everywhere) polynomial $\Gamma_{0\alpha}$ covariants, $\mathbf{p}^{(\alpha)}(\mathbf{x})$ any other polynomial $\Gamma_{0\alpha}$ covariant. The equation

$$\Delta_p(\mathbf{x})\mathbf{p}^{(\alpha)}(\mathbf{x})=\sum_{i=1}^{\sigma_\alpha}\Delta_p(\mathbf{x})F_i^{(1)}(\mathbf{x})\mathbf{p}_i^{(\alpha)}(\mathbf{x}),$$

where $\Delta_p(\mathbf{x})=\mathrm{Det}\,|\mathbf{p}_i^{(\alpha)}(\mathbf{x})|$, $\Delta_p(\mathbf{x})F_i^{(1)}(\mathbf{x})$ are polynomials, is a linear equation for $(\sigma_\alpha+1)$ $\Gamma_{0\alpha}$ covariants. Yet it is generally impossible to express any $\mathbf{p}^{(\alpha)}(\mathbf{x})$ as a linear combination of $\mathbf{p}_i^{(\alpha)}(\mathbf{x})$ remaining within the polynomial algebra for the $F_i^{(1)}(\mathbf{x})$ are generally rational functions.

If $\mathcal{J}$ is an algebra, then $\mathcal{J}^{(1)}$ is an algebra, if $\mathcal{J}$ is a field, then $\mathcal{J}^{(1)}$ is also a field since the polynomial or rational functions of polynomial or rational invariants are again polynomial or rational invariants. Therefore in the latter case holds Theorem 1, whereas in the former case holds Theorem 2.

It seems, that Theorem 1 provides a way of constructing covariants of the general functional form. This is also somewhat stressed by Lax,[2] when mentioning the Noether's theorem[8] and the integrity basis as a ground for construction of $\mathcal{J}^{(1)}$. The integrity basis is, however, good just for constructing the polynomial invariants, whereas to use Theorem 1 we need the rational ones. To determine them we must proceed beyond the simple knowledge of the integrity basis.

Let us show that the extended integrity basis provides a set of covariants from which we can embark in constructing covariants at least in terms of functions which could be expanded as power series in $\mathbf{x}\in L_n$. We shall illustrate it on the simplest possible example of the group $C_{2z}=2_z$. The group has two one-dimensional irreducible representations and its extended integrity basis is

Γ_1	Γ_2
z	$x\,y$
$x^2\ y^2$	
xy	

As a carrier space L_n we took here the three-dimensional Cartesian vector space in which C_{2z} is understood as a crystallographic group. Notice that there are two Γ_2 covariants independent modulo the algebra of invariants. To pass from Theorem 2 to Theorem 1 we have to use the fact that the ratios x/y or y/x or generally $(ax+by)/(cx+dy)$ are invariants—then one Γ_2 covariant will suffice.

The functions z, x^2, y^2, xy constitute the integrity basis of polynomial algebra $\mathcal{P}_1$, i.e., any invariant polynomial in x, y, z is expressible as a polynomial in z, x^2, y^2, xy. The elements of this integrity basis are also polynomially independent, i.e., no one of them is expressible as a polynomial in the others (otherwise such polynomial would be redundant from the viewpoint of the integrity basis). There is, however, an algebraic relation between them: $(xy)^2=x^2y^2$.

Let us consider now an invariant function $f(x,y,z)$ which can be expanded into power series. Collecting the terms in this expansion up to nth order we can express the function as

$$f(x,y,z)=P_1(z,x^2,y^2)+xyP_2(z,x^2,y^2)+\text{terms of higher orders than } n,$$

where P_1, P_2 are polynomials of orders n, $n-2$, respectively. In the limit we shall have

$$f(x,y,z)=f_1(z,x^2,y^2)+xyf_2(z,x^2,y^2),$$

where f_1, f_2 are functions which can be expanded into power series. Any Γ_2 covariant will be expressed as

$$xf(x,y,z)+yf'(x,y,z)=xf_1(z,x^2,y^2)+yf_1'(z,x^2,y^2)$$
$$+x^2yf_2(z,x^2,y^2)+xy^2f_2'(z,x^2,y^2)$$
$$=xg(z,x^2,y^2)+yh(z,x^2,y^2).$$

Quite generally, to get any $\Gamma_{0\alpha}$ covariant from a set $\mathcal{J}$, it suffices to find a general form of an invariant. The covariants will then be expressed by (4), where instead of $P_i^{(1)}(x)$ we have to write the general form of invariant. This approach has the advantage that the properties of functions $f(x,y,z)$ can be easily correlated with those of $f_1(z,x^2,y^2), f_2(z,x^2,y^2)$; the functions will be, for example, rational, analytic continuous etc. simultaneously.

Let us, in conclusion, mention that, for a special case of vector representations of the crystal point groups, the general functional form of invariants has been found by Döring[9,10] with use of another approach.

[1]J.J. Hopfield, J. Phys. Chem. Solids 15, 97 (1960).
[2]M. Lax, *Symmetry Principles in Solid State and Molecular Physics* (Wiley, New York, 1974).
[3]V. Kopský, J. Phys. C 8, 3251 (1975).
[4]V. Kopský, J. Phys. A (to be published).
[5]H. Weyl, *Classical Groups* (Princeton University, Princeton, N.J., 1946).

[6]This is true for finite groups, where there is a finite number of irreducible representations of finite dimensions. It is not, for example, possible to find such sets for the group SO(3), where there is just one polynomial basis for each irreducible representation $D^{(J)}$ (J integer), namely—the set of harmonics Y_{Jm}, $m=-J, \ldots, J$.

[7]V. Kopský, J. Phys. C 9, 3391, 3405 (1976).
[8]E. Noether, Math. Ann. 77, 89 (1916).
[9]W. Döring, Ann. Phys. (Leipz.) 7, 102 (1958).
[10]W. Döring and G. Simon, Ann. Phys. (Leipz.) 7, 144 (1961).

Scattering theory for Stark Hamiltonians involving long-range potentials

J. Zorbas[a)]

Department of Mathematics, University of British Columbia, Vancouver, B. C., Canada V6T 1W5
(Received 2 May 1977)

A time-dependent and stationary scattering theory is developed for operators of the form $H = H_0 + V$, $H_0 = -\Delta + \mathbf{E}\cdot\mathbf{x}$ with V a long-range potential having the asymptotic form $V(\mathbf{x}) = O(|\mathbf{x}|^{-l})$ as $|\mathbf{x}| \to \infty$, $0 < l \le 1/2$.

I. INTRODUCTION

The "modified" or "renormalized" wave operators $\Omega_\pm$ corresponding to the self-adjoint operators H_2 and H_1 with H_1 spectrally absolutely continuous are defined via the following strong limits

$$\Omega_\pm = \operatorname*{s-lim}_{t\to\pm\infty} \exp(iH_2 t)\exp[-iH_1 t - iG(H_1, t)], \tag{1.1}$$

where $G(H_1, t)$ is an appropriate self-adjoint function of H_1 and t. For the particular choice of self-adjoint operators $H_1 = -\Delta$ and $H_2 = -\Delta + V$ the existence of renormalized wave operators with $G(-\Delta, t) \ne 0$ has been shown[1–6] for a general class of long-range potentials $V(\mathbf{x})$ having the asymptotic form

$$V(\mathbf{x}) = O(|\mathbf{x}|^{-l}), \text{ as } |\mathbf{x}| \to \infty, \ 0 < l \le 1.$$

In Sec. II of this paper we show the existence of renormalized wave operators for a general class of Stark Hamiltonians: $H_2 = H = H_0 + V_l$, $H_1 = H_0 = -\Delta + \mathbf{E}\cdot\mathbf{x}$ with $V_l(\mathbf{x}) = \eta|\mathbf{x}|^{-l} + V_s(\mathbf{x})$, $\eta \in R^1$, $0 < l \le \frac{1}{2}$ and $V_s(\mathbf{x})$ satisfying condition $\mathcal{A}$ of Sec. II. For short-range potentials the spectral and scattering properties of H have been studied in detail by Avron and Herbst[7] and Herbst.[8] In particular Avron and Herbst[7] have shown the existence of wave operators $W_\pm(H, H_0)$ for potentials V_s satisfying $\mathcal{A}$ [note when the limits (1.1) exist with $G(H_1, t) = 0$ we term the resulting limits wave operators and denote them by $W_\pm(H_2, H_1)$].

When $G(H_1, t)$ can be chosen zero the resulting wave operators can be shown to have the following Riemann–Stieltjes integral representations[9–11]

$$W_\pm(H_2, H_1) = \operatorname*{s-lim}_{\epsilon\to+0} W_{\pm\epsilon},$$

$$W_{\pm\epsilon} = \int_{-\infty}^{+\infty} \frac{\pm i\epsilon}{H_2 - \lambda \pm i\epsilon}\, d_\lambda E_\lambda^{H_1} = \int_{-\infty}^{+\infty} d_\lambda E_\lambda^{H_2} \frac{\pm i\epsilon}{\lambda - H_1 \pm i\epsilon}, \tag{1.2}$$

where $E_\lambda^{H_2}$ and $E_\lambda^{H_1}$ are the spectral functions for H_2 and H_1 respectively. The representations (1.2) lead to Hilbert space versions[9–11] of the Lippmann–Schwinger equations and to integral representations for the T operator. The derivation of (1.2) is not valid for the renormalized wave operators with $G(H_1, t) \ne 0$ since in general $\exp[-iG(H_1, t)]$ does not form a one-parameter group in t.

One way to circumvent the above difficulties is to obtain the renormalized wave operators as strong limits of modified time-evolution operators $\Omega^Z(t)$, i.e.,

$$\Omega_\pm\psi = \operatorname*{s-lim}_{t\to\pm\infty} \Omega^Z(t)\psi, \qquad \Omega^Z(t) = \exp(iH_2 t) Z \exp(-iH_1 t), \tag{1.3}$$

with Z a densely defined operator and ψ in a dense subset of the Hilbert space $\mathcal{H}$. Under various assumptions on the domains of H_2, H_1, and Z stationary representations similar to (1.2) can be derived. The existence of operators Z such that (1.3) is valid with $H_2 = -\Delta + V$, $H_1 = -\Delta$ has been shown[12–19] for a general class of long-range potentials.

In Sec. III of this paper unitary operators Z_l are constructed such that the renormalized wave operators corresponding to $H = H_0 + V_l$, $H_0 = -\Delta + \mathbf{E}\cdot\mathbf{x}$ have the representations (1.3). The resulting time-dependent formalism is applied in Sec. IV to obtain a stationary scattering theory for Stark Hamiltonians involving long-range potentials.

II. RENORMALIZED WAVE OPERATORS

We will assume $\mathbf{E} = (\epsilon_0, 0, 0)$, $\epsilon_0 > 0$, $\mathbf{x} = (x, \mathbf{x}_\perp)$. In general we follow the notation of Avron and Herbst.[7]

The following condition has been shown[7] to be sufficient for the existence of wave operators $W_\pm(H, H_0)$, $H = H_0 + V$, $H_0 = -\Delta + \epsilon_0 x$.

$\mathcal{A}$: $V = V_1 + V_2$ where

(a) $V_1 \in \mathcal{L}^2(R^3)$

(b) $(1 + \mathbf{x}^2)^{-N} V_2 \in \mathcal{L}^2(R^3)$ for some N and for $x \le 0$

$$|V_2(\mathbf{x})| \le C(1 + |\mathbf{x}_\perp|)^n (1 + |x|)^{-(n+1+\epsilon)/2}$$

for some $n \ge 0$, $\epsilon > 0$, and constant C.

The above condition excludes potentials having the asymptotic behavior $V(\mathbf{x}) = O(|\mathbf{x}|^{-l})$ as $|\mathbf{x}| \to \infty$, $0 < l \le \frac{1}{2}$. The existence of renormalized wave operators for such long-range potentials is shown in the following theorem.

Theorem 2.1: Assume $V_l(\mathbf{x}) = \eta|\mathbf{x}|^{-l} + V_s(\mathbf{x})$ where $\eta \in R^1$, $0 < l \le \frac{1}{2}$, and V_s satisfies $\mathcal{A}$. Let H be any self-adjoint extension of

$$(H_0 + V_l(\mathbf{x})) \restriction C_0^\infty(R^3), \quad H_0 = -\Delta + \epsilon_0 x, \quad \epsilon_0 > 0.$$

Then the renormalized wave operators

$$\Omega_\pm = \operatorname*{s-lim}_{t\to\pm\infty} \exp(iHt)\exp[-iH_0 t - iG_l(t)] \tag{2.1}$$

[a)]Present address: Department of Mathematics, University of Alberta, Edmonton, Canada T6G 2G1.

exist where

$$G_l(t) = \begin{cases} \epsilon(t)\epsilon_0^{-1/2}\eta\log|t| & \text{for } l=\frac{1}{2}, \\ \epsilon(t)\frac{\epsilon_0^{-l}\eta|t|^{1-2l}}{1-2l} & \text{for } 0<l<\frac{1}{2}, \end{cases}$$

$$\epsilon(t) = \begin{cases} 1 & t>0, \\ -1 & t<0. \end{cases} \tag{2.2}$$

The case $\epsilon_0>0$ considered in the above theorem differs from $\epsilon_0=0$ in several respects. For $\epsilon_0>0$ Avron and Herbst[7] have shown that the functions $G_l(t)$ can be chosen zero for $V_l(\mathbf{x})=O(|\mathbf{x}|^{-l})$ as $|\mathbf{x}|\to\infty$, $\frac{1}{2}<l\leq 1$. In contrast when $\epsilon_0=0$ this is not possible. Furthermore for $\epsilon_0>0$ the functions $G_l(t)$ do not depend on H_0 while for $\epsilon_0=0$ the functions $G(-\Delta,t)$ depend explicitly on $-\Delta$.

Theorem (4.1) of Prugovečki[20] shows that the renormalized wave operators (2.1) satisfy the intertwining properties

$$\exp(iHs)\Omega_\pm=\Omega_\pm\exp(iH_0s), \quad s\in R^1.$$

It follows from these relations that $H\restriction\Omega_\pm\mathcal{H}$ are unitarily equivalent to H_0. Thus $\sigma_{\text{a.c.}}(H)=\sigma_{\text{a.c.}}(H_0)=(-\infty,+\infty)$ [see Theorem (1.1) of Ref. 7] where $\sigma_{\text{a.c.}}(A)$ is the absolutely continuous spectrum of A.

Proof of Theorem 2.1: Following the proof of Theorem 3.2 given in Ref. 7 we must show $f_\psi^l(t)$ is integrable on $[t_0,+\infty)$ for some $t_0>0$ and $\hat{\psi}\in C_0^\infty(R^3)$ where

$$f_\psi^l(t)=\|\{V_l(x-\epsilon_0t^2,\ \mathbf{x}_\perp)-\eta\epsilon_0^{-l}t^{-2l}\}\exp(i\Delta t)\psi\|$$

for $0<l\leq\frac{1}{2}$ (we consider the case $t>0$ only since $t<0$ is analogous).

The function $f_\psi^l(t)$ can be bound as follows:

$$f_\psi^l(t)\leq I_1(t)+I_2(t)+I_3(t),$$

where

$$I_1(t)=\|V_s(x-\epsilon_0t^2,\mathbf{x}_\perp)\exp(i\Delta t)\psi\|,$$

$$I_2(t)=|\eta|\ \|\{[(x-\epsilon_0t^2)^2+|\mathbf{x}_\perp|^2]^{-l/2} - \epsilon_0^{-l}t^{-2l}\}\chi_{ct}\exp(i\Delta t)\psi\|,$$

$$I_3(t)=|\eta|\ \|\{[(x-\epsilon_0t^2)^2+|\mathbf{x}_\perp|^2]^{-l/2} - \epsilon_0^{-l}t^{-2l}\}(1-\chi_{ct})\exp(i\Delta t)\psi\|,$$

where $\chi_{ct}(|\mathbf{x}|)$ is 1 for $(|\mathbf{x}|\leq ct$ and 0 for $|\mathbf{x}|>ct$.

The proof of Theorem (3.2) given in Ref. 7 shows $I_1(t)$ is integrable on $[t_0,+\infty)$.

The standard estimate (Ref. 9, p. 414)

$$\|\exp(i\Delta t)\psi\|_\infty\leq\hat{c}t^{-3/2}$$

together with

$$\{[(x-\epsilon_0t^2)^2+|\mathbf{x}_\perp|^2]^{-l/2}-\epsilon_0^{-l}t^{-2l}\} = \frac{-\epsilon_0^{-l}t^{-2l}l}{2}\int_0^{(\mathbf{x}^2-2x\epsilon_0t^2)/\epsilon_0^2t^4}(1+u)^{-l/2-1}du$$

leads to the following inequality

$$I_2(t)\leq C_1t^{-2l-3/2}\left\{\int_{|\mathbf{x}|\leq ct}d\mathbf{x}\left|\int_0^{(\mathbf{x}^2-2x\epsilon_0t^2)/\epsilon_0^2t^4}(1+u)^{-l/2-1}du\right|^2\right\}^{1/2}$$

for some constant C_1. There exists a $t_0>0$ such that $(\mathbf{x}^2-2x\epsilon_0t^2)/\epsilon_0^2t^4>-\frac{1}{2}$ for all $t>t_0$ and $|\mathbf{x}|\leq ct$. Thus for constants C_2 and C_3

$$I_2(t)\leq C_2t^{-2l-11/2}\left\{\int_{|\mathbf{x}|\leq ct}d\mathbf{x}(\mathbf{x}^2-2x\epsilon_0t^2)^2\right\}^{1/2}\leq C_3t^{-2l-1}$$

which is integrable on $[t_0,+\infty)$ for $l>0$.

The integrability of $I_3(t)$ follows immediately from the estimate [Ref. 7, Eq. (3.9)]

$$|[\exp(i\Delta t)\psi](\mathbf{x})|\leq D_N(1+\mathbf{x}^2+t^2)^{-N}$$

valid for any positive integer N and $|\mathbf{x}|\geq ct$ with c an appropriate constant depending on ψ.

III. MODIFIED TIME-EVOLUTION OPERATORS

The operator $H_0=-\Delta+\epsilon_0x$ is unitarily equivalent [Ref. 7, Theorem (1.1)] to the multiplication operator $\tilde{H}_0=|\mathbf{q}|^2+\epsilon_0x$, $|\mathbf{q}|^2=q_1^2+q_2^2$, acting in $\mathcal{L}^2(R^3,d\mathbf{q}dx)$, i.e., $H_0=U\tilde{H}_0U^{-1}$.

We define the family of operators $\tilde{Z}_l$ acting in $\mathcal{L}^2(R^3, d\mathbf{q}dx)$ as follows:

$$\tilde{Z}_l=W^{-1}U_lW, \tag{3.1}$$

where W is the one-dimensional Fourier transform

$$(W\psi)(\mathbf{q},t)=\text{l.i.m.}(2\pi)^{-1/2}\int_{-\infty}^{+\infty}\exp(ix't)\psi(\mathbf{q},x')dx' \tag{3.2}$$

for $\psi\in\mathcal{L}^2(R^3,d\mathbf{q}dx')$ and U_l is the multiplication operator in $\mathcal{L}^2(R^3,d\mathbf{q}dt)$ given by

$$(U_l\psi)(\mathbf{q},t)=\exp[-iG_l(t/\epsilon_0)]\psi(\mathbf{q},t). \tag{3.3}$$

Theorem 3.1: Suppose that the renormalized wave operaotrs (2.1) exist with the real function $G_l(t)$ satisfying

$$\lim_{t\to\pm\infty}\exp[iG_l(t)-iG_l(\tau+t)]=1 \tag{3.4}$$

for almost all $\tau\in R^1$ and $dG_l(t)/dt$ a bounded continuous function of t. Then

$$\Omega_\pm=\text{s-}\lim_{t\to\pm\infty}\exp(iHt)Z_l\exp(-iH_0t) = W_\pm(H,\hat{H}_0)Z_l=Z_lW_\pm(\hat{H},H_0) \tag{3.5}$$

where $\hat{H}_0=Z_lH_0Z_l^{-1}$, $\hat{H}=Z_l^{-1}HZ_l$, and $Z_l=U\tilde{Z}_lU^{-1}$.

Proof: Once the first equality in (3.5) is verified the second and third equalities follow immediately via the unitarity of Z_l.

In order to verify the first equality we must show for $\psi\in S(R^3)$

$$\lim_{t\to\pm\infty}B(t)=0,$$

$$B(t)=\|\{\exp[iH_0t+iG_l(t)]Z_l\exp(-iH_0t)-I\}\psi\|.$$

The function $B(t)$ can be bound as follows:

$$B(t)\leq B_1(t)+B_2(t)+B_3(t),$$

where

$$B_1(t)=\|\chi_R\{\exp[i\tilde{H}_0t+iG_l(t)]\tilde{Z}_l\exp(-i\tilde{H}_0t)-I\}\psi\|,$$

$$B_2(t)=\|(1-\chi_R)\exp[i\tilde{H}_0t+iG_l(t)]\tilde{Z}_l\exp(-i\tilde{H}_0t)\psi\|,$$

$$B_3(t)=\|(1-\chi_R)\psi\|,$$

with $\chi_R(x)=1$ for $|x|\leq R$ and $\chi_R(x)=0$ for $|x|>R$.

Integrating by parts yields

$$B_2^2(t)=(2\pi)^{-1}\int_{|x|>R} d\mathbf{q}dx|x|^{-2}\left|\int_{-\infty}^{+\infty} du\left(\frac{d\exp(-ixu)}{du}\right)\right.$$
$$\left.\times\exp[-iG_l(u/\epsilon_0+t)](W\psi)(\mathbf{q},u)\right|^2$$
$$\leq D\int_{|x|>R} dx|x|^{-2},$$

where the constant D depends on ψ. Thus by choosing R large enough $B_2(t)$ and $B_3(t)$ can be made arbitrarily small independent of t.

To complete the proof we must show $\lim_{t\to\pm\infty}B_1(t)=0$ for each fixed R where

$$B_1^2(t)=(2\pi)^{-1}\int_{|x|\leq R} d\mathbf{q}dx\left|\int_{-\infty}^{+\infty}dt'\left\{\exp\left[-iG_l\left(\frac{t'}{\epsilon_0}+t\right)\right.\right.\right.$$
$$\left.\left.\left.+iG_l(t)\right]-1\right\}\exp(-ixt')(W\psi)(\mathbf{q},t')\right|^2.$$

It is straightforward to see that the limits $t\to\pm\infty$ can be take within the t' integral which via (3.4) shows $\lim_{t\to\pm\infty}B_1(t)=0$.

Remark: The functions (2.2) are not unique since any real function $\hat{G}_l(t)$ which satisfies

$$\lim_{t\to\pm\infty}\|\{\exp[-iG_l(t)+i\hat{G}_l(t)]-1\}\psi\|=0$$

yields the same renormalized wave operators as $G_l(t)$. In particular the functions (2.2) in the definition of the renormalized wave operators can be replaced by

$$G_l(t)=\begin{cases}\eta(t)\epsilon_0^{-1/2}\eta\log(1+t^2)^{1/2}, & l=\frac{1}{2},\\ \eta(t)\dfrac{\epsilon_0^{-l}\eta(1+t^2)^{(1-2l)/2}}{1-2l}, & 0<l<\frac{1}{2},\end{cases}\qquad(3.6)$$

where $\eta(t)$ is a real C^∞ function such that $\eta(t)=1$ for $t>t_0$, $t_0>0$, and $\eta(t)=-1$ for $t<-t_0$. The assumptions in Theorem (3.1) concerning $G_l(t)$ are satisfied for $G_l(t)$ given by (3.6).

IV. STATIONARY SCATTERING THEORY

A stationary two-Hilbert space scattering theory has been derived by Chandler and Gibson[21] and Prugovečki[22] for wave operators based on the modified time-evolution operators (1.3) with H_1 and H_2 self-adjoint operators acting in the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively and Z is a bounded operator from $\mathcal{H}_1$ to $\mathcal{H}_2$. In order to apply these results to the scattering problem considered in this paper we require the following technical lemma.

Lemma 4.1: Assume $dG_l(t)/dt$ is a continuous bounded function of t, then Z_l maps $\mathcal{D}(H_0)$ onto $\mathcal{D}(H_0)$.

Proof: For $\psi\in S(R^3)$ we have

$$(\tilde{H}_0\tilde{Z}_l\psi)(\mathbf{q},x)=(\tilde{Z}_l\tilde{H}_0\psi)(\mathbf{q},x)+(\tilde{K}_l\psi)(\mathbf{q},x)$$

where

$$(\tilde{K}_l\psi)(\mathbf{q},x)=-\epsilon_0(2\pi)^{-1/2}\int_{-\infty}^{+\infty}dt'\exp[-ixt'-iG_l(t'/\epsilon_0)]$$
$$\times\frac{dG_l(t'/\epsilon_0)}{dt'}(W\psi)(\mathbf{q},t').$$

Since $dG_l(t)/dt$ is bounded we obtain for some constant α independent of $\psi\in S(R^3)$

$$\|\tilde{H}_0\tilde{Z}_l\psi\|\leq\|\tilde{H}_0\psi\|+\alpha\|\psi\|.\qquad(4.1)$$

Since $S(R^3)$ is a core for the closed operator $\tilde{H}_0$ if follows that for any $\phi\in\mathcal{D}(\tilde{H}_0)$ there exists $\phi_n\to\phi$ as $n\to\infty$, $\phi_n\in S(R^3)$. Thus by (4.1), $\{\tilde{H}_0\tilde{Z}_l\phi_n\}$ is Cauchy which implies $\tilde{Z}_l\phi\in\mathcal{D}(\tilde{H}_0)$, i.e., Z_l maps $\mathcal{D}(H_0)$ into $\mathcal{D}(H_0)$. A similar argument shows Z_l^{-1} maps $\mathcal{D}(H_0)$ into $\mathcal{D}(H_0)$ which verifies the lemma.

For simplicity we assume in this section that H is self-adjoint with $\mathcal{D}(H)=\mathcal{D}(H_0)$ (see Ref. 7, Corollary 4.3). If in addition $dG_l(t)/dt$ is a continuous bounded function of t, i.e., by Lemma (4.1), $Z_l\mathcal{D}(H_0)=\mathcal{D}(H_0)$, then the results contained in Sec. 3B and 3C of Chandler and Gibson[21] and Sec. 2 of Prugovečki[22] are applicable to the renormalized wave operators (3.5). We summarize these results in the following two theorems.

Theorem 4.2: Assume H is self-adjoint with $\mathcal{D}(H)=\mathcal{D}(H_0)$. Furthermore assume the representations (3.5) for the renormalized wave operators are valid with the real functions $G_l(t)$ satisfying (3.4) and $dG_l(t)/dt$ a continuous bounded function of t. Then:

(a)

$$\Omega_\pm=\operatorname*{s-lim}_{\epsilon\to+0}W_{\pm\epsilon}^{Z_l},\qquad(4.2)$$

$$W_{\pm\epsilon}^{Z_l}=\int_{-\infty}^{+\infty}\frac{\pm i\epsilon}{H-\lambda\pm i\epsilon}Z_l d_\lambda E_\lambda^{H_0},\quad\epsilon>0.\qquad(4.3)$$

(b) For $\epsilon>0$

$$W_{\pm\epsilon}^{Z_l}=Z_l-\int_{-\infty}^{+\infty}\frac{1}{H-\lambda\pm i\epsilon}Vd_\lambda E_\lambda^{H_0},\qquad(4.4)$$

$$(W_{\pm\epsilon}^{Z_l})^*=Z_l^{-1}-\int_{-\infty}^{+\infty}\frac{1}{\lambda-H_0\mp i\epsilon}V^*d_\lambda E_\lambda^{H},\qquad(4.5)$$

where $V=HZ_l-Z_lH_0$.

(c)

$$\Omega_\pm=Z_l-\operatorname*{s-lim}_{\epsilon\to+0}\int_{-\infty}^{+\infty}\frac{1}{H-\lambda\pm i\epsilon}Vd_\lambda E_\lambda^{H_0},\qquad(4.6)$$

$$\Omega_\pm=Z_l-\operatorname*{w-lim}_{\epsilon\to+0}Z_l\int_{-\infty}^{+\infty}\frac{1}{H_0-\lambda\pm i\epsilon}V^*d_\lambda E_\lambda^{H}\Omega_\pm.\qquad(4.7)$$

Theorem 4.3: Under the same assumptions as Theorem 4.2 the T operator $T=\Omega_+^*\Omega_--I$ has the following stationary representations:

$$\text{(a)}\quad T=2i\operatorname*{w-lim}_{\epsilon\to+0}\int_{-\infty}^{+\infty}d_\lambda E_\lambda^{H_0}V^*\Omega_-\frac{\epsilon}{(H_0-\lambda)^2+\epsilon^2},\qquad(4.8)$$

$$\text{(b)}\quad T=(-2\pi i)\operatorname*{w-lim}_{\epsilon\to0}\int_{-\infty}^{+\infty}d_\lambda E_\lambda^{H_0}\int_{-\infty}^{+\infty}\delta_\epsilon(\lambda-\mu)$$
$$\times T([\lambda+\mu+i\epsilon]/2)d_\mu E_\mu^{H_0}\qquad(4.9)$$

where $\delta_\epsilon(\beta)=(\epsilon/\pi)(\beta^2+\epsilon^2)^{-1}$ and

$$T(z)=(z-H_0)Z_l^{-1}(z-H)^{-1}Z_l(z-H_0)-(z-H_0).\qquad(4.10)$$

Remarks: Due to the unitarity of Z_l the results of Theorems 4.2 and 4.3 can be rewritten in various ways. For example, the operator (4.10) can be rewritten in terms of $\hat{H}$ and $\hat{H}_0$ of Theorem (3.1) as follows: lows:

$$\begin{aligned}T(z)&=(z-H_0)(z-\hat{H})^{-1}(z-H_0)-(z-H_0)\\&=Z_l^{-1}\{(z-\hat{H}_0)(z-H)^{-1}(z-\hat{H}_0)-(z-\hat{H}_0)\}Z_l.\end{aligned}\qquad(4.11)$$

When there are no long-range potentials present, i.e., $\eta=0$ then $Z_l=I$ and Theorems 4.2 and 4.3 provide a stationary Hilbert space formalism for the wave operators $W_{\pm}(H, H_0)$.

ACKNOWLEDGMENTS

I would like to thank Ira Herbst for a helpful conversation and Lon Rosen for informing me of the work of Avron and Herbst.

[1]J.D. Dollard, J. Math. Phys. **5**, 729–38 (1974).
[2]W.O. Amrein, Ph.A. Martin, and B. Misra, Helv. Phys. Acta **43**, 313–44 (1970).
[3]V.S. Buslaev and V.B. Matveev, Theor. Math. Phys. **2**, 266–74 (1970).
[4]P.K. Alsholm, " Wave operators for long-range scattering," mimeographed manuscript (1974), Department of Math., U. of California, Berkeley.
[5]P.K. Alsholm and T. Kato, "Scattering with long-range potentials," in *Partial Differential Equations*, *Proceedings of Symposia in pure Mathematics*, Vol. 23, (Berkeley, 1971) (American Math. Soc., Providence, R.I., 1973), pp. 393–9.
[6]L. Hörmander, Math. Z. **146**, 69–71 (1976).
[7]J.E. Avron and I.W. Herbst, Commun. Math. Phys. **52**, 239–54 (1977).
[8]I.W. Herbst, "Unitary equivalence of Stark Hamiltonians," preprint, Phys. Dept., Princeton University.
[9]E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1970).
[10]W.O. Amrein, V. Georgescu, and J.M. Jauch, Helv. Phys. Acta **44**, 407–34 (1971).
[11]D.B. Pearson, Nuovo Cimento A **2**, 853–80 (1971).
[12]D. Mulherin and I.I. Zinnes, J. Math. Phys. **11**, 1402–8 (1970).
[13]V.B. Matveev and M.M. Skriganov, Theor. Math. Phys. **10**, 156–164 (1972).
[14]E. Prugovečki and J. Zorbas, J. Math. Phys. **14**, 1398–1409 (1973); Erratum **15**, 268 (1974).
[15]E. Prugovečki and J. Zorbas, Nucl. Phys. A **213**, 541–69 (1973).
[16]J. Zorbas, "Stationary scattering theory for long-range potentials," thesis, Univ. of Toronto (1974).
[17]C. Chandler and A.G. Gibson, J. Math. Phys. **15**, 291–4 (1974).
[18]V. Georgescu, "Methodes stationnaires pour des potentiels a longue portee a symetrie spherique," thesis, Universite de Geneve (1974).
[19]D. Masson and E. Prugovečki, J. Math. Phys. **17**, 297–302 (1976).
[20]E. Prugovečki, Nuovo Cimento B **4**, 105–123.
[21]C. Chandler and A.G. Gibson, J. Math. Phys. **14**, 1328–35 (1973).
[22]E. Prugovečki, J. Math. Phys. **14**, 957–62 (1973).

Matrix element expansion of a spin wavefunction

George B. Smith

Department of Physics, University of New Orleans, New Orleans, Louisiana 70122
(Received 22 June 1977)

An expansion of the wavefunction for a free, massive particle with spin is obtained in terms of the matrix elements for the unitary, irreducible representations of SL(2,c) by a method based on the theory of induced representations. It is further shown that this expansion is equivalent to the Shapiro integral transformation for the wavefunction.

I. INTRODUCTION

An expansion of the wavefunction for a free, massive, spinless particle in terms of unitary, irreducible representations (unirreps) of the Lorentz group was first obtained by Shapiro.[1] Later, Chou and Zastavenko[2] generalized this work so that it applied to arbitrary spin. The method used by these authors was to assume the existence of an integral transformation relating the wavefunction to elements of the support spaces of the unirreps of the Lorentz group. After obtaining an explicit form for the kernel of this transformation the expansion of the wavefunction was known, if it existed. There was no guarantee of this existence, however.

Finally, Popov[3] obtained this same transformation by a method which utilized the power of the representation theory of SL(2,c).[4,5] The essential merit of this approach is that the existence of the transformation is guaranteed.

In this paper we will use the concept of a covariant projection, which derives from the theory of induced representations,[6,7] to obtain an expansion of the wavefunction in terms of the matrix elements, relative to a canonical basis, of the unirreps of SL(2,c). In contrast to the expansions previously obtained, this results in a very compact form for the expansion.

The method employed in this paper clearly shows that the expansion of the wavefunction is essentially a relation between representations of SL(2,c) which are induced by two different subgroups. This produces a neat, transparent derivation of the expansion of the wavefunction which utilizes current group theoretic techniques.

We will also obtain the Shapiro transformation in a straightforward manner and in a form closely resembling the form given in Ref. 3, thereby demonstrating the equivalence of our expansion to the Shapiro integral transformation.

In Sec. II, we give a brief discussion of the irreducible representations of SL(2,c), and we set up the formalism of covariant projections. In Sec. III, we establish the canonical basis and obtain a closed expression for the matrix elements.

In Sec. IV, we apply the formalism to obtain the desired expansion of the wavefunction; and in Sec. V, we summarize our results. We derive the Shapiro transformation in an Appendix.

II. COVARIANT PROJECTIONS

The (left) regular representation of SL(2,c) is realized on the space L of square-integrable functions defined over the six-dimensional parameter space of SL(2,c). If $F\in L$ then the regular representation is given by the action of its operators on elements of L

$$\{\Gamma(l_0)F\}(l)\equiv F(l_0^{-1}l)\quad \text{for all } l,\ l_0\in \mathrm{SL}(2,c). \tag{1}$$

The support space L decomposes into orthogonal subspaces $L_{n\rho}$ which are irreducible under the action of the set $\{\Gamma(l)\colon l\in \mathrm{SL}(2,c)\}$. The subspace $L_{n\rho}$ supports the (n,ρ)th irreducible representation (irrep) of SL(2,c). The labels n and ρ are associated with the eigenvalues of the two irreducible operators of SL(2,c).

We define the (n,ρ)th irrep of SL(2,c) by the action of the set $\{\Gamma(l):l\in \mathrm{SL}(2,c)\}$ on elements $\phi^{n\rho}(z,z^*)$ of $L_{n\rho}$ according to

$$\{\Gamma(l)\phi^{n\rho}\}(z,z^*)=(\delta-\beta z)^{\eta}\,(\delta^*-\beta^* z^*)^{\xi}\phi^{n\rho}(z',z'^*), \tag{2}$$

with the complex numbers η and ξ being given by

$$\eta\equiv \tfrac{1}{2}(n+i\rho)-1,\quad \xi\equiv\tfrac{1}{2}(-n+i\rho)-1, \tag{3}$$

and where z denotes either the complex parameter or the corresponding element of SL(2,c)'s subgroup $Z\equiv\{z=(\begin{smallmatrix}1&0\\ z&1\end{smallmatrix})\colon z\in C\}$. In Eq. (2), z' is the complex number $(\alpha z-\gamma)/(\delta-\beta z)$, with $l=(\begin{smallmatrix}\alpha&\beta\\ \gamma&\delta\end{smallmatrix})$ being a typical element of SL(2,c).

If n is an integer and ρ is a real number then Eq. (2) defines an irreducible representation of SL(2,c) which is unitary with respect to the scalar product

$$(\phi,\psi)=(i/2)\int dz\,dz^*\ \phi^*(z,z^*)\,\psi(z,z^*)\quad \phi,\psi\in L_{n\rho}.$$

Unirreps of SL(2,c) satisfying this criterion are known as representations of the principal series.

There is another series of unirreps of SL(2,c), the complementary series,[5] which we will not be concerned with since the principal series is complete in the sense that the regular representation can be decomposed in terms of the principal series representations except on a set of measure zero.[6] We will therefor obtain our expansion of the wavefunction wholly in terms of the principal series representations.

To this end, the concept of a covariant projection will be essential. First however, let us define what we mean by a covariant function.

 0022-2488/78/1903-0581$1.00

Suppose that H is some proper subgroup of SL$(2,c)$ and that $D^r(H)$ is a representation of H. A function $F^r \in L$ is said to be covariant with respect to the rth representation of the subgroup H if it satisfies the condition

$$F^r(lh) = D^r(h^{-1})F^r(l) \tag{4}$$

for all $l \in$ SL$(2,c)$ and all $h \in H$.

Let M be the subset SL$(2,c)/H$ so that any $l \in$ SL$(2,c)$ can be uniquely decomposed as $l = mh$ with $m \in M$ and $h \in H$. Hence, Eq. (4) implies the decomposition

$$F^r(l) = D^r(h^{-1})f^r(m), \tag{5}$$

which defines f^r with $l = mh$.

The function f^r belongs to a subspace L_r of L which is invariant under the action of the operator representation defined by Eq. (1). That is,

$$\{\Gamma(l)f^r\}(m) = D^r(h_w^{-1})f^r(m'), \tag{6}$$

where $l^{-1}m = m'h_w$, with $m, m' \in M$ and $h_w \in H$. This defines a (generally reducible) representation of SL$(2,c)$ which is said to be induced by the representation $D^r(H)$ of the subgroup H. Equation (2) is an example of this construction with the subgroup

$$K \equiv \left\{k = \begin{pmatrix}\lambda & \mu\\ 0 & \delta\end{pmatrix} : \delta = \lambda^{-1};\ \lambda, \mu, \delta \in C\right\}$$

as the inducing subgroup.

Generally, a function $F \in L$ will not possess any particular covariance properties. However, given an arbitrary function $F \in L$ we can construct a function $F^r \in L$ which is covariant with respect to the rth representation of H via the integral operator P^r as

$$F^r(l) = \{P^r F\}(l) \equiv \int d_l h\, D^r(h)\, F(lh). \tag{7}$$

In this, $d_l h$ is a left-invariant measure for the subgroup H.

If H is compact then the integral in Eq. (7) is guaranteed to exist, and the operator P^r is idempotent. For this reason we refer to $\{P^r F\}$ as the covariant projection of the function F. If H is noncompact then we can modify F so that it is modulated by a bounded function with compact support on the subgroup H, which will guarantee the existence of the integral in (7).[7]

III. MATRIX ELEMENTS FOR SL $(2,c)$

It is essential to our purpose that we obtain a basis in the irreducible space $L_{n\rho}$ which decomposes this space into orthogonal subspaces, each of which supports a unirrep of the unitary subgroup SU(2). That is, we want a basis $\{\phi^{n\rho}_{sm}(z,z^*)\}$ in $L_{n\rho}$ which satisfies the two criterion

$$\{\Gamma(l)\phi^{n\rho}_{sm}\}(z,z^*) = \sum \phi^{n\rho}_{s'm'}(z,z^*)\Gamma^{n\rho}_{s'm',sm}(l), \tag{8}$$

$$\{\Gamma(u)\phi^{n\rho}_{sm}\}(z,z^*) = \sum \phi^{n\rho}_{sm'}(z,z^*)D^s_{m'm}(u) \tag{9}$$

for $u \in$ SU(2). Equation (8) defines the matrix elements $\Gamma^{n\rho}_{s'm',sm}(l)$, relative to the basis $\phi^{n\rho}_{sm}(z,z^*)$, of the (n,ρ)th irrep of SL$(2,c)$. The $D^s_{m'm}(u)$ appearing in (9) are the usual matrix elements for the s'th unirrep of SU(2). A basis satisfying (9) will be an eigenbasis of the square of the total angular momentum and its z-axis projection.

The sum in Eq. (9) is over the index m' which runs the range $-s, -s+1, \ldots, s$. The sum in Eq. (8) is a double sum over the indices s' and m'. These indices run the ranges $|\frac{1}{2}n|,\ |\frac{1}{2}n|+1, \ldots$ and $-s, -s+1, \ldots, s$, respectively. These limits on the summation indices are to be understood throughout this paper.

The conditions embodied in equations (2) and (9) are sufficient to determine explicitly[8] the form of the basis $\phi^{n\rho}(z,z^*)$. It is

$$\phi^{n\rho}_{sm}(z,z^*) = (1+z^*z)^{i\rho/2-1}D^s_{-n/2,m}(u^{-1}(z,z^*)), \tag{10}$$

where $u(z,z^*)$ is the unitary matrix given by

$$u(z,z^*) = (1+z^*z)^{-1/2}\begin{pmatrix}1 & -z^*\\ z & 1\end{pmatrix}. \tag{11}$$

The various elements of the basis $\phi^{n\rho}_{sm}(z,z^*)$ are orthogonal to each other, satisfying an orthogonality relation of the form

$$\int dz\,dz^*\,(1+z^*z)^{\mathrm{Im}(\rho)}\phi^{*n\rho}_{s'm'}(z,z^*)\,\phi^{n\rho}_{sm}(z,z^*) = \frac{2\pi i}{2s+1}\delta_{s's}\,\delta_{m'm}. \tag{12}$$

This relation can be directly verified using the orthogonality of the $D^s_{m'm}(u)$.

We can utilize this orthogonality, along with Eqs. (2) and (8), to obtain a closed form for the matrix elements. It is

$$\Gamma^{n\rho}_{sm,s'm'}(l) = \frac{2s+1}{2\pi i}\int dz\,dz^*\,(1+z^*z)^{\mathrm{Im}(\rho)}\phi^{*n\rho}_{sm}(z,z^*) \times(\delta-\beta z)^{\eta}(\delta^*-\beta^*z^*)^{\xi}\,\phi^{n\rho}_{s'm'}(z',z'^*), \tag{13}$$

where η and ξ are defined in Eq. (3), and again where $l = \left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ is any element of SL$(2,c)$. The transformed quantity z' is again given by the expression $(\alpha z-\gamma)/(\delta-\beta z)$.

When we restrict l to the unitary subgroup we get

$$\Gamma^{n\rho}_{sm,s'm'}(u) = \delta_{ss}\,D^s_{mm'}(u) \tag{14}$$

for $u \in$ SU(2), which is the block-diagonal form we sought.

An arbitrary element $\phi^{n\rho}(z,z^*)$ of $L_{n\rho}$ can be expanded in terms of the basis (10) as

$$\phi^{n\rho}(z,z^*) = \sum A^{n\rho}_{sm}\phi^{n\rho}_{sm}(z,z^*), \tag{15}$$

where the $A^{n\rho}_{sm}$ are the appropriate complex expansion coefficients. The sum is again a double sum over the indices s and m.

Conversely, the set of all such functions (15) forms the linear space $L_{n\rho}$. That the $\phi^{n\rho}_{sm}(z,z^*)$ are complete in $L_{n\rho}$ can be inferred directly from their definition in terms of the complete set of functions, $D^s_{m'm}(u)$.

The complex coefficients $A^{n\rho}_{sm}$ can be used to form a spinor $(\ldots, A^{n\rho}_{sm}, \ldots)$ which corresponds to the function defined by Eq. (15). The set of all such spinors formed from functions in $L_{n\rho}$ also forms a linear vector space which we will call the associated spinor space.
This spinor space is clearly isomorphic to $L_{n\rho}$, and

therefore also supports the (n,ρ)th unirrep of SL$(2,c)$. That is, the realization of the (n,ρ)th unirrep of SL$(2,c)$ on the associated spinor space is given by

$$A^{n\rho}_{sm} \to A'^{n\rho}_{sm} = \sum \Gamma^{n\rho}_{sm,s'm'}(l) A^{n\rho}_{s'm'}. \quad (16)$$

IV. EXPANSION OF THE WAVEFUNCTION

The wavefunction $\psi^s_m(p)$ of a free particle with spin s, z-axis projection m, nonzero rest mass m_0, and four-momentum[9] p transforms under a Lorentz transformation $l: p \to p' = lpl^\dagger$ according to[3,10]

$$\psi^s_m(p) \to \psi'^s_m(p') = \sum_{m'} D^s_{mm'}(u_w^{-1}(l,p))\psi^s_{m'}(p), \quad (17)$$

where $u_w(l,p)$ is the Wigner rotation defined by

$$u_w^{-1}(l,p) = b^{-1}(p')lb(p). \quad (18)$$

In this, $b(p)$ is the pure Lorentz boost from the rest frame of the particle to the rest frame of the observer. It has the representation $b(p) = (p/m_0)^{1/2}$ in terms of the Hermitian matrix p. Equation (17) defines a reducible representation of SL$(2,c)$ which is induced by the subgroup SU(2). It is therefore a special case.

From the wavefunction $\psi^s_m(p)$ we can define a function

$$\chi(l) \equiv -(m_0/2\pi^2)\sum_m D^s_{\alpha m}(u^{-1})\,\psi^s_m(p),$$

with $l = b(p)u$, $u \in$ SU(2), and with α fixed. This function transforms according to the regular representation of SL$(2,c)$ and has no manifest covariance properties. We can utilize a projection operator $P^{n\rho}$, of the type defined in equation (7), to project out that part of $\chi(l)$ which is covariant with respect to the subgroup K. That is,

$$Q^{n\rho}(l) \equiv \{P^{n\rho}\chi\}(l) = \int d_l k'\, \Delta^{n\rho}(k')\chi(lk'), \quad (19)$$

where $d_l k'$ is a left-invariant measure for K, and $\Delta^{n\rho}(K)$ is the representation of K given by

$$\Delta^{n\rho}(k) = \lambda^{-\eta}\,\lambda^{*-\xi}, \quad (20)$$

η, ξ, and λ having been defined earlier.

Decomposing lk' as $b(p')u'$, $u' \in$ SU(2), we can write (19) as

$$Q^{n\rho}(l) = -(m_0/2\pi^2)\sum_m \int d_l k'\, \Delta^{n\rho}(k')\, D^s_{\alpha m}(u'^{-1})\psi^s_m(p'). \quad (21)$$

Now we factor from the right a matrix of the form

$$\tau \equiv \begin{pmatrix} \exp(i\tau/2) & 0 \\ 0 & \exp(-i\tau/2) \end{pmatrix} \text{ with } \tau \in \text{Re} \quad (22)$$

from $lk' = b(p')u'$ leaving $l\underline{k} = b(p')\underline{u}$, where $\underline{k}$ has real diagonal elements. The measure $d_l k'$ becomes a product of measures $d_l\underline{k}\, d\tau$, and (21) becomes

$$Q^{n\rho}(l) = -(m_0/\pi)\sum_m \int d_l\underline{k}\, \Delta^{n\rho}(\underline{k})\, D^s_{-n/2,m}(\underline{u}^{-1})\,\psi^s_m(p') \quad (23)$$

since the τ integration gives $2\pi\delta_{-n/2,\alpha}\delta_{\alpha,m'}$. In this equation, the relation $l\underline{k} = b(p')\underline{u}$ determines $b(p')$ and $\underline{u}$.

Equation (23) is fundamental, and from it we shall derive an expression for the coefficients in the expansion of the wave-function. In the Appendix we shall derive the inverse Shapiro transformation from Eq. (23).

If we take $l = z \in Z$, Eq. (23) gives the projection of the wavefunction $\psi^s_m(p)$ into the space $L_{n\rho}$. Thus, the function $Q^{n\rho}(z)$ transforms according to Eq. (2) under a Lorentz transformation. This is easy to demonstrate using the invariance of the measure $d_l\underline{k}$.

By parametrizing the rotation $\underline{u}$ as $u(z', z'^*)\tau'$, with τ' and $u(z', z'^*)$ being given by Eqs. (22) and (11), respectively, we can make use of the identity $u(z', z'^*) = z'k(z', z'^*)$, where

$$k(z', z'^*) \equiv (1 + z'^* z')^{-1/2}\begin{pmatrix} 1 & -z'^* \\ 0 & 1 + z'^* z' \end{pmatrix}, \quad (24)$$

to write $l\underline{k} = z\underline{k} = b(p')\underline{u}$ in the form $b^{-1}(p')z = z'k_w$ with $k_w = k(z', z'^*)\tau'\underline{k}^{-1}$ and with $l = z \in Z$. Furthermore, since

$$\Delta^{n\rho}(k(z', z'^*)) = (1 + z'^* z')^{(i\rho/2)-1} \quad (25)$$

we can utilize Eq. (10) to write (23) as

$$Q^{n\rho}(z) = \frac{1}{4\pi m_0}\sum_m \int \frac{d^3P'}{P'_0}\, \Delta^{n\rho}(k_w^{-1})\phi^{n\rho}_{sm}(z', z'^*)\psi^s_m(p'), \quad (26)$$

with l chosen to be $z \in Z$. To obtain this last equation we have also used

$$d_l\underline{k} = \frac{-1}{4m_0^2}\,\frac{d^3P'}{P'_0} \quad (27)$$

which is calculated from $z\underline{k} = b(p')\underline{u}$.

Now $Q^{n\rho}$ is covariant with respect to the subgroup K. We can therefore choose $l = z \in Z$ and expand $Q^{n\rho}(z)$ in accordance with equation (15) since $Q^{n\rho}(z)$ is an element of $L_{n\rho}$. Having expanded $Q^{n\rho}(z)$ in this manner we can use the orthogonality of the $\phi^{n\rho}_{sm}(z, z^*)$ to isolate the expansion coefficients. We obtain

$$A^{n\rho}_{sm} = \frac{1}{4\pi m_0}\sum_{m'}\int \frac{d^3P'}{P'_0}\left\{\frac{2s+1}{2\pi i}\int dz\,dz^*\,(1 + z^* z)^{\mathrm{Im}(\rho)} \times \phi^{*n\rho}_{sm}(z, z^*)\Delta^{n\rho}(k_w^{-1})\phi^{n\rho}_{sm'}(z', z'^*)\right\}\psi^s_{m'}(p'), \quad (28)$$

with $b^{-1}(p')z = z'k_w$.

Comparing Eq. (28) with Eq. (13) we see that the expression in curly brackets is the matrix element $\Gamma^{n\rho}_{sm,sm'}(b(p))$. Thus, we can write (28) in the final form

$$A^{n\rho}_{sm} = \frac{1}{4\pi m_0}\sum_{m'}\int \frac{d^3P}{P_0}\,\Gamma^{n\rho}_{sm,sm'}(b(p))\,\psi^s_{m'}(p). \quad (29)$$

This equation gives us the recipe for calculating the projection of the wavefunction onto any of the irreducible subspaces $L_{n\rho}$. It is particularly important to us because our expansion of the wavefunction will contain the $A^{n\rho}_{sm}$ as expansion coefficients.

To obtain the expansion of the wavefunction we proceed in a less direct fashion than we did above. Using the function $Q^{n\rho}(l)$ defined by Eq. (19) we form the function

$$\{P^s_{m\alpha}Q^{n\rho}\}(l) \equiv \int du'\, D^s_{m\alpha}(u')\,Q^{n\rho}(lu') \quad (30)$$

which is covariant with respect to SU(2). In this, du' is an invariant measure over SU(2), and the $D^s_{m\alpha}(u')$ are the usual SU(2) matrix elements. The index α will again be fixed to the value $-n/2$ by an integration over the z-axis rotations.

Using the wavefunction $\psi^s_m(p)$ we can define another function

$$\chi^s_m(l) \equiv \sum_{m'} D^s_{mm'}(u^{-1})\,\psi^s_{m'}(p) \quad (31)$$

with $l = b(p)u$. This function is also covariant with respect to SU(2).

Now χ^s_m and $\{P^s_{m\alpha} Q^{n\rho}\}$ both transform according to the regular representation of SL(2,c) and they are both covariant with respect to the sth unirrep of SU(2). Hence, $\chi^s_m(b(p))$ and $\{P^s_{m\alpha} Q^{n\rho}\}$ $(b(p))$ both belong to the unitary, reducible subspace L_s defined by Eq. (17) so that a mapping from one to the other must exist.

The covariance properties of χ^s_m and $\{P^s_{m\alpha} Q^{n\rho}\}$ with respect to SU(2) and the Lorentz covariance of the mapping restrict the mapping to the form

$$\chi^s_m(b(p)) = \sum_n \int d\rho\, \alpha^{n\rho}\{P^s_{m\alpha} Q^{n\rho}\}(b(p)), \tag{32}$$

where the constants $\alpha^{n\rho}$ are the undetermined part of the kernel of the mapping.

Since the regular representation of SL(2,c) is unitary, the expansion (32) need only include the unitary, non-equivalent representations of the principal series. That is, $\sum_n \int d\rho$ is restricted to the range $n \in$ Integers, $\rho \in$ Reals with $\rho \in (0,\infty)$.

Using (30) and (31) in (32) we get the relation

$$\psi^s_m(p) = \sum_n \int d\rho\, \alpha^{n\rho} \int du'\, D^s_{m\alpha}(u') Q^{n\rho}(b(p)u') \tag{33}$$

between the wavefunction and its irreducible components. This relation is fundamental, and from it we shall obtain the expansion of the wavefunction in terms of the matrix elements (13). In the Appendix we shall derive the Shapiro transformation from Eq. (33).

To evaluate the u' integration in Eq. (33) we parametrize u' as $u' = u(z', z'^*)\tau' = z'k(z', z'^*)\tau'$, where τ' and $k(z', z'^*)$ are given by Eqs. (22) and (24), respectively. This parametrization transforms $b(p)u' = zk$ (which defines z and k) into $b^{-1}(p)z = z'k_w$ with $k_w = k(z', z'^*)\tau' k^{-1}$ so that (33) becomes

$$\psi^s_m(p) = \sum_n \int d\rho\, \alpha^{n\rho} \int du(z', z'^*)\, d\tau' D^s_{m\alpha}(u(z', z'^*)\tau') \times \Delta^{n\rho}(\tau'^{-1} k^{-1}(z', z'^*) k_w)\, Q^{n\rho}(z). \tag{34}$$

To write this equation we have utilized the covariance of $Q^{n\rho}$.

The τ' integration can be performed directly, giving $2\pi\delta_{\alpha,-n/2}$, and the parametrization $u' = u(z', z'^*)\tau'$ gives

$$du(z', z'^*) = \frac{dz'\, dz'^*}{4\pi^2 i(1 + z'^* z')^2} \tag{35}$$

by the usual method. Using this, along with Eqs. (25), (10), and the expansion (15) for $Q^{n\rho}$ we can bring (34) into the form

$$\psi^s_m(p) = \sum_n \int d\rho (2\pi i)^{-1} \alpha^{n\rho} \sum_{s'} \sum_{m'} A^{n\rho}_{s'm'} \int dz'\, dz'^* \times \phi^{*n\rho}_{sm}(z', z'^*) \Delta^{n\rho}(k_w) \phi^{n\rho}_{s'm'}(z, z^*), \tag{36}$$

with $b(p)z' = zk_w^{-1}$.

Finally, invoking Eq. (13) for the matrix elements [with $\mathrm{Im}(\rho) = 0$] gives the desired expansion

$$\psi^s_m(p) = \sum_n \int d\rho (2s+1)^{-1} \alpha^{n\rho} \sum_{s'} \sum_{m'} \Gamma^{n\rho}_{sm,s'm'}(b^{-1}(p))\, A^{n\rho}_{s'm'}. \tag{37}$$

This expansion, as well as the expression (29) for the expansion coefficients, is Lorentz covariant. This is easily demonstrated using Eqs. (16) and (17).

To calculate the $\alpha^{n\rho}$ we eliminate $\psi^s_m(p)$ between Eqs. (29) and (37). This gives the orthogonality condition

$$\sum_{m''} \int \frac{d^3P}{P_0} \left[\frac{\alpha^{n\rho}}{4\pi m_0(2s+1)}\right] \Gamma^{n\rho}_{sm'',s'm'}(b^{-1}(p)) \Gamma^{n'\rho'}_{sm,sm''}(b(p)) = \delta_{nn'}\delta(\rho - \rho')\delta_{ss'}\delta_{mm'}, \tag{38}$$

from which a lengthy calculation leads to the result

$$\alpha^{n\rho} = (n^2 + \rho^2)/4\pi m_0. \tag{39}$$

For the details of this calculation see Ref. 2.

V. CONCLUSION

We have shown that by relating functions that transform according to representations of SL(2,c) which are induced by two different subgroups we can obtain an expansion of the wavefunction for a free, massive particle with spin s in terms of the matrix elements for the unirreps of the principal series of SL(2,c). This expansion, which is given explicitly by Eqs. (37) and (39), is obtained by means of a mapping in the space of SU(2)-covariant functions. Its inverse, Eq. (29) for the expansion coefficients, is obtained by projecting out that part of an SU(2)-covariant function which is covariant with respect to the subgroup K. In the Appendix we have shown that these equations are equivalent to the Shapiro integral transformation [Eqs. (A6) and (A8)] for the wavefunction.

Our equations have several advantages over the conventional form of the Shapiro transformation. One advantage is that our equations do not involve an integration over SU(2). Another advantage is that they are expressed in terms of more concrete quantities than are Eqs. (A6) and (A8); namely, the matrix elements as opposed to the elements of the irreducible spaces.

These things lend our equations a very compact and symmetric appearance with the advantage that all of the momentum dependence is grouped together in the matrix elements.

ACKNOWLEDGMENTS

The author wishes to thank Dr. John R. Higgins for countless discussions and many useful suggestions concerning several fundamental aspects of this work.

APPENDIX

In Sec. IV, we obtained the expansion (37) and its inverse (29) from Eqs. (33) and (23), respectively. Here we will show that Eqs. (33), (23), and (39) comprise the Shapiro integral transformation for the wavefunction of a free, massive particle with spin s.

We begin with Eq. (33) in which we let $b(p)u' = uk$, which defines u and k to within a rotation of the form (22). This gives

$$\psi^s_m(p) = \sum_n \int d\rho \left[\frac{n^2 + \rho^2}{4\pi m_0}\right] \int du'\, D^s_{m\alpha}(u') Q^{n\rho}(uk), \tag{A1}$$

where we have used Eq. (39) for the $\alpha^{n\rho}$. We now utilize the covariance of $Q^{n\rho}$ with respect to the subgroup K, and we set $\alpha = -n/2$ to obtain

$$\psi^s_m(p) = \sum_n \int d\rho \left[\frac{n^2+\rho^2}{4\pi m_0}\right] \int du' \, D^s_{m,-n/2}(u') \Delta^{n\rho}(k^{-1}) Q^{n\rho}(u). \quad \text{(A2)}$$

The relation $b(p)u' = uk$ gives the results

$$du' = \{(P_0 - \mathbf{P}\cdot\mathbf{N})/m_0\}^{-2} du \quad \text{(A3)}$$

and

$$\Delta^{n\rho}(k^{-1}) = \exp(in\tau/2)\{(P_0 - \mathbf{P}\cdot\mathbf{N})/m_0\}^{1-(i\rho/2)} \quad \text{(A4)}$$

with $u' = \underline{u}\tau$ and with

$$\mathbf{N} \equiv \{-2\mathrm{Re}(uv),\ 2\mathrm{Im}(uv),\ u^*u - v^*v\} \quad \text{(A5)}$$

being the unit vector obtained by rotating the z-axis vector $\{0, 0, 1\}$ with the rotation

$$u = \begin{pmatrix} u & v \\ -v^* & u^* \end{pmatrix} \in \mathrm{SU}(2).$$

Using these results, Eq. (A2) is readily manipulated into the form

$$\psi^s_m(p) = \sum_n \int d\rho \left[\frac{n^2+\rho^2}{4\pi m_0}\right] \int du \left\{\frac{(P_0 - \mathbf{P}\cdot\mathbf{N})}{m_0}\right\}^{-1-(i\rho/2)} \times D^s_{m,-n/2}(\underline{u})\, Q^{n\rho}(u). \quad \text{(A6)}$$

To obtain the inverse to this we set $l = u$ in Eq. (23) to get

$$Q^{n\rho}(u) = -(m_0/\pi)\sum_m \int d_l\underline{k}\, \Delta^{n\rho}(\underline{k})\, D^s_{-n/2,m}(\underline{u}^{-1})\psi^s_m(p) \quad \text{(A7)}$$

with $b(p)\underline{u} = u\underline{k}$. Utilizing Eq. (A4), along with

$$d_l\underline{k} = \frac{-1}{4m_0^2}\frac{d^3P}{P_0},$$

we bring (A7) into the desired form

$$Q^{n\rho}(u) = \frac{1}{4\pi m_0}\sum_m \int \frac{d^3P}{P_0}\left\{\frac{(P_0 - \mathbf{P}\cdot\mathbf{N})}{m_0}\right\}^{(i\rho/2)-1} \times D^s_{-n/2,m}(\underline{u}^{-1})\psi^s_m(p). \quad \text{(A8)}$$

Equation (A6), together with its inverse (A8), comprise the Shapiro integral transformation in a form similar to that given in Ref. 3.

[1] I. S. Shapiro, dissertation, Moscow State University (1955); Dokl. Akad. Nauk SSSR **106**, 647 (1956) [Sov. Phys.-Dokl. **1**, 91 (1956)].

[2] Chou Kuang-Chao and L. G. Zastavenko, Zh. Eksp. Teor. Fiz. **35**, 1417 (1958) [Sov. Phys. JETP **8**, 990 (1959)].

[3] V. S. Popov, Zh. Eksp. Teor. Fiz. **37**, 1116 (1959) [Sov. Phys. JETP **10**, 794 (1960)].

[4] I. M. Gel'fand and M. A. Naimark, J. Phys. (USSR) **10**, 93 (1946).

[5] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).

[6] G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).

[7] S. A. Gaal, *Linear Analysis and Representation Theory* (Springer-Verlag, New York, 1973), see Proposition 3 on p. 350.

[8] Y. Frishman and C. Itzykson, Phys. Rev. **180**, 1556 (1969), Appendix B.

[9] By p we mean the 2×2 Hermitian matrix

$$p \equiv \begin{pmatrix} P_0 + P_3 & P_1 - iP_2 \\ P_1 + iP_2 & P_0 - P_3 \end{pmatrix}$$

with the 4-momentum P_μ being given by $P_0 = E$ and $\{P_i\} = -\mathbf{P}$. In this, E is the energy of the particle and P its spatial momentum in units with $c = \hbar = 1$.

[10] E. P. Wigner, Ann. Math. **40**, 149 (1939).

The interaction of the gravitational and electromagnetic fields

David Lovelock

Department of Mathematics, University of Arizona, Tucson, Arizona 85721
and Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada
(Received 5 August 1977)

Various sets of conditions are presented which a reasonable theory might be expected to satisfy, in attempting to explain the interaction of the gravitational and electromagnetic fields. It is shown that each of these sets leads inevitably to the Einstein–Maxwell field equations. Attention is also drawn to the fact that these equations must be modified if it is furthermore demanded that Maxwell's equation in flat space–time be an exact solution of them.

1. INTRODUCTION

It is well known that, in regions devoid of sources, the interaction of the gravitational and electromagnetic fields (characterized by a metric tensor g_{ij} and a covariant vector ψ_i, respectively) is assumed to be governed by the Einstein—Maxwell field equations

$$aG^{ij} = b[F^{ih}F^j{}_h - \tfrac{1}{4}g^{ij}(F^{rs}F_{rs})] \tag{1.1}$$

and

$$F^{hj}{}_{|j} = 0, \tag{1.2}$$

where

$$F_{ij} = \psi_{i,j} - \psi_{j,i}, \tag{1.3}$$

and $a, b,$ are constants. Here G^{ij} is the Einstein tensor, the vertical bar and comma denoting covariant and partial differentiation respectively. It is also well known[1] that the identity (1.3) is equivalent to

$$\epsilon^{habc}F_{ab|c} = 0, \tag{1.4}$$

so that the full set of the Einstein—Maxwell equations is either (1.1)—(1.3) or (1.1), (1.2), and (1.4). This system of equations is to be solved for the pair (g_{ij}, F_{ab}) subject to appropriate boundary conditions.

In flat space—time, i.e., when

$$(g_{ij}) = (\eta_{ij}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1.5}$$

it is obvious that (1.2) and (1.3) [or (1.4)] reduce precisely to Maxwell's equations, these being experimentally confirmed to a high degree of accuracy. Indeed, it is for this reason that (1.2)—(1.4) have been proposed as acceptable field equations in curved space—time.[2] Although the justification for (1.1) does not enjoy a similar firm experimental foundation, there exist a variety of mathematical arguments[3] in support of it, perhaps the primary one being based on the identity

$$[aG^{ij} - b(F^{ih}F^j{}_h - \tfrac{1}{4}g^{ij}F^{rs}F_{rs})]_{|j} = bF^i{}_hF^{hj}{}_{|j} + (b/4)g^{ij}\epsilon_{hjke}F^{ke}\epsilon^{habc}F_{ab|c}, \tag{1.6}$$

which clearly indicates that the divergence of (1.1) vanishes whenever Maxwell's equations (1.2) and (1.4) [or (1.3)] are satisfied.[4]

However, although it is true that if $g_{ij} = \eta_{ij}$ then (1.2) and (1.3) reduce to Maxwell's equations in flat space—time, it is also true that *the system of equations* (1.1)—(1.3) *does not reduce to Maxwell's equations in flat space—time unless* $b = 0$. More precisely, the pair (η_{ij}, F_{ab}), where F_{ab} is a nonzero solution of Maxwell's equations in flat space—time, is an *exact* solution of (1.1)—(1.3) if and only if $b = 0$. [This is verified by substituting (1.5) in (1.1) and deducing that

$$b[F^{ih}F^j{}_h - \tfrac{1}{4}\eta^{ij}(F^{rs}F_{rs})] = 0,$$

from which it can easily be shown[5] that either $b = 0$ or $F_{ij} = 0$.] Consequently, if we demand that the experimentally verified Maxwell's equations be an exact solution of (1.1)—(1.3) in flat space—time, then (1.1) reduces to

$$G^{ij} = 0, \tag{1.7}$$

and the electromagnetic-gravitational interaction would then be governed by (1.7), (1.2), and (1.4), i.e., there would be no electromagnetic contribution to the gravitational field equations. *The set of Eqs.* (1.7), (1.2), *and* (1.4) *we shall call the modified Einstein—Maxwell equations.*

In this paper we seek alternative equations to (1.1)—(1.4), which are to be obtained from reasonable assumptions. Here five different approaches are taken to this problem, and we arrive at the same conclusion, viz., *reasonable assumptions lead inevitably to the Einstein—Maxwell equations* (with cosmological term). Consequently, if to these reasonable assumptions is added the condition that the corresponding field equations should admit Maxwell's equations in flat space—time as an exact solution, we would be forced to consider the modified Einstein—Maxwell equations.[6] As far as we are aware, there is no classical experimental evidence, either in support of or in conflict with, the Einstein—Maxwell equations, whether modified or not.

2. APPROACH 1: ALTERNATIVES TO (1.1), RETAINING (1.2) AND ASSUMING (1.3)

In this approach we shall seek an alternative equation with which to replace (1.1), while still retaining (1.2) and (1.3). In order to attempt to find such an equation, to be provisionally denoted by

$$B^{ij} = 0, \tag{2.1}$$

 0022-2488/78/1903-0586$1.00

we shall be concerned with the following problem. To find all tensor densities B^{ij} satisfying:

(a) B^{ij} is a concomitant of g_{ab} (and its first and second partial derivatives) together with F_{ab}, i.e.,

$$B^{ij}=B^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab}), \tag{2.2}$$

where F_{ab} is defined by (1.3);

(b) B^{ij} is symmetric, i.e.,

$$B^{ij}=B^{ji}; \tag{2.3}$$

(c) The divergence of B^{ij} vanishes whenever Maxwell's equations (1.2) and (1.3) are satisfied in the sense that

$$B^{ij}{}_{|j}=\alpha^{ih}F_h{}^j{}_{|j}, \tag{2.4}$$

where α^{ih} is an unspecified tensor density for which

$$\alpha^{ih}=\alpha^{ih}(g_{ab};g_{ab,c};g_{ab,cd};g_{ab,cde};F_{ab}). \tag{2.5}$$

The motivation behind (2.2) is (1.1), while (2.3) is motivated by the fact that the "Einstein equation" (2.1) is usually assumed to be symmetric. Condition (2.4) is motivated by comparison with (1.6), while (2.5) is suggested by (2.2) and (2.4).

We introduce the two tensor densities

$$B^{ij;ab,cd}=\frac{\partial B^{ij}}{\partial g_{ab,cd}}, \quad\text{and}\quad B^{ij;ab}=\frac{\partial B^{ij}}{\partial F_{ab}}, \tag{2.6}$$

which will then satisfy the following identities[7]:

$$B^{ij;ab,cd}=B^{ji;ab,cd}=B^{ij;ba,cd}=B^{ij;ab,dc}=B^{ij;cd,ab},$$

$$B^{ij;ab,cd}+B^{ij;ad,bc}+B^{ij;ac,db}=0,$$

$$B^{ij;ab}=B^{ji;ab}=-B^{ij;ba}.$$

These identities will be used frequently in the sequel without specific mention.

When written out in full, Eq. (2.4) reads

$$B^{ij;ab,cd}g_{ab,cdj}+2B^{ij;ab}\psi_{a,bj}+\frac{\partial B^{ij}}{\partial g_{ab,c}}g_{ab,cj}+\frac{\partial B^{ij}}{\partial g_{ab}}g_{ab,j}+\Gamma^i_{hj}B^{hj}$$

$$=\alpha^{ih}g^{jk}[\psi_{h,kj}-\psi_{k,hj}-\Gamma^a_{hj}F_{ak}-\Gamma^a_{kj}F_{ha}]. \tag{2.7}$$

If we differentiate (2.7) with respect to $\psi_{r,st}$ we find

$$B^{it;rs}+B^{is;rt}=\alpha^{ir}g^{st}-\tfrac{1}{2}\alpha^{is}g^{rt}-\tfrac{1}{2}\alpha^{it}g^{rs}. \tag{2.8}$$

In (2.8) we interchange i with s, and i with t to obtain

$$B^{st;ri}+B^{si;rt}=\alpha^{sr}g^{it}-\tfrac{1}{2}\alpha^{si}g^{rt}-\tfrac{1}{2}\alpha^{st}g^{ri} \tag{2.9}$$

and

$$B^{it;rs}+B^{ts;ri}=\alpha^{tr}g^{si}-\tfrac{1}{2}\alpha^{ts}g^{ri}-\tfrac{1}{2}\alpha^{ti}g^{rs}. \tag{2.10}$$

Adding (2.8) and (2.9) and subtracting (2.10), we see that

$$2B^{is;rt}=\alpha^{ir}g^{st}+\alpha^{sr}g^{it}-\alpha^{tr}g^{si}+\tfrac{1}{2}(\alpha^{ti}-\alpha^{it})g^{rs}+\tfrac{1}{2}(\alpha^{ts}-\alpha^{st})g^{ri}-D^{is}g^{rt}, \tag{2.11}$$

where

$$D^{is}=\tfrac{1}{2}(\alpha^{is}+\alpha^{si}).$$

In view of the fact that the left-hand side of (2.11) is skew-symmetric in rt we thus have

$$D^{ir}g^{st}+D^{sr}g^{it}-2D^{tr}g^{si}+D^{it}g^{rs}+D^{st}g^{ri}-2D^{is}g^{rt}=0,$$

from which it is easily shown that

$$D^{is}=0,$$

i.e.,

$$\alpha^{is}=-\alpha^{si}. \tag{2.12}$$

From (2.11) and (2.12) we thus find

$$2B^{is;rt}=\alpha^{ir}g^{st}+\alpha^{sr}g^{it}+\alpha^{rt}g^{si}+\alpha^{ti}g^{rs}+\alpha^{ts}g^{ri}. \tag{2.13}$$

If we multiply (2.13) by g_{ri} we see that

$$\alpha^{ts}=\tfrac{2}{3}B^{is;rt}g_{ir}, \tag{2.14}$$

from which, by (2.2) and (2.5), we conclude that α^{ij} is independent of $g_{ab,cde}$, i.e.,

$$\alpha^{ij}=\alpha^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab}). \tag{2.15}$$

By virtue of the fact that

$$B^{is;rt;ab}=\frac{\partial(B^{is;rt})}{\partial F_{ab}}=B^{is;ab;rt},$$

it is easily established from (2.13) that

$$\alpha^{ir;ab}g^{st}+\alpha^{sr;ab}g^{it}+\alpha^{rt;ab}g^{si}+\alpha^{ti;ab}g^{rs}+\alpha^{ts;ab}g^{ri}$$

$$=\alpha^{ia;rt}g^{sb}+\alpha^{sa;rt}g^{ib}+\alpha^{ab;rt}g^{si}+\alpha^{bi;rt}g^{as}+\alpha^{bs;rt}g^{ai}, \tag{2.16}$$

where

$$\alpha^{ir;ab}=\frac{\partial\alpha^{ir}}{\partial F_{ab}}.$$

If we introduce the tensor density

$$\mu^{br}=\alpha^{bs;rt}g_{st}$$

and multiply (2.16) by g_{st}, we find

$$3\alpha^{ir;ab}+\alpha^{ia;br}+\alpha^{bi;ar}+\alpha^{ab;ir}=\mu^{br}g^{ai}-\mu^{ar}g^{ib}. \tag{2.17}$$

By multiplying (2.17) by g_{ib} and g_{rb} we may conclude that

$$\mu^{ai}=\lambda g^{ai}, \tag{2.18}$$

where λ is a scalar density and

$$\lambda=\lambda(g_{ab};g_{ab,c};g_{ab,cd};F_{ab}).$$

When (2.18) is substituted in (2.17) we see that

$$3\alpha^{ir;ab}+\alpha^{ia;br}+\alpha^{bi;ar}+\alpha^{ab;ir}=\lambda(g^{br}g^{ai}-g^{ar}g^{ib}). \tag{2.19}$$

In (2.19) we cycle on $a\ b\ i$ to find

$$\alpha^{ir;ab}+\alpha^{ar;bi}+\alpha^{br;ia}=-(\alpha^{ia;br}+\alpha^{bi;ar}+\alpha^{ab;ir}). \tag{2.20}$$

If we substitute the right-hand side of (2.20) into the left-hand side of (2.19), we then have

$$2\alpha^{ir;ab}-\alpha^{ar;bi}-\alpha^{br;ia}=\lambda(g^{br}g^{ai}-g^{ar}g^{ib}). \tag{2.21}$$

In (2.21) we interchange a and i, and add the resulting equation to (2.21) to find

$$3(\alpha^{ar;ib}+\alpha^{ir;ab})=\lambda(2g^{br}g^{ai}-g^{ar}g^{bi}-g^{ir}g^{ba}). \tag{2.22}$$

In (2.22) we interchange i and r and subtract the resulting equation from (2.22) to find

$$2\alpha^{ir;ab}-\alpha^{ar;bi}-\alpha^{ai;rb}=\lambda(g^{br}g^{ai}-g^{ar}g^{bi}).$$

A comparison of the latter equation with (2.21) shows that

$$\alpha^{ai;br}=\alpha^{br;ai},$$

which by (2.20), implies

$$\alpha^{ia;br}+\alpha^{bi;ar}+\alpha^{br;ia}=0. \tag{2.23}$$

When (2.23) is applied to (2.19) we have

$$\alpha^{ir;ab}=\tfrac{1}{3}\lambda(g^{br}g^{ai}-g^{ar}g^{ib}). \tag{2.24}$$

In view of the fact that

$$\alpha^{ir;ab;cd}=\frac{\partial(\alpha^{ir;ab})}{\partial F_{cd}}=\alpha^{ir;cd;ab},$$

it is easy to show that (2.24) implies that

$$\frac{\partial\lambda}{\partial F_{ab}}=0,$$

i.e.,

$$\lambda=\lambda(g_{ab};g_{ab,c};g_{ab,cd}). \tag{2.25}$$

By virtue of (2.25), Eq. (2.24) may be integrated to yield

$$\alpha^{ir}=\tfrac{2}{3}\lambda F^{ir}+\beta^{ir}, \tag{2.26}$$

where β^{ir} is a tensor density satisfying

$$\beta^{ir}=-\beta^{ri}, \tag{2.27}$$

and

$$\beta^{ir}=\beta^{ir}(g_{ab};g_{ab,c};g_{ab,cd}). \tag{2.28}$$

We now return to (2.7) and differentiate it with respect to $g_{ab,cdj}$, noting (2.15), to find

$$B^{ij;ab,cd}+B^{id;ab,jc}+B^{ic;ab,dj}=0. \tag{2.29}$$

The latter equation, together with (2.3), implies in the usual way,[8] that

$$B^{ij;ab,cd}=B^{cd;ab,ij}$$

and

$$B^{ij;ab,cd}=B^{ij;ab,cd}(g_{rs};F_{rs}). \tag{2.30}$$

From (2.29) it is clear that

$$(B^{ij;rs;tu;ab,cd}+B^{id;rs;tu;ab,jc}+B^{ic;rs;tu;ab,dj})g_{ir}=0,$$

which, by (2.14), (2.25), (2.26), and (2.28), implies that[9]

$$\lambda=a\sqrt{g}, \tag{2.31}$$

where a is a constant. From (2.29) it is also clear that

$$(B^{ij;rs;ab,cd}+B^{id;rs;ab,jc}+B^{ic;rs;ab,dj})g_{ir}=0,$$

which, by (2.14), (2.26), and (2.31), gives rise to

$$\beta^{sj;ab,cd}+\beta^{sd;ab,jc}+\beta^{sc;ab,dj}=0. \tag{2.32}$$

However, in view of (2.28) and the fact that β^{ij} is a tensor density, (2.32) is equivalent[10] to

$$\beta^{ij}{}_{|j}=0. \tag{2.33}$$

Elsewhere,[11] it has been shown that (2.27), (2.28), and (2.33) imply that

$$\beta^{ij}=0. \tag{2.34}$$

From (2.26), (2.31), and (2.34) we finally find

$$\alpha^{ij}=b\sqrt{g}\,F^{ij}, \tag{2.35}$$

where b is an arbitrary constant.

We now introduce the tensor density

$$A^{ij}=B^{ij}+b\sqrt{g}\,[F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}(F^{ab}F_{ab})]. \tag{2.36}$$

It is clear that A^{ij} is a tensor density for which

$$A^{ij}=A^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab}), \tag{2.37}$$

$$A^{ij}=A^{ji}, \tag{2.38}$$

and, by virtue of (2.4) and (2.35),

$$A^{ij}{}_{|j}=0. \tag{2.39}$$

Differentiation of (2.39) with respect to $\psi_{a,bc}$ [compare (2.2), (2.4), (2.7), (2.8) with (2.37) and (2.39)] yields

$$A^{it;rs}+A^{is;rt}=0,$$

which, together with (2.38), implies

$$A^{it;rs}=0,$$

i.e.,

$$A^{ij}=A^{ij}(g_{ab};g_{ab,c};g_{ab,cd}). \tag{2.40}$$

However, all tensor densities satisfying (2.38), (2.39), and (2.40) have been constructed,[12] the result being

$$A^{ij}=a\sqrt{g}\,G^{ij}+c\sqrt{g}\,g^{ij}, \tag{2.41}$$

where a, c are constants. A comparison of (2.41) with (2.36) establishes the following[13] theorem.

Theorem: The only tensor density which satisfies (2.2)–(2.5) is

$$B^{ij}=a\sqrt{g}\,G^{ij}+c\sqrt{g}\,g^{ij}-b\sqrt{g}\,[F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}(F^{rs}F_{rs})]. \tag{2.42}$$

Consequently, (2.1), (1.2), and (1.3) are the Einstein–Maxwell field equations (with cosmological term).

3. APPROACH 2: ALTERNATIVES TO (1.1), RETAINING (1,2) AND ASSUMING (1.3)

Some authors[14] have suggested that the energy-momentum tensor should be asymmetric which would be inconsistent with (2.3). Guided by this observation we shall again seek an alternative to (1.1) [while still retaining (1.2) and (1.3)], to be denoted by

$$C^{ij}=0, \tag{3.1}$$

where C^{ij} is a tensor density satisfying the following conditions:

(a) $C^{ij}=C^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab}),$ (3.2)

where F_{ab} is defined by (1.3);

(b) $C^{ij}{}_{|j}=\lambda^{ih}F_{h}{}^{j}{}_{|j}$ (3.3)

and

$$C^{ji}{}_{|j}=\mu^{ih}F_{h}{}^{j}{}_{|j}, \tag{3.4}$$

where λ^{ih}, μ^{ih} are tensor densities, unspecified except for the conditions

$$\lambda^{ih} = \lambda^{ih}(g_{ab}; g_{ab,c}; g_{ab,cd}; g_{ab,cde}; F_{ab}), \tag{3.5}$$

and

$$\mu^{ih} = \mu^{ih}(g_{ab}; g_{ab,c}; g_{ab,cd}; g_{ab,cde}; F_{ab}). \tag{3.6}$$

If we introduce the tensor densities:

$$B^{ij} = \tfrac{1}{2}(C^{ij} + C^{ji}), \quad D^{ij} = \tfrac{1}{2}(C^{ij} - C^{ji}),$$
$$\alpha^{ih} = \tfrac{1}{2}(\lambda^{ih} + \mu^{ih}), \quad \beta^{ih} = \tfrac{1}{2}(\lambda^{ih} - \mu^{ih}) \tag{3.7}$$

we see that (3.2)—(3.6) imply that B^{ij} and α^{ih} satisfy (2.1)—(2.5), in which case B^{ij} and α^{ih} are completely determined by the theorem of Sec. 2. The problem of determining C^{ij} satisfying (3.2)—(3.6) thus reduces to finding all D^{ij} satisfying

$$D^{ij} = D^{ij}(g_{ab}; g_{ab,c}; g_{ab,cd}; F_{ab}), \tag{3.8}$$

$$D^{ij} = -D^{ji}, \tag{3.9}$$

$$D^{ij}{}_{|j} = \beta^{ih} F_h{}^j{}_{|j}, \tag{3.10}$$

where β^{ih} is a tensor density and

$$\beta^{ih} = \beta^{ih}(g_{ab}; g_{ab,c}; g_{ab,cd}; g_{ab,cde}; F_{ab}). \tag{3.11}$$

In a manner similar to that from which (2.8) was obtained from (2.2)—(2.5), we find from (3.8)—(3.11) that

$$D^{ij;ab} + D^{ib;aj} = \beta^{ia} g^{bj} - \tfrac{1}{2}\beta^{ib} g^{ja} - \tfrac{1}{2}\beta^{ij} g^{ba}. \tag{3.12}$$

We now consider the tensor density E^{ijkh} defined by

$$E^{ijkh} = D^{ij;kh} + D^{ki;jh} + D^{hi;kj} + D^{jk;ih} + D^{jh;ki} + D^{kh;ij}. \tag{3.13}$$

By virtue of (3.2) and (3.9) it is easily seen that E^{ijkh} is totally skew-symmetric in $ijkh$, in which case

$$E^{ijkh} = \gamma \epsilon^{ijkh}, \tag{3.14}$$

where γ is a scalar and

$$\gamma = \gamma(g_{ab}; g_{ab,c}; g_{ab,cd}; F_{ab}). \tag{3.15}$$

However, (3.13) can also be expressed in the form

$$E^{ijkh} = 6D^{ij;kh} + (D^{ki;jh} + D^{hi;jk}) + 3(D^{hi;kj} + D^{ji;kh}) + (D^{jk;ih} + D^{ji;kh}) + (D^{jh;ki} + D^{ji;kh}) + (D^{kh;ij} + D^{ih;kj}).$$

To each of the terms in brackets on the right-hand side of the latter equation we apply (3.12) to find

$$6D^{ij;kh} = E^{ijkh} - \tfrac{1}{2} g^{hk}\beta^{ij} + \tfrac{1}{2} g^{jk}(\beta^{hi} - 4\beta^{ih}) + \tfrac{5}{2} g^{jh}\beta^{ik} + \tfrac{1}{2} g^{ki}(3\beta^{jh} - 2\beta^{hj}) - \tfrac{3}{2}\beta^{jk} g^{hi} + \tfrac{1}{2}\beta^{hk} g^{ji}. \tag{3.16}$$

Multiplication of (3.16) by g_{jh}, attention being paid to (3.14), leads to

$$6g_{jh} D^{ij;kh} = (13\beta^{ik} + \beta^{ki} + \beta g^{ki})/2, \tag{3.17}$$

where

$$\beta = g_{ij}\beta^{ij} = \beta(g_{ab}; g_{ab,c}; g_{ab,cd}; g_{ab,cde}; F_{ab}). \tag{3.18}$$

However, from (3.12), we see that

$$6g_{jh} D^{ij;kh} = 9\beta^{ik},$$

which, when combined with (3.17), gives rise to

$$\beta^{ik} = \tfrac{1}{4}\beta g^{ik}.$$

From the latter, (3.14) and (3.16), we thus have

$$6D^{ij;kh} = \gamma \epsilon^{ijkh} + \tfrac{3}{4}\beta(g^{jh} g^{ik} - g^{jk} g^{hi}). \tag{3.19}$$

In view of (3.8) and the fact that

$$D^{ij;kh;ab} = D^{ij;ab;kh},$$

it is not difficult to show that (3.15), (3.18), and (3.19) imply

$$\beta = \beta(g_{ab}; g_{ab,c}; g_{ab,cd}) \tag{3.20}$$

and

$$\gamma = \gamma(g_{ab}; g_{ab,c}; g_{ab,cd}).$$

By virtue of (3.19) and (3.20), (3.10) now implies

$$D^{ij;ab,cd} + D^{id;ab,jc} + D^{ic;ab,dj} = 0,$$

which when applied to (3.19) and (3.20) gives

$$\beta = 8\kappa\sqrt{g}, \quad \alpha = 12\tau, \tag{3.21}$$

where κ, τ are arbitrary constants. Substitution of (3.21) in (3.19) and integration thus yields

$$D^{ij} = \tfrac{1}{6}\tau \epsilon^{ijke} F_{ke} + \tfrac{1}{4}\kappa\sqrt{g} F^{ij} + E^{ij}, \tag{3.22}$$

where E^{ij} is a tensor density,

$$E^{ij} = E^{ij}(g_{ab}; g_{ab,c}; g_{ab,cd}), \tag{3.23}$$

$$E^{ij} = -E^{ji}, \tag{3.24}$$

and

$$E^{ij}{}_{|j} = 0, \tag{3.25}$$

the latter following from (3.10), (3.19), (3.21), and (3.22). However, (3.23)—(3.25) imply[15] that

$$E^{ij} = 0,$$

in which case (3.22) reduces to

$$D^{ij} = 2\tau \epsilon^{ijkh} F_{kh} + 2\kappa\sqrt{g} F^{ij}. \tag{3.26}$$

We are now in a position to prove the following theorem.

Theorem: The only tensor density C^{ij} satisfying (3.2)—(3.4) is

$$C^{ij} = a\sqrt{g}\, G^{ij} + c\sqrt{g}\, g^{ij} - b\sqrt{g}[F^{ih}F^j{}_h - \tfrac{1}{4} g^{ij}(F^{rs}F_{rs})] + \tau \epsilon^{ijkh} F_{kh} + \kappa\sqrt{g} F^{ij}. \tag{3.27}$$

Furthermore, the field equations (3.1), (1.2), and (1.3) are equivalent to the Einstein—Maxwell field equations.

Proof: Equation (3.27) is an immediate consequence of (2.42), (3.7), and (3.26). Thus we need only show that (3.1), (1.2), and (1.3) are equivalent to the Einstein—Maxwell field equations.

From (3.7) we see that (3.1) is equivalent to

$$B^{ij} = 0 \tag{3.28}$$

and

$$D^{ij} = 0. \tag{3.29}$$

Equations (3.28), (1.2), and (1.3) are clearly equivalent to (1.1)—(1.3) so we restrict our considerations to (3.29), which, by (3.26) implies that

$$\tau\epsilon^{ijkh}F_{kh}=-\kappa\sqrt{g}F^{ij}. \tag{3.30}$$

If we multiply (3.30) by F_{ij} and observe the identity[16]

$$\epsilon^{ijkh}F_{kh}F_{ej}=\tfrac{1}{4}\delta^i_e(\epsilon^{abcd}F_{ab}F_{cd}),$$

we find

$$\tfrac{1}{4}\delta^i_e(\tau\epsilon^{abcd}F_{ab}F_{cd})=-\kappa\sqrt{g}F^{ij}F_{ej}, \tag{3.31}$$

from which we deduce that

$$\tau\epsilon^{abcd}F_{ab}F_{cd}=-\kappa\sqrt{g}F^{ab}F_{ab}.$$

Substitution of the latter in (3.31) implies that

$$\kappa\sqrt{g}[F^{ij}F_{ej}-\tfrac{1}{4}\delta^i_e(F^{ab}F_{ab})]=0,$$

from which we conclude that

$$\kappa=\tau=0$$

(if F_{ij} is not identically zero) in which case (3.29) is identically satisfied.

4. APPROACH 3: ALTERNATIVES TO (1.1), RETAINING (1.2) AND (1.4)

As has been pointed out elsewhere[17] the existence of magnetic monopoles would have a drastic effect on the equivalence of (1.3) and (1.4), because, in the presence of sources, (1.1), (1.2), and (1.4) are all augmented by appropriate source terms. Under these circumstances, (1.3) is no longer a consequence of the augmented (1.4), which immediately implies that we cannot infer the existence of a vector field ψ_i for which (1.3) is valid.

Guided by these comments we shall again seek an alternative to (1.1), to be denoted by

$$H^{ij}=0, \tag{4.1}$$

where H^{ij} is a tensor density satisfying the following conditions:

(a) $H^{ij}=H^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab})$, (4.2)

where F_{ab} is any antisymmetric tensor field, i.e.,

$$F_{ab}=-F_{ba}; \tag{4.3}$$

(b) $H^{ij}=H^{ji}$; (4.4)

(c) $H^{ij}{}_{|j}=\sqrt{g}\,\alpha^{ih}F_h{}^j{}_{|j}+\beta^i{}_h\epsilon^{hjab}F_{ab|j}$, (4.5)

where α^{ih}, $\beta^i{}_h$ are tensors and

$$\begin{aligned}\alpha^{ih}&=\alpha^{ih}(g_{ab};g_{ab,c};g_{ab,cd};g_{ab,cde};F_{ab}),\\ \beta^i{}_h&=\beta^i{}_h(g_{ab};g_{ab,c};g_{ab,cd};g_{ab,cde};F_{ab}).\end{aligned} \tag{4.6}$$

[Compare (4.5) with (1.6).]

Written out in detail, (4.5) reads

$$\begin{aligned}H^{ij;ab}F_{ab,j}+H^{ij;ab,cd}g_{ab,cdj}+\frac{\partial H^{ij}}{\partial g_{ab,c}}g_{ab,cj}\\+\frac{\partial H^{ij}}{\partial g_{ab}}g_{ab,j}-\Gamma^i_{hj}H^{hj}\\=\sqrt{g}\,\alpha^{ih}g^{jk}F_{hk|j}+\beta^i{}_h\epsilon^{hjab}F_{ab|j}.\end{aligned} \tag{4.7}$$

Differention of (4.7) with respect to $F_{ab,j}$ gives rise to

$$H^{ij;ab}=\tfrac{1}{2}\sqrt{g}\,\alpha^{ia}g^{bj}-\tfrac{1}{2}\sqrt{g}\,\alpha^{ib}g^{ai}+\beta^i{}_h\epsilon^{hjab}, \tag{4.8}$$

which, by (4.4) implies that

$$\begin{aligned}\sqrt{g}\,\alpha^{ia}g^{bj}-\sqrt{g}\,\alpha^{ib}g^{aj}+2\beta^i{}_h\epsilon^{hjab}\\=\sqrt{g}\,\alpha^{ja}g^{bi}-\sqrt{g}\,\alpha^{jb}g^{ai}+2\beta^j{}_h\epsilon^{hiab}.\end{aligned} \tag{4.9}$$

If we multiply (4.19) by g_{bj} we find

$$\sqrt{g}\,\alpha^{ia}=-\tfrac{1}{2}\sqrt{g}(g^{bj}\alpha_{bj})g^{ai}+\beta_{jh}\epsilon^{hiaj},$$

from which it follows that

$$\sqrt{g}\,\alpha^{ia}=\beta_{jh}\epsilon^{hiaj}. \tag{4.10}$$

This equation clearly implies that α^{ia} is skew-symmetric. We now return to (4.9) and multiply it by ϵ_{rjab} to obtain, by virtue of (4.10),

$$\beta^{ir}=-\beta^{ri}, \tag{4.11}$$

which, when taken in conjunction with (4.10), gives rise to

$$\beta_{rs}=-\tfrac{1}{4}\sqrt{g}\,\epsilon_{iars}\alpha^{ia}. \tag{4.12}$$

From (4.8) and (4.12) we thus find

$$\begin{aligned}H^{ij;ab}=\tfrac{1}{2}\sqrt{g}(\alpha^{ia}g^{bj}-\alpha^{ib}g^{aj}\\+\alpha^{bj}g^{ai}-\alpha^{ba}g^{ij}+\alpha^{ja}g^{bi}).\end{aligned} \tag{4.13}$$

This equation is formally the same as (2.13), and it is now possible to parallel the arguments presented in Sec. 2 to establish the following.

Theorem: The only tensor density which satisfies (4.2)–(4.6) is[18]

$$H^{ij}=a\sqrt{g}\,G^{ij}+c\sqrt{g}\,g^{ij}-b\sqrt{g}[F^{ih}F_{jh}-\tfrac{1}{4}\delta^i_j(F^{rs}F_{rs})]. \tag{4.14}$$

Consequently, (4.1), (1.2), and (1.4) are the Einstein–Maxwell field equations.

5. APPROACH 4: ALTERNATIVES TO (1.1), RETAINING (1.2) AND (1.4)

For the reasons mentioned in Secs. 3 and 4, we shall now solve the following problem. Find all tensor densities K^{ij} which satisfy

(a) $K^{ij}=K^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab})$, (5.1)

where F_{ab} satisfies (4.3);

(b) $K^{ij}{}_{|j}=\sqrt{g}\,\lambda^{ih}F_h{}^j{}_{|j}+\eta^i{}_h\epsilon^{hjab}F_{ab|j}$

and (5.2)

$$K^{ji}{}_{|j}=\sqrt{g}\,\mu^{ih}F_h{}^j{}_{|j}+\xi^i{}_h\epsilon^{hjab}F_{ab|j},$$

where λ^{ih}, $\eta^i{}_h$, μ^{ih}, $\xi^i{}_h$ are tensors, all functions of g_{ab} and its first three partial derivatives together with F_{ab}, e.g.,

$$\lambda^{ih}=\lambda^{ih}(g_{ab};g_{ab,c};g_{ab,cd};g_{ab,cde};F_{ab}).$$

If we introduce the tensors:

$$\begin{aligned}H^{ij}&=\tfrac{1}{2}(K^{ih}+K^{ji}),\quad J^{ij}=\tfrac{1}{2}(K^{ij}-K^{ji}),\\ \alpha^{ih}&=\tfrac{1}{2}(\lambda^{ih}+\mu^{ih}),\quad \beta^i{}_h=\tfrac{1}{2}(\eta^i{}_h+\xi^i{}_h),\\ \rho^{ih}&=\tfrac{1}{2}(\lambda^{ih}-\mu^{ih}),\quad \theta^i{}_h=\tfrac{1}{2}(\eta^i{}_h-\xi^i{}_h),\end{aligned} \tag{5.3}$$

we see that (5.1)–(5.3) imply that H^{ij}, α^{ih}, $\beta^i{}_h$ satisfy (4.2)–(4.6), in which case they are completely determined by the theorem of Sec. 4. The problem of determining K^{ij} satisfying (5.1), (5.2) thus reduces to finding all J^{ij} satisfying

$$J^{ij}=J^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab}), \tag{5.4}$$

$$J^{ij}=-J^{ji}, \tag{5.5}$$

and

$$J^{ij}{}_{|j}=\sqrt{g}\,\rho^{ih}F_h{}^j{}_{|j}+\theta^i{}_h\epsilon^{hjab}F_{ab|j}. \tag{5.6}$$

From (5.6) we find

$$J^{ic;ab}=\tfrac{1}{2}\sqrt{g}(\rho^{ia}g^{cb}-\rho^{ib}g^{ca})+\theta^i{}_h\epsilon^{hcab}, \tag{5.7}$$

which, by (5.5), gives rise to

$$\tfrac{1}{2}\sqrt{g}\;(\rho^{ia}g^{cb}-\rho^{ib}g^{ca}+\rho^{ca}g^{ib}-\rho^{cb}g^{ia})$$
$$=\theta^i{}_h\epsilon^{chab}+\theta^c{}_h\epsilon^{ihab}. \tag{5.8}$$

Multiplication of (5.8) by g_{cb} yields

$$4\sqrt{g}\rho^{ia}-\sqrt{g}\,(g_{cb}\rho^{cb})g^{ia}=2\theta_{bh}\epsilon^{ihab}, \tag{5.9}$$

from which we obtain

$$\tfrac{1}{2}\sqrt{g}\,\epsilon_{rsia}\rho^{ia}=(\theta_{rs}-\theta_{sr})/2. \tag{5.10}$$

However, by multiplying (5.8) by $g_{cs}\epsilon_{rbia}$, we find

$$\tfrac{1}{2}\sqrt{g}\,\epsilon_{rsia}\rho^{ia}=g_{sr}\theta^a{}_a-4\theta_{sr}. \tag{5.11}$$

A comparison of (5.10) and (5.11) yields

$$\theta_{sr}=\tfrac{1}{4}\theta^a{}_a g_{sr} \tag{5.12}$$

which, when substituted in (5.9) implies that

$$\rho^{ia}=\tfrac{1}{4}\rho^b{}_b g^{ia}. \tag{5.13}$$

If we substitute (5.12) and (5.13) in (5.7) we obtain an expression which is formally equivalent to (3.19). It is thus possible to parallel the arguments presented in Sec. 3 to establish the following.

Theorem: The only tensor density K^{ij} satisfying (5.1) and (5.2) is

$$K^{ij}=a\sqrt{g}G^{ij}+c\sqrt{g}\,g^{ij}-b\sqrt{g}[F^{ih}F^j{}_h$$
$$-\tfrac{1}{4}g^{ij}(F^{rs}F_{rs})]+\tau\epsilon^{ijkh}F_{kh}+\kappa\sqrt{g}F^{ij}, \tag{5.14}$$

where a, b, c, τ, κ, are constants. Furthermore, the field equations

$$K^{ij}=0$$

together with (1.2) and (1.4) are equivalent to the Einstein—Maxwell field equations.

6. APPROACH 5: ALTERNATIVES TO (1.1), (1.2) AND (1.4)

In this section we shall we shall find all tensors L^{ij}, A^i, B^i for which:

(a) $L^{ij}=L^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab})$,

$$A^i=A^i(g_{ab};g_{ab,c};g_{ab,cd};F_{ab|c}), \tag{6.1}$$

$$B^i=B^i(g_{ab};g_{ab,c};g_{ab,cd};F_{ab|c}),$$

where

$$F_{ab}=-F_{ba}; \tag{6.2}$$

(b) If $R_{ijkh}=0$, then $A^i=\sqrt{g}F^{ij}{}_{|j}$ and $B^i=\epsilon^{iabc}F_{ab|c}$; (6.3)

(c) $A^i{}_{|i}=0$ and $B^i{}_{|i}=0$; (6.4)

(d) $L^{ij}{}_{|j}=\alpha^i{}_hA^h+\beta^i{}_hB^h$, $L^{ji}{}_{|j}=\gamma^i{}_hA^h+\lambda^i{}_hB^h$, (6.5)

where $\alpha^i{}_h$, $\beta^i{}_h$, $\gamma^i{}_h$, $\lambda^i{}_h$ are tensors and are functions of g_{rs} and F_{rs}. The source-free field equations are then assumed to be of the form:

$$L^{ij}=0,\quad A^i=0,\quad B^i=0. \tag{6.6}$$

Condition (b) is motivated by the experimentally accepted validity of Maxwell's equations in special relativity. Condition (c) is motivated by, and interpreted as, conservation of charge, both electric and magnetic. Condition (d) is motivated by the requirement that the divergence of L^{ij} should vanish whenever "Maxwell's equations" $A^i=0$, $B^i=0$, are satisfied [compare (6.5) with (1.6)].

If we define

$$M^{ij}=\tfrac{1}{2}(L^{ij}+L^{ji})=M^{ji}, \tag{6.7}$$

$$N^{ij}=\tfrac{1}{2}(L^{ij}-L^{ji})=-N^{ji}, \tag{6.8}$$

$$a^i{}_h=\tfrac{1}{2}(\alpha^i{}_h+\gamma^i{}_h),\quad b^i{}_h=\tfrac{1}{2}(\beta^i{}_h+\lambda^i{}_h),$$
$$c^i{}_h=\tfrac{1}{2}(\alpha^i{}_h-\gamma^i{}_h),\quad d^i{}_h=\tfrac{1}{2}(\beta^i{}_h-\lambda^i{}_h), \tag{6.9}$$

then (6.5) implies

$$M^{ij}{}_{|j}=a^i{}_hA^h+b^i{}_hB^h, \tag{6.10}$$

$$M^{ij}{}_{|j}=c^i{}_hA^h+d^i{}_hB^h, \tag{6.11}$$

while (6.6) is equivalent to

$$M^{ij}=0,\quad A^i=0,\quad B^i=0, \tag{6.12}$$

and

$$N^{ij}=0. \tag{6.13}$$

It is known[19] that (6.3), (6.4), (6.7), and (6.10) imply that

$$A^i=\sqrt{g}\,F^{ij}{}_{|j},\quad B^i=\epsilon^{iabc}F_{ab|c}, \tag{6.14}$$

while (6.12) are precisely the Einstein—Maxwell field equations. Furthermore, by applying the analysis of the previous section, we see that (6.8), (6.11), and (6.14) imply that (6.13) gives no further conditions. We thus have the following theorem.

Theorem: If conditions (6.1)—(6.5) are satisfied, then (6.6) are precisely the Einstein—Maxwell field equations (1.1), (1.2), and (1.4).

ACKNOWLEDGMENT

I would like to thank my colleagues, Hanno Rund and Marlan O. Scully, for valuable discussions. This work was supported in part by NSF MCS 76-06680 and NRC A7342.

[1]See, for example, D. Lovelock and H. Rund, *Tensors, Differential Forms and Variational Principles* (Wiley-Interscience, New York, 1975).

[2]A. Trautman, "Foundations and Current Problems of General Relativity" in *Lectures on General Relativity* (Prentice-Hall, Englewood Cliffs, New Jersey, 1964).

[3]D. Lovelock, Proc. Royal Soc. Lond. A **341**, 285 (1974); Gen. Rel. Grav. **5**, 399 (1974); Quaest. Math. **1**, 101 (1976); J. Math. Phys. **18**, 1491 (1977).

[4]J. L. Synge, "Introduction to General Relativity" in *Relativity, Groups and Topology* (Gordon and Breach, New York, 1964), p. 87.

[5]See Ref. 1, p. 326. This conclusion is not affected if the cosmological term is included in the left-hand side of (1.1).

[6]Some comments on the consequences of adopting the modified Einstein—Maxwell equations are in order. Firstly, (1.7) does not imply that there should never be an energy-momentum tensor on the right-hand side of (1.7), merely that in the electromagnetic-gravitational interaction case the conventional electromagnetic energy-momentum tensor is absent. Secondly, there is a solution of the modified Einstein—Maxwell equations which could correspond to a spherically symmetric, charged, massive body at rest at the origin [and would be the counterpart of the Reissner—Nordstrom solution of (1.1)—(1.3)], viz., the Schwarzschild solution with the classical ψ_i. Thirdly, the modified Einstein—Maxwell equations would imply that the trajectory of an uncharged test particle is unaffected by the charge of the source, in marked contrast to the situation predicted by the Einstein—Maxwell equations. Finally, if the modified Einstein—Maxwell equations are correct, then attempts at a unified field theory would have to be abandoned.

[7]See Ref. 1, Chap. 8.

[8]See Ref. 7.

[9]D. Lovelock, Arch. Ration. Mech. Anal. 33, 54 (1969).

[10]D. Lovelock, Aequat. Math. 4, 127 (1970).

[11]D. Lovelock, J. Math. Phys. 13, 874 (1972).

[12]See Ref. 11.

[13]This theorem could also be established by appealing to the significant work of I.M. Anderson, "Mathematical Foundations of the Einstein Field Equations," unpublished Ph.D. thesis, University of Arizona, 1976.

[14]O. Costa de Beauregard, "Translational Inertia Spin Effect," in *Perspectives in Geometry and Relativity, Hlavaty Festschrift* (University of Indiana Press, Bloomington, 1966), p. 44; D.W. Sciama, Proc. Cambridge Philos. Soc. **54**, 72 (1958).

[15]See Ref. 11.

[16]See Ref. 1, p. 326.

[17]J. Schwinger, Phys. Rev. D **12**, 3105 (1975).

[18]This theorem is also true if (4.2) is replaced by $H^{ij} = H^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab};F_{ab,c})$.

[19]See Ref. 3, fourth reference.

Multidimensional wave radiation from a source with Gaussian time variation and Gaussian-approximated distribution about a spherical sheet

Lim Chee-Seng

Department of Mathematics, University of Singapore, Singapore
(Received 16 February 1977)

This paper concerns the wave field of a source with the title-indicated space–time function which, additionally, possesses an arbitrary directional variation. The multivariate solution obtained comprises an estimated error plus peak-induced spherical harmonics that are hyperconically confined, i.e., bounded by diverging and converging spherical fronts. Such fronts are not necessarily singular. Compliance with the radiation principle ensues, through contour integration, from Cauchy initial conditions. For an odd number of spatial dimensions, an inner zone created after a focusing phenomenon exhibits an analogy with a Petrowsky's lacuna. Naturally, the wave field varies with direction, but only because its source does so. Spherically as well as axially symmetric cases constitute major corollaries. Asymptotic developments, evolving ultimately into steady limits, are also deducible. An indirect application is illustrated for magnetoacoustic flow parallel to a magnetic field; on induction by a cylindrical Gaussian-approximated current distribution, weak effects appear everywhere during the steady state and are superposed upon strong stationary wave effects bounded by cone sheets which project either (i) downstream for a supersonic–super-Alfvénic flow, or (ii) upstream for a restricted subsonic–sub-Alfvénic flow. Finally, the main results are directly applied to elastic wave propagation from a two-component Gaussian body force concentrated about a spherical base; a spherically symmetric radial component generates a strong irrotational wave field normally involving an instantaneous point singularity; an axisymmetric azimuthal component generates a strong solenoidal wave field.

I. INTRODUCTION

Consider the Cauchy-type[1] radiation problem governed by the inhomogeneous wave equation

$$\phi_{tt}=\nabla^2\phi+(\tau\sqrt{\pi})^{-1}\exp(-t^2/\tau^2)f(\mathbf{x}), \tag{1.1}$$

together with zero initial conditions at instant $t=T$:

$$\phi(\mathbf{x},T)=0,\quad \phi_t(\mathbf{x},T)=0. \tag{1.2}$$

Here, the position vector $\mathbf{x}=(x_1,x_2,\ldots,x_n)\in R_n$, the infinite n-dimensional space whose Laplacian $\nabla^2=\partial^2/\partial x_1^2+\cdots+\partial^2/\partial x_n^2$. Unless otherwise specified, the integer $n\geq 2$. In accordance with (1.2), emission proceeds from a state of rest during which the radiating source is abruptly switched on when $t=T$, subsequently taken at $-\infty$. Thereafter, the time variation of the source is Gaussian, attaining its peak at $t=0$. This peak can be substantially raised and sharpened by making its time scale τ appropriately small; the present paper is essentially concerned with such a situation. Furthermore, the spatial distribution $f(\mathbf{x})$ of the source is originally defined to be a convolution, eventually interpretable as an approximate radial Gaussian with a high and sharp concentration about a spherical (circular if $n=2$) sheet $|\mathbf{x}|=r(>0)$. It is also directionally dependent on an arbitrary density factor.

To determine the scalar field ϕ during the unsteady phase, (1.1) and (1.2) are first combined into a single radiation equation within the class envisaged by Lighthill[2,3], i.e., incorporating a partial zero mode. Multiple Fourier synthesis is then applied.

The peaking of the source generates strong hyperconical fields, resolvable into spherical harmonics and physically contained by expanding as well as contracting concentric spherical fronts. Superposed is a weak estimable "error" effect traversing the entire $(\mathbf{x},t)$ hyperspace. Dominant quantities in each case represent the solution for the instantaneous impulse over the spherical sheet $|\mathbf{x}|=r$. Trailing terms arise from the Gaussian stretch beyond $|\mathbf{x}|=r$ and the peak instant $t=0$. The present paper effectively examines a situation wherein such an impulsive source sheet becomes dissipated in both space and time.

By certain modifications of the basic analysis, explicit solutions are deduced for magnetoacoustic flow past a cylindrical Gaussian-approximated current distribution; anisotropy in the magnetoacoustic wave system prevents a direct application to an initial value problem; hence a modified approach is necessary to accommodate a radiation condition in place of initial conditions. However a direct application is possible for an isotropic elastic medium.

II. FOURIER TRANSFORMATION

To tackle the problem posed, we first define the function

$$\Phi=\phi H(t-T)\equiv\phi\ \ (t>T),\quad 0(t<T), \tag{2.1}$$

where H denotes the Heaviside unit function. Thus, by a law of generalized functions,[4] we get, on multiplying (1.1) by $H(t-T)$ and using (1.2),

$$\Phi_{ss}=\nabla^2\Phi+(\tau\sqrt{\pi})^{-1}\exp[-(s+T)^2/\tau^2]f(\mathbf{x})H(s), \tag{2.2}$$

with $s=t-T$. We now introduce in $(\mathbf{x},s)$ hyperspace, the Fourier transform

$$\text{H.F.T.}[\Phi]=(2\pi)^{-n-1}\int_{R_n}\exp(-i\boldsymbol{\alpha}\cdot\mathbf{x})d\mathbf{x}\int_{-\infty}^{\infty}\Phi\exp(-i\omega s)ds, \tag{2.3}$$

 0022-2488/78/1903-0593$1.00

whose inverse

$$\Phi = \int_{R_n} \exp(i\boldsymbol{\alpha}\cdot\mathbf{x})d\boldsymbol{\alpha} \int_{-\infty-i\epsilon}^{\infty-i\epsilon} \text{H.F.T.}[\Phi]\exp(i\omega s)d\omega, \tag{2.4}$$

with the outer integral in (2.4) ranging, with respect to the wave vector $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\ldots,\alpha_n)$, over R_n whose typical volume element $d\boldsymbol{\alpha}=d\alpha_1 d\alpha_2\cdots d\alpha_n$; also, the scalar product $\boldsymbol{\alpha}\cdot\mathbf{x}=\alpha_1x_1+\cdots+\alpha_nx_n$. Since $\Phi\equiv 0$ when $s<0$, then regarding the ω path$(-\infty-i\epsilon,\infty-i\epsilon)$, one requires that for each $\boldsymbol{\alpha}\in R_n$ (see, e.g., Ref. 2, Appendix B and Ref. 3):

$$-\epsilon<\text{Im}\{\text{lowest complex } \omega\text{-singularity of H.F.T.}[\Phi]\}. \tag{2.5}$$

The H.F.T. of the source term in (2.2) involves, within the physical $\mathbf{x}$ space,

$$\text{F.T.}[f]=(2\pi)^{-n}\int_{R_n} f(\mathbf{x})\exp(-i\boldsymbol{\alpha}\cdot\mathbf{x})d\mathbf{x}, \tag{2.6}$$

combined with

$$(2\pi)^{-1}(\tau\sqrt{\pi})^{-1}\int_{-\infty}^{\infty} H(s)\exp[-i\omega s-(s+T)^2/\tau^2]ds.$$

Whereupon, the H.F.T. of (2.2) produces

$$(\alpha^2-\omega^2)\text{H.F.T.}[\Phi] = \frac{\text{F.T.}[f]}{2\pi\tau\sqrt{\pi}}\int_T^{\infty}\exp[-i\omega(u-T)-u^2/\tau^2]du. \tag{2.7}$$

Applying this to (2.4):

$$\Phi=\int_{R_n}\text{F.T.}[f]L(t\,|\,\alpha)\exp(i\boldsymbol{\alpha}\cdot\mathbf{x})d\boldsymbol{\alpha}, \tag{2.8}$$

where, writing $\alpha=|\boldsymbol{\alpha}|=(\alpha_1^2+\alpha_2^2+\cdots+\alpha_n^2)^{1/2}$,

$$L(t\,|\,\alpha)=(2\pi)^{-1}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}\frac{\exp(i\omega t)}{\alpha^2-\omega^2}d\omega(\tau\sqrt{\pi})^{-1} \times\int_T^{\infty}\exp(-i\omega u-u^2/\tau^2)du. \tag{2.9}$$

The right side of (2.7) involves the factor $\exp(i\omega T)$, which stays analytic over $\text{Im}\omega\le 0$ as $T\to-\infty$. In this limit approach,

$$(\tau\sqrt{\pi})^{-1}\int_T^{\infty}\exp(-i\omega u-u^2/\tau^2)du\to\exp(-\tfrac{1}{4}\omega^2\tau^2), \tag{2.10}$$

which (holds for complex as well as real ω and) is analytic throughout $|\omega|<\infty$. Hereafter, we assume the limit $T=-\infty$ and that, correspondingly, $t>-\infty$. In particular, $\phi\equiv\Phi$. Also, from (2.7), the two real poles of H.F.T.$[\Phi]$ at $\omega=|\boldsymbol{\alpha}|,-|\boldsymbol{\alpha}|$ are its lowest singularities. So, according to (2.5), $\epsilon>0$. Furthermore, (2.9) becomes

$$L(t\,|\,\alpha)=(2\pi)^{-1}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}\frac{\exp(i\omega t-\frac{1}{4}\omega^2\tau^2)}{\alpha^2-\omega^2}d\omega. \tag{2.11}$$

For our purposes, we envisage the spatial source distribution $f(\mathbf{x})$ as a convolution, over R_n, of a function $\rho(\mathbf{x})$ with the Gaussian, viz.,

$$f(\mathbf{x})=(\kappa\sqrt{\pi})^{-n}\int_{R_n}\rho(\mathbf{y})\exp[-(\mathbf{x}-\mathbf{y})^2/\kappa^2]d\mathbf{y}, \tag{2.12}$$

κ being a length scale. Applying (2.6):

$$\begin{aligned}\text{F.T.}[f]&=(2\pi)^{-n}\int_{R_n}\rho(\mathbf{y})\exp(-i\boldsymbol{\alpha}\cdot\mathbf{y})d\mathbf{y}(\kappa\sqrt{\pi})^{-n}\\&\quad\times\int_{R_n}\exp[-i\boldsymbol{\alpha}\cdot(\mathbf{x}-\mathbf{y})-(\mathbf{x}-\mathbf{y})^2/\kappa^2]d\mathbf{x},\\&=(2\pi)^{-n}\int_{R_n}\rho(\mathbf{y})\exp(-\boldsymbol{\alpha}\cdot\mathbf{y}-\tfrac{1}{4}\alpha^2\kappa^2)d\mathbf{y},\end{aligned}$$

via an n-dimensional extension of (2.10). So, from (2.8),

$$\phi\equiv\Phi=\int_{R_n}\rho(\mathbf{y})K(\mathbf{x},t\,|\,\mathbf{y})d\mathbf{y}, \tag{2.13}$$

where

$$\begin{aligned}K(\mathbf{x},t\,|\,\mathbf{y})&=(2\pi)^{-n}\int_{R_n}L(t\,|\,\alpha)\exp[i\boldsymbol{\alpha}\cdot(\mathbf{x}-\mathbf{y})-\tfrac{1}{4}\alpha^2\kappa^2]d\boldsymbol{\alpha},\\&=(2\pi)^{-n}\int_0^{\infty}L(t\,|\,\alpha)\exp(-\tfrac{1}{4}\alpha^2\kappa^2)\alpha^{n-1}d\alpha\\&\quad\times\int_{\Omega_n}\exp[i\alpha\boldsymbol{\xi}\cdot(\mathbf{x}-\mathbf{y})]d\Omega_{\xi},\end{aligned} \tag{2.14}$$

with the inner integration performed by letting the unit position vector $\boldsymbol{\xi}=\boldsymbol{\alpha}\alpha^{-1}$ range over Ω_n, the n-dimensional unit sphere (circle if $n=2$) with surface element $d\Omega_{\xi}$ located about $\boldsymbol{\xi}$. Now a law of spherical mean[5] stipulates

$$\int_{\Omega_n}g(\mathbf{x}\cdot\boldsymbol{\xi})d\Omega_{\xi}=\frac{2\pi^{(1/2)(n-1)}}{\Gamma(\frac{1}{2}n-\frac{1}{2})}\int_{-1}^{1}(1-\xi^2)^{(1/2)(n-3)}g(|\mathbf{x}|\,\xi)d\xi. \tag{2.15}$$

Also[6] if $\nu>-\frac{1}{2}$,

$$\frac{(\frac{1}{2}z)^{\nu}}{\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-\xi^2)^{\nu-(1/2)}\exp(iz\xi)d\xi=J_{\nu}(z), \tag{2.16}$$

the Bessel function of order ν. Consequently, (2.14) reduces to

$$\begin{aligned}K(\mathbf{x},t\,|\,\mathbf{y})&=[(2\pi)^{-n/2}/|\mathbf{x}-\mathbf{y}|^{(n/2)-1}]\int_0^{\infty}L(t\,|\,\alpha)\\&\quad\times J_{(1/2)n-1}(\alpha\,|\mathbf{x}-\mathbf{y}|)\exp(-\tfrac{1}{4}\alpha^2\kappa^2)\alpha^{(1/2)n}d\alpha.\end{aligned} \tag{2.17}$$

III. THE SPATIAL SOURCE DISTRIBUTION

According to generalized function theory[4] a limit interpretation of (2.12) is

$$\lim_{\kappa\to 0}f(\mathbf{x})=\rho(\mathbf{x}). \tag{3.1}$$

Let us choose

$$\rho(\mathbf{x})=\chi(\hat{\mathbf{x}})\delta(|\mathbf{x}|-r)|\mathbf{x}|^{1-n}\quad(r>0), \tag{3.2}$$

where the unit vector $\hat{\mathbf{x}}=\mathbf{x}|\mathbf{x}|^{-1}$, and δ denotes the one-dimensional Dirac delta function. It then follows immediately from (2.13) that

$$\phi=\int_{\Omega_n}\chi(\boldsymbol{\zeta})K(\mathbf{x},t\,|\,r\boldsymbol{\zeta})d\Omega_{\zeta}, \tag{3.3}$$

with the unit vector $\boldsymbol{\zeta}$ ranging over Ω_n. Clearly $K(\mathbf{x},t\,|\,r\boldsymbol{\zeta})$ plays the role of a kernel. Its dependence on $\mathbf{x}$ and $r\boldsymbol{\zeta}$ arises solely through the argument [see (2.17)]

$$|\mathbf{x}-r\boldsymbol{\zeta}| = (\mathbf{x}^2+r^2-2|\mathbf{x}|r\hat{\mathbf{x}}\cdot\boldsymbol{\zeta})^{1/2}$$
$$= [(r\hat{\mathbf{x}})^2+(|\mathbf{x}|\boldsymbol{\zeta})^2-2|\mathbf{x}|r\hat{\mathbf{x}}\cdot\boldsymbol{\zeta}]^{1/2}$$
$$= |r\hat{\mathbf{x}}-|\mathbf{x}|\boldsymbol{\zeta}|.$$

So,

$$K(\mathbf{x},t|r\boldsymbol{\zeta}) = K(r\hat{\mathbf{x}},t|\ |\mathbf{x}|\boldsymbol{\zeta}). \tag{3.4}$$

Hence, by virtue of (3.3), we assert the following *reciprocity principle*: *for a fixed observational direction* $\hat{\mathbf{x}}$, *the solution* ϕ *remains unchanged if* $|\mathbf{x}|$ *and* r *are interchanged.*

Hereafter we restrict the length scale κ to being small and approximate accordingly. We shall subsequently prove that

$$f(\mathbf{x}) \sim \chi(\hat{\mathbf{x}})(\kappa\sqrt{\pi})^{-1}\exp[-(|\mathbf{x}|-r)^2/\kappa^2](r|\mathbf{x}|)^{1/2-n/2}, \tag{3.5}$$

corresponding to *a Gaussian distribution loaded* (for small κ) *about the spherical sheet* $|\mathbf{x}|=r$ *and coupled to a directionally dependent density* $\chi(\hat{\mathbf{x}})$. (Throughout, substitute "circular" for "spherical" whenever $n=2$.) Note a consistency of (3.1), (3.2), and (3.5) with an attainment of the delta function via a Gaussian sequence. Also, applying (3.2) to (2.12):

$$f(\mathbf{x}) = (\kappa\sqrt{\pi})^{-n}\exp[-(\mathbf{x}^2+r^2)/\kappa^2]F(\mathbf{x}), \tag{3.6}$$

where, letting $\lambda = 2r|\mathbf{x}|/\kappa^2$,

$$F(\mathbf{x}) = \int_{\Omega_n}\chi(\boldsymbol{\zeta})\exp(\lambda\hat{\mathbf{x}}\cdot\boldsymbol{\zeta})d\Omega_{\zeta}. \tag{3.7}$$

Write $\boldsymbol{\zeta}=(\zeta_1,\ldots,\zeta_n)$. In terms of $n-1$ angular coordinates,[7]

$$\zeta_1=\cos\psi_1 (n\ge 2),\quad \zeta_2=\sin\psi_1\cos\psi_2\quad (n\ge 3), \tag{3.8}$$

$$\zeta_3=\sin\psi_1\sin\psi_2\cos\psi_3\quad (n\ge 4),\ldots, \tag{3.9}$$

$$\zeta_{n-1}=\sin\psi_1\cdots\sin\psi_{n-2}\cos\psi_{n-1}\quad (n\ge 3), \tag{3.10}$$

$$\zeta_n=\sin\psi_1\cdots\sin\psi_{n-2}\sin\psi_{n-1}\quad (n\ge 2). \tag{3.11}$$

Also

$$d\Omega_{\zeta}=\prod_{\nu=1}^{n-1}\sin^{n-1-\nu}\psi_\nu d\psi_\nu. \tag{3.12}$$

Consider the recurrence relations:

$$F_{n-2}=\int_0^{2\pi}X(\psi_1,\ldots,\psi_{n-1})\exp(\lambda\hat{\mathbf{x}}\cdot\boldsymbol{\zeta})d\psi_{n-1} \tag{3.13}$$

with $n\ge 2$; while for $n\ge 3$,

$$F_{n-\mu-1}=\int_0^{\pi}F_{n-\mu}\sin^{\mu-1}\psi_{n-\mu}d\psi_{n-\mu}\quad (\mu=2,\ldots,n-1). \tag{3.14}$$

Suppose

$$\chi(\boldsymbol{\zeta})=X(\psi_1,\ldots,\psi_{n-1}). \tag{3.15}$$

Then corresponding to $\mu=n-1$, (3.14) [or (3.13) if $n=2$] can be identified with (3.7):

$$F(\mathbf{x})=F_0. \tag{3.16}$$

An angular system $(\theta_1,\ldots,\theta_{n-1})$ can be similarly assigned to the physical direction $\hat{\mathbf{x}}=(\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_n)$, e.g., by substituting $\hat{x}_\nu$, θ_ν for, respectively, ζ_ν, ψ_ν throughout (3.8)–(3.11). At present, we assume that the $\boldsymbol{\zeta}$ frame has already been orientated in such a way that, for the given $\hat{\mathbf{x}}$ direction relative to it, $\theta_\nu\ne 0$ or π, but $\in(0,\pi)$ $(\nu=1,\ldots,n-2)$ while $\theta_{n-1}\ne 0$ or 2π, but $\in(0,2\pi)$. Now define

$$A_{n-\mu}=\sum_{\nu=1}^{n-\mu}\hat{x}_\nu\zeta_\nu\ (\mu=2,\ldots,n-1), \tag{3.17}$$

$$B_{n-\mu}=\prod_{\nu=1}^{n-\mu}\sin\psi_\nu\sin\theta_\nu\ \ (\mu=2,\ldots,n-1). \tag{3.18}$$

Evidently, for $\mu=2,\ldots,n-2$,

$$B_{n-\mu}=B_{n-\mu-1}\sin\psi_{n-\mu}\sin\theta_{n-\mu}, \tag{3.19}$$

$$A_{n-\mu}=A_{n-\mu-1}+B_{n-\mu-1}\cos\psi_{n-\mu}\cos\theta_{n-\mu}, \tag{3.20}$$

via (3.8)–(3.11); so

$$A_{n-\mu}+B_{n-\mu}=A_{n-\mu-1}+B_{n-\mu-1}\cos(\psi_{n-\mu}-\theta_{n-\mu}). \tag{3.21}$$

Also,

$$A_1+B_1=\cos(\psi_1-\theta_1); \tag{3.22}$$

$$\hat{\mathbf{x}}\cdot\boldsymbol{\zeta}=A_{n-2}+B_{n-2}\cos(\psi_{n-1}-\theta_{n-1}). \tag{3.23}$$

Note from (3.17) and (3.18) that $A_{n-\mu}$ and $B_{n-\mu}$ are independent of $\psi_{n-\mu+1}$. Also, within the ranges of integration for $F_0,F_1,\ldots,F_{n-\mu-1}$, viz., $\psi_\nu\in(0,\pi)$ with $\nu=1,\ldots,n-\mu$, the quantity $B_{n-\mu}>0$.

An appropriate approximation of $F(\mathbf{x})$ is achieved through successive asymptotic approximations for large λ. Just for this purpose alone, a basic hypothesis is that $X(\psi_1,\ldots,\psi_{n-1})$ is analytic over all $n-1$ ranges of integration indicated in (3.13) and (3.14). We start from (3.13), expressible via (3.23) as

$$F_{n-2}=\exp(\lambda A_{n-2})\int_0^{2\pi}X(\psi_1,\ldots,\psi_{n-1})$$
$$\times\exp[\lambda B_{n-2}\cos(\psi_{n-1}-\theta_{n-1})]d\psi_{n-1},$$

to which we apply the well known method of steepest descents.[8] Among an infinity of saddle points (all real, of order one, and) determined by $\sin(\psi_{n-1}-\theta_{n-1})=0$, only three are actually relevant, viz., θ_{n-1}, $\theta_{n-1}+\pi$, $\theta_{n-1}-\pi$. The saddle point θ_{n-1} invariably lies on the given path $(0,2\pi)$. Suppose $\theta_{n-1}\in(0,\pi)$; then $\theta_{n-1}+\pi\in(0,2\pi)$ and so contributes to F_{n-2}, but not $\theta_{n-1}-\pi$. Alternatively, if $\theta_{n-1}\in(\pi,2\pi)$, then $\theta_{n-1}-\pi\in(0,2\pi)$ and contributes to F_{n-2}, but not $\theta_{n-1}+\pi$. If $\theta_{n-1}=\pi$, then $\theta_{n-1}-\pi=0$ and $\theta_{n-1}+\pi=2\pi$ so that they both contribute. In each case, the contribution from $\theta_{n-1}-\pi$ or $\theta_{n-1}+\pi$, being governed by the factor $\exp(-\lambda B_{n-2})$, is negligible compared with the saddle point contribution of θ_{n-1} which dominates:

$$F_{n-2}\sim(2\pi/\lambda B_{n-2})^{1/2}\exp[\lambda(A_{n-2}+B_{n-2})]$$
$$\times X(\psi_1,\ldots,\psi_{n-2},\theta_{n-1}).$$

Next assume

$$F_{n-\mu}\sim(2\pi/\lambda B_{n-\mu})^{(\mu-1)/2}\exp[\lambda(A_{n-\mu}+B_{n-\mu})]$$
$$\times X(\psi_1,\ldots,\psi_{n-\mu},\theta_{n-\mu+1},\ldots,\theta_{n-1}), \tag{3.24}$$

which evidently holds for $\mu=2$. This validity may be extended, by induction, to cover $\mu=3,\ldots,n-1$. Thus a first approximation of (3.14) via (3.24), incorporating

(3.19) and (3.21), yield

$$F_{n-\mu-1} \sim (2\pi/\lambda B_{n-\mu-1})^{(\mu-1)/2} \exp(\lambda A_{n-\mu-1}) \times \int_0^\pi X(\psi_1,\ldots,\psi_{n-\mu},\theta_{n-\mu+1},\ldots,\theta_{n-1}) \times \exp[\lambda B_{n-\mu-1}\cos(\psi_{n-\mu}-\theta_{n-\mu})] \times (\sin\psi_{n-\mu}/\sin\theta_{n-\mu})^{(\mu-1)/2} d\psi_{n-\mu}. \quad (3.25)$$

Again, we encounter the same type of saddle points. Among these, only $\theta_{n-\mu}$ lies on the present (shorter) path $(0,\pi)$ and provides the dominant contribution. When evaluated, it confirms the anticipated consistency of (3.25) with (3.24), which therefore holds for $\mu = 2,\ldots,n-1$. In particular, with $\mu = n-1$, and (3.18), (3.22) accounted for, (3.24) enables (3.14) to be approximated for $F_0 = F(\mathbf{x})$ by the same steepest descent procedure. Whereupon,

$$F(\mathbf{x}) \sim (2\pi/\lambda)^{(n-1)/2} \exp(\lambda) X(\theta_1,\ldots,\theta_{n-1}) = (\kappa\sqrt{\pi})^{n-1}(\gamma|\mathbf{x}|)^{1/2-n/2}\exp(2\gamma|\mathbf{x}|/\kappa^2)\chi(\hat{\mathbf{x}}). \quad (3.26)$$

Substitution into (3.6) proves (3.5).

IV. THE L FUNCTION

To compute the solution from (3.3), one needs the K kernel. Its derivation from (2.17) must be preceded by that of the L function via (2.11). This involves the integrand

$$\gamma(\omega) = [\exp(i\omega t - \tfrac{1}{4}\omega^2\tau^2)]/(\alpha^2-\omega^2), \quad (4.1)$$

integrated, for $\alpha \in (0,\infty)$, along the horizontal contour $\mathrm{Im}\omega = -\epsilon$ (<0) from one infinite end to the other. Our present goal is the evaluation of the L function to an estimable error trailing a recognizable quantity that will eventually enable a convenient management of (2.17). Henceforth, we also assume the time scale τ to be small.

First, we wish to deform a portion of the prescribed contour $(-\infty - i\epsilon, \infty - i\epsilon)$ until part of it rests on a horizontal line, say $\omega = \sigma + i2t_0/\tau^2$, where σ and t_0 denote, respectively, a real variable and a real constant. So, when $t \neq 0$, then in (4.1), the exponent factor rather desirably approaches zero along this line as $\tau \to 0$ if $t_0(2t - t_0) > 0$, which is satisfied by $t_0 = t$. Thus we select $L^*: \omega = \sigma + i2t/\tau^2$, a horizontal path within $\mathrm{Im}\omega \gtrless 0$ according as $t \gtrless 0$; tentatively, we allow $\sigma \in (-N/\tau^2, N/\tau^2)$ with $N > 0$. Then

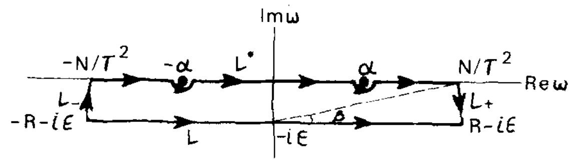


FIG. 2. Case: $t = 0$.

$$\int_{L^*} \gamma(\omega)d\omega = \exp(-t^2/\tau^2)\int_{-N/\tau^2}^{N/\tau^2} \frac{\exp(-\frac{1}{4}\sigma^2\tau^2)d\sigma}{\alpha^2 - (\sigma + i2t/\tau^2)^2}. \quad (4.2)$$

Since

$$|\alpha^2 - (\sigma + i2t/\tau^2)^2| = [(\alpha-\sigma)^2 + 4t^2/\tau^4]^{1/2}[(\alpha+\sigma)^2 + 4t^2/\tau^4]^{1/2} > 4t^2/\tau^4,$$

so

$$\left|\int_{L^*}\gamma(\omega)d\omega\right| < \tfrac{1}{4}(\tau^4/t^2)\exp(-t^2/\tau^2)\int_{-N/\tau^2}^{N/\tau^2}\exp(-\tfrac{1}{4}\sigma^2\tau^2)d\sigma = \tfrac{1}{2}\sqrt{\pi}(\tau^3/t^2)\exp(-t^2/\tau^2)\mathrm{erf}(\tfrac{1}{2}N/\tau), \quad (4.3)$$

whenever $t \neq 0$. If $t = 0$, (4.3) fails; however, we can still use L^*, which is now a real path: $\mathrm{Im}\omega = 0$; in this case (4.2) remains valid but must, if necessary, be envisaged in the sense of a Cauchy principal value.

For each $t \in (-\infty,\infty)$ and each $\alpha \in (0,\infty)$, we restrict the positive parameters ϵ and N to being sufficiently small and large:

$$\epsilon < 2|t|/\tau^2 \ (\text{if } t \neq 0), \quad N > \max(\alpha\tau^2, 2|t| + \epsilon\tau^2). \quad (4.4)$$

Now, with $\omega = -i\epsilon$ as center, construct two circular arcs $L_\pm$ having the same radius R, and joining the end points of L^* at $\omega = \pm N/\tau^2 + i2t/\tau^2$ to the original path $\mathrm{Im}\omega = -\epsilon$. On it, let L denote the resultant intercept. According as $t \gtreqless 0$, we refer to Figs. 1, 2, or 3. Evidently,

$$R = [N^2/\tau^4 + (2t/\tau^2 + \epsilon)^2]^{1/2},$$

and so $\to \infty$ when $N \to \infty$. Furthermore, $L_\pm$ subtend the same acute angle

$$\beta = \tan^{-1}\left(\frac{2t + \epsilon\tau^2}{N}\right)$$

at the center. In view of (4.4),

$$0 < \beta < \pi/4 \text{ for } t \geq 0, \quad -\pi/4 < \beta < 0 \text{ for } t < 0; \quad (4.5)$$

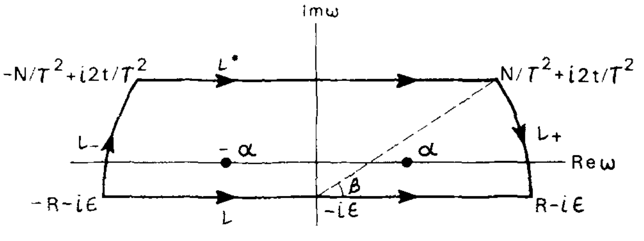


FIG. 1. Case: $t > 0$. Deformation of the contour L for the integration of $\gamma(\omega)$.

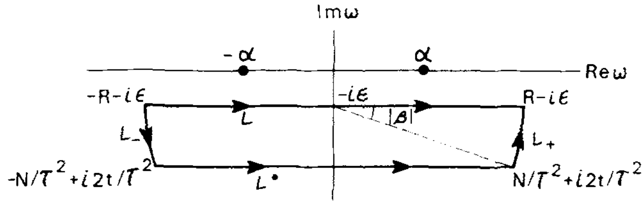


FIG. 3. Case: $t < 0$.

furthermore, the only singularities of $\gamma(\omega)$, viz. two real simple poles at $\omega=\pm\alpha$ are well inside (or outside) the domain bounded by L, L_-, L^*, L_+ for $t>0$ (or <0), but are originally threaded by L^* when $t=0$. In the latter case, L^* may, for a principal value (P.V.) interpretation, be indented below both poles via infinitesimally small semicircles described anticlockwise. With L_-, L^*, L_+ directed and arranged as shown (Figs. 1, 2, 3), they constitute a clockwise (or anticlockwise) deformation of the path L for $t\geq 0$ (or <0), appropriately indented when $t=0$. Thence, according to residue theory,

$$\int_L \gamma(\omega)d\omega = 2\pi i H(t)[\underset{\omega=\alpha}{\text{residue}}\,\gamma(\omega) + \underset{\omega=-\alpha}{\text{residue}}\,\gamma(\omega)] + \left(\int_{L_-} + \int_{L^*} + \int_{L_+}\right)\gamma(\omega)d\omega \tag{4.6}$$

provided $t\neq 0$; but, if $t=0$, then

$$\int_L \gamma(\omega)d\omega = \pi i[\underset{\omega=\alpha}{\text{residue}}\,\gamma(\omega) + \underset{\omega=-\alpha}{\text{residue}}\,\gamma(\omega)] + \left(\int_{L_-} + \text{P.V.}\int_{L^*} + \int_{L_+}\right)\gamma(\omega)d\omega. \tag{4.7}$$

We next examine behaviors along the circular arcs $L_\pm : \omega + i\epsilon = Re^{i\psi}$, where

$$\text{on } L_+: 0<\psi<\beta(t\geq 0), \quad -|\beta|<\psi<0(t<0), \tag{4.8}$$

$$\text{on } L_-: \pi-\beta<\psi<\pi(t\geq 0), \quad \pi<\psi<\pi+|\beta|(t<0). \tag{4.9}$$

Now,

$$[\text{Re}(i\omega t - \tfrac{1}{4}\omega^2\tau^2)]_{L_\pm} = \epsilon t + \tfrac{1}{4}\epsilon^2\tau^2 - \tfrac{1}{2}[\tfrac{1}{2}\tau^2R^2\cos 2\psi + (2t+\epsilon\tau^2)R\sin\psi] < \epsilon t + \tfrac{1}{4}\epsilon^2\tau^2, \tag{4.10}$$

since $(2t+\epsilon\tau^2)\sin\psi>0$ and $\cos 2\psi>0$ by virtue of (4.4), (4.5), (4.8), (4.9). Let us assume that N, as restricted under (4.4), is also large enough to permit $R>2(\alpha^2+\epsilon^2)^{1/2}$. Then

$$\begin{aligned}|\alpha^2-\omega^2|_{L_\pm} &= |\alpha+i\epsilon - Re^{i\psi}|\;|\alpha - i\epsilon + Re^{i\psi}| \\ &\geq |R - |\alpha+i\epsilon||\;|R-|\alpha-i\epsilon|| \\ &= [R-(\alpha^2+\epsilon^2)^{1/2}]^2 > \tfrac{1}{4}R^2.\end{aligned} \tag{4.11}$$

Let $\Delta\psi$ represent the relevant positive ψ interval among those four defined by (4.8) and (4.9). Whence,

$$\begin{aligned}\left|\int_{L_\pm}\gamma(\omega)d\omega\right| &\leq \int_{L_\pm}|\gamma(\omega)|\;|d\omega| \\ &= R\int_{\Delta\psi}\frac{\exp[\text{Re}(i\omega t - \frac{1}{4}\omega^2\tau^2)]_{L_\pm}}{|\alpha^2-\omega^2|_{L_\pm}}d\psi \\ &< \frac{4|\beta|}{R}\exp(\epsilon t + \tfrac{1}{4}\epsilon^2\tau^2)\to 0 \quad \text{as } N\to\infty,\end{aligned}$$

after accounting for (4.1), (4.10), and (4.11). Whereupon, after evaluating from (4.1), the specified residues in (4.6), the latter's limit application to (2.11) yields, for $t\neq 0$,

$$\begin{aligned}L(t|\alpha) &= \lim_{N\to\infty}(2\pi)^{-1}\int_L\gamma(\omega)d\omega \\ &= H(t)\alpha^{-1}\sin(\alpha t)\exp(-\tfrac{1}{4}\alpha^2\tau^2) + E_L(t|\alpha),\end{aligned} \tag{4.12}$$

where

$$E_L(t|\alpha) = \lim_{N\to\infty}(2\pi)^{-1}\int_{L^*}\gamma(\omega)d\omega = (2\pi)^{-1}\exp(-t^2/\tau^2) \times \int_{-\infty}^{\infty}\frac{\exp(-\frac{1}{4}\sigma^2\tau^2)d\sigma}{\alpha^2-(\sigma+i2t/\tau^2)^2} \tag{4.13}$$

and, is in view of (4.3), bounded:

$$|E_L(t|\alpha)| < \tfrac{1}{4}(\tau^3/t^2\sqrt{\pi})\exp(-t^2/\tau^2). \tag{4.14}$$

However, in the case $t=0$, both residues in (4.7) eliminate each other; consequently, and via (4.2),

$$\begin{aligned}L(0|\alpha) &= \lim_{N\to\infty}(2\pi)^{-1}\text{P.V.}\int_{L^*}\gamma(\omega)d\omega \\ &= (2\pi)^{-1}\text{P.V.}\int_{-\infty}^{\infty}\frac{\exp(-\frac{1}{4}\sigma^2\tau^2)}{\alpha^2-\sigma^2}d\sigma,\end{aligned} \tag{4.15}$$

i.e., with $E_L(t|\alpha)$ formally expressed by (4.13),

$$L(0|\alpha) = \text{P.V.}E_L(0|\alpha). \tag{4.16}$$

V. THE K KERNEL

The expression (2.17) for the K kernel can be developed with the aid of (4.12) and (4.16). First, define

$$E_K(\mathbf{x},t|\mathbf{y}) = \frac{(2\pi)^{-(1/2)n}}{|\mathbf{x}-\mathbf{y}|^{(1/2)n-1}}\int_0^\infty E_L(t|\alpha)J_{(n/2)-1}(\alpha|\mathbf{x}-\mathbf{y}|) \times \exp(-\tfrac{1}{4}\alpha^2\kappa^2)\alpha^{(1/2)n}d\alpha, \tag{5.1}$$

with P.V.$E_L(0|\alpha)$ replacing $E_L(t|\alpha)$ when $t=0$. In particular, then,

$$K(\mathbf{x},0|\mathbf{y}) = E_K(\mathbf{x},0|\mathbf{y}). \tag{5.2}$$

Note also that [cf. (3.4)], if $\hat{\mathbf{y}} = \mathbf{y}|\mathbf{y}|^{-1}$,

$$E_K(\mathbf{x},t|\mathbf{y}) = E_K(|\mathbf{y}|\hat{\mathbf{x}},t||\mathbf{x}|\hat{\mathbf{y}}). \tag{5.3}$$

Application of (4.12) to (2.17) leads to an α-integral involving the integrand factor

$$\exp[-\tfrac{1}{4}\alpha^2(\kappa^2+\tau^2)]\sin(\alpha t) = \sum_{l=0}^{\infty}\frac{(\kappa^2+\tau^2)^l}{4^l l!}\frac{\partial^{2l}}{\partial t^{2l}}\sin(\alpha t). \tag{5.4}$$

We now quote two results from Ref. 6 [viz. Sec. 3.4(3)]

$$\sin(\alpha t) = (\tfrac{1}{2}\pi\alpha t)^{1/2}J_{1/2}(\alpha t) \tag{5.5}$$

[applicable to (5.4)], and the Neumann series (Sec. (16.1)

$$\frac{J_\nu[(Z^2+z^2-2Zz\cos\psi)^{1/2}]}{(Z^2+z^2-2Zz\cos\psi)^{\nu/2}} = \frac{2^\nu\Gamma(\nu)}{(Zz)^\nu}\sum_{m=0}^{\infty}(m+\nu)J_{m+\nu}(Z)J_{m+\nu}(z)C_m^\nu(\cos\psi). \tag{5.6}$$

Here, $C_m^\nu(\cos\psi)$ denotes the Gegenbauer function of degree m and order ν; its argument $\cos\psi$ may be regarded as the scalar product $\hat{\mathbf{x}}\cdot\hat{\mathbf{y}}$. Thus, we eventually arrive at

$$K(\mathbf{x},t|\mathbf{y}) = E_K(\mathbf{x},t|\mathbf{y}) + H(t)\frac{(|\mathbf{x}|\,|\mathbf{y}|)^{1-(1/2)n}}{2^{3/2}\pi^{(1/2)n-1/2}} \times \sum_{l=0}^{\infty}\frac{(\kappa^2+\tau^2)^l}{4^l l!}\frac{\partial^{2l}}{\partial t^{2l}}\sqrt{t} \times \sum_{m=0}^{\infty}(m+\tfrac{1}{2}n-1)\Gamma(\tfrac{1}{2}n-1)C_m^{(1/2)n-1}(\hat{\mathbf{x}}\cdot\hat{\mathbf{y}})G_m, \tag{5.7}$$

provided $t \neq 0$, and where

$$G_m = \int_0^\infty J_{1/2}(\alpha t) J_{m+(1/2)n-1}(\alpha|\mathbf{x}|) J_{m+(1/2)n-1}(\alpha|\mathbf{y}|)\alpha^{1/2} d\alpha,$$

$$= \begin{cases} 0 : t^2 < (|\mathbf{x}| - |\mathbf{y}|)^2, & (5.8) \\ (2\pi|\mathbf{x}|\,|\mathbf{y}|t)^{-1/2} P_{m+(n/2)-3/2}\left(\frac{\mathbf{y}^2+\mathbf{x}^2-t^2}{2|\mathbf{x}|\,|\mathbf{y}|}\right) : (|\mathbf{x}| - |\mathbf{y}|)^2 < t^2 < (|\mathbf{x}| + |\mathbf{y}|)^2, & (5.9) \\ \cos[(m+\tfrac{1}{2}n-1)\pi](\tfrac{1}{2}\pi^3|\mathbf{x}|\,|\mathbf{y}|t)^{-1/2} Q_{m+(n/2)-3/2}\left(\frac{t^2-\mathbf{x}^2-\mathbf{y}^2}{2|\mathbf{x}|\,|\mathbf{y}|}\right) : (|\mathbf{x}| + |\mathbf{y}|)^2 < t^2, & (5.10) \end{cases}$$

[see Ref. 6, Sec. 13.46, Eqs. (1), (4), (5)] with $P_{m+(1/2)n-3/2}$ and $Q_{m+(n/2)-3/2}$ denoting Legendre functions of the first and second kinds.

A bound can be readily formulated for the remainder term E_K if $t \neq 0$. Now [Ref. 6, Sec. 3.31, Eq. (1)]

$$|J_\nu(z)| \leq |(\tfrac{1}{2}z)^\nu| \exp(|\mathrm{Im}z|)/\Gamma(\nu+1) \quad (\nu > -\tfrac{1}{2}).$$

When incorporated together with (4.14) into (5.1):

$$|E_K(\mathbf{x}, t|\mathbf{y})| < \frac{\tau^3 \exp(-t^2/\tau^2)}{t^2\Gamma(\frac{1}{2}n)(2\sqrt{\pi})^{n+1}} \int_0^\infty \exp(-\tfrac{1}{4}\alpha^2\kappa^2)\alpha^{n-1} d\alpha,$$

$$\leq \frac{\tau^{3-n}\exp(-t^2/\tau^2)}{4t^2\pi^{(1/2)n+1/2}} \quad (t \neq 0) \tag{5.11}$$

provided $\tau \leq \kappa$, assumed hereafter.

VI. THE RADIATION FIELD

The radiation field is essentially secured by applying to (3.3) the results (5.7)–(5.10), or just (5.2) if $t = 0$. Thus, throughout $-\infty < t < \infty$:

$$\phi = \begin{cases} E \quad (-\infty < t < \big||x| - r\big|), & (6.1) \\ P + E \quad (\big||\mathbf{x}| - r\big| < t < |\mathbf{x}| + r), & (6.2) \\ Q + E \quad (|\mathbf{x}| + r < t < \infty). & (6.3) \end{cases}$$

Here,

$$E = \int_{\Omega_n} \chi(\boldsymbol{\zeta}) E_K(\mathbf{x}, t|r\boldsymbol{\zeta}) d\Omega_\zeta. \tag{6.4}$$

Also,

$$P = \frac{1}{\pi^{(1/2)n}(r|\mathbf{x}|)^{(1/2)(n-1)}} \sum_{l=0}^\infty \frac{(\kappa^2+\tau^2)^l}{2^{2l+2}l!} \frac{\partial^{2l}P^*}{\partial t^{2l}}, \tag{6.5}$$

$$Q = \frac{\cos(\frac{1}{2}n\pi)}{\pi^{(1/2)n+1}(r|\mathbf{x}|)^{(1/2)(n-1)}} \sum_{l=0}^\infty \frac{(\kappa^2+\tau^2)^l}{2^{2l+1}l!} \frac{\partial^{2l}Q^*}{\partial t^{2l}}, \tag{6.6}$$

where, in terms of arguments

$$\theta = \cos^{-1}\left(\frac{r^2+\mathbf{x}^2-t^2}{2r|\mathbf{x}|}\right), \quad A = \cosh^{-1}\left(\frac{t^2-\mathbf{x}^2-r^2}{2r|\mathbf{x}|}\right) \tag{6.7}$$

which are real over the respective domains supporting P and Q, we have

$$P^* = \sum_{m=0}^\infty \chi_m(\hat{\mathbf{x}}) P_{m+(n/2)-3/2}(\cos\theta), \tag{6.8}$$

$$Q^* = \sum_{m=0}^\infty (-1)^{m+1} \chi_m(\hat{\mathbf{x}}) Q_{m+(n/2)-3/2}(\cosh A), \tag{6.9}$$

with coefficients

$$\chi_m(\hat{\mathbf{x}}) = (m+\tfrac{1}{2}n-1)\Gamma(\tfrac{1}{2}n-1) \int_{\Omega_n} \chi(\boldsymbol{\zeta}) C_m^{(1/2)n-1}(\hat{\mathbf{x}}\cdot\boldsymbol{\zeta}) d\Omega_\zeta. \tag{6.10}$$

Both P and Q acquire their directional dependence on $\hat{\mathbf{x}}$ through these χ_m coefficients. Note the important fact that

$$Q \equiv 0 \quad \text{for odd } n. \tag{6.11}$$

Regarding (6.10), the following rules (Ref. 6, Sec. 11.41) are relevant:

$$\lim_{\nu\to 0} (m+\nu)\Gamma(\nu) C_m^\nu(\cos\psi) = 2\cos m\psi \quad (m \neq 0),$$

$$\lim_{\nu\to 0} \nu\Gamma(\nu) C_0^\nu(\cos\psi) \equiv C_0^0(\cos\psi) \equiv 1.$$

Thus, for the two-dimensional problem,

$$\chi_0(\hat{\mathbf{x}}) = \int_0^{2\pi} X(\psi) d\psi, \tag{6.12}$$

$$\chi_m(\hat{\mathbf{x}}) = 2\int_0^{2\pi} X(\psi)\cos[m(\psi-\theta_1)] d\psi \quad (m \geq 1), \tag{6.13}$$

where $X(\psi) = \chi(\cos\psi, \sin\psi)$ and θ_1 signifies the polar angle: $\hat{\mathbf{x}} = (\cos\theta_1, \sin\theta_1)$ (cf. Sec. 3).

We now use (6.4) to establish an upper bound for $|E|$ when $t \neq 0$. If $\boldsymbol{\zeta}(\in\Omega_n)$ is related to $\psi_1, \ldots, \psi_{n-1}$ by (3.8)–(3.11), then $\sin\psi_\nu = |\sin\psi_\nu|$ for $\nu = 1, \ldots, n-2$, and leads to

$$|E| < \frac{\tau^{3-n}\exp(-t^2/\tau^2)}{4t^2\pi^{(1/2)n+1/2}} \int_{\Omega_n} |\chi(\boldsymbol{\zeta})| d\Omega_\zeta, \tag{6.14}$$

after accounting for (3.12) and (5.11). Moreover, since (Ref. 5, Sec. 1)

$$\int_{\Omega_n} d\Omega_\zeta = 2\pi^{(1/2)n}/\Gamma(\tfrac{1}{2}n) = \text{surface area of } \Omega_n, \tag{6.15}$$

therefore if $\chi(\boldsymbol{\zeta})$ is bounded over Ω_n,

$$|E| < \frac{\tau^{3-n}\exp(-t^2/\tau^2)}{2t^2\sqrt{\pi}\Gamma(\frac{1}{2}n)} \max_{\hat{\mathbf{x}}\in\Omega_n} |\chi(\hat{\mathbf{x}})| \quad (t \neq 0). \tag{6.16}$$

Hence, when $t \neq 0$, the "error" E can be made as small as desired by choosing a sufficiently small time scale τ. The quantity E normally depends on $\mathbf{x}$ and r, but not its bound in (6.16). Its exact representation of (6.4) can be expanded via (5.1) and (5.6) to yield

$$E = \frac{1}{2\pi^{(1/2)n}(r|\mathbf{x}|)^{(1/2)n-1}} \sum_{m=0}^\infty \chi_m(\hat{\mathbf{x}}) E_m(|\mathbf{x}|, t), \tag{6.17}$$

valid $\forall t \in (-\infty, \infty)$ and wherein

$$E_m(|\mathbf{x}|,t) = \int_0^\infty J_{m+(n/2)-1}(\alpha r) J_{m+(n/2)-1}(\alpha|\mathbf{x}|) E_L(t|\alpha) \times \exp(-\tfrac{1}{4}\alpha^2\kappa^2)\alpha\, d\alpha; \qquad (6.18)$$

for $t=0$, the E_L factor must be replaced by its P.V. So, like P and Q, directional dependence through the χ_m coefficients is also experienced by E, and hence by ϕ as well. Note [from (4.13), (6.17), and (6.18)] that the source has imparted its Gaussian factor $\exp(-t^2/\tau^2)$ to E.

At the center $\mathbf{x}=\mathbf{0}$, our results simplify considerably. First, since[9]

$$Q_\nu(z) = \tfrac{1}{2}\int_{-\infty}^{\infty} [z+(z^2-1)^{1/2}\cosh\psi]^{-\nu-1} d\psi \quad (\nu > -1), \qquad (6.19)$$

therefore $\lim_{\mathbf{x}\to\mathbf{0}} (r|\mathbf{x}|)^{(1/2)(1-n)} Q_{m+(n/2)-3/2}(\cosh A) = 0$ if $m \geq 1$. Furthermore, when $t^2 > r^2$, then via the transformation: $\xi = \tanh(\frac{1}{2}\psi)$ of the integral variable in (6.19) and an appeal to (2.15), we get

$$\lim_{\mathbf{x}\to\mathbf{0}} (r|\mathbf{x}|)^{(1-n)/2} Q_{(n/2)-3/2}(\cosh A) = \frac{\Gamma(\frac{1}{2}n-\frac{1}{2})\int_{\Omega_n} d\Omega_\zeta}{2\pi^{(n-1)/2}(t^2-r^2)^{(n-1)/2}}.$$

So, accounting for (6.15) and the fact $C_0^{(n/2)-1}(\hat{\mathbf{x}}\cdot\zeta) \equiv 1$, (6.6) combine with (6.9) to yield

$$Q|_{\mathbf{x}=0} = -\cos(\tfrac{1}{2}n\pi)\frac{\Gamma(\frac{1}{2}n-\frac{1}{2})}{\pi^{(n+1)/2}} \int_{\Omega_n} \chi(\zeta) d\Omega_\zeta \times \sum_{l=0}^{\infty} \frac{(\kappa^2+\tau^2)^l}{2^{2l+1} l!} \frac{\partial^{2l}}{\partial t^{2l}} (t^2-r^2)^{(1-n)/2} \quad (t^2 > r^2). \qquad (6.20)$$

Now, (6.1)–(6.3) must be interpreted for $\mathbf{x}=\mathbf{0}$ as follows:

$$\phi|_{\mathbf{x}=0} = \begin{cases} E|_{\mathbf{x}=0} & (-\infty < t < r), \quad (6.21) \\ Q|_{\mathbf{x}=0} + E|_{\mathbf{x}=0} & (r < t < \infty). \quad (6.22) \end{cases}$$

Also, from (5.3) and (6.4)

$$E|_{\mathbf{x}=0} = E_K(r\hat{\mathbf{x}}, t|0) \int_{\Omega_n} \chi(\zeta) d\Omega_\zeta, \qquad (6.23)$$

which, incidentally, obeys the same inequality [viz. (6.14) or (6.16)] as E when $\mathbf{x} \neq \mathbf{0}$. [N.B., in actual fact, the result (6.23) should be independent of $\hat{\mathbf{x}}$. The expressions (6.20)–(6.23) fully define the solution at $\mathbf{x} = 0$.]

If r is now substituted by $|\mathbf{x}|$, this solution converts, by virtue of our reciprocity principle (Sec. 3), into the solution at any $\mathbf{x}$ position, but for $r=0$, i.e., corresponding to [cf. (3.6) and (3.7)]

$$f(\mathbf{x}) = (\kappa\sqrt{\pi})^{-n} \exp(-\mathbf{x}^2/\kappa^2) \int_{\Omega_n} \chi(\zeta) d\Omega_\zeta, \qquad (6.24)$$

which describes a point concentrated Gaussian spatial distribution weighted by a directionally independent density $\int_{\Omega_n}\chi(\zeta)d\Omega_\zeta$ encountered in both (6.20) and (6.23). The radiation field associated with (6.24) comprises an E-perturbation traversing the entire $\mathbf{x}$-space since activation time $t=-\infty$, and superposed upon a Q field. The latter, which is nontrivial only for even n, emerges after the peak instant $t=0$ of the source and progresses behind an expanding, possibly singular, spherical front viz. $|\mathbf{x}| = t$.

Suppose $\tau = 0$. Then from (6.14) or (6.16),

$$E \equiv 0 \text{ provided } t \neq 0. \qquad (6.25)$$

Actually, it is implicit that the contour deformation described in Sec. 4 does not really apply if t vanishes (simultaneously with τ), in which event E is not properly defined. However, from (2.11)

$$L(0|\alpha) = (2\pi)^{-1}\int_{-\infty-i\epsilon}^{\infty-i\epsilon} \frac{d\omega}{\alpha^2-\omega^2} \equiv 0;$$

the vanishing follows from the Cauchy–Goursat theorem after closing, within $\mathrm{Im}\,\omega < -\epsilon$, the given path with an infinite semicircle which can be shown to yield no integral contribution. Consequently, via (2.17) and (3.3), $\phi \equiv 0$ when $t=0$, and, in fact, over the entire range for (6.1). The emission process is *impulsively started* by means of an *instantaneous activation*, at time $t=0$, of the source $f(\mathbf{x})\delta(t)$ with spatial distribution $f(\mathbf{x})$ approximated by (3.5). Suppose this spatial distribution is now *singularly confined to the spherical sheet* $|\mathbf{x}| = r$, i.e., adopting the limit of $f(\mathbf{x})$ as $\kappa \to 0$ in the sense of (3.1) accompanied by (3.2). Correspondingly, the infinite series of (6.5) and (6.6) degenerate into their leading terms. Whereupon,

$$\phi = \begin{cases} 0 \quad (-\infty < t < ||\mathbf{x}| - r|), & (6.26) \\ \dfrac{P^*}{4\pi^{(1/2)n}(r|\mathbf{x}|)^{(1/2)(n-1)}} \quad (||\mathbf{x}|-r| < t < |\mathbf{x}|+r), & (6.27) \\ \dfrac{Q^*\cos(\frac{1}{2}n\pi)}{2\pi^{(1/2)n+1}(r|\mathbf{x}|)^{(1/2)(n-1)}} \quad (|\mathbf{x}|+r < t < \infty), & (6.28) \end{cases}$$

the field of the impulsive, spherical sheet source.

Evidently, the succeeding terms $l = 1, 2, \cdots$ in (6.5) and (6.6) are mainly due to the spatial, almost-Gaussian spread of the source beyond its spherical base $|\mathbf{x}| = r$. Moreover, its Gaussian time stretch beyond $t=0$ is primarily responsible for the "error" E. The latter represents a relatively weak component, strong effects being retained by P, as well as Q (if $\neq 0$).

VII. THE PROPAGATION PATTERN

Consider the two hypersheets $C_\pm : \mathbf{x}^2 = (t-r)^2$. These are hypercones convertexed end-to-end at $|\mathbf{x}| = 0$, $t=r$ about which their generator spins at a constant inclination of 45° to the time t axis which is, in a sense, one of symmetry. We define C_+ as being that member projecting indefinitely ahead of its vertex into $t > r$. Its complement C_- projects rearwards and is assumed to terminate finitely upon the hyperplane $t=0$. The cross section here is the spherical base $|\mathbf{x}| = r$ supporting our radiating distribution $f(\mathbf{x})$. Issuing forth from this base and indefinitely into $t > 0$ may be envisaged yet another (partial) hyperconical sheet $C : \mathbf{x}^2 = (t+r)^2$, again with generator revolving at 45° to the t axis. Evidently, C covers both C_+ and C_- (see Fig. 4). Let us denote: the infinite hyperconical domain ahead of C_+ by D_Q; the hyperdomain enveloped between C_+, C_- and C by D_P; the entire hyperdomain behind C (i.e., inclusive of $t \leq 0$) and infiltrating past $t=0$ into the finite hyperconical zone behind C_- by D. It can then be shown from (6.1)–(6.3) that

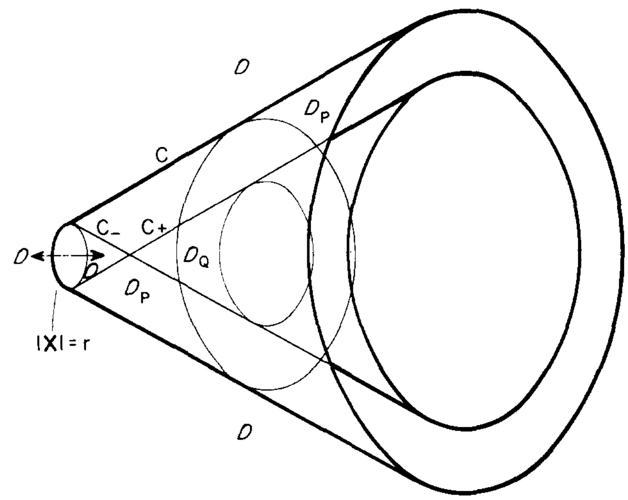


FIG. 4. The hyperconical evolution of the radiation field relative to the supporting base $|\mathbf{x}| = r$ at instant $t = 0$. From here, the right/left directed arrow points toward $t \gtrless 0$.

$$\phi = \begin{cases} E \quad \forall(\mathbf{x}, t) \in D, & (7.1) \\ P + E \quad \forall(\mathbf{x}, t) \in D_P, & (7.2) \\ Q + E \quad \forall(\mathbf{x}, t) \in D_Q, & (7.3) \end{cases}$$

This scheme is geometrically portrayed with the aid of Fig. 4. Normally, due to its variation with $\hat{\mathbf{x}}$, ϕ is not symmetric about the t axis.

To translate into physical terms, we observe the changing pattern (representing the actual wave propagation) produced upon an intersection of the Fig. 4 configuration with a hyperplane travelling normal to the t axis at unit velocity from $t = -\infty$ to $t = \infty$.

Ever since the source is triggered at instant $t = -\infty$, it emits a weak effect E. This has completely permeated the infinite surrounding medium by the time the source peaks during $t = 0$. The peaking causes two spherical fronts to separate concentrically from the spherical source base $|\mathbf{x}| = r$. One front expands with unit speed, its path being the characteristic hypersurface C; we shall refer to it as the C-front. The other,

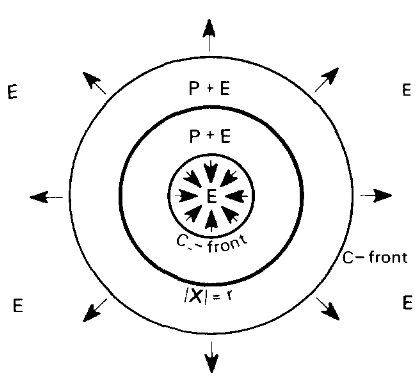


FIG. 5. The physical propagation scheme during the interval $0 < t < r$. Radial arrows indicate expansion or contraction of the spherical fronts.

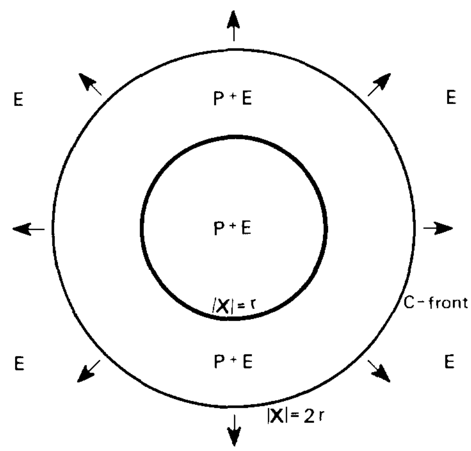


FIG. 6. The outline at instant $t = r$.

a C_--front contracts with unit speed along the characteristic hypercone C_-. Between them, a strong P field diverges into and is, thereby, superposed upon the weak E effect. Figure 5 depicts the propagation pattern before time $t = r$.

The retreat of both C- and C_--fronts away from the source base $|\mathbf{x}| = r$ agrees with *Sommerfeld's radiation principle* (see, e.g., Stoker, Ref. 10 and also Refs. 2, 3, and 11). It represents a natural consequence of our preliminary postulate (2.1), equivalently, that radiation never precedes, but succeeds source activation. The phenomenon is also compatible with the fact that the hypercone fields, obviously induced by peaking, are recorded only after the latter's attainment (see Fig. 4). Such an aftereffect stems again from (2.1), in this instance, through contour integration.

The C_--front converges at the center $\mathbf{x} = 0$ when $t = r$. By this time, the C-front has achieved a radius of $2r$ (see Fig. 6) and continues to expand with further time increase. However the shrinking C_--front is then replaced by a C_+-front. This originates at the exact instant $t = r$ and focus $\mathbf{x} = 0$ of convergence of the C_--front, and thereafter expands with unit speed along the characteristic hypercone C_+. The strong field Q appears inside the C_+-front. If n is even, $Q \not\equiv 0$, and the convertex $(\mathbf{x}, t) = (0, r)$ corresponds to *an instantaneous point filter through which a converging part of P converts into Q*; in fact, via the rule[12]

$$\tfrac{1}{2}\pi \sin(\nu\pi) P_\nu(\cos\theta) = Q_\nu(-\cos\theta) + \cos(\nu\pi) Q_\nu(\cos\theta),$$

it can be shown from (6.5)–(6.9) that

$$P = -Q \quad \forall(\mathbf{x}, t) \in D_P \quad (n \text{ even}),$$

and so the actual conversion is, precisely, a sign change. On the other hand, if n is odd, $Q \equiv 0$, i.e., $(\mathbf{x}, t) = (0, r)$ corresponds to *an instantaneous point sink through which the converging part of P disappears*; furthermore, *the C_+-front encloses a spherical zone somewhat analogous to a Petrowsky's lacuna of silence.*[13-20] Normally, if singularities arise, they would be confined to the C- and $C_\pm$-fronts. Thus, in particu-

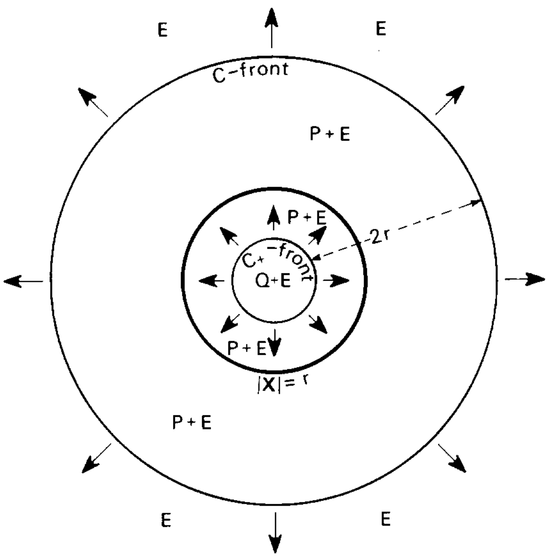


FIG. 7. Case: $r<t<2r$.

lar, the possibly singular filter or sink action at $\mathbf{x}=0$ occurs only at the simultaneous instant of C_- annihilation and C_+ creation. Before and after this instant, the solution at $\mathbf{x}=0$ [i.e., off the covertex $(0,r)$ but, otherwise, along the t axis in Fig. 4] is analytic and given by, respectively, (6.21) and (6.22).

The outer C-front remains intact throughout its progress. Together with the C_+-front, it bounds an advancing spherical layer of constant thickness $2r$, within which P now gets transported. The full radiation field thus described is superposed upon the weak E effect (see Fig. 7 for $r<t<2r$). The C_+-front crosses the source base $|\mathbf{x}|=r$ when $t=2r$, after which the overall emission picture maintains a permanent development *ad infinitum*.

VIII. THE SPHERICALLY SYMMETRIC PROBLEM

Suppose, regarding the source, its spatial density $\chi(\hat{\mathbf{x}})\equiv 1$, which is independent of the direction $\hat{\mathbf{x}}$ or *spherically symmetric* about the center $\mathbf{x}=0$. Then by applying the rule (2.15) to (6.10), we get

$$\chi_m(\hat{\mathbf{x}})=2\pi^{(n-1)/2}(m+\tfrac{1}{2}n-1)\frac{\Gamma(\frac{1}{2}n-1)}{\Gamma(\frac{1}{2}n-\frac{1}{2})}\times\int_{-1}^{1}(1-\xi^2)^{(n-3)/2}C_m^{n/2-1}(\xi)d\xi. \quad (8.1)$$

But if $\nu>-\frac{1}{2}$ [Ref. 6, Sec. 11.5(8)],

$$\int_0^\pi e^{iz\cos\psi}C_m^\nu(\cos\psi)\sin^{2\nu}\psi\, d\psi=\frac{i^m 2^\nu\sqrt{\pi}\Gamma(\nu+\frac{1}{2})\Gamma(2\nu+m)}{m!\,\Gamma(2\nu)}\frac{J_{\nu+m}(z)}{z^\nu}. \quad (8.2)$$

Now put $z=0$, and substitute the integration variable to enable comparison with (8.1):

$$\chi_0(\hat{\mathbf{x}})\equiv 2\pi^{(1/2)n},\quad \chi_m(\hat{\mathbf{x}})\equiv 0\quad (m=1,2,\cdots). \quad (8.3)$$

Thus, (6.8) and (6.9) reduce to

$$P^*=2\pi^{n/2}P_{(n/2)-3/2}(\cos\theta),$$
$$Q^*=-2\pi^{n/2}Q_{(n/2)-3/2}(\cosh A); \quad (8.4)$$

whilst from (6.17),

$$E=(r|\mathbf{x}|)^{1-n/2}E_0(|\mathbf{x}|,t) \quad (8.5)$$

and obeys (6.16) with $\max_{\hat{\mathbf{x}}\in\Omega_n}|\chi(\hat{\mathbf{x}})|=1$. The solution ϕ is determined by employing these results in (6.1)–(6.3), (6.5), and (6.6). Evidently, ϕ is spherically symmetric about $\mathbf{x}=0$; so it is symmetrical about the t axis in hyperspace. Otherwise, the discussions in Sec. 7 hold.

If n is odd, (8.4) and (6.7) reveal that P^* is a polynomial in t of degree $n-3$; hence (6.5) degenerates to a finite series:

$$P=\frac{1}{(r|\mathbf{x}|)^{(n-1)/2}}\sum_{l=0}^{(n-3)/2}\frac{(\kappa^2+\tau^2)^l}{2^{2l+1}l!}\frac{\partial^{2l}P_{(n/2)-3/2}(\cos\theta)}{\partial t^{2l}}. \quad (8.6)$$

The generating spatial distribution is approximated by (3.5). In the present situation, its exact form can be conveniently established as follows. Again, use (2.15), this time, to transform (3.7) to an ξ-line integral involving the integrand factor $\exp(\lambda\xi)$. This integral can be easily tackled by applying (2.16) together with the fact (cf. Ref. 6, Secs. 3.62 and 3.7) that over $-\pi<\arg z\le\frac{1}{2}\pi$,

$$e^{(1/2)\nu\pi i}J_\nu(ze^{-(1/2)\pi i})=e^{-(1/2)\nu\pi i}J_\nu(ze^{(1/2)\pi i})=I_\nu(z), \quad (8.7)$$

the modified Bessel function of the first kind. Thereupon, substitution of the subsequent result for $F(\mathbf{x})$ into (3.6) yields the exact value

$$f(\mathbf{x})=2\kappa^{-2}I_{(1/2)n-1}(2r|\mathbf{x}|/\kappa^2)\times\exp[-(\mathbf{x}^2+r^2)/\kappa^2](r|\mathbf{x}|)^{1-(1/2)n} \quad (8.8)$$

which, as expected, is spherically symmetric. Consistency with (3.5) exists by virtue of the asymptotic approximation (Ref. 6, Sec. 7.23)

$$I_{(1/2)n-1}(2r|\mathbf{x}|/\kappa^2)\sim(4\pi r|\mathbf{x}|/\kappa^2)^{-1/2}\exp(2r|\mathbf{x}|/\kappa^2). \quad (8.9)$$

IX. THE AXISYMMETRIC PROBLEM

Suppose the spatial density is axisymmetric about the x_1 axis in $\mathbf{x}$ space or, equivalently,

$$\chi(\hat{\mathbf{x}})=X(\theta_1), \quad (9.1)$$

θ_1 being the colatitude: $\hat{x}_1=\cos\theta_1$ $(0\le\theta_1\le\pi)$. In the subsequent analysis, $n\ge 3$.

We shall first establish a preliminary result. Consider the $(n-1)$-dimensional vector $\hat{\mathbf{z}}$ satisfying $\hat{\mathbf{z}}\sin\theta_1=(\hat{x}_2,\hat{x}_3,\ldots,\hat{x}_n)$. Then [cf. (3.8)–(3.11)], $\hat{\mathbf{z}}$ is a unit vector in the $(x_2,x_3,\ldots,x_n)$ frame. Likewise, if η is related to the unit vector $\boldsymbol{\zeta}=(\zeta_1,\ldots,\zeta_n)$ by $\eta\sin\psi_1=(\zeta_2,\zeta_3,\ldots,\zeta_n)$, then η is also an $(n-1)$-dimensional unit vector. Its end ranges over the $(n-1)$-dimensional unit sphere Ω_{n-1} with surface element $d\Omega_\eta$, say. So, following the method of Sec. 3, we have

$$\int_{\Omega_n} X(\psi_1) Y(\mathbf{x}\cdot\boldsymbol{\zeta}) d\Omega_\zeta$$

$$= \int_0^\pi \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} X(\psi_1) Y(\mathbf{x}\cdot\boldsymbol{\zeta}) \prod_{\nu=1}^{n-1} \sin^{n-1-\nu}\psi_\nu d\psi_\nu \qquad (9.2)$$

$$= \int_0^\pi X(\psi) \sin^{n-2}\psi\, d\psi$$

$$\times \int_{\Omega_{n-1}} Y(|\mathbf{x}|\cos\theta_1\cos\psi + |\mathbf{x}|\hat{\mathbf{z}}\cdot\boldsymbol{\eta}\sin\theta_1\sin\psi) d\Omega_\eta$$

$$= \frac{2\pi^{(n/2)-1}}{\Gamma(\frac{1}{2}n-1)} \int_0^\pi X(\psi)\sin^{n-2}\psi\, d\psi \int_{-1}^{1} (1-\eta^2)^{(n-4)/2}$$

$$\times Y(|\mathbf{x}|\cos\theta_1\cos\psi + |\mathbf{x}|\eta\sin\theta_1\sin\psi) d\eta, \qquad (9.3)$$

after applying the rule (2.15) to the integral over Ω_{n-1}.

The spherical integral in (6.10) is of the type (9.2) on account of (9.1) and therefore, in accordance with (9.3), becomes

$$\frac{2\pi^{(n/2)-1}}{\Gamma(\frac{1}{2}n-1)} \int_0^\pi X(\psi)\sin^{n-2}\psi\, d\psi$$

$$\times \int_{-1}^{1} (1-\eta^2)^{(n-4)/2}\, C_m^{(n/2)-1}(\cos\theta_1\cos\psi + \eta\sin\theta_1\sin\psi) d\eta,$$

wherein the inner integral can be resolved via a certain formula (Ref. 6, Sec. 11.5). Thereupon the coefficient

$$\chi_m(\hat{\mathbf{x}}) = 2^{n-2}\pi^{(n/2)-1} m!\,(m+\tfrac{1}{2}n-1)[\Gamma(\tfrac{1}{2}n-1)]^2$$

$$\times [\Gamma(m+n-2)]^{-1}\, C_m^{(n/2)-1}(\cos\theta_1)$$

$$\times \int_0^\pi X(\psi) C_m^{(n/2)-1}(\cos\psi)\sin^{n-2}\psi\, d\psi. \qquad (9.4)$$

The solution ϕ is now fully described by (6.1)–(6.3) with the aid of (6.5)–(6.9) and either (6.14), (6.16), or (6.17). As (9.4) indicates, $\chi_m(\hat{\mathbf{x}})$ is axisymmetric; hence, so is ϕ.

As with spherical symmetry, the source distribution can be exactly represented. We start from (3.7) and use (9.1)–(9.3) to obtain

$$F(\mathbf{x}) = \frac{2\pi^{(n/2)-1}}{\Gamma(\frac{1}{2}n-1)} \int_0^\pi X(\psi)\exp(\lambda\cos\theta_1\cos\psi)\sin^{n-2}\psi\, d\psi$$

$$\times \int_{-1}^{1} (1-\eta^2)^{1/2(n-4)/2} \exp(\lambda\eta\sin\theta_1\sin\psi) d\eta,$$

$$= \frac{(2\pi)^{(n-1)/2}}{(\lambda\sin\theta_1)^{(n-3)/2}} \int_0^\pi X(\psi) I_{(n/2)-3/2}(\lambda\sin\theta_1\sin\psi)$$

$$\times \exp(\lambda\cos\theta_1\cos\psi)(\sin\psi)^{1/2(n-1)/2}\, d\psi, \qquad (9.5)$$

by virtue of (2.16) and (8.7). The distribution $f(\mathbf{x})$ is then determined from (3.6). Like its density $X(\theta_1)$, $f(\mathbf{x})$ is axisymmetric. The factor $F(\mathbf{x})$ can be expanded by first applying to (9.5), the series [Ref. 6, Sec. 11.5, Eq. (9)]

$$\frac{J_{\nu-1/2}(z\sin\psi\sin\psi')}{(z\sin\psi\sin\psi')^{\nu-1/2}} \exp(iz\cos\psi\cos\psi')$$

$$= \frac{2^{2\nu}[\Gamma(\nu)]^2}{\sqrt{2\pi}} \sum_{m=0}^{\infty} \frac{i^m m!\,(m+\nu)}{\Gamma(m+2\nu)} \frac{J_{m+\nu}(z)}{z^\nu} C_m^\nu(\cos\psi) C_m^\nu(\cos\psi') \qquad (9.6)$$

after reverse accommodation of (8.7), and then identifying each subsequent coefficient with (9.4). Thus, we finally arrive at

$$F(\mathbf{x}) = (2/\lambda)^{(n/2)-1} \sum_{m=0}^{\infty} \chi_m(\hat{\mathbf{x}}) I_{m+(n/2)-1}(\lambda). \qquad (9.7)$$

So, the expansions of both $f(\mathbf{x})$ and ϕ involve the same set of χ_m coefficients computable from (9.4).

X. CERTAIN FEATURES OF THE FUNCTIONS P^* AND Q^*

Within the hyperdomain D_P (Fig. 4), the infinite series (6.5) for P involves the function P^*. We are interested in its behavior on the D_P side of the boundary hypersurfaces $C, C_\pm$. With θ defined by (6.7), we have

$\cos\theta = 1$ along C or C_-, $= -1$ along C_+.

A crucial point in our analysis is the expansion (Ref. 12, Sec. 4.8):

$$(1-2xy+y^2)^{-\nu} = \sum_{m=0}^{\infty} C_m^\nu(x) y^m. \qquad (10.1)$$

Since $C_m^{1/2}(x) = P_m(x)$, therefore $P_m(1) \equiv 1$ and $P_m(-1) \equiv (-1)^m$, which are known facts. We now concentrate only on the case where n is odd. Whence by (6.8) and (6.10), from the D_P side, when $\cos\theta = \pm 1$:

$$P^* = (\pm 1)^{(n-3)/2}\Gamma(\tfrac{1}{2}n-1)$$

$$\times \int_{\Omega_n} \chi(\boldsymbol{\zeta}) \left[\sum_{m=0}^{\infty} (\pm 1)^m (m+\tfrac{1}{2}n-1) C_m^{(n/2)-1}(\hat{\mathbf{x}}\cdot\boldsymbol{\zeta})\right] d\Omega_\zeta. \qquad (10.2)$$

Now, by combining (10.1) with its y derivative multiplied by y, we get

$$\sum_{m=0}^{\infty} (m+\nu) C_m^\nu(x) y^m = \nu(1-2xy+y^2)^{-\nu}$$

$$\times [1 + 2y(x-y)(1-2xy+y^2)^{-1}], \qquad (10.3)$$

and this vanishes identically if $y = \pm 1$. Whereupon, from (10.2),

$$P^* \equiv 0 \text{ on the } D_P \text{ side of } C, C_\pm. \qquad (10.4)$$

An important corollary may be drawn from the results (6.26)–(6.28), viz., *for the spherical sheet impulse, its radiation field in an odd n-dimensional space is continuous across the C-, $C_\pm$-fronts and, in fact, vanishes along both sides of each of these fronts*. So, generally, the C- and $C_\pm$-fronts need not convey singularities.

Suppose the observer is sufficiently far from and within the D_Q side of the hypercone C_+, i.e., with reference to (6.7), $\cosh A \gg 1$. The quantity Q^* recorded may then be asymptotically approximated. Now (Ref. 9, Sec. 15.31),

$$Q_\nu(z) = 2^{-\nu} z^{-\nu-1} \int_0^1 (1-\xi^2)^\nu d\xi + O(z^{-\nu-3})$$

$$(|z| \gg 1; \nu > -1). \qquad (10.5)$$

Hence, via (6.9),

$$Q^* \sim -\chi_0(\hat{\mathbf{x}}) Q_{(n/2)-3/2}(\cosh A), \qquad (10.6)$$

$$\sim -\sqrt{\pi}\,\Gamma(\tfrac{1}{2}n-\tfrac{1}{2})(2\cosh A)^{(1-n)/2} \int_{\Omega_n} \chi(\boldsymbol{\zeta}) d\Omega_\zeta, \qquad (10.7)$$

after accounting for (6.10), (2.15), and (6.15). In particular, for even n, (6.28) implies that at each $\mathbf{x}$-position deep inside the expanding C_+-front, the field of the spherical sheet impulse develops asymptotically:

$$\phi \sim \frac{(-1)^{(n/2)+1}\Gamma(\frac{1}{2}n-\frac{1}{2})\int_{\Omega_n}\chi(\zeta)d\Omega_\zeta}{2\pi^{(n+1)/2}(t^2-\mathbf{x}^2-r^2)^{(n-1)/2}} \qquad (10.8)$$

whenever $t \gg |\mathbf{x}|+r$. Evidently this result approaches zero as $t\to\infty$, corresponding to *a gradual and ultimate steady state attainment of silence*. For odd n, but in the general situation with the fully Gaussian source, a somewhat similar physical phenomenon is experienced through a different mathematical route, viz., that everywhere inside the C_+-front,

$$\phi \equiv E \to 0 \text{ gradually as } t\to\infty; \qquad (10.9)$$

however, with impulsive generation but not necessarily restricted to a sheet (i.e., $\kappa \geq \tau = 0$): $E\equiv 0$, so that once the reception point is crossed by the C_+-front, *an immediate steady state of silence prevails.*

XI. MAGNETOACOUSTIC FLOW PAST A CYLINDRICAL GAUSSIAN-APPROXIMATED CURRENT DISTRIBUTION

In Lighthill's paper,[2] there is an introductory discussion on pulsatory magnetoacoustic (or MGD) excitations generated within a stationary gas by a fluid injecting source with *Gaussian strength loaded about a point.* The principles established in the present paper can be *indirectly* applied to examine magnetoacoustic flow past a current source with a different Gaussian distribution.

Consider a uniform state wherein an infinitely conducting gas with unit magnetic permeability and density ρ_0 flows with velocity $\mathbf{v}$ in the presence of a magnetic field $\mathbf{H}$; suppose c denotes the sound speed, while $\mathbf{a} = \mathbf{H}(4\pi\rho_0)^{-1/2}$ the Alfvén velocity whose magnitude $|\mathbf{a}| = a$. Small disturbances are being induced by a weak azimuthal electric current of density $\rho_0\mathbf{J}|\mathbf{H}|^{-1}$. Let $|\mathbf{H}|\mathbf{h}, \mathbf{q}, \rho_0 p$ denote perturbations in, respectively, the magnetic field, gas velocity, and pressure. Nonrelativistic linearized equations governing the perturbed motion are then

$$Dp/Dt = -c^2 \operatorname{div}\mathbf{q}, \quad \operatorname{div}\mathbf{h} = 0, \qquad (11.1)$$

$$\partial\mathbf{h}/\partial t = \operatorname{curl}(\mathbf{q}\times\mathbf{a}a^{-1}) + \operatorname{curl}(\mathbf{v}\times\mathbf{h}), \qquad (11.2)$$

$$D\mathbf{q}/Dt + \operatorname{grad}p + a\mathbf{a}\times\operatorname{curl}\mathbf{h} = \mathbf{J}\times\mathbf{a}a^{-1}, \qquad (11.3)$$

where $D/Dt \equiv \partial/\partial t + \mathbf{v}\cdot\operatorname{grad}$. We refer to a three-dimensional Cartesian frame with typical position $(x_1, x_2, z) = (\mathbf{x}, z)$, $\mathbf{x} = (x_1, x_2)$ denoting a two-dimensional position in R_2, while $-\infty < z < \infty$. The positive z-direction is aligned with the Alfvén velocity:

$$\mathbf{a} = (0, 0, a). \qquad (11.4)$$

The flow velocity is chosen to be aligned with or opposed to this direction:

$$\mathbf{v} = (0, 0, v) \text{ with } v > 0 \text{ or } v < 0. \qquad (11.5)$$

The azimuthal source current, circulating anticlockwise, say, exerts a *transverse* electromagnetic body force measured, per unit mass, by the vector

$$\mathbf{J}\times\mathbf{a}a^{-1} \equiv |\mathbf{J}|(\hat{\mathbf{x}}, 0) = (\mathrm{Z}, 0), \text{ say}, \qquad (11.6)$$

$\hat{\mathbf{x}} = (\cos\theta_1, \sin\theta_1)$ being the unit radial vector in R_2; here $\mathrm{Z} = |\mathbf{J}|\hat{\mathbf{x}}$, a two-dimensional vector. We shall employ a source current with a cylindrical Gaussian-approximated distribution that is fairly concentrated about the mean circle $|\mathbf{x}| = r (>0), z = 0$:

$$|\mathbf{J}| \sim j(\hat{\mathbf{x}})(l\kappa\pi)^{-1}\exp[-z^2/l^2 - (|\mathbf{x}|-r)^2/\kappa^2](r|\mathbf{x}|)^{-1/2}, \qquad (11.7)$$

the approximation being only for a sufficiently small length-scale κ. When $\kappa = 0$, the representation becomes exact and corresponds to a cylindrical singular current sheet with a longitudinal Gaussian distribution. The length scale l is also assumed small, and $j(\hat{\mathbf{x}})$ denotes an azimuthally dependent scalar. The rapid decay, from the mean circle, of the current strength may, for example, be caused by electrical resistance along the conducting element.

We seek, ultimately, a steady state solution. This may be derived from an unsteady solution which has incorporated appropriate initial conditions. Unfortunately, the magnetoacoustic time-dependent problem is anisotropic and cannot be resolved by merely applying the principal results of this paper. Nonetheless, an indirect application is possible if we start differently. In the method we shall adopt, zero initial conditions will be avoided. In compensation, we shall impose, instead, a radiation condition. This corresponds physically to Sommerfeld's radiation principle (Sec. 7) which was not applied to the main problem, but follows naturally from initial conditions. Both are equivalent to saying that only the source emits radiation.

First let us propose a two dimensional vector $\mathbf{A}$ such that

$$p + a\mathbf{a}\cdot\mathbf{h} = -\operatorname{div}(\mathbf{A}, 0). \qquad (11.8)$$

Whence, it is easily seen that (11.3) may be interpreted via (11.6) as

$$D\mathbf{q}/Dt - a\mathbf{a}\cdot\operatorname{grad}\mathbf{h} = \operatorname{grad}\operatorname{div}(\mathbf{A}, 0) + (\mathrm{Z}, 0). \qquad (11.9)$$

The three-dimensional vector $(\mathbf{A}, 0)$ is analogous to the vector potential of classical electromagnetic theory.

The radiation condition is accommodated in accordance with Lighthill.[2,3,11] During an unsteady development, an exponential growth is imparted to the source, whereby the pair $\{\mathbf{J}, \mathrm{Z}\}$ becomes

$$\{\mathbf{J}, \mathrm{Z}\}e^{\epsilon t} = e^{\epsilon t}\int_{R_2} d\boldsymbol{\alpha}\int_{-\infty}^{\infty}\{\mathbf{J}^*, \mathrm{Z}^*\}\exp[i(\boldsymbol{\alpha}, \omega)\cdot(\mathbf{x}, z)]d\omega, \qquad (11.10)$$

in terms of Fourier transforms resembling H.F.T.'s [see (2.3)]:

$$\{\mathbf{J}^*, \mathrm{Z}^*\} = (2\pi)^{-3}\int_{R_2}\exp(-i\boldsymbol{\alpha}\cdot\mathbf{x})d\mathbf{x}\int_{-\infty}^{\infty}\{\mathbf{J}, \mathrm{Z}\}\exp(-i\omega z)dz. \qquad (11.11)$$

Throughout, $\epsilon > 0$. All induced perturbations are allowed to acquire in-phase exponential intensifications. Precisely, the set

$$\{p, \mathbf{q}, \mathbf{h}, \mathbf{A}\} = e^{\epsilon t}\int_{R_2} d\boldsymbol{\alpha}\int_{-\infty}^{\infty}\{p^*, \mathbf{q}^*, \mathbf{h}^*, \mathbf{A}^*\} \times\exp[i(\boldsymbol{\alpha}, \omega)\cdot(\mathbf{x}, z)]d\omega. \qquad (11.12)$$

By the time free perturbations from infinity arrive within observation range, they become negligibly small compared with source generated $O(e^{\epsilon t})$ quantities and therefore tend to escape detection near the steady state. Equations are now transformed accordingly, and with (11.4) and (11.5) accounted for. Observe that D/Dt transforms into $i\omega v_\epsilon$ with $v_\epsilon = v - i\epsilon\omega^{-1}$. Thus, if h_3 denotes the z-component of $\mathbf{h}$, (11.8) and (11.9) become

$$p^* + a^2 h_3^* = -i(\boldsymbol{\alpha}, \omega)\cdot(\mathbf{A}^*, 0) \equiv -i\mathbf{A}^*\cdot\boldsymbol{\alpha}, \qquad (11.13)$$

$$\omega(v_\epsilon \mathbf{q}^* - a^2\mathbf{h}^*) = i(\boldsymbol{\alpha}, \omega)\mathbf{A}^*\cdot\boldsymbol{\alpha} - i(\mathbf{Z}^*, 0). \qquad (11.14)$$

From (11.13), the third components of (11.14) and (11.2), as well as both equations in (11.1), we have

$$v_\epsilon^3 \omega h_3^* = (c^2 - v_\epsilon^2)(\boldsymbol{\alpha}, \omega)\cdot\mathbf{q}^*. \qquad (11.15)$$

Accounting for (11.1) again, it can then be proven that (11.13)–(11.15) are compatible for every $\boldsymbol{\alpha}$ in R_2 if

$$[\alpha^2 - \omega^2 m^2(v_\epsilon)]\mathbf{A}^* = \mathbf{Z}^*, \qquad (11.16)$$

wherein

$$m(v_\epsilon) = m(v - i\epsilon\omega^{-1}) = \left[\frac{(v_\epsilon^2 - a^2)(v_\epsilon^2 - c^2)}{v_\epsilon^2(a^2 + c^2) - a^2c^2}\right]^{1/2}, \qquad (11.17)$$

and $\alpha = |\boldsymbol{\alpha}|$. Hereafter, we confine attention to

either $v > \max(a, c)$, (11.18)

or $-\min(a, c) < v < -ac(a^2 + c^2)^{-1/2}$, (11.19)

corresponding, respectively, to either a *supersonic–super-Alfvênic* flow in the *positive z* direction, or a *restricted subsonic–sub-Alfvênic* flow in the *negative z* direction. So by (11.17), $\lim_{\epsilon\to 0} m(v_\epsilon) = m(v) = m$, say, is *real and positive*. By (11.16), (11.12) yields

$$\mathbf{A} = e^{\epsilon t}\int_{R_2} d\boldsymbol{\alpha}\int_{-\infty}^{\infty} \frac{\mathbf{Z}^*\exp[i(\boldsymbol{\alpha}, \omega)\cdot(\mathbf{x}, z)]}{\alpha^2 - \omega^2 m^2(v - i\epsilon\omega^{-1})}\, d\omega. \qquad (11.20)$$

Bearing in mind the arguments associated with (2.12) and (3.2), we introduce a two-dimensional vector

$$\mathbf{f}(\mathbf{x}) = (\kappa\sqrt{\pi})^{-2}\int_{R_2} \boldsymbol{\rho}(\mathbf{y})\exp[-(\mathbf{x} - \mathbf{y})^2/\kappa^2]d\mathbf{y}, \qquad (11.21)$$

and select

$$\boldsymbol{\rho}(\mathbf{x}) = \hat{\mathbf{x}} j(\hat{\mathbf{x}})\delta(|\mathbf{x}| - r)|\mathbf{x}|^{-1}. \qquad (11.22)$$

Then, by virtue of (3.5), (11.7) is consistent with (11.6) if

$$\mathbf{Z} = (l\sqrt{\pi})^{-1}\exp(-z^2/l^2)\mathbf{f}(\mathbf{x}). \qquad (11.23)$$

According to Lighthill, each perturbation function becomes unique in the steady state if, prior to this, we approximate its multiple Fourier integral for small ϵ, evaluate the approximation, and then let $\epsilon \to 0$; in practice, the limit $\epsilon = 0$ may be taken after any innermost integration, e.g., the ω integration with regards to (11.20). Thus, via (11.11), (11.20), and (11.23), we assert that, in the steady state

$$m\mathbf{A} = \int_{R_2}\left[(2\pi)^{-2}\int_{R_2}\mathbf{f}(\mathbf{y})\exp(-i\boldsymbol{\alpha}\cdot\mathbf{y})d\mathbf{y}\right] \times \lim_{\epsilon\to 0} I(z|\alpha)\exp(i\boldsymbol{\alpha}\cdot\mathbf{x})d\boldsymbol{\alpha}, \qquad (11.24)$$

where $I(z|\alpha)$ may be accepted in the form

$$I(z|\alpha) = (2\pi)^{-1}\int_{-\infty}^{\infty}\frac{\exp(i\omega z m^{-1} - \frac{1}{4}\omega^2\tau^2)d\omega}{(\alpha + i\epsilon m' - \omega)(\alpha - i\epsilon m' + \omega)} \qquad (11.25)$$

wherein the integration variable ω derives from that in (11.20) via a scale change; also the new length-scale $\tau = lm^{-1}$, while the derivative

$$m' = m'(v) = \frac{v^3[a^2(v^2 - c^2) + c^2(v^2 - a^2)]}{m[v^2(a^2 + c^2) - a^2c^2]^2}, \qquad (11.26)$$

which stays *positive* under (11.18) or (11.19).

Regarding (11.25), integration is performed along a *real* path. There are two integrand singularities, precisely, simple poles; these are *complex* and occur at

$$\omega = \alpha + i\epsilon m', \quad -\alpha + i\epsilon m' \quad \text{slightly above } \operatorname{Im}\omega = 0. \qquad (11.27)$$

Evidently, there are essential differences between representations (2.11) and (11.25). In both situations however, the poles appear above the respective integral paths. For our present real path, we may therefore employ a contour deformation similar to that indicated in Fig. 1 or Fig. 3. If $z \gtreqless 0$, the present deformation is directed into $\operatorname{Im}\omega \gtreqless 0$ and crosses/avoids both complex poles. When appropriately extended to infinity, the deformed contour comprises two circular arcs (which contribute nothing to the subsequent integration) joined to the horizontal path $(-\infty + i2zm^{-1}\tau^{-2},\ \infty + i2zm^{-1}\tau^{-2})$. Eventually, we arrive at

$$I(z|\alpha) = iH(z)(\text{relevant residues}) + (2\pi)^{-1}\exp(-z^2m^{-2}\tau^{-2}) \times\int_{-\infty}^{\infty}\frac{\exp(-\frac{1}{4}\sigma^2\tau^2)d\sigma}{\alpha^2 - (\sigma + i2zm^{-1}\tau^{-2} - i\epsilon m')^2}, \qquad (11.28)$$

where, in particular, it can be easily demonstrated that

$$\lim_{\epsilon\to 0}(\text{relevant residues}) = (i\alpha)^{-1}\sin(\alpha z m^{-1})\exp(-\tfrac{1}{4}\alpha^2\tau^2). \qquad (11.29)$$

For $z = 0$, no deformation is necessary, and

$$I(0|\alpha) = (2\pi)^{-1}\int_{-\infty}^{\infty}\frac{\exp(-\frac{1}{4}\omega^2\tau^2)d\omega}{\alpha^2 - (\omega - i\epsilon m')^2}, \qquad (11.30)$$

which must be interpreted in the sense of a principal value at the limit $\epsilon = 0$. By comparing (11.28)–(11.30) with (4.12), (4.13), (4.15), and (4.16), we deduce that the limiting value

$$\lim_{\epsilon\to 0} I(z|\alpha) \equiv L(zm^{-1}|\alpha) \quad \text{over } -\infty < z < \infty, \qquad (11.31)$$

with $L(t|\alpha)$ defined by (4.12) or (4.15) depending on whether $t \neq 0$ or $t = 0$.

In view of (11.31) and (2.6), formulas (11.24) and (2.8) are related. The exact relationship can be identified by comparing (11.21) and (11.22) with (2.12) and (3.2). Whereupon, we deduce that the steady state value of $m\mathbf{A}$ can be derived from the earlier unsteady R_2 ϕ-solution to our main problem by merely substituting zm^{-1} and $\hat{\mathbf{x}}j(\hat{\mathbf{x}})$ for t and $\chi(\hat{\mathbf{x}})$, respectively. Suppose the three-dimensional physical domains D, D_P, and D_Q are derived from the respective hyperdomains $\mathcal{D}$, $\mathcal{D}_P$, and

D_Q illustrated in Fig. 4 by substituting zm^{-1} for t in the equations governing the hyperconical boundaries C, C_+, C_-. Clearly, D, D_P, and D_Q maintain a proportionate similarity to D, D_P, and D_Q, respectively. Via (6.6), (6.7), (6.9), (6.12)–(6.14), (7.1)–(7.3), and (7.5), we draw the following conclusion:

$$m\mathbf{A} = \begin{cases} \mathbf{E} \quad \forall(\mathbf{x}, z) \in D, & (11.32) \\ -\mathbf{Q} + \mathbf{E} \quad \forall(\mathbf{x}, z) \in D_P, & (11.33) \\ \mathbf{Q} + \mathbf{E} \quad \forall(\mathbf{x}, z) \in D_Q, & (11.34) \end{cases}$$

where

$$\mathbf{Q} = \frac{1}{\pi^2(r\,|\mathbf{x}|)^{1/2}} \sum_{\mu=0}^{\infty} \frac{(\kappa^2+\tau^2)^\mu m^{2\mu}}{2^{2\mu+1}\mu!} \times \sum_{\nu=0}^{\infty} (-1)^\nu \mathbf{j}_\nu(\hat{\mathbf{x}}) \frac{\partial^{2\mu}}{\partial z^{2\mu}} Q_{\nu-1/2}\left(\frac{z^2 m^{-2} - \mathbf{x}^2 - r^2}{2r\,|\mathbf{x}|}\right), \quad (11.35)$$

with

$$\mathbf{j}_0(\hat{\mathbf{x}}) = \int_0^{2\pi} \boldsymbol{\zeta} j(\boldsymbol{\zeta}) d\psi: \quad \boldsymbol{\zeta} = (\cos\psi, \sin\psi), \quad (11.36)$$

$$\mathbf{j}_\nu(\hat{\mathbf{x}}) = 2\int_0^{2\pi} \boldsymbol{\zeta}\, j(\boldsymbol{\zeta}) \cos[\nu(\psi - \theta_1)] d\psi \quad (\nu \geq 1), \quad (11.37)$$

while

$$|\mathbf{E}| < \frac{\|j\|\tau \exp(-z^2 m^{-2}\tau^{-2})}{2^{3/2} z^2 m^{-2}\pi} \quad (z \neq 0), \quad (11.38)$$

with

$$\|j\| = \left(\int_0^{2\pi} |j(\boldsymbol{\zeta})|^2 d\psi\right)^{1/2}. \quad (11.39)$$

Here, (11.38) follows from (6.14) through the fact that

$$\left(\int_0^{2\pi} |j(\boldsymbol{\zeta})|\,|\cos\psi|\, d\psi\right)^2 + \left(\int_0^{2\pi} |j(\boldsymbol{\zeta})|\,|\sin\psi|\, d\psi\right)^2 < 2\pi\|j\|^2,$$

by the Cauchy–Schwarz inequality.

Let us discuss the constituents of $m\mathbf{A}$. Bounded by conical sheets (of the type C, C_+, C_-—see Fig. 4) are two adjacent opposing strong fields, viz., $-\mathbf{Q}$ in D_P and $+\mathbf{Q}$ in D_Q. Superposed is a weak $\mathbf{E}$ field permeating all space and satisfying (11.38). The containing domains D_P and D_Q project *ad infinitum* in the positive z direction, i.e., *downstream* for a supersonic–super-Alfvénic flow, but *upstream* for a restricted subsonic–sub-Alfvénic flow. This phenomenon follows mathematically from the fact that both poles expressed by (11.27) lie within $\mathrm{Im}\omega > 0$. It is therefore a consequence of the applied radiation condition and can be explained in terms of wave propagation.

Consider an associated free motion in the absence of any source, in particular

$$\mathbf{J} \equiv 0, \quad \mathbf{Z} \equiv 0. \quad (11.40)$$

Then, as a basis within the context of (11.8) and (11.9),

$$\mathbf{A} = \mathbf{A}^* \exp\{i[(\boldsymbol{\alpha}, \omega)\cdot(\mathbf{x}, z) - \sigma t]\} \quad (11.41)$$

constitutes an admissible travelling wave function provided [via comparison of (11.41) with (11.12)] (11.16) remains satisfied with $\mathbf{Z}^* \equiv \mathbf{0}$ and ϵ replaced by $-i\sigma$. Hence, for $\mathbf{A} \neq \mathbf{0}$,

$$\omega^2 m^2(v - \sigma\omega^{-1}) = \alpha^2, \quad (11.42)$$

a dispersion relation. With reference to a three-dimensional cylindrical frame, the group velocity of wave energy propagation has a z component $\partial\sigma/\partial\omega$, a transverse (radial) component $\partial\sigma/\partial\alpha$, and, denoting the azimuthal angle in $(\boldsymbol{\alpha}, \omega)$-space by ψ_1, an azimuthal component $\alpha^{-1}\partial\sigma/\partial\psi_1$ which is clearly zero by (11.42). It is easily established that

$$\frac{\partial\sigma}{\partial\omega} = \frac{m(v - \sigma\omega^{-1})}{m'(v - \sigma\omega^{-1})} + \frac{\sigma}{\omega}, \quad (11.43)$$

$$\frac{\partial\sigma}{\partial\alpha} = \frac{-\alpha}{\omega m(v - \sigma\omega^{-1}) m'(v - \sigma\omega^{-1})} = \mp\frac{1}{m'(v - \sigma\omega^{-1})} \quad (11.44)$$

at a $\pm$ root to (11.42). For a stationary wave at the steady state, $\sigma = 0$, we get

$$\partial\sigma/\partial\omega = m(v)/m'(v) > 0 \quad \text{and} \quad \partial\sigma/\partial\alpha = \mp 1/m'(v), \quad (11.45)$$

implying that the particular group velocity vector is inclined to the positive z direction and is parallel to generators of the conical boundaries for D_P and D_Q. Therefore, $-\mathbf{Q}$, $+\mathbf{Q}$ *represent stationary wave fields sustained by wave energy flux streaming steadily from the current source into* D_P *and* D_Q. The longitudinal Gaussian spread of the source distribution beyond $z = 0$, gives rise to a weak and diffusive $\mathbf{E}$ field that receives no wave energy.

XII. APPLICATION IN ELASTODYNAMICS

To demonstrate a relatively straightforward application, we consider an elastic displacement $\mathbf{u}$ produced from some initially uniform state:

$$\mathbf{u}(\mathbf{x}, T) = 0, \quad \mathbf{u}_t(\mathbf{x}, T) = 0 \quad (\mathbf{x} \in R_3) \quad (12.1)$$

within an isotropic medium. Suppose the generating agent is a body force $\mathbf{Z}$ having Gaussian time variation:

$$\mathbf{Z} = (\tau\sqrt{\pi})^{-1} \exp(-t^2/\tau^2)(\mathbf{f}_1 + \mathbf{f}_2) \quad (\tau > 0), \quad (12.2)$$

with

$$\mathbf{f}_1 = f_1(|\mathbf{x}|)\hat{\mathbf{x}}, \quad \mathbf{f}_2 = f_2(|\mathbf{x}|, \theta_1)\mathbf{i}_2, \quad (12.3)$$

where, relative to the R_3 Cartesian frame with typical position $\mathbf{x} = (x_1, x_2, x_3)$,

$$\hat{\mathbf{x}} = (\cos\theta_1, \sin\theta_1\cos\theta_2, \sin\theta_1\sin\theta_2) = \text{radial unit vector}, \quad (12.4)$$

while

$$\mathbf{i}_2 = (0, -\sin\theta_2, \cos\theta_2) = \text{azimuthal unit vector}, \quad (12.5)$$

θ_1 being the colatitude. Hence $\mathbf{Z}$ involves a *spherically symmetric radial component* $f_1(|\mathbf{x}|)$ plus an *axisymmetric azimuthal component* $f_2(|\mathbf{x}|, \theta_1)$. It is easily verified from spherical polar representations that

$$\nabla\times\mathbf{f}_1 = 0, \quad (12.6)$$

$$\nabla\cdot\mathbf{f}_2 = 0. \quad (12.7)$$

The relevant equation of motion is

$$\mathbf{u}_{tt} = (c_1^2 - c_2^2)\nabla(\nabla\cdot\mathbf{u}) + c_2^2\nabla^2\mathbf{u} + \mathbf{Z}, \quad (12.8)$$

c_1 and c_2 being, respectively, the dilatational and equivoluminal wave speeds. In particular, accounting for (12.2) and (12.7):

$$(\partial^2/\partial t^2 - c_1^2\nabla^2)\nabla\cdot\mathbf{u} = (\tau\sqrt{\pi})^{-1}\exp(-t/\tau^2)\nabla\cdot\mathbf{f}_1. \qquad (12.9)$$

So if $\mathbf{u}^{(1)}$ satisfies

$$\mathbf{u}_{tt}^{(1)} - c_1^2\nabla^2\mathbf{u}^{(1)} = (\tau\sqrt{\pi})^{-1}\exp(-t^2/\tau^2)\mathbf{f}_1, \qquad (12.10)$$

and

$$\mathbf{u}^{(1)}(\mathbf{x},T)=0,\ \mathbf{u}_t^{(1)}(\mathbf{x},T)=0, \qquad (12.11)$$

then

$$\mathbf{u}=\mathbf{u}^{(1)}+\mathbf{u}^{(2)}, \qquad (12.12)$$

for some $\mathbf{u}^{(2)}$ satisfying

$$\nabla\cdot\mathbf{u}^{(2)}\equiv 0. \qquad (12.13)$$

In view of (12.6), (12.10) and (12.11) imply that

$$\nabla\times\mathbf{u}^{(1)}\equiv 0,\quad \text{so that } \nabla(\nabla\cdot\mathbf{u}^{(1)})\equiv\nabla^2\mathbf{u}^{(1)}. \qquad (12.14)$$

Consequently, incorporating (12.2) and (12.10), (12.8) yields

$$\mathbf{u}_{tt}^{(2)} - c_2^2\nabla^2\mathbf{u}^{(2)} = (\tau\sqrt{\pi})^{-1}\exp(-t^2/\tau^2)\mathbf{f}_2. \qquad (12.15)$$

By (12.1) and (12.11), we observe that

$$\mathbf{u}^{(2)}(\mathbf{x},T)=0,\quad \mathbf{u}_t^{(2)}(\mathbf{x},T)=0. \qquad (12.16)$$

Now, (12.10) with (12.11) and (12.15) with (12.16) constitute two independent Cauchy-type problems in R_3 within the class covered by (1.1) and (1.2). To extract their solutions from our main results, we assume that the initial time $T=-\infty$ and employ a Gaussian-approximated distribution for $\mathbf{Z}$ about the spherical sheet $|\mathbf{x}|=r$: For a small length-scale κ,

$$f_1(|\mathbf{x}|)\sim(\kappa\sqrt{\pi})^{-1}\exp[-(|\mathbf{x}|-r)^2/\kappa^2](r|\mathbf{x}|)^{-1}, \qquad (12.17)$$

$$f_2(|\mathbf{x}|,\theta_1)\sim X_2(\theta_1)(\kappa\sqrt{\pi})^{-1}\exp[-(|\mathbf{x}|-r)^2/\kappa^2](r|\mathbf{x}|)^{-1}, \qquad (12.18)$$

$X_2(\theta_1)$ being an arbitrary density function. Hence (12.3) expresses vectorial forms within the class of (3.5). Evidently the solution for $c_\nu\mathbf{u}^{(\nu)}$ ($\nu=1,2$) can be derived from the ϕ solution of (7.1)–(7.3) in R_3 by substituting $c_\nu t$ and $c_\nu\tau$ for t and τ, respectively, as well as $\hat{\mathbf{x}}$ ($\nu=1$) or $\mathbf{i}_2X_2(\theta_1)$ ($\nu=2$) for $\chi(\hat{\mathbf{x}})$. Thus, via (6.5), (6.7), (6.8), (6.10), and (6.11), we deduce that if, for $\nu=1$ or 2,

$$\mathbf{p}_\nu=\frac{1}{\pi^{3/2}r|\mathbf{x}|}\sum_{l=0}^{\infty}\frac{(\kappa^2+c_\nu^2\tau^2)^l}{2^{2l+2}l!\,c_\nu^{2l}}\times\sum_{m=0}^{\infty}\chi_m^{(\nu)}(\hat{\mathbf{x}})\frac{\partial^{2l}}{\partial t^{2l}}P_m\left(\frac{r^2+\mathbf{x}^2-c_\nu^2t^2}{2r|\mathbf{x}|}\right), \qquad (12.19)$$

where

$$\chi_m^{(1)}(\hat{\mathbf{x}})=\pi^{1/2}(m+\tfrac{1}{2})\int_{\Omega_3}\zeta C_m^{1/2}(\hat{\mathbf{x}}\cdot\zeta)d\Omega_\zeta, \qquad (12.20)$$

$$\chi_m^{(2)}(\hat{\mathbf{x}})=\pi^{1/2}(m+\tfrac{1}{2})\int_{\Omega_3}X_2(\psi_1)C_m^{1/2}(\hat{\mathbf{x}}\cdot\zeta)(0,-\sin\psi_2,\cos\psi_2)d\Omega_\zeta, \qquad (12.21)$$

with

$$d\Omega_\zeta=\sin\psi_1 d\psi_1 d\psi_2\quad(0\le\psi_1\le\pi,\ 0\le\psi_2\le 2\pi), \qquad (12.22)$$

and

$$\zeta=(\cos\psi_1,\ \sin\psi_1\cos\psi_2,\ \sin\psi_1\sin\psi_2), \qquad (12.23)$$

then

$$c_\nu\mathbf{u}^{(\nu)}=\begin{cases}\mathbf{e}_\nu & \forall(\mathbf{x},t)\in D^{(\nu)}, & (12.24)\\ \mathbf{p}_\nu+\mathbf{e}_\nu & \forall(\mathbf{x},t)\in D_P^{(\nu)}, & (12.25)\\ \mathbf{e}_\nu & \forall(\mathbf{x},t)\in D_Q^{(\nu)}; & (12.26)\end{cases}$$

$D^{(\nu)}$, $D_P^{(\nu)}$, and $D_Q^{(\nu)}$ are four-dimensional hyperconical domains proportionally similar to, and derived from D, D_P, D_Q (see Fig. 4) by substituting $c_\nu t$ for t into the equations for the hyperconical boundaries C, C_+, and C_-. Therefore, $c_\nu\mathbf{u}^{(\nu)}$ possesses a propagation pattern resembling that described in Sec. 7 (with the aid of Figs. 5–7) for ϕ in the odd n situation. The quantity $\mathbf{p}_\nu$ represents a strong field while $\mathbf{e}_\nu$ represents a weak field: In accordance with (6.14),

$$|\mathbf{e}_1|<\frac{\sqrt{3}\exp(-t^2/\tau^2)}{2c_1^2t^2\pi}\quad(t\ne 0), \qquad (12.27)$$

$$|\mathbf{e}_2|<\frac{\|X_2\|\exp(-t^2/\tau^2)}{c_2^2t^2\pi^{3/2}}\quad(t\ne 0), \qquad (12.28)$$

where

$$\|X_2\|=\left(\int_0^\pi|X_2(\psi)|^2d\psi\right)^{1/2}. \qquad (12.29)$$

Here, (12.27) and (12.28) follow respectively from

$$\left(\int_{\Omega_3}|\cos\psi_1|d\Omega_\zeta\right)^2+\left(\int_{\Omega_3}|\sin\psi_1||\cos\psi_2|d\Omega_\zeta\right)^2+\left(\int_{\Omega_3}|\sin\psi_1||\sin\psi_2|d\Omega_\zeta\right)^2=12\pi^2,$$

$$\left(\int_{\Omega_3}|X_2(\psi_1)||\sin\psi_2|d\Omega_\zeta\right)^2+\left(\int_{\Omega_3}|X_2(\psi_1)||\cos\psi_2|d\Omega_\zeta\right)^2=32\left(\int_0^\pi|X_2(\psi_1)|\sin\psi_1 d\psi_1\right)^2\le 16\pi\|X_2\|^2.$$

The expressions (12.20) and (12.21) can be substantially simplified. First, we note that for any function Y,

$$\int_0^{2\pi}Y[\cos(\psi-\theta_2),\ \sin(\psi-\theta_2)]d\psi=\int_0^\pi[Y(\cos\psi,\ \sin\psi)+Y(\cos\psi,\ -\sin\psi)]d\psi. \qquad (12.30)$$

Thence it can be shown from (12.20) and (12.21) that if

$$A(\theta_1,\psi_1)=\int_0^\pi C_m^{1/2}(\cos\theta_1\cos\psi_1+\sin\theta_1\sin\psi_1\cos\psi)d\psi, \qquad (12.31)$$

$$B(\theta_1,\psi_1)=\int_0^\pi C_m^{1/2}(\cos\theta_1\cos\psi_1+\sin\theta_1\sin\psi_1\cos\psi)\cos\psi\, d\psi, \qquad (12.32)$$

then

$$\frac{\chi_m^{(1)}(\hat{\mathbf{x}})}{\pi^{1/2}(2m+1)}=(1,0,0)\int_0^\pi A(\theta_1,\psi_1)\sin\psi_1\cos\psi_1\, d\psi_1+(0,\ \cos\theta_2,\sin\theta_2)\int_0^\pi B(\theta_1,\psi_1)\sin^2\psi_1\, d\psi_1, \qquad (12.33)$$

$$\frac{\chi_m^{(2)}(\hat{\mathbf{x}})}{\pi^{1/2}(2m+1)}=\mathbf{i}_2\int_0^\pi X_2(\psi_1)B(\theta_1,\psi_1)\sin\psi_1\, d\psi_1. \qquad (12.34)$$

Hereafter, we appeal to the following rules (Ref. 12, Secs. 4.2, 4.3, 4.8):

$$C_m^{1/2}(\xi)=P_m(\xi),\quad P_0(\xi)=1,\quad P_1(\xi)=\xi, \tag{12.35}$$

$$P_m(\cos\theta_1\cos\psi_1+\sin\theta_1\sin\psi_1\cos\psi)$$
$$=P_m(\cos\theta_1)P_m(\cos\psi_1)$$
$$+2\sum_{l=1}^{m}\frac{(m-l)!}{(m+l)!}P_m^l(\cos\theta_1)P_m^l(\cos\psi_1)\cos(l\psi), \tag{12.36}$$

with

$$P_m^l(\xi)=(-1)^l(1-\xi^2)^{1/2}\,d^lP_m(\xi)/d\xi^l, \tag{12.37}$$

an associated Legendre function which, in particular, satisfies the orthogonality relations

$$\int_{-1}^{1}P_m^l(\xi)P_k^l(\xi)d\xi=\frac{2(m+l)!}{(2m+1)(m-l)!}$$
$$(k=m),\ O(k\neq m). \tag{12.38}$$

From (12.31)–(12.35), we immediately note that

$$\chi_0^{(1)}(\hat{\mathbf{x}})\equiv 0,\quad \chi_0^{(2)}(\hat{\mathbf{x}})\equiv 0. \tag{12.39}$$

Henceforth, we assume that the integer $m\geq 1$. Using (12.35) and (12.36), (12.31) and (12.32) yield

$$A(\theta_1,\psi_1)=\pi P_m(\cos\theta_1)P_m(\cos\psi_1),$$
$$B(\theta_1,\psi_1)=\pi m^{-1}(m+1)^{-1}P_m^1(\cos\theta_1)P_m^1(\cos\psi_1). \tag{12.40}$$

Whereupon, via (12.35), (12.37), and (12.38), (12.33) and (12.34) reduce to

$$\chi_1^{(1)}(\hat{\mathbf{x}})=2\pi^{3/2}\hat{\mathbf{x}},\quad \chi_m^{(1)}(\hat{\mathbf{x}})\equiv 0\quad (m\geq 2), \tag{12.41}$$

$$\chi_m^{(2)}(\hat{\mathbf{x}})=\mathbf{i}_2\frac{\pi^{3/2}(2m+1)}{m(m+1)}P_m^1(\cos\theta_1)$$
$$\times\int_0^{\pi}X_2(\psi)P_m^1(\cos\psi)\sin\psi\,d\psi\quad (m\geq 1). \tag{12.42}$$

On account of (12.39) and (12.41), (12.19) leads to

$$\mathbf{p}_1=\frac{\hat{\mathbf{x}}}{8(r|\mathbf{x}|)^2}[2(r^2+\mathbf{x}^2-c_1^2t^2)-\kappa^2-c_1^2\tau^2], \tag{12.43}$$

a radial vector with spherically symmetric magnitude; it is generally irrotational. The final solution for $c_1\mathbf{u}_1$ is determined from (12.24)–(12.26) accompanied by (12.27) and (12.43). In view of (12.14), the weak field $\mathbf{e}_1$ must also be irrotational. Now, the strong field $\mathbf{p}_1$ exists only within the hyperconical layer $D_p^{(1)}$, wherein $\mathbf{x}=0$ is normally never encountered (Fig. 4) except when $t=c_1^{-1}r$ (Fig. 6). *At this instant*, (12.43) reveals that $\mathbf{p}_1$ *acquires an inverse square singularity at* $\mathbf{x}=0$. No singularity occurs, however, if $\kappa=0=\tau$, corresponding to a singular radial force component that acts impulsively from an infinitesimally thin spherical sheet.

Owing to (12.39), the inner series for $v=2$ in (12.19) effectively starts from $m=1$; the coefficients $\chi_m^{(2)}(\hat{\mathbf{x}})$ should be directly computed from (12.42). The latter indicates that the strong field $\mathbf{p}_2$ acts azimuthally with an axisymmetric magnitude. Therefore, $\mathbf{p}_2$ is solenoidal, and so is the weak field $\mathbf{e}_2$ by virtue of (12.13). The solution for $c_2\mathbf{u}_2$ is now complete and is expressed by (12.24)–(12.26) accompanied by (12.19) and (12.28).

[1]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. 2.
[2]M. J. Lighthill, Philos. Trans. R. Soc. London A **252**, 397–430 (1960).
[3]M. J. Lighthill, J. Fluid Mech. **27**, 725–752 (1967).
[4]M. J. Lighthill, *Fourier Analysis and Generalized Functions* (Cambridge University, London, 1958).
[5]F. John, *Plane Waves and Spherical Means* (Interscience, New York, 1955).
[6]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University, London, 1944), 2nd ed., paperback, 1966.
[7]S. G. Mikhlin, *Multidimensional Singular Integrals and Integral Equations* (Pergamon, London, 1965).
[8]E. T. Copson, *Asymptotic Expansions* (Cambridge University, London, 1965).
[9]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University, London, 1927), 4th ed., paperback, 1965.
[10]J. J. Stoker, Comm. Pure Appl. Math. **9**, 577–595 (1956).
[11]M. J. Lighthill, J. Inst. Math. Its Appl. **1**, 1–28 (1965).
[12]W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Chelsea, New York, 1949).
[13]I. G. Petrowsky, Rec. Math. [Mat. Sbornik] **17**, 289–368 (1945).
[14]L. Gårding, Acta. Math. **85**, 1–62 (1951).
[15]L. Gårding, *Theory of Lacunas*, *Hyperbolic Equations and Waves*, edited by M. Froissart (Springer-Verlag, Berlin, 1970), pp. 13–21.
[16]G. F. D. Duff, Philos. Trans. R. Soc. London A **252**, 249–273 (1960).
[17]H. Weitzner, Phys. Fluids **4**, 1238–1245 (1961).
[18]R. Burridge, Proc. Cambridge Philos. Soc. **63**, 819–825 (1967).
[19]J. Bazer and D. H. Y. Yen, Comm. Pure Appl. Math. **22**, 279–333 (1969).
[20]R. G. Payton, Proc. Cambridge Philos. Soc. **70**, 191–210 (1971).

Invariant properties of n-point functions and n-point functionals connected with the translational invariance of the formal measure

Jerzy Hańćkowiak

Institute of Theoretical Physics, University of Wrocław, Wrocław, Poland
(Received 21 March 1977)

We discuss invariant properties of the generating functionals resulting from the translationally invariant formal measure used to define these functionals. The functionals considered depend on functions defined on five-dimensional space, and we relate to them n-point functionals and n-point functions. We derive equations for the above quantities, and we consider the connection with four-dimensional n-point quantities.

1. INTRODUCTION

Many observations point out that enlarging dimensions of considered quantities usually leads to the simplification of the formalism. For example, using the generating function which depends on two variables is the important step in the construction of the Hermitian polynomials dependent on one variable only. As a second example see Refs. 1 and 2, where the five-dimensional formalism enables one to use a statistical description of classical fields to the quantum field theory.

In this work we consider the functionals dependent on the functions defined on the five-dimensional space. This enlarging of dimensions is due to the fact that a local interaction is included in the arguments of the considered functionals. In this way we obtain the functionals in which translational invariance of the formal measure used to define the considered functionals leads to direct consequences.

In Sec. 2 we show that the two infinitely dimensional Abelian groups appear in this context. We construct here the functional $G_j[U]$ which has invariant properties similar to Bloch's theorem in solid state. In quantum field theory such a functional leads to the n-point functions without vacuum divergences.

In Sec. 3 we derive equations for n-point functions and n-point functionals connected with the functional $J[U]$ which leads to formulas with vacuum divergences. However, in this case relations between n-point quantities defined on R^4 and R^5 spaces are most simple. We also give in this section a schedule of a self-consistent method that can be used to calculate the n-point functions.

In Sec. 4 we consider the structure of the generating functional $G_j[U]$, which is related to n-point functions without vacuum divergences.

2. THE TRANSLATIONAL INVARIANT FORMAL MEASURE

In Ref. 1 we have introduced, instead of the usual generating functional $T[j]$, a more general one which may be used simultaneously to describe the different, self-interacting scalar fields

$$J[U]=\int \delta\varphi \exp\{\tfrac{1}{2}i\int dx\,dy\,\varphi(x)K(x,y)\varphi(y) + i\int d\hat{x}\,U(\omega,x)\delta(\omega-\varphi(x))\}. \tag{2.1}$$

Here $x\in R^4$, φ is one scalar field, $\delta\varphi \propto \Pi_x d\varphi(x)$. We see, therefore, that J is the Gaussian type functional defined on functions $U(\hat{x})$ with $\hat{x}$ from the five-dimensional space R^5, $\hat{x}=(\omega,x)$. After transformation of the variable φ,

$$\varphi(x)\to\varphi(x)+h(x), \tag{2.2}$$

taking into account the translational invariance of the formal measure $\delta\varphi$, we get the formula

$$J[U]=\exp(ihKh/2)\int\delta\varphi\exp\{\tfrac{1}{2}i\varphi K\varphi + i\int d\hat{x}[\omega\int dyK(x,y)h(y) + U(\omega+h(x),x)]\delta(\omega-\varphi(x))\}, \tag{2.3}$$

which, after introducing the operator

$$(\Lambda_h U)(\hat{x})=U(\omega+h(x),x)+\omega\int dy\,K(x,y)h(y), \tag{2.4}$$

can be described as

$$J[U]=\exp(ihKh/2)J[\Lambda_h U] \tag{2.5}$$

with any h.

This relation describes the property of the functional J connected with the translational invariance of the formal measure used in the definition of (2.1). It is worthwhile to notice that in the case of the five-dimensional formalism the above invariance can be described by means of transformations of functions U only.

Now we introduce the two infinite dimensional Abelian groups of transformations depending on functions h,

$$(\Gamma_h U)(\hat{x})=U(\omega+h(x),x) \tag{2.6}$$

and

$$(B_h U)(\hat{x})=U(\hat{x})+\omega\int dy\,K(x,y)h(y), \tag{2.7}$$

with the help of these one can express the transformations (2.4) as

$$\Lambda_h=B_h\Gamma_h. \tag{2.8}$$

 0022-2488/78/1903-0608$1.00

The inverse operator

$$\Gamma_h^{-1} = \Gamma_{-h} \tag{2.9}$$

and

$$B_h^{-1} = B_{-h}. \tag{2.10}$$

Hence

$$\Lambda_h^{-1} = \Gamma_{-h} B_{-h}. \tag{2.11}$$

Now the relation (2.5) is

$$J[U] = \exp(ihKh/2) J[B_h \Gamma_h U] \tag{2.12a}$$

or

$$J[\Gamma_{-h} U] = \exp(ihKh/2) J[B_h U]. \tag{2.12b}$$

The functional representations of the corresponding groups of transformations are

$$\hat{\Gamma}_h J[U] = J[\Gamma_h U] \tag{2.13}$$

and

$$\hat{B}_h J[U] = J[B_h U]. \tag{2.14}$$

Because transformations Γ_h and B_h are Abelian

$$\Gamma_{h'} \Gamma_h = \Gamma_{h'+h}, \quad B_{h'} B_h = B_{h'+h}, \tag{2.15}$$

their representations are also Abelian

$$\hat{\Gamma}_{h'} \hat{\Gamma}_h = \hat{\Gamma}_{h'+h}, \quad \hat{B}_{h'} \hat{B}_h = \hat{B}_{h'+h}. \tag{2.16}$$

With the help of generators of transformations

$$\hat{\Gamma}_h = \exp(h, \hat{\Gamma}), \quad \hat{B}_h = \exp(h, \hat{B}) \tag{2.17}$$

where

$$(h, \hat{\Gamma}) \equiv \int dx\, h(x) \hat{\Gamma}(x)$$
$$= \int dx\, h(x) \int d\omega\, \omega (\partial U(\omega, x)/\partial\omega) \delta/\delta U(\omega, x) \tag{2.18a}$$

and

$$(h, \hat{B}) = \int dx\, h(x) \int d\omega\, dy\, \omega K(x, y) \delta/\delta U(\omega, y). \tag{2.18b}$$

The hat over quantities means here that they act in the space of functionals.

The equality (2.5) may be now described as

$$\hat{\Gamma}_{-h} J = \exp(ihKh/2) \hat{B}_h J \tag{2.19}$$

or

$$J = \exp(ihKh/2) \hat{\Gamma}_h \hat{B}_h J. \tag{2.20}$$

These relations exhibit the structure of transformations describing invariant properties of the functional J resulting from translational invariance of the formal measure $\delta\varphi$. Concerning the possible solutions of Eqs. (2.5) or (2.20) one can say that if the functional G solves equation

$$G[U] = G[\Lambda_h U] \tag{2.21}$$

and J solves (2.5), then the new functional

$$J'[U] = G[U] J[U] \tag{2.22}$$

also solves (2.5). Also any superposition of two solutions of (2.5) is again solution of (2.5). The functional G can be constructed from J in different ways, e.g.,

$$G[U] = f(J^*[U] J[U]), \tag{2.23}$$

where f is any function.

By means of any operator A which commutes with operators Λ_h

$$[A, \Lambda_h] = 0, \tag{2.24}$$

one can construct the functional G as follows

$$G[U] = J[U]/J[AU]. \tag{2.25}$$

Now with the help of the functional J defined by (2.1) and transformations defined for any U as follows,

$$(A_j U)(\omega, x) = U(\omega, x) - \omega j(x), \tag{2.26}$$

we introduce the new functional G_j which parametrically depends on functions j and which leads to the normalized generating functional for time ordered Green's functions in QFT,

$$G_j[U] = J[U]/J[A_j U]. \tag{2.27}$$

To obtain the behavior of this functional with respect to transformations Λ_h we first calculate the commutator $[A_j, \Lambda_h]$,

$$(A_j \Lambda_h U)(\omega, x)$$
$$= U(\omega + h(x), x) + \omega \int K(x, y) h(y)\, dy - \omega j(x),$$

$$(\Lambda_h A_j U)(\omega, x)$$
$$= \Lambda_h (U(\omega, x) - \omega j(x))$$
$$= U(\omega + h(x), x) - (\omega + h(x)) j(x)$$
$$+ \omega \int K(x, y) h(y)\, dy.$$

It means that

$$([A_j, \Lambda_h] U)(\omega, x) = h(x) j(x). \tag{2.28}$$

Taking into account the property of the functional J,

$$J[U + g] = \exp[i \int dx\, g(x)] J[U], \tag{2.29}$$

that is true for any g which does not depend on ω, we obtain from (2.28) and (2.5)

$$G_j[\Lambda_h U] = J[\Lambda_h U]/J[\Lambda_h A_j U] = \exp[-i(h, j)] G_j[U] \tag{2.30}$$

with

$$(h, j) = \int dx\, h(x) j(x).$$

The equality (2.30) resembles in some way Bloch's theorem in solid state, where wave vectors $\mathbf{k}$ are replaced by functions $j(x)$ and translations are replaced by transformations Λ_h(2.4), describing the different symmetry of the infinite dimensional "crystal" by means of different K. For

$$U = \omega j(x) + \mathcal{L}_{\text{int}}(\omega), \tag{2.31}$$

$$G_j[U]\big|_{U=(2.31)} \equiv T_N[j] = T[j]/T[0] \quad [\text{see } (3.9)], \tag{2.32}$$

is the above-mentioned normalized generating functional describing one, scalar, quantum field with the action integral

$$L=\frac{1}{2}\int dx\,dy\,\varphi(x)K(x,y)\varphi(y)+\int dx\mathcal{L}_{\rm int}(\varphi(x)). \quad (2.33)$$

3. n-POINT FUNCTIONALS AND n-POINT FUNCTIONS

From the physical point of view, n-point functions generated by considered functionals are much more interesting than functionals alone.

n-point functionals connected with the functional J are

$$J[\hat{x}_1,\ldots,\hat{x}_n|U]\equiv\delta^n J[U]/\delta U(\hat{x}_1)\cdots\delta U(\hat{x}_n). \quad (3.1)$$

The relation (2.5) may be described with the help of them as follows:

$$J[\hat{x}_1,\ldots,\hat{x}_n/U]$$
$$=\exp(ihKh/2)\int d\hat{y}_{(n)}J[\hat{y}_1,\ldots,\hat{y}_n/\Lambda_h U]$$
$$\times\delta(\Lambda_h U)(\hat{y}_1)/\delta U(\hat{x}_1)\cdots\delta(\Lambda_h U)(\hat{y}_n)/\delta U(x_n)$$
$$=\exp(ihKh/2)J[\omega_1+h(x_1),x_1;\ldots;\omega_n+h(x_n),x_n/\Lambda_h U], \quad (3.2)$$

where we have used the formula

$$\delta(\Lambda_h U)(\omega',y)/\delta U(\omega,x)=\delta(\omega'-h(y)-\omega)\delta(y-x). \quad (3.3)$$

After differentiating (3.2) with respect to $h=0$ we get relations between n-point and $(n+1)$-point functionals connected with the functional J,

$$\sum_{k=1}^{n}\delta(y-x_k)\frac{\partial}{\partial\omega_k}J[\hat{x}_1,\ldots,\hat{x}_n/U]$$
$$=-\int d\hat{z}\,J[\hat{x}_1,\ldots,\hat{x}_n,\hat{z}/U]\delta(\Lambda_h U)(\hat{z})/\delta h(y)\big|_{h=0}$$
$$=-\int d\hat{z}\,J[\hat{x}_1,\ldots,\hat{x}_n,\hat{z}\,|\,U]\{U'(\hat{z})\delta(z-y)+\omega K(z,y)\}.$$

Hence we get

$$\sum_{k=1}^{n}\delta(y-x_k)\frac{\partial}{\partial\omega_k}J[\hat{x}_1,\ldots,\hat{x}_n/U]$$
$$=-\int d\hat{z}\,J[\hat{x}_1,\ldots,\hat{x}_n,\hat{z}/U]\{U'(\hat{z})\delta(z-y)+\omega K(z,y)\}. \quad (3.4)$$

Here we remember again that $\hat{x}$ means vector from R^5, $\hat{z}=(\omega,z)$, $U'(\hat{z})=(\partial/\partial\omega)U(\hat{z})$. Further restrictions for n-point functionals will be obtained if we calculate higher derivatives of (3.2).

Now we consider the special case of the n-point functionals, namely

$$T[\hat{x}_1,\ldots,\hat{x}_n/j]\equiv J[\hat{x}_1,\ldots,\hat{x}_n/U]\big|_{U=\omega j(x)+\mathcal{L}_{\rm int}(\omega)}, \quad (3.5)$$

where $\mathcal{L}_{\rm int}$ is fixed function and j plays the role of an external source appearing in the usual formulation of QFT. We get from (3.4)

$$\sum_{k=1}^{n}\delta(y-x_k)\frac{\partial}{\partial\omega_k}T[\hat{x}_1,\ldots,\hat{x}_n/j]$$
$$=-\int d\hat{z}\,T[\hat{x}_1,\ldots,\hat{x}_n,\hat{z}/j]\{(j(y)+\mathcal{L}'_{\rm int}(\omega))\delta(z-y)+\omega K(z,y)\}. \quad (3.6)$$

n-point functions connected with given n-point functionals are generally defined by putting the function variable equal to zero.

We have, e.g.,

$$T(\hat{x}_1,\ldots,\hat{x}_n)\equiv T[\hat{x}_1,\ldots,\hat{x}_n/j]\big|_{j=0}. \quad (3.7)$$

From (3.6) the following equations are obtained for them:

$$\sum_{k=1}^{n}\delta(y-x_k)\frac{\partial}{\partial\omega_k}T(\hat{x}_1,\ldots,\hat{x}_n)$$
$$=-\int d\hat{z}\,T(\hat{x}_1,\ldots,\hat{x}_n,\hat{z})\{\mathcal{L}'_{\rm int}(\omega)\delta(z-y)+\omega K(z-y)\}. \quad (3.8)$$

The unnormalized functional T, generating time-ordered Green's functions with vacuum divergences, is defined by means of the functional J [see (2.1)] as follows,

$$T[j]\equiv J[U]\big|_{U=\omega j(x)+\mathcal{L}_{\rm int}(\omega)}$$
$$=\int\delta\varphi\exp(i\varphi K\varphi/2)\exp(i\mathcal{L}_{\rm int}[\varphi])\exp[i(\varphi,j)]. \quad (3.9)$$

The n-point functionals connected with T are, as usual, defined as

$$T[x_1,\ldots,x_n/j]=\delta^n T[j]/\delta j(x_1)\cdots\delta j(x_n). \quad (3.10)$$

Hence and from (3.5) we get

$$T[x_1,\ldots,x_n/j]=\int d\omega_{(n)}\omega_1\cdots\omega_n T[\hat{x}_1,\ldots,\hat{x}_n/j]. \quad (3.11)$$

We obtain a similar expression for n-point functions,

$$T(x_1,\ldots,x_n)=\int d\omega_{(n)}\omega_1\cdots\omega_n T(\hat{x}_1,\ldots,\hat{x}_n). \quad (3.12)$$

The last equality reveals the true meaning of the n-point functions $T(\hat{x}_1,\ldots,\hat{x}_n)$ defined on the Cartesian product of n five-dimensional spaces R^5. n-point functions $T(x_1,\ldots,x_n)$, after removing vacuum divergences, amputating, and passing to momentum space, describe scattering processes.

From the definition of the n-point functions $T(\hat{x}_1,\ldots,\hat{x}_n)$ one can derive the other's relations to the n-point functions $T(x_1,\ldots,x_n)$, e.g.,

$$T(\underbrace{x,\ldots,x}_{n},x_1,\ldots,x_n)=i^{n-1}\int d\omega\,d\omega_{(n)}\omega^n\omega_1\cdots\omega_n\times T(\omega,x;\hat{x}_1,\ldots,\hat{x}_n), \quad (3.13a)$$

$$T(\hat{x}_1,\ldots,\hat{x}_n)=-i\int d\omega_{n+1}T(\hat{x}_1,\ldots,\hat{x}_{n+1}). \quad (3.13b)$$

It is easy to prove with the help of (3.13) that (3.8) leads to Schwinger's equations for the n-point functions $T(x_1, \ldots, x_n)$.

The equations of the type (3.12)—(3.13) may be treated as a kind of consistency condition which does not depend on dynamics imposed on the functions $T(\hat{x}_1, \ldots, \hat{x}_n)$ fulfilling Eqs. (3.8). However, in the first approximation one may try to solve the Eqs. (3.8) without these conditions. For lowest n-point functions we have the following equations

$$\delta(y-x_1)\frac{\partial}{\partial\omega_1}T(\hat{x}_1) = -\int d\hat{z}\, T(\hat{x}_1,\hat{z})\{\mathcal{L}'_{\text{int}}(\omega)\delta(z-y)+\omega K(z,y)\}. \qquad (3.14)$$

Because of Poincare's invariance

$$T(\hat{x}_1) \equiv i\int \delta\varphi\, \delta(\omega_1 - \varphi(x_1))\exp(i\varphi K\varphi/2) + i\mathcal{L}_{\text{int}}[\varphi]$$

does not depend on x_1. Introducing

$$\frac{\partial}{\partial\omega_1}T(\hat{x}_1) \equiv f(\omega_1) \qquad (3.15)$$

we rewrite (3.14),

$$\delta(y-x_1)f(\omega_1) = -\int d\hat{z}\, T(\hat{x}_1,\hat{z})\{\mathcal{L}_{\text{int}}(\omega)\delta(z-y)+\omega K(z,y)\}. \qquad (3.14')$$

Here $\hat{x}_1 = (\omega_1, x_1)$, $\hat{z} = (\omega, z)$, and $f(\omega)$ is the unknown function of ω. Fixing f in some way, one can try to find from (3.14′) the two-point function $T(\hat{x}_1, \hat{x}_2)$ in agreement with relativistic invariance, with spectral condition and symmetric condition

$$T(\hat{x}_1, \hat{x}_2) = T(\hat{x}_2, \hat{x}_1), \qquad (3.16)$$

and with the constraint [see (3.13b)]

$$\frac{-i\partial}{\partial\omega_1}\int d\omega_2 T(\hat{x}_1,\hat{x}_2) = f(\omega_1). \qquad (3.17)$$

The last property suggests that a kind of self-consistent method can be used to calculate the two-point function $T(\hat{x}_1, \hat{x}_2)$. We postulate $f^{(j)}(\omega)$ and calculate from (3.14′), $T^{(j)}(\hat{x}_1, \hat{x}_2)$. After that we calculate

$$f^{(j+1)}(\omega_1) \equiv \frac{-i\partial}{\partial\omega_1}\int d\omega_2 T^{(j)}_{(\hat{x}_1,\hat{x}_2)}, \qquad (3.18)$$

and from (3.14′) we calculate $T^{(j+1)}_{(\hat{x}_1,\hat{x}_2)}$. If such a procedure is convergent we see that

$$\lim_{j\to\infty} T^{(j)}(\hat{x}_1, \hat{x}_2) = T(\hat{x}_1, \hat{x}_2) \qquad (3.19)$$

fulfills Eq. (3.14′) and condition (3.17) with

$$f(\omega_1) \equiv -i\lim_{j\to\infty}\frac{\partial}{\partial\omega_1}\int d\omega_2\, T^{(j)}(\hat{x}_1,\hat{x}_2). \qquad (3.20)$$

4. THE STRUCTURE OF THE FUNCTIONAL G_j

The functional $G_j[U]$ introduced in Sec. 2 defines the generating functional $T_N[j]$ [see (2.32)] which does not contain the vacuum divergences. Because of (2.30), the n-point functionals

$$G_j[\hat{x}_1, \ldots, \hat{x}_n/U] \equiv \delta^n G_j[U]/\delta U(\hat{x}_1)\cdots\delta U(\hat{x}_n) \qquad (4.1)$$

fulfill [analogous to Eq. (3.2)],

$$G_j[\hat{x}_1, \ldots, \hat{x}_n/U] = \exp[i(h,j)]G_j[\omega_1 + h(x_1), x_1; \ldots; \omega_n + h(x_n), x_n/\Lambda_h U]. \qquad (4.2)$$

Differentiation with respect to $h=0$ now gives

$$ij(y)G_j[\hat{x}_1, \ldots, \hat{x}_n/U] + \sum_{k=1}^{n}\delta(y-x_k)\frac{\partial}{\partial\omega_k}G_j[\hat{x}_1, \ldots, \hat{x}_n/U] = -\int dz\, G_j[\hat{x}_1, \ldots, \hat{x}_n, \hat{z}/U] \times\{U'(\hat{z})\delta(z-y) + \omega K(z,y)\}. \qquad (4.3)$$

However the n-point functionals related to the functional G_j have no simple connection with n-point functionals and functions of the functional $T_N[j]$. Another new functional Z defined by the relation

$$J[U] = \exp(Z[U]), \qquad (4.4)$$

exhibits the other features of the functional

$$G_j[U] \equiv J[U]/J[A_jU] = \exp(Z[U])/\exp(Z[A_jU]). \qquad (4.5)$$

We can expand the functional $Z[A_jU]$ in a Volterra series at U, because from (2.26)

$$(A_jU)(\omega, x) = U(\omega, x) - \omega j(x).$$

We get

$$G_j[U] = \exp\left\{-\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\times\int d\hat{x}_{(n)} Z[\hat{x}_1, \ldots, \hat{x}_n/U]\omega_1 j(x_1)\cdots\omega_n j(x_n)\right\}. \qquad (4.6)$$

Since Z from (4.4) and (2.5) has the following symmetry property,

$$Z[U] = \frac{i}{2}hKh + Z[\Lambda_h U], \qquad (4.7)$$

n-point functionals $Z[\hat{x}_1, \ldots, \hat{x}_n/U]$ fulfill the same equations as (3.4). The n-point functionals

$$z[\hat{x}_1, \ldots, \hat{x}_n/j] \equiv Z[\hat{x}_1, \ldots, \hat{x}_n/U]\big|_{U=\omega j(x)+\mathcal{L}_{\text{int}}(\omega)} \qquad (4.8)$$

fulfill the analogous Eq. (3.6)

$$\sum_{k=1}^{n}\delta(y-x_k)\frac{\partial}{\partial\omega_k}z[\hat{x}_1,\ldots,\hat{x}_n/j]$$

$$=-\int d\hat{z}\,z[\hat{x}_1,\ldots,\hat{x}_n,\hat{z}/j]\{(j(y)+\mathcal{L}'_{\text{int}}(\omega))\delta(z-y)+\omega K(z,y)\}. \tag{4.9}$$

Now we see that

$$T_N[j]=G_j[U]\big|_{U=\omega j+\mathcal{L}_{\text{int}}}$$
$$=\exp\left\{-\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\int d\hat{x}_{(n)}\,z[\hat{x}_1,\ldots,\hat{x}_n/j]\times\omega_1 j(x_1)\cdots\omega_n j(x_n)\right\}. \tag{4.10}$$

[1]J. Hańćkowiak, J. Math. Phys. 16, 1524 (1975).
[2]J. Hańćkowiak, Preprint No. 333, University of Wrocław (1975).

A geometric theory of charge and mass

Richard Shurtleff

Physics Department, University of Massachusetts, Amherst, Massachusetts 01002
(Received 26 January 1977)

A geometric model of a charge is constructed by defining several geometries on the same spacetime manifold. A Riemannian geometry describes the vacuum. On the same spacetime, two Weyl geometries are constructed for the charge description. The geometries are constrained by a variational principle. Energy conservation requires the equality of active and passive mass. Chargeless particles have essentially no mass. The treatment of radiation relies on the approximate nature of the wave equation. Variable mass terms in the wave equation cause the $2S-2P$ levels in hydrogen to separate by 30 000 Mhz. This unobserved transition together with the lack of spin sets a limit to the correspondence of the model to real electrons.

I. INTRODUCTION

The general theory of relativity provided support for the viewpoint that the course of events which physics describes can most simply be expressed as the result of geometric constraints on a spacetime manifold. This report extends the geometry and the constraints so that the geometry itself provides the physical description of charges.

Basic to the model is the observation that several different geometries can be constructed on the same manifold. On spacetime a Riemannian geometry is defined to describe the classical gravitational vacuum. Several Weyl[1] geometries can then be defined to describe the charges. We use two Weyl geometries per charge. The metrics of the Weyl geometries are conformal to the metric of the Riemannian geometry. The conformal factor is essentially the density of the charge. Since charges in matter are localized, nonzero lengths in the Weyl geometries occur locally. Thus the Weyl geometries are trivial except in the neighborhood of the charge. The only geometry to have long range effects is the vacuum Riemannian geometry.

The number of Weyl geometries per charge was determined by the effects in our theory caused by gauge dependence. The notion of gauge and of gauge invariance was introduced by Weyl and is implicit in his geometry. Our theory relies heavily on the gauge dependence of the field equations. This dependence requires a unique gauge to be determined by the physical interpretation. Gauge variables are present in both the form of the electromagnetic current and the form for the mass. The current is linear in the gauge terms and the mass is quadratic. The physical interpretation requires the gauge terms in the mass formula to remain, but the terms in the current must vanish. To do this, we use two Weyl geometries to describe the charge, their gauge terms being additive inverses. This device gets rid of the unwanted terms in the current since they are linear, and cancel, yet keeps the required quadratic terms in the mass formula. Therefore, to have the proper physical interpretation, two Weyl geometries are required to describe a single charge.

The conformal scalar curvatures of the Weyl geometries must be modified to use in the variational principle. The field equations consist of a Klein–Gordon type equation for the charge motion and the source equations for gravity and electromagnetism. Conservation of energy and charge follow from identities. All the equations are covariant, but none are gauge invariant.

Inherent in the construction of a Weyl geometry is an electromagnetic vector potential. This is assumed to contribute additively to the total potential of the vacuum. To avoid self-energy problems, the potentials of the charge geometry is assumed to be due to other sources. Thus radiation is carried away by a different potential. These assumptions are not time reversal invariant. Furthermore, radiation and conservation of energy together require a change in the state of the charge, since energy radiated must be lost by the charge. By assumption, only the external vector potential can change the state of the charge. Therefore, to treat radiation, we must assume the wave equation is not exact, relying instead on the source equation and conservation of energy equation to describe the radiation.

The usual concept of mass includes two separate notions: mass as the source of gravity and mass as inertia, known as active and passive mass, respectively.[2] Passive mass enters in the wave equation; active mass occurs in the energy equation of gravity. Each type has a rest mass which is a constant in the theory. Their equality arises as follows. The energy an atomic electron absorbs from the external field changes its active mass. The amount of energy lost in radiation can be found from the conservation of energy equation and the electromagnetic source equation. These two energies must be equal if an atom that absorbs radiation and subsequently emmits radiation is not a source or sink of energy. The formula for the radiated energy contains the ratio of active rest mass to passive. This ratio must be one if the atom is not a source or sink of energy.

This is a theory of electromagnetic charges, i.e., electrons. But the theory is spinless, so the charges do not reproduce the behavior of electrons. The question of many charge statistics is tied to spin, so we treat only the single charge. The fine structure of spectra is also linked to spin so the details of spectroscopy cannot be reproduced. Further evidence of this failure is the prediction of a 30 000 Mhz shift in the $2S-2P$ levels of hydrogen. This shift is due to variable mass terms in the wave equation.

 0022-2488/78/1903-0613$1.00

The difference with previous geometric models of charge are apparent. We avoid singularities by spreading out the charge as in wave mechanics. Geometric models of point charges have used unusual topologies to represent the inherent singularities of point sources, for example, the multiconnected topology due to Wheeler.[3] Breaking gauge invariance is basic to our approach. Weyl[4] required this invariance in his theory of electromagnetism. Flint[5] used the conformal factor in a Weyl geometry as the square of the wavefunction, but failed to employ the concept of many geometries on the same surface.

Weyl's geometry has appeared to conflict with fundamental atomic phenomena. Given an atomic clock and the speed of light, a well-defined unit of length is determined. Weyl's geometry rests on a concept of indeterminate length. The way to avoid this conflict is to introduce several geometries: In one atomic lengths are fundamental; in the others Weyl's geometry holds. In our theory, the Riemannian geometry of the vacuum measures atomic lengths, and the Weyl geometries are localized to atomic dimensions, as discussed above. Dirac[6] has used two metrics: one measuring length with the atomic standard, one to which Weyl's theory applies. He uses the Weyl geometry to describe effects of the large numbers hypothesis.

Since Weyl's geometry may be unfamiliar, we provide a quick derivation of the results needed here. Following this is the variational principle and the discussion of the field equations.

II. GEOMETRY

Riemannian geometry in the limit of zero curvature reduces to Euclidean geometry. In particular, vectors of equal components have equal lengths. Weyl's geometry retains curvature in the limit of a Euclidean metric. The lengths of two vectors located at points on the manifold separated by coordinate differences, dx^σ, differ according to the formula

$$dl = l(a \cdot dx), \tag{1}$$

where a^σ is a vector and l is the Riemannian length of the vectors. Thus, even when the Riemannian curvature is zero, the affine connections cannot be null. The vector a^σ was interpreted by Weyl as the electromagnetic vector potential.

To derive the affine connections, recall that, in Riemannian geometry with metric $g_{\alpha\beta}$, the equation $d(l^2)=0$ suffices. For a vector with components V^σ, this means

$$(\partial_\sigma g_{\alpha\beta})V^\alpha V^\beta dx^\sigma + g_{\alpha\beta}[\partial_\sigma(V^\alpha V^\beta)]dx^\sigma = 0. \tag{2}$$

In Weyl's theory,

$$d(l^2) = l^2(2a \cdot dx) = 2a_\sigma g_{\alpha\beta} V^\alpha V^\beta dx^\sigma. \tag{3}$$

If $(-2a_\sigma + \partial_\sigma)g_{\alpha\beta}$ replaces $\partial_\sigma g_{\alpha\beta}$ in the formula for the Christoffel connections $C^\alpha_{\beta\gamma}$, we have the affine connections $\Gamma^\alpha_{\beta\gamma}$ of the Weyl geometry.

Weyl did not want a conformal transformation of the metric to affect the intrinsic geometry of the manifold. It is clear from the above derivation that the following leaves the affine connections unchanged,

$$\bar{g}_{\alpha\beta} = U g_{\alpha\beta} \quad \text{and} \quad \bar{a}_\sigma = a_\sigma + \tfrac{1}{2}(\partial_\sigma \ln U). \tag{4}$$

That follows because

$$U(-2a_\sigma + \partial_\sigma)g_{\alpha\beta} = (-2\bar{a}_\sigma + \partial_\sigma)\bar{g}_{\alpha\beta}, \tag{5}$$

and the factor U cancels out of (3).

Calculating the Weyl scalar curvature $\bar{W}$, using the affine connections $\Gamma^\alpha_{\beta\gamma}$ with the metric $\bar{g}_{\alpha\beta}$ yields

$$\bar{W} = \bar{R} + 6\bar{a}^2 - 6((-\bar{g})^{1/2}\bar{a}^\sigma)_{,\sigma} / (-\bar{g})^{1/2} \tag{6}$$

where[7] $\bar{g} = \det(\bar{g}_{\alpha\beta})$ and $\bar{R}$ is the Riemannian scalar curvature of $\bar{g}_{\alpha\beta}$. In terms of the metric $g_{\alpha\beta}$ the quantities become

$$\bar{\Gamma}^\alpha_{\beta\gamma} = \Gamma^\alpha_{\beta\gamma}, \quad \bar{W} = W/U, \quad \bar{a}_\sigma = a_\sigma + \tfrac{1}{2}(\partial_\sigma \ln U)$$

and

$$_0\bar{f}_{\alpha\beta} = \bar{a}_{\alpha,\beta} - \bar{a}_{\beta,\alpha} = a_{\alpha,\beta} - a_{\beta,\alpha} = {}_0 f_{\alpha\beta}.^8 \tag{7}$$

Combining equations to obtain the conformal scalar curvature, we have

$$\bar{R} = \frac{W}{U} - \frac{6\bar{a}^2}{U} + \frac{6((-\bar{g})^{1/2}\bar{a}^\sigma)_{,\sigma}}{(-\bar{g})^{1/2}}. \tag{8}$$

The last term on the right becomes a divergence in the action principle integral; thus it has no affect on the field equations and is dropped from the following equations.

Assuming the vector a^σ has pure imaginary components, and a^σ has an imaginary part, R is kept real by replacing the following:

$$a^2 \to |a|^2, \quad \bar{a}^2 \to |\bar{a}|^2, \quad \text{and} \quad a^\sigma{}_{;\sigma} = 0. \tag{9}$$

Altogether,

$$\bar{R} = (1/U)(R + 6|a|^2 - 6|\bar{a}|^2), \tag{10}$$

where R is the scalar curvature of the metric $g_{\alpha\beta}$.

The last formula for $\bar{R}$ must be modified to serve as Lagrangian. To do this, use the replacement,

$$\bar{a}_\sigma = a_\sigma + \partial_\sigma \ln|u| \to \bar{a}_\sigma = a_\sigma + \partial_\sigma \ln u, \tag{11}$$

where $|u|^2 = U$.

Therefore, $\bar{R}$ with this change is the scalar curvature when u is real.

The modification can be described in another way. Notice that $\bar{a}_\sigma$ is the result of a conformal transformation from the vacuum metric to the charge metric. The inverse should give the vector potential that the charge would predict for the vacuum, call it $a_{v\sigma}$,

$$a_{v\sigma} = \bar{a}_\sigma - \partial_\sigma \ln|u| = a_\sigma + \partial_\sigma \ln u - \partial_\sigma \ln|u| = a_\sigma + i \mathrm{Im}\, \partial_\sigma \ln u,$$

$$\text{where} \quad i = \sqrt{-1} \text{ and } \mathrm{Im} X = (X - X^*)/2i. \tag{12}$$

If we take $a_{v\sigma}$ as the vector potential of the vacuum, the modification is to replace the a_σ in the term $6a^2/U$ by $(a_{v\sigma} - i\mathrm{Im}\,\partial_\sigma \ln u)$, in (10). From this point of view it is clear $\bar{R}$ is no longer the scalar curvature since we use two different gauges for the same potential in the formula for $\bar{R}$. To emphasize the change, we define

$$S = (1/U)(R + 6|a_v|^2 - 6|\bar{a}|^2). \tag{13}$$

Notice that if a Weyl geometry and the Lagrangian S are used for the vacuum, then the constraints on the vacuum metric are unchanged if $u=1$, since then $S=R$. Furthermore, this is a rigorous way of introducing a vacuum vector potential to the geometry. The total vector potential in the vacuum ϕ_σ is assumed to be the sum of the charge geometry potential plus a vacuum geometry contribution γ_σ

$$\phi_\sigma = a_{v\sigma} + \gamma_\sigma. \tag{14}$$

The field equations depend on which form of a_σ is varied in the Lagrangian; we chose $a_{v\sigma}$.

The vector a_σ of Weyl's theory must be allowed only purely imaginary numerical components. To justify this assumption, consider a point charge electron in circular orbit about a point charge proton.[9] Calculate the change in length after one revolution. Select those orbits for which there is no change,

$$\oint dl/l = 2\pi n i, \text{ where } n \text{ is an integer.} \tag{15}$$

For the case considered, a_σ has only a time component, $-ke/r$, where k is constant, e is the proton charge, and r is the distance from the proton. Then, applying classical mechanics to cancel radial forces and using (1), we find

$$r = -n^2/mk^2. \tag{16}$$

If $k=\pm ie/\hbar$,[10] then the radii selected are those of Bohr's model of the hydrogen atom. This is the justification for assuming a_σ is purely imaginary.

III. FIELD EQUATIONS

The field equations result from a variational principle constructed from the scalar curvature R of the Riemannian geometry, the modified form[11] S of the curvatures in the Weyl geometries, and the square of the electromagnetic field. Only the case of one charge is derived; thus two Weyl geometries are needed, labeled by a preindex. The Lagrangian is

$$L = (R + cf^2)(-g)^{1/2} + \sum_{j=1}^{2} b\, {}_jS(-{}_j\bar{g})^{1/2}, \tag{17}$$

where c and b are constants, and $f_{\alpha\beta} = \phi_{\alpha,\beta} - \phi_{\beta,\alpha}$, where ϕ^σ is the total potential in vacuum.

The field equations are covariant, but not gauge invariant. This last property is used in the physical interpretation. Specifically, the equations are written with these substitutions:

$${}_ju = {}_jv \exp(i\, {}_jp\cdot x) \quad\text{and}\quad {}_ja_v = a'_\sigma - i\, {}_jp_\sigma \text{ for } j=1,2, \tag{18}$$

where the vectors ${}_jp$ are constant, and a'_σ is the same for both particle component geometries. A more complete description may require a more general transformation. Each ${}_ju$ may be thought of as an amplitude modulated plane wave. The field equations are labeled by the function varied,

$$\begin{aligned} {}_ju^*Eq.:\ 0 &= (R/6 + |{}_ja_v|^2)\, {}_jv \\ &\quad + (-g)^{-1/2}(a'_\alpha + \partial_\alpha)g^{\alpha\beta}(-g)^{1/2}(a'_\beta + \partial_\beta)\, {}_jv, \end{aligned} \tag{19}$$

$$a\,Eq.:\ j^\alpha = \frac{1}{4\pi}F^{\alpha\sigma}{}_{;\sigma} = q\sum_{j=1}^{2}[{}_jp^\alpha\, {}_jU + \mathrm{Im}({}_jv^*\partial^\alpha\, {}_jv)],$$

where $q = 3bi/4\pi ck$, (20)

$$\begin{aligned} g\,Eq.:\ 0 &= \Big(1 + \sum_{j=1}^{2} b\, {}_jU\Big)G^{\alpha\beta} - 8\pi ck^2 T^{\alpha\beta}_{\mathrm{em}} \\ &\quad + b\sum_{j=1}^{2}\{(g^{\alpha\beta}\Box\, {}_jU - {}_jU^{;\alpha;\beta}) \\ &\quad + 6\, {}_jU({}_jp^\alpha\, {}_jp^\beta + 2ia'^{(\alpha}\mathrm{Im}(\partial^{\beta)}\ln {}_jv) \\ &\quad - \partial^{(\alpha}\ln {}_jv^*\partial^{\beta)}\ln {}_jv) - 3\, {}_jUg^{\alpha\beta} \\ &\quad \times[({}_jp)^2 + 2ia'\cdot\mathrm{Im}\partial\ln {}_jv - |\partial\ln {}_jv|^2]\}, \end{aligned} \tag{21}$$

where

$$\Box U = U^{;\sigma}{}_{;\sigma},\ F^{\alpha\beta} = f^{\alpha\beta}/k,\ G^{\alpha\beta} = R^{\alpha\beta} - \tfrac{1}{2}Rg^{\alpha\beta}, \text{ and}$$

$$T^{\alpha\beta}_{\mathrm{em}} = -(1/4\pi)(F^{\alpha\beta}F^\beta{}_\sigma - \tfrac{1}{4}g^{\alpha\beta}F^2). \tag{22}$$

We have kept the notation ${}_jU$ because ${}_jU = |{}_jv|^2$. The ${}_ju^*$ equation may be rewritten in a simpler form,

$$0 = [R/6 + ({}_jp)^2 - 2i\, {}_jp\cdot a' + a'^\sigma{}_{;\sigma}]\, {}_jv + 2a'\cdot\partial\, {}_jv + \Box\, {}_jv. \tag{23}$$

The charge density of the source is spread out over a volume of space as a glance at (20) shows. If one section of this density were to be repelled by another section of the charge density, there are no external forces which could hold it together. It must be assumed that the charge reacts to that part of the vector potential which has other charges as its source. Accordingly, a'_σ is assumed to be the external vector potential and γ_σ is assumed to be the vector potential arising from the charge itself. This assumption destroys the time inversion symmetry of the theory. If a charge absorbs energy from the field corresponding to a'_σ, the inverse process is the emission of energy from the charge to the field due to a'_σ. This means γ_σ and a'_σ are the same. But then a'_σ has the charge as part of its source and the charge blows up. Therefore, time inversion symmetry fails.

There are two identities which must be satisfied. Since $F^{\alpha\beta}$ is antisymmetric, the divergence of the current j^σ is identically zero,

$$j^\sigma{}_{;\sigma} = 0 = \sum_{j=1}^{2}\mathrm{Im}({}_ju^*\Box\, {}_ju). \tag{24}$$

Together with the ${}_ju^*$ equation, this implies

$$\sum_{j=1}^{2}({}_jU\, {}_ja^\sigma_v)_{;\sigma} = 0. \tag{25}$$

Care must be exercised so that ${}_ju$ and ${}_ja_{v\sigma}$ satisfy this equation.

The divergence of $G^{\alpha\beta}$ is identically zero, so

$$\begin{aligned} 0 &= -8\pi ck^2 T^{\alpha\beta}_{\mathrm{em};\beta} + b\sum_{j=1}^{2}[{}_jU_{,\beta}(G^{\alpha\beta} + R^{\alpha\beta} + {}_jp^\alpha\, {}_jp^\beta \\ &\quad - \tfrac{1}{2}g^{\alpha\beta}({}_jp)^2) + 6i\, {}_0f^{\alpha\beta}\,\mathrm{Im}({}_jv^*\partial_\beta\, {}_jv) \\ &\quad - 6i\, {}_jU^{,\alpha}a'\cdot\mathrm{Im}(\partial\ln {}_jv) - 6(\partial^{(\alpha}\, {}_jv^*\partial^{\beta)}\, {}_jv \\ &\quad - \tfrac{1}{2}g^{\alpha\beta}|\partial\, {}_jv|^2)_{;\beta}]. \end{aligned} \tag{26}$$

IV. MASS AND CHARGE

The mass M and charge $-e$ are defined as the volume integrals of the time components of the particle stress–energy tensor $T^{\alpha\beta}_p$ and the current vector j^σ, respectively:

$$M=\int T_p^{00}d^3x \quad\text{and}\quad -e=\int j^0\,d^3x, \tag{27}$$

where

$$-8\pi T_p^{00}=G^{00}+8\pi T_{\rm em}^{00}. \tag{28}$$

Since the amount of mass and charge is defined in terms of one component of a tensor and vector, the result depends on the coordinate system of the observer. The only well-defined coordinate system is the rest frame of the charge. Therefore, we assume the expectation value of momentum is zero. However, the kinetic energy of the charge need not be zero, for example, if the charge is the electron in a one-electron atom. The treatment is further simplified by considering the wavefunctions ${}_jv$ to be eigenstates of energy. Equation (25) will then be satisfied if a'_σ has only a time component. The mass is spread out over a volume much larger than the Schwarzschild limit. This is obvious, since nuclei are the tightest binders, and atomic electrons are spread out over volumes of radius many times bigger than the Schwarzschild radius. Thus we assume $g^{\alpha\beta}$ is the flat space metric and

$$\partial_0\,{}_jv=i\,{}_jw\,{}_jv \quad\text{and}\quad |\partial_n\,{}_jv|\ll|{}_jw|. \tag{29}$$

Before evaluating the mass and charge, certain problems with the form of the current must be eliminated. If the particle at rest is to have no current, the following terms of (20) must be zero:

$$0={}_1p\,{}_1U+{}_2p\,{}_2U. \tag{30}$$

Assuming equal normalization of components, we have

$$0={}_1p+{}_2p. \tag{31}$$

By thinking of ${}_ju$ as amplitude modulated plane waves, the carrier waves are complex conjugates of each other. Thus,

$$\int{}_1u^*\,{}_2u\,d^3x=\int{}_1v^*\,{}_2v\exp(2i\,{}_2p\cdot x)\,d^3x. \tag{32}$$

The wavelength of the plane wave will turn out to be roughly $\frac{1}{2}\hbar/m$, which is 10^{-2} Å for electrons. Changes in ${}_1v^*\,{}_2v$ are on the order of angstroms, so the integral is approximately zero. This near orthogonality mimics that of the spin components in elementary quantum mechanics. Here, however, the wave equation (19) does not mix spin components, so there are no spin effects.

Returning to the mass and charge evaluation, using (31), and assuming $b\,{}_jU$ and R negligible, one finds

$$M=(3bV/4\pi)(p^n)^2+(3b/4\pi)\int{}_jv^*\Delta\,{}_jv\,d^3x. \tag{33}$$

where $\frac{1}{2}V=\int{}_jU\,d^3x$, $(p^n)^2=({}_jp^n)^2$, $\Delta v=v^{;n}{}_{;n}$, and the surface of the volume of integration is far from the charge so that divergences in T_p^{00} integrate out. The constant term on the right in (33) is the active rest mass M_0,

$$M_0=(3bV/4\pi)(p^n)^2. \tag{34}$$

Keeping only the lowest order terms in (19) implies

$${}_jw^2={}_jp^2=p^2. \tag{35}$$

The charge is approximately

$$-e=\tfrac{1}{2}V({}_1w+{}_2w). \tag{36}$$

Multiplying by $({}_1w-{}_2w)$ implies ${}_1w={}_2w=w$.

To evaluate the constants of the theory, assume (19) is an approximate form of the Klein–Gordon equation

$$[m^2/\hbar^2+(-ieA_\sigma/\hbar+\partial^\sigma)(-ieA^\sigma/\hbar+\partial^\sigma)]v=0, \tag{37}$$

where A_σ is the electromagnetic vector potential. Therefore,

$$p^2=m^2/\hbar^2 \quad\text{and}\quad k=-ie/\hbar. \tag{38}$$

Note that this value for k agrees with the Bohr atom treatment in Sec. II. Comparison of Eq. (21) with the general relativity result for electromagnetism, shows $ck^2=-1$. Now the constants can be determined:

$$w=\pm m/\hbar,\ bV=\pm4\pi\hbar^2/3m,\ (p^n)^2=\pm mM_0/\hbar^2,\ ck^2=-1,$$
$$k=-ie/\hbar,\ \text{and}\ (p^0)^2=w^2+(p^n)^2=(m^2/\hbar^2)(1\pm M_0/m),$$
$$\text{where } m>0. \tag{39}$$

Notice that negative w implies negative rest mass M_0, since $(p^n)^2>0$. Antiparticles are well known to exist, and can be interpreted as negative energy states. To include such species, we need only expand the two Weyl geometry theory to four, two having positive w and two negative. Wave components with opposite w's can be considered orthogonal in many cases.

$$\int{}_1v^*\,{}_3v\,d^3x=I\exp(2i\,{}_3wt), \tag{40}$$

where I is a function of position and time, t. If the function I does not change appreciably in the span of time on the order of $1/w=10^{-20}$s, then, averaged over a few such time spans, (40) shows the effective orthogonality of components with opposite w's.

A four component wavefunction is most natural to describe charge. The need for a two-component theory can now be seen to arise because the constant vector p^σ is needed so that the rest mass of the charge is nonzero, (34), but yet the electromagnetic source cannot have a constant part essentially independent of the wavefunction, (20). The mass depends on the square of p^σ, and the constant part of the current is linear in p^σ. Thus, having two components with vectors ${}_jp^\sigma$ of equal magnitudes, but opposite directions, cancels the unwanted terms in the current, but retains nonzero rest mass. Two more components are needed to include antiparticles in the description.

Two rest masses, m and M_0, appear in the equations; one is the source of a gravitational field, the other measures the inertial resistance to applied force in the wave equation. The difference in these two masses has nothing to do with the Eötvös–Dicke experiment and the equivalence principle. In the absence of electromagnetic and gravitational forces, it is evident from (19) that charges follow straight line paths. Covariance of the equations requires that in the absence of electromagnetic forces, charges follow geodesics. Thus the path of a charge unaffected by electromagnetism is independent of any intrinsic characteristic of the charge. These predictions are the content of the equivalence principle supported by experiment. Thus, first, the theory obeys the equivalence principle. Secondly, the "gravitational mass" m_g arises from Newton's equation

$$m_ia=m_gK/r^2, \tag{41}$$

where K is constant and r is the distance from the gravitational source. It may be argued that m_i, the inertial mass, is the mass m. But the active mass M is a factor in K, and is definitely not m_i.

For a bound charge, (33) can be put in the form

$$M = M_0 - 2\langle \text{k.e.}\rangle,$$

$$\text{where } \langle \text{k.e.}\rangle = \frac{1}{V}\sum_{j=1}^{2}\int_j v^*\left(-\frac{\hbar^2}{2m}\Delta\right)_j v d^3x. \quad (42)$$

For a one electron atom, the vector potential may be assumed to have only a time component, which is inversely proportional to the distance from the nucleus. Application of the virial theorem implies

$$M = M_0 - 2E, \quad \text{where } E \text{ is the binding energy.} \quad (43)$$

Discussion of this effect will be delayed until radiation has been treated.

The discussion of mass would be incomplete without noticing that chargeless particles are essentially massless in this theory. This follows because charge null implies k null implies ${}_ja_{v\sigma}$ null, so

$$m = \hbar(R/6)^{1/2}. \quad (44)$$

Mesons are "strongly" charged even if neutral electromagnetically. A discussion of strong charges lies beyond the scope of this report, so (44) does not contradict observation.

V. RADIATION

The conservation of energy equation is (26) with $\alpha = 0$, which we use in the nonrelativistic limit, (29). By using (20), (22), and (30), two of the terms combine,

$$8\pi T^{0\beta}_{\text{em};\beta} + \sum_{j=1}^{2} 6bi\,{}_0f^{0\beta}\text{Im}({}_jv^*\partial_\beta\,{}_jv) = (8\pi/k)j_\beta(f^{0\beta} - {}_0f^{0\beta})$$

$$= (8\pi/k)j_\beta\,{}_\gamma f^{0\beta} = 8\pi\,{}_\gamma T^{0\beta}_{\text{em};\beta}, \quad (45)$$

where ${}_\gamma f^{\alpha\beta} = \gamma^{\alpha,\beta} - \gamma^{\beta,\alpha}$, and ${}_\gamma T^{\alpha\beta}_{\text{em}}$ is the electromagnetic stress—energy tensor (22) calculated with ${}_\gamma F^{\alpha\beta} = (1/k)_\gamma f^{\alpha\beta}$. In the derivation of (45), we used the result that a'_σ is sourceless, as previously discussed. The cancellation involves the rate of doing work on the charge by the field due to a'_σ, and the rate that energy is lost by the field. To see this notice that the term ${}_0f^{0\beta}{}_{j\beta}$ is the Lorentz force. The energy balance between charge and field due to a'_σ occurs in detail.

Of more interest is the energy emitted from the region of space containing the charge. So we integrate over space, assuming the volume of integration large enough so that certain surface integrals vanish. The expression for the rate of change of energy in the field due to γ_σ, in the volume of integration becomes

$${}_\gamma\dot{P}^{0'} = -\frac{b}{8\pi}\sum_{j=1}^{2}\frac{d}{dt}\int [{}_jU(G^{00} + R^{00} + 6({}_jp^0)^2 - 3g^{00}p^2$$

$$- 6ig^{00}a'\cdot\text{Im}(\partial\ln{}_jv)) - 6\partial^0{}_jv^*\partial^0{}_jv + 3g^{00}|\partial{}_jv|^2]d^3x. \quad (46)$$

To simplify this, use (23) multiplied by $-3\,{}_jv^*$, (30), assume ${}_jv$ is negligible on the surface of the volume of integration, and

$${}_jv^{,0} \approx iw\,{}_jv, \quad \text{where } w = m/\hbar > 0. \quad (47)$$

All this implies

$${}_\gamma\dot{P}^0 = -\frac{b}{8\pi}\sum_{j=1}^{2}\frac{d}{dt}\int [6\,{}_jU((p^0)^2 - w^2)]d^3x$$

$$= -\frac{M_0}{V}\sum_{j=1}^{2}\frac{d}{dt}\int_j U d^3x. \quad (48)$$

If ${}_jv$ satisfy (19), then $d/dt(\int_j U d^3x) = 0$, implying no radiation. This argument requires more faith in the wave equation than is justified. Terms for the spin must be added before the equation can be regarded as exact. We do not consider it accurate enough to deny the delicate process of radiation to proceed. With the assumption that the coefficients in an eigenfunction expansion of the wavefunction depend on time, we explore the consequences of the source equation (20).

Assume a two-level system,

$${}_jv = [c_1y_1\exp(i\epsilon_1t) + c_2y_2\exp(i\epsilon_2t)]\exp(iwt), \quad (49)$$

where c_n are real and y_n are orthogonal energy eigenstates of (19) normalized to $V/2$. Then $c_1^2 + c_2^2 = 1$, so define θ such that

$$c_1 = \sin\theta, \quad c_2 = \cos\theta, \quad \dot{c}_1 = \dot{\theta}c_2, \quad \text{and } \dot{c}_2 = -\dot{\theta}c_1. \quad (50)$$

Even when radiating, the charge of the electron in an atomic system should have a constant part,

$$-\dot{e} = 0 = \frac{d}{dt}\langle\int j^0 d^3x\rangle_t$$

$$= q\frac{d}{dt}(w\int U d^3x + \epsilon_1c_1^2V + \epsilon_2c_2^2V). \quad (51)$$

Therefore,

$$\frac{d}{dt}\int U d^3x = -\frac{V}{w}2c_1c_2\dot{\theta}(\epsilon_1 - \epsilon_2) \quad (52)$$

and

$${}_\gamma\dot{P}^0 = \frac{M_0}{m}2c_1c_2\dot{\theta}\hbar(\epsilon_1 - \epsilon_2). \quad (53)$$

Integrating over all time, the energy liberated is $(M_0/m)\hbar(\epsilon_1 - \epsilon_2)$. In order that the atom not be a source or sink of energy, this value must agree with (43). Therefore, noticing that V changes in (34),

$$M_0 = m. \quad (54)$$

An evaluation of the current shows the emitted radiation has frequency $|\epsilon_1 - \epsilon_2|$. It is well known that absorption of energy from radiation occurs at the same frequency. The total energy absorbed and the total emitted in radiative transitions divided by the frequency of the radiation is constant, h. Electromagnetic radiation from atomic electrons appears in quantized energy packets, in agreement with this result.

VI. AN ENERGY LEVEL SHIFT

The effect of the extra mass terms in (19), $p\cdot a'$ and a'^2, may be treated as a small perturbation. Their effect is greatest when a'_σ is the largest. Therefore, assume a one-electron atom. To first order the energy change will be

$$\Delta E = -\frac{h^2}{2mV}\sum_{j=1}^{2}\int_j U a'^2 d^3x. \quad (55)$$

The term linear in p^σ is zero by (30). If a'_σ has only a time component, $-ke/r$, then

$$\Delta E = \frac{e^4 Z^2}{2m}\left\langle\frac{1}{r^2}\right\rangle, \quad \text{where} \left\langle\frac{1}{r^2}\right\rangle = \frac{1}{V}\sum_{j=1}^{2}\int \frac{{}_jU}{r^2}\, d^3x. \tag{56}$$

Assuming both wave components are the eigenfunction belonging to the same level in hydrogen yields a difference in ΔE for the $2S-2P$ levels

$$\Delta E_S - \Delta E_p = mZ^4\alpha^4/12, \quad \text{where } \alpha = e^2/\hbar. \tag{57}$$

The equivalent frequency shift is 3×10^4 Mhz. A shift of this magnitude is not observed. We conclude that the theory does not represent the details of atomic spectroscopy accurately. This is not such a bad failure for a spinless geometric theory of charge.

VII. CONCLUSION

This paper has discussed a specific geometric theory of charge, conceived from the viewpoint that geometry itself most simply describes physical events. The simplicity of the geometry is compromised by the necessary modification of the Weyl scalar curvatures. The physics lacks spin and related concepts. Yet the theory describes, with some accuracy, phenomena ranging from the astronomical, with general relativity as a limit, to the minute with a treatment of electromagneitc radiation.

[1]H. Weyl, *Space, Time, Matter*, transl. by H. L. Brose (Methuen, London, 1922); A.S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U. P., London, 1960), 2nd ed., Chap. VII; R. Adler, *et al.*, *Introduction to General Relativity* (McGraw-Hill, New York, 1965), 1st ed., Chap. 13, p. 401.
[2]H. Bondi, "Negative mass in general relativity," Rev. Mod. Phys. **29**, 423 (1957).
[3]J.A. Wheeler, "Geons," Phys. Rev. **97**, 511 (1955); C.W. Misner and J. A. Wheeler, "Classical physics as geometry: Gravitation, electromagnetism, unquantized charge, and mass as properties of curved empty space," Ann. Phys. (N.Y.) **2**, 525 (1957).
[4]H. Weyl, Ref. 1, Sec. 36, p. 295; A.S. Eddington, Ref. 1, Sec. 90, p. 209.
[5]H. T. Flint and J.W. Fisher, "The fundamental equations of wave mechanics and the metrics of space," Proc. Roy. Soc. (London) A **117**, 625, 630 (1928).
[6]P. A. M. Dirac, "Long range forces and broken symmetries," Proc. Roy. Soc. (London) A **333**, 403 (1973); "Cosmological models and the Large Numbers hypothesis," Proc. Roy. Soc. (London) A **338**, 439 (1974).
[7]Commas denote partial derivatives, semicolons for covariant derivatives. All covariant derivatives are respect to $C^\alpha_{\beta\gamma}$.
[8]The invariance of ${}_0f^{\alpha\beta}$ under conformal transformations is the origin of the term "gauge invariance" of electrodynamics.
[9]R. Adler, *et al.*, Ref. 1, p. 416.
[10]Units throughout have $G = c = 1$, where G is the gravitational constant and c is the speed of light.
[11]The term S is not gauge invariant. The use of a gauge noninvariant Lagrangian is discussed by P. A. M. Dirac, "A new classical theory of electrons," Proc. Roy. Soc. (London) A **209**, 293 (1951).

Large time behavior of the superfluorescent decay

V. Benza and E. Montaldi

Istituto di Scienze Fisiche dell'Università—Milano
and Istituto Nazionale di Fisica Nucleare—Sezione di Milano, 20133 Milano, Italy
(Received 29 July 1977)

The large time behavior of the superfluorescent decay is investigated by the inverse scattering technique.

1. INTRODUCTION

This paper is devoted to the study of the large time behavior of the electric field emitted by a completely excited finite length sample of nondegenerate two-level atoms decaying in the vacuum (superfluorescence). This process is described, at the semiclassical level, by the equation

$$\varphi_{tt} + \varphi_{xt} = \sin\varphi, \tag{1.1}$$

satisfied by the Bloch angle, when the cooperation time τ_c and the cooperation length l_c are taken equal to 1. We assume a small initial perturbation ϵ of the unstable equilibrium configuration, which turns out to be a key step towards the solution of the problem. Indeed, solving Eq. (1.1) by the inverse spectral transform, the Marchenko equations can be treated in a perturbative way, the smallness parameter being $\epsilon t^{1/4}$. Our result expresses the electric field leaving the sample as $t^{-3/4}$ times an oscillating function of t, which appears to be in reasonable agreement with numerical computations.

2. THE INVERSE SPECTRAL TRANSFORM

We consider the interaction of an electric field with a two-level atomic system, described by the Maxwell–Bloch equations

$$\mathcal{P}_t=\mathcal{E}\mathcal{N}, \quad \mathcal{N}_t=-\mathcal{E}\mathcal{P}, \quad \mathcal{E}_t+\mathcal{E}_x=\mathcal{P}, \tag{2.1}$$

where $\mathcal{P}$, $\mathcal{N}$, and $\mathcal{E}$ are the polarization, the population inversion, and the electric field envelope respectively, and the third row is the wave equation in the well-known slowly-varying envelope approximation, describing one-sided (from left to right) propagation.[1]

By putting $\mathcal{P}=\sin\varphi$, $\mathcal{N}=\cos\varphi$, and $\mathcal{E}=\varphi_t$, Eq. (2.1) reduces to the sine-Gordon equation in the coordinate system relevant for nonlinear optics,

$$\varphi_{tt} + \varphi_{xt} = \sin\varphi. \tag{2.2}$$

We impose the physical requirement $\varphi \to \pi \pmod{2\pi}$ for $|x| \to \infty$, which means that the atoms are assumed to occupy the ground level at infinity.

Equation (2.2) may be solved by the inverse scattering method, with the associated linear problem

$$v_x=\begin{bmatrix} -i\left(\xi+\frac{1}{4\xi}\cos\varphi\right) & -\frac{i}{4\xi}\sin\varphi-\frac{1}{2}(\varphi_x+\varphi_t) \\ -\frac{i}{4\xi}\sin\varphi+\frac{1}{2}(\varphi_x+\varphi_t) & i\left(\xi+\frac{1}{4\xi}\cos\varphi\right) \end{bmatrix} v, \tag{2.3a}$$

$$v_t=\frac{i}{4\xi}\begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix} v. \tag{2.3b}$$

This system follows, by the appropriate change of variables, from the analogous one given by Ablowitz *et al.*[2] for the sine-Gordon equation written in light-cone coordinates. It looks very similar to the one considered by Kaup[3] in his treatment of the same equation in laboratory coordinates. We now introduce, for real ξ, the usual Jost solutions of (2.3a)

$$\begin{aligned} &\phi(\xi,x)\to\begin{bmatrix}1\\0\end{bmatrix}\exp(-ikx), \quad x\to-\infty, \\ &\bar\phi(\xi,x)=i\sigma_2\phi^*(\xi^*,x)\to\begin{bmatrix}0\\-1\end{bmatrix}\exp(ikx), \end{aligned} \tag{2.4a}$$

$$\begin{aligned} &\psi(\xi,x)\to\begin{bmatrix}0\\1\end{bmatrix}\exp(ikx), \quad x\to+\infty, \\ &\bar\psi(\xi,x)=i\sigma_2\psi^*(\xi^*,x)\to\begin{bmatrix}1\\0\end{bmatrix}\exp(-ikx), \end{aligned} \tag{2.4b}$$

where

$$k=k(\xi)=\xi-1/4\xi, \tag{2.5}$$

and

$$\sigma_2=\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}$$

is a Pauli spin matrix.

We note that, with our choice of the asymptotic conditions, $\phi(\xi,x)\exp(ikx)$ and $\psi(\xi,x)\exp(-ikx)$ are analytic functions of ξ in the upper half ξ-plane. The proof essentially follows the same lines of Ref. 3 and will not be repeated here. For some remarks concerning an alternative choice of the asymptotic conditions, see Ref. 4.

Let us assume, for $\psi(\xi,x)$, the triangular representation[3]

$$\psi(\xi,x)=\begin{bmatrix}0\\1\end{bmatrix}\exp(ikx)+\int_x^\infty ds\left[K(x,s)+\frac{i}{\xi}M(x)L(x,s)\right]\exp(iks), \tag{2.6}$$

where

$$M(x)=\begin{bmatrix} -\sin\frac{\varphi}{2} & \cos\frac{\varphi}{2} \\ \cos\frac{\varphi}{2} & \sin\frac{\varphi}{2} \end{bmatrix}. \tag{2.7}$$

Proceeding along the lines of Ref. 3, we get

0022-2488/78/1903-0619$1.00

$$\begin{bmatrix} \frac{\partial}{\partial x}-\frac{\partial}{\partial y} & \frac{1}{2}(\varphi_t+\varphi_x) \\ -\frac{1}{2}(\varphi_t+\varphi_x) & \frac{\partial}{\partial x}+\frac{\partial}{\partial y} \end{bmatrix} K(x,y) = \begin{bmatrix} 0 & 2\cos\frac{\varphi}{2} \\ -2\cos\frac{\varphi}{2} & 0 \end{bmatrix} L(x,y), \tag{2.8a}$$

$$\begin{bmatrix} \frac{\partial}{\partial x}-\frac{\partial}{\partial y} & -\frac{1}{2}\varphi_t \\ \frac{1}{2}\varphi_t & \frac{\partial}{\partial x}+\frac{\partial}{\partial y} \end{bmatrix} L(x,y) = \begin{bmatrix} 0 & \frac{1}{2}\cos\frac{\varphi}{2} \\ -\frac{1}{2}\cos\frac{\varphi}{2} & 0 \end{bmatrix} K(x,y), \tag{2.8b}$$

together with the conditions

$$\lim_{y\to+\infty} K(x,y)=\lim_{y\to+\infty} L(x,y)=0, \tag{2.9}$$

$$K_1(x,x)=\tfrac{1}{4}(\varphi_t+\varphi_x), \tag{2.10}$$

$$L_1(x,x)=-\tfrac{1}{4}\cos\frac{\varphi}{2}. \tag{2.11}$$

The scattering data are introduced, in the usual way, by

$$\begin{aligned} \phi(\xi,x)&=a(\xi)\bar{\psi}(\xi,x)+b(\xi)\psi(\xi,x),\\ \bar{\phi}(\xi,x)&-\bar{a}(\xi)\psi(\xi,x)+\bar{b}(\xi)\bar{\psi}(\xi,x). \end{aligned} \tag{2.12}$$

The Marchenko equations are now readily written down by

$$\begin{aligned} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} K^*(x,y)+\begin{bmatrix} 0 \\ 1 \end{bmatrix} F^{(0)}(x+y) &+\int_x^\infty ds[K(x,s)F^{(0)}(s+y)\\ &+\bar{L}(x,s)F^{(1)}(s+y)]=0, \end{aligned} \tag{2.13}$$

$$\begin{aligned} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \bar{L}^*(x,y)+\begin{bmatrix} 0 \\ 1 \end{bmatrix} F^{(1)}(x+y) &+\int_x^\infty ds[K(x,s)F^{(1)}(s+y)+\bar{L}(x,s)F^{(2)}(s+y)]=0, \end{aligned} \tag{2.14}$$

where

$$\bar{L}(x,y)=2M(x)L(x,y), \tag{2.15}$$

and

$$\begin{aligned} F^{(n)}(z)&=\frac{1}{2\pi}\int_c d\xi\left(\frac{i}{2\xi}\right)^n \rho(\xi)\exp[ik(\xi)z],\\ \rho(\xi)&=\frac{b(\xi)}{a(\xi)}, \end{aligned} \tag{2.16}$$

the contour C starting from $\xi=-\infty+i0^+$, passing over all zeros of $a(\xi)$, and ending at $\xi=+\infty+i0^+$.

3. THE SUPERFLUORESCENCE INITIAL CONDITIONS

We now solve the direct scattering problem (2.3a) under the initial data

$$\varphi(x,0)=\begin{cases} \pi, & |x|>l,\\ 0<\epsilon\ll\pi, & |x|<l \end{cases}$$
$$=\pi-(\pi-\epsilon)\theta(l^2-x^2), \tag{3.1}$$

$$\varphi_t(x,0)=0. \tag{3.2}$$

Physically, this choice is intended to simulate the spontaneous cooperative decay of a finite sample of completely excited atoms in the electric field vacuum ($\varphi_t=0$). The small initial Bloch angle ϵ provides a perturbation of the equilibrium position, necessary for the decay process to start. We want to obtain some information about the behavior of the electric field leaving the sample, i.e., of $\varphi_t(l,t)$. As a matter of fact, this quantity does not depend on the values of the solution for $x>l$, because—as already emphasized—Eq. (2.2) describes only propagation from left to right.

We quote here only the relevant results (for details of derivation, see Appendix A):

$$\begin{aligned} a(\xi)=\frac{\exp(2ikl)}{2\omega(\omega+\xi+\cos\epsilon/4\xi)}\Bigg[&\left(\omega+\xi+\frac{1}{4\xi}\right)^2 \times\exp(2i\omega l)\cos^2\frac{\epsilon}{2}+\left(\omega+\xi-\frac{1}{4\xi}\right)^2\\ &\times\exp(-2i\omega l)\sin^2\frac{\epsilon}{2}\Bigg], \end{aligned} \tag{3.3}$$

$$b(\xi)=-i\left(\omega+\xi+\frac{1}{4\xi}\right)\left(\omega+\xi-\frac{1}{4\xi}\right)\times\frac{\sin(2\omega l)\sin\epsilon}{2\omega(\omega+\xi+\cos\epsilon/4\xi)}, \tag{3.4}$$

where

$$\omega=\omega(\xi)=\left(\xi^2+\frac{1}{16\xi^2}+\tfrac{1}{2}\cos\epsilon\right)^{1/2}. \tag{3.5}$$

It turns out that $a(\xi)$ has no zeros, because Eq. (3.5) is not consistent with the requirement $a(\xi)=0$. Indeed, by solving Eq. (3.5) for ξ, it is easily seen that the product of the roots is ϵ independent; on the other hand, by writing

$$\begin{aligned} \cos^2\frac{\epsilon}{2}&=\omega^2-\left(\xi-\frac{1}{4\xi}\right)^2,\\ \sin^2\frac{\epsilon}{2}&=\left(\xi+\frac{1}{4\xi}\right)^2-\omega^2, \end{aligned} \tag{3.6}$$

the condition $a(\xi)=0$ factorizes into the equations

$$\begin{aligned} &2\xi^2+2\omega\xi+\tfrac{1}{2}\cos\epsilon=0,\\ &4\xi^2\cos\epsilon-4i\xi\omega\cot2\omega l+1=0, \end{aligned} \tag{3.7}$$

for both of which the product of the roots depends on ϵ, thus contradicting our previous statement.

Equation (2.16), recalling[5] that $\rho(\xi,t)=\rho(\xi,0)\exp(it/2\xi)$, then takes the form

$$F^{(n)}(z,t)=-\frac{\sin\epsilon}{2\pi}\int_{-\infty}^{\infty} d\xi\left(\frac{i}{2\xi}\right)^n \times\frac{\exp[ik(\xi)(z-2l)+it/2\xi]}{1/4\xi^2+\cos\epsilon-(i\omega/\xi)\cot2\omega l}. \tag{3.8}$$

Before concluding this section, let us consider the following choice of initial conditions:

$$\varphi(x,0)=\pi\theta(x^2-l^2),\quad \varphi_t(x,0)=\epsilon\theta(l^2-x^2), \tag{3.9}$$

which describe the decay from the unstable equilibrium position stimulated by a small initial uniform electric field. We point out that these conditions give rise to a rather different structure of the scattering data.

Indeed (see Appendix A) one finds in this case

$$a(\xi)=\frac{\exp(2ikl)}{(\Omega+\xi+1/4\xi)^2+\epsilon^2/4}\left[\frac{\epsilon^2}{4}\exp(-2i\Omega l)+\left(\Omega+\xi+\frac{1}{4\xi}\right)^2\exp(2i\Omega l)\right], \tag{3.10}$$

$$b(\xi)=\frac{\epsilon(\Omega+\xi+1/4\xi)}{(\Omega+\xi+1/4\xi)^2+\epsilon^2/4}\sin(2\Omega l), \tag{3.11}$$

where

$$\Omega=\left[\left(\xi+\frac{1}{4\xi}\right)^2+\frac{\epsilon^2}{4}\right]^{1/2}, \tag{3.12}$$

and it is now possible to show that, for any ϵ and l, the equation $a(\xi)=0$ is consistently satisfied. The solution of the corresponding Marchenko equations [compare (2.16)] thus becomes considerably more complicated, due to the appearance of multisoliton contributions together with the continuous spectrum. At the present time, we are not able to explain the physical origin of such a striking difference, as compared with the preceding case. Nonetheless, bearing in mind that we are interested in the large time behavior of the electric field leaving the sample, we cannot *a priori* exclude that, for $x-l\ll 1$, the situation is in fact similar to the one originated from the conditions (3.1) and (3.2), which will be discussed in the next section.

4. DERIVATION OF THE ASYMPTOTIC SOLUTION

It is known that, when the continuous spectrum is involved, the Marchenko equations are in general not exactly solvable; in this case, however, it is possible to obtain the behavior of the solution for large t.

As shown in Appendix B, the leading term of the asymptotic expansion of Eq. (3.8) for large t is

$$F^{(n)}(z,t)\sim-\frac{\sin\epsilon}{\sqrt{2\pi}}\frac{i^n(z-2l)^{n-1}}{\nu^{n-1/2}}\frac{\cos\epsilon\cos(\nu+\pi/4)-\cot\nu l/(z-2l)\sin(\nu+\pi/4)}{\cos^2\epsilon+\cot^2\nu l(z-2l)},\quad n=0,2, \tag{4.1}$$

$$F^{(1)}(z,t)\sim+\frac{\sin\epsilon}{\sqrt{2\pi\nu}}\frac{\cos\epsilon\sin(\nu+\pi/4)+\cot\nu l/(z-2l)\cos(\nu+\pi/4)}{\cos^2\epsilon+\cot^2\nu l/(z-2l)},$$

where

$$\nu=\nu(z,t)=\sqrt{(z-2l)(2t-z+2l)}. \tag{4.2}$$

We remark that the condition $z>2l$ is not a restrictive one, since—as already pointed out—we are interested in the solution at $x\geq l$.

Equation (4.1) is in fact valid for finite z; for large z, with $2t/z=1+\alpha$, α being a positive finite quantity, we have instead

$$F^{(n)}(z,t)\sim-\sin\epsilon\left(\frac{i}{2}\right)^n\alpha^{1/4-n/2}(\pi z)^{-1/2}\times\frac{(1/4\alpha+\cos\epsilon)\cos\theta-\alpha^{-1/2}\omega(\alpha^{1/2})\cot[2\omega(\alpha^{1/2})l]\sin\theta}{(1/4\alpha+\cos\epsilon)^2+(1/\alpha)\omega^2(\alpha^{1/2})\cot^2[2\omega(\alpha^{1/2})l]},\quad n=0,2, \tag{4.3}$$

$$F^{(1)}(z,t)\sim\tfrac{1}{2}\sin\epsilon\,\alpha^{-1/4}(\pi z)^{-1/2}\times\frac{(1/4\alpha+\cos\epsilon)\sin\theta+\alpha^{-1/2}\omega(\alpha^{1/2})\cot[2\omega(\alpha^{1/2})l]\cos\theta}{(1/4\alpha+\cos\epsilon)^2+(1/\alpha)\omega^2(\alpha^{1/2})\cot^2[2\omega(\alpha^{1/2})l]},$$

$$\theta=2(z-2l)\alpha^{1/2}+\pi/4.$$

In both Eqs. (4.1) and(4.3), we have the behavior of a power times an oscillating function. On the contrary, the last case ($z\to\infty$, with $2t/z=1-\beta$, $0<\beta<1$) is characterized by powers of z times $\exp(-2\beta^{1/2}z)$, so that, in Eqs. (2.13) and (2.14), the contributions to $K(x,y)$ and $\bar{L}(x,y)$ arising from this region can be neglected.

According to this last remark, the integration in the Marchenko equations can be restricted to the interval $(x,2t)$. For $s\sim 2t$, the kernels are given by expression (4.3), which is of a lower order as compared with (4.1). This enables us to replace, in Eqs. (2.13) and (2.14), the upper integration limit by a function $\alpha(t)$ such that $t<\alpha(t)<2t$, and the $F^{(n)}(z,t)$ by (4.1).

We point out that, as shown in Appendix B, the asymptotic estimate (4.1) is already a good onefor $t\gtrsim 25(x-l)^{-1}$, the unit time being the cooperation time τ_c,[6] according to Eq. (2.2), and the unit length being l_c.

Although we are not able to solve Marchenko equations even in this limit, we can take a sufficiently small ϵ, and try a perturbative solution. We note that the expansion parameter is actually $\epsilon t^{1/4}$, as it can be seen from (4.1) for $n=0$. From Eqs. (2.13) and (2.14) we easily recognize that the lowest order solution is already given by the inhomogeneous terms. From (2.11) and (2.15), we have

$$\tan\frac{\varphi}{2}=\frac{1+2\bar{L}_2(x,x)}{2\bar{L}_1(x,x)}, \tag{4.4}$$

whence we obtain

$$\varphi=2\tan^{-1}\frac{1}{2F_0^{(1)}(2x,t)}, \tag{4.5}$$

where $F_0^{(1)}(2x,t)$, the first order approximation of $F^{(1)}(2x,t)$, is given by

$$F_0^{(1)}(2x,t)=\epsilon\nu^{-1/2}A(\nu,x), \tag{4.6}$$

with

$$A(\nu,x)=\frac{1}{\sqrt{2\pi}}\sin\frac{\nu l}{2x-2l}\cos\left(\frac{\nu l}{2x-2l}-\nu-\frac{\pi}{4}\right),\quad \nu=\nu(2x,t). \tag{4.7}$$

Finally, for the electric field φ_t, we get

$$\varphi_t=\frac{2\epsilon}{t-x+l}\left[\frac{1}{2\sqrt{\nu}}A(\nu,x)-\sqrt{\nu}\,\frac{\partial A(\nu,x)}{\partial\nu}\right],\quad t\ll\epsilon^{-4}, \tag{4.8}$$

where x has to be taken at the right of l.

5. CONCLUDING REMARKS

It is well known, from numerical computations[6] that for short times Eq. (1.1), with the initial conditions (3.1) and (3.2), gives rise to a sharp pulse of the outgoing electric field (superfluorescent pulse). The times involved in our analysis are too large in order to include the description of such a pulse; what we see is just the queue of the decay process. Our result states on analytical grounds that, independently of the value of the initial tipping angle ϵ ($\epsilon\lesssim 10^{-1}$) and of the length l of the sample, one has always a damped oscillatory behavior. The scaling factor $t^{-3/4}$ appears to be the same previously found by Burnham and Chiao in a paper on coherent resonance fluorescence.[7] These authors investigate the response of a system of two-level atoms, driven by a resonant deltalike light pulse, by studying Eq. (1.1) under the boundary conditions $\varphi(x,t=x)=\epsilon$ and $\varphi_t(x=0,t)=0$. This problem, translated in the new variables $z=x$ and $\tau=t-x$, takes the form

$$\varphi_{z\tau}=\sin\varphi,\quad \varphi(z=0,\tau)=\varphi(z,\tau=0)=\epsilon, \tag{5.1}$$

and these boundary conditions suggest that we search for a similarity solution $\varphi(z,\tau)=Y(q)$, $q=2(\tau z)^{1/2}$. Such an ansatz is not consistent with our initial conditions for any value of the parameter l; in fact, the asymptotic solution (4.5) exhibits dependence both on $(\tau z)^{1/2}$ and $(\tau/z)^{1/2}$. It is not surprising that field (4.8) is qualitatively analogous to the one obtained in Eq. (19) of Ref. 7 by a linearization procedure, since for large t we are dealing essentially with the linearized version of Eq. (1.1); on the other hand, it is noteworthy to recall that Eq. (4.8) heavily keeps track of the whole nonlinear evolution of the initial data.

ACKNOWLEDGMENTS

We are indebted to Professor R. Bonifacio, Professor L. Lugiato, Professor M. Gronchi, and Professor L. Girardello for helpful discussions.

APPENDIX A

The direct scattering problem (2.3a), with the initial data (3.1) and (3.2), can be written as

$$v_x=[R\theta(l^2-x^2)-ik\sigma_3\theta(x^2-l^2)-i(\pi-\epsilon)x\delta(l^2-x^2)\sigma_2]v, \tag{A1}$$

where

$$R=-\frac{i}{4\xi}\sigma_1\sin\epsilon-\left(\frac{i}{4\xi}\cos\epsilon+i\xi\right)\sigma_3. \tag{A2}$$

By putting

$$v=\exp\left(-\frac{i}{2}(\pi-\epsilon)\,\theta(x^2-l^2)\sigma_2\right)w, \tag{A3}$$

Eq. (A1) becomes

$$w_x=[ik\theta(x^2-l^2)(\sigma_1\sin\epsilon+\sigma_3\cos\epsilon)+R\theta(l^2-x^2)]w. \tag{A4}$$

The solutions of Eq. (A4), corresponding to the Jost solutions $\phi(\xi,x)$ and $\psi(\xi,x)$ of (A1), are

$$w_L=\begin{bmatrix}\sin\epsilon/2\\ -\cos\epsilon/2\end{bmatrix}\exp(-ikx),\quad x<-l,$$
$$w_R=\begin{bmatrix}\cos\epsilon/2\\ \sin\epsilon/2\end{bmatrix}\exp(ikx),\quad x>l. \tag{A5}$$

The general solution of (A4) for $|x|<l$ is given by

$$w_I=\alpha\begin{bmatrix}-\dfrac{\sin\epsilon}{4\xi}\\ \omega+\xi+\dfrac{\cos\epsilon}{4\xi}\end{bmatrix}\exp(i\omega x)+\beta\begin{bmatrix}\omega+\xi+\dfrac{\cos\epsilon}{4\xi}\\ \dfrac{\sin\epsilon}{4\xi}\end{bmatrix}\exp(-i\omega x), \tag{A6}$$

$$\alpha=-\frac{\omega+\xi+1/4\xi}{2\omega(\omega+\xi+\cos\epsilon/4\xi)}\times\exp[i(\omega+k)l]\cos\frac{\epsilon}{2},$$
$$\beta=\frac{\omega+\xi-1/4\xi}{2\omega(\omega+\xi+\cos\epsilon/4\xi)}\times\exp[-i(\omega-k)l]\sin\frac{\epsilon}{2}, \tag{A7}$$

for the ϕ solution, and

$$\alpha'=\beta, \quad \beta'=-\alpha, \tag{A8}$$

α' and β' being the coefficients relative to ψ.

The expressions (3.3) and (3.4) are now obtained by recalling that $a=\phi_1\psi_2-\phi_2\psi_1$, $b=\bar{\psi}_1\phi_2-\bar{\psi}_2\phi_1$, where the Jost solutions can be taken in $|x|<l$.

The derivation of Eqs. (3.10) and (3.11) is quite analogous.

APPENDIX B

We are concerned with the asymptotic behavior of integrals having the general form

$$\begin{aligned} I(\alpha,\beta) &= \int_{-\infty}^{\infty} d\xi\, h(\xi) \exp[i(\alpha\xi+\beta/\xi)] \\ &= \int_0^{\infty} d\xi \{h(\xi)\exp[i(\alpha\xi+\beta/\xi)] \\ &\quad + h(-\xi)\exp[-i(\alpha\xi+\beta/\xi)]\}. \end{aligned} \tag{B1}$$

Let us assume α and β to have the same sign. By the substitution $\xi=(\beta/\alpha)^{1/2}\exp u$, we have

$$\exp[\pm i(\alpha\xi+\beta/\xi)] \to \exp[\pm 2i\sqrt{\alpha\beta}\cosh u], \tag{B2}$$

and, according to standard arguments,[8] the leading term of the asymptotic expansion for $\beta\to\infty$ is obtained by the replacement $\cosh u \to 1+u^2/2$. This gives

$$\begin{aligned} I(\alpha,\beta) \sim (\pi^2\beta/\alpha^3)^{1/4} [&h(\sqrt{\beta/\alpha})\exp(2i\sqrt{\alpha\beta}+i\pi/4) \\ &+ h(-\sqrt{\beta/\alpha})\exp(-2i\sqrt{\alpha\beta}-i\pi/4)]. \end{aligned} \tag{B3}$$

We observed, furthermore, that the kernel (B2) is the same appearing in a well-known integral representation of the Bessel function $J_0(2\sqrt{\alpha\beta})$,[9] for which numerical computations show that the asymptotic formula is quite reliable already for $\sqrt{\alpha\beta}\geq 5$. This remark, as well as the fact that our $h(\xi)$ is a bounded function, justifies the assertion made in the text that the asymptotic estimate (4.1) is already a good one for $t>25(x-l)^{-1}$.

When α and β have opposite sign, Eq. (B2) must be modified according to

$$\exp[\pm i(\alpha\xi+\beta/\xi)] \to \exp[\pm 2i\sqrt{|\alpha\beta|}\sinh u], \tag{B4}$$

and this kernel is typical of a standard integral representation of the modified Bessel function $K_0(2\sqrt{|\alpha\beta|})$,[9] which is known to decay exponentially for large argument, thus justifying the assertion made below Eq. (4.3).

[1]See, for instance, G. L. Lamb, Jr., Phys. Rev. A **12**, 2052 (1975).

[2]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973).

[3]D. J. Kaup, Stud. Appl. Math. LIV, 165–179 (1975).

[4]The interaction is switched off also if $\varphi\to 0 \pmod{2\pi}$ as $|x|\to\infty$; in this case, however, one has

$$\phi(\xi,x)\to\begin{bmatrix}1\\0\end{bmatrix}\exp(-ik'x),\ x\to-\infty,$$

$$\psi(\xi,x)\to\begin{bmatrix}0\\1\end{bmatrix}\exp(ik'x),\ x\to+\infty,$$

$$k'=k'(\xi)=\xi+1/4\xi,$$

and, as it is easily seen, these Jost solutions diverge for $x\to-\infty$ and $x\to+\infty$ respectively, if ξ is in the half-circle $|\xi|<\frac{1}{2}$, $\mathrm{Im}\,\xi>0$. One is then led to continue $\bar{\phi}(\xi,x)$ and $\bar{\psi}(\xi,x)$ in this domain, and the discussion becomes somewhat involved.

[5]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **L111**, 249–315 (1974).

[6]R. Saunders and R. K. Bullough, "Theory of FIR Superfluorescence," University of Manchester, preprint (1977).

[7]D. C. Burnham and R. Y. Chiao, Phys. Rev. **188**, 667 (1969).

[8]See, for instance, E. T. Copson, *Asymptotic Expansions* (Cambridge U. P., Cambridge, 1965).

[9]A. Erdelyi *et al.*, *Bateman Manuscript Project, Higher Transcendental Functions* (McGraw-Hill, New York, 1954), Vol. 2, p. 82.

Homogeneous and isotropic world models in the Yang–Mills dynamics of gravity. The structure of the adiabats[a)]

M. Camenzind

I. Institut für Theoretische Physik, Universität Hamburg, 2000 Hamburg 36, Federal Republic of Germany
(Received 8 September 1976)

The time evolution of homogeneous and isotropic matter distributions is analyzed for the restricted Yang–Mills curvature dynamics of gravity. This theory of gravity is a tidal dynamics for which relativistic matter in detailed balancing cannot produce tidal forces. It defines a dynamical system on the curvature plane spanned by the two components of the Riemann curvature of Robertson–Walker space–times; the essential features of the cosmological solutions are presented by means of their phase portraits in the curvature plane. In the asymptotic limit ($S \to \infty$) the phase portrait, which in general depends on the equation of state and on the change of the entropy per particle, is structurally stable under the transition from Einstein's dynamics to the Yang–Mills dynamics for any realistic equation of state. The phase portraits are explicitly constructed for the equation of state $p = n\rho$, $0 \leq n \leq 1$, and constant entropy per particle. A criterion for the existence of regular trajectories is given for the full Yang–Mills dynamics including entropy production. Finally, we discuss the relations between the observational parameters.

1. INTRODUCTION

The Yang–Mills approach to gravity has mainly been created for two reasons: first, to incorporate the structures of a Yang–Mills gauge theory into the description and dynamics of gravitation—since these constructions have successfully been applied to cover the interactions in special relativity—and secondly, to test an alternative to overcome the difficulties arising in general relativity from the existence of singularities; the "issue of the final state for collapsing matter" has not yet been solved in a physically and mathematically attractive way. The Yang–Mills dynamics for the Lorentz connection (or the spin connection) on space–time is, by definition, an extension of general relativity[1]; the basic dynamical variables are the connection coefficients instead of the metric components, they act as the "gauge potentials of gravity"; this extension of Einstein's dynamics offers indeed a new issue of the final state.

A simple and at the same time illustrative example for the behavior of (nonradiative) geometries under the Yang–Mills dynamics is the time evolution of homogeneous and isotropic world models; because of the symmetries of the space–time, the expansion parameter of the timelike matter congruence is the only dynamical variable, and, consequently, the coupling between matter and geometry of the space–time is governed by a single dynamical equation of second order for the expansion. The whole theory is based on the structures of the well-known Yang–Mills gauge theories: The Lorentz frame bundle on the space–time acts as gauge bundle, the Lorentz connection as gauge connection, and the curvature components as gauge fields. Because of the symmetries, the curvature of homogeneous and isotropic space–times has only two independent components; they span the so-called curvature plane $E_2(\Omega)$ for these space–times. The Bianchi equation and the Yang–Mills equation define then a dynamical system on this curvature plane E_2. The main characteristics of the corresponding phase portrait depend on the equation of state; essentially, we can distinguish between three types of phase portraits: the phase portrait for nonrelativistic matter, $v_S^2 = dp/d\rho < \frac{1}{3}$, the phase portrait for relativistic matter, $v_S^2 = \frac{1}{3}$, and the phase portrait for superrelativistic matter, $\frac{1}{3} < v_S^2 \leq 1$.

There is a further element which determines the phase portrait; the matter of the cosmic fluid is given in terms of a two-component theory consisting of the massive component and the massless part (the photon gas and the neutrino background). As long as there is no energy transfer between these two constituents of the cosmic matter, the time evolution is uniquely governed by the equation of state; the time evolution with energy transfer (which has to be included for the discussion of the lepton era, e.g.) is far more complicated and will not be discussed in the following. Of particular importance is the relationship between Einstein's trajectories and the Yang–Mills trajectories in the phase plane; in some cases, the solutions of Einstein's equations form trajectories in E_2 which span a dynamically closed subset of E_2. In this sense, the radiation-dominated solutions of general relativity (with $p = \frac{1}{3}\rho$ for all constituents of the cosmic matter) move along trajectories which coincide with the *free* trajectories of the Yang–Mills dynamics; the Yang–Mills current vanishes exactly if $dp/d\rho = \frac{1}{3}$ and if the entropy per particle remains constant. Furthermore, $dp/d\rho = 0$ generates the Friedmann trajectory in the phase plane which coincides with the trajectory spanned by the solutions of the Friedmann equation. Because of this latter property, the time evolution of the late epoch of the thermal history of the universe shows the same functional dependence as general relativity.

The Yang–Mills dynamics without energy transfer has solutions of the oscillatory type, the motions represent oscillations around a suitable equilibrium point in E_2. These oscillating models for $p \ll \rho$ differ only in their early phase from the singular Friedmann type, while in the late epoch the Friedmann trajectory acts as an attractor. This is a general property of the Yang–Mills

[a)]Work supported by Schweizerischer Nationalfonds.

dynamics: In a neighborhood of the critical points of the free dynamical system, the phase portrait is in some sense structurally stable under the transition from Einstein's dynamics to the Yang–Mills dynamics, and this for all equations of state. Therefore, observations based on the present state of the universe can actually not distinguish between a singular mode of evolution and a regular periodic mode of evolution with a high relative amplitude in the radius.

The paper is arranged as follows: Sections 2 and 3 describe the gauge bundle and the gauge dynamics for the torsionless connection on Robertson–Walker space–times; in Sec. 4 we discuss the corresponding phase plane E_2 and the dynamical system on E_2. The meaning of the critical points and the classification of the free solutions of this dynamical system are given in Sec. 5. Sections 6 and 7 concern properties of the phase portrait, and finally we discuss the relationship between the observational parameters.

2. THE GAUGE BUNDLE AND THE GAUGE CONNECTION ON ROBERTSON-WALKER SPACE-TIMES

We base our description of homogeneous and isotropic world models on the Robertson–Walker line element[2]

$$ds^2 = dt^2 - S^2(t)\, d\sigma^2, \tag{2.1}$$

$$d\sigma^2 = d\chi^2 + \Sigma^2\, d\Omega^2, \tag{2.2}$$

Σ is given according to the curvature type of 3-space

$$\Sigma = \begin{cases} \sin\chi, & k=+1, \\ \chi, & k=0, \\ \sinh\chi, & k=-1. \end{cases} \tag{2.3}$$

This metric (2.1), $g_{\rm RW}$, defines the gauge bundle $P_4(V_4, \mathrm{SO}(1,3), g_{\rm RW}, \pi)$ over the homogeneous and isotropic space–time V_4, consisting of all Lorentz frames defined by $g_{\rm RW}$. As a particular cross section of $P_4^{\rm RW}$ we choose the comoving observer system σ, $\sigma = \{X_0, X_1, X_2, X_3\}$

$$X_0 = \partial_t \qquad X_1 = S^{-1}\partial_\chi, \qquad X_2 = (S\Sigma)^{-1}\partial_\vartheta, \quad X_3 = (S\Sigma\sin\vartheta)^{-1}\partial_\phi. \tag{2.4}$$

The torsionless connection Γ on $P_4^{\rm RW}$ follows from the first structure equation for the 1-forms θ^a dual to (2.4) (for the notation, see Ref. 3)[4]:

$$\tilde\omega = \tilde\Gamma_a \theta^a\text{: connection form} \tag{2.5}$$

$$\tilde\Gamma_{\underline{0}} = 0 \quad \text{(comoving condition)} \tag{2.6}$$

$$\tilde\Gamma_{\underline{1}} = \theta K_1. \tag{2.7}$$

$$\tilde\Gamma_{\underline{2}} = \theta K_2 + \Sigma'/(S\Sigma) J_3, \tag{2.8}$$

$$\tilde\Gamma_{\underline{3}} = \theta K_3 + \cot\vartheta/(S\Sigma) J_1 - \Sigma'/(S\Sigma) J_2. \tag{2.9}$$

$\theta = \dot S/S$ is the expansion parameter of the fluid lines tangent to X_0; because of the spherical symmetry the expansion is isotropic, shear and rotation vanish.

From the second structure equation for the Lorentz connection Γ we obtain the expressions for the curvature 2-form $\tilde\Omega$

$$\tilde\Omega = \tfrac12 \tilde R_{ab}\, \theta^a \wedge \theta^b, \tag{2.10}$$

with the following components

$$\tilde R_{0i} = (\dot\theta + \theta^2) K_i, \tag{2.11}$$

$$\tilde R_{ik} = -(\theta^2 + k/S^2)\epsilon_{ikl} J_l, \quad i,k,l = 1,2,3. \tag{2.12}$$

The timelike components $\tilde R_{0i}$ describe the tidal forces acting between neighboring fluid lines, and, since

$$\dot\theta + \theta^2 = \ddot S/S, \tag{2.13}$$

they are a measure for the curvature of the curve $S = S(t)$.

The geometry of our space–times V_4 is therefore completely determined by the above two curvature components, $\tilde R_{ik}$ and $\tilde R_{0i}$; for this reason, we introduce the two curvature functions

$$x = x(t) = -(\theta^2 + k/S^2), \tag{2.14}$$

$$y = y(t) = \dot\theta + \theta^2 = \ddot S/S; \tag{2.15}$$

the minus sign in the definition of x has its origin in the relation between x and y; this will come out later on. x and y span a plane $E_2(\Omega)$, which will be called the "curvature plane" because of the geometric meaning of the two functions x and y.

3. THE GAUGE DYNAMICS AND MATTER CONSERVATION

The Lorentz gauge dynamics for gravity is a generalized metric theory of gravity modeled according to the structure of general gauge theories (applied, e.g., to the electromagnetic, weak and strong interaction in special relativity).[1] The dynamical content of the theory is involved in the

Bianchi equations

$$\partial_{[\mu} \tilde R_{\rho\sigma]} + [\tilde\Gamma_{[\mu}, \tilde R_{\rho\sigma]}] = 0, \tag{3.1}$$

Lorentz–Yang–Mills equations

$$(-g)^{-1/2}\partial_\mu[(-g)^{1/2}\tilde R^{\mu\rho}] + [\tilde\Gamma_\mu, \tilde R^{\mu\rho}] = \kappa \tilde J^\rho, \tag{3.2}$$

current conservation

$$(-g)^{-1/2}\partial_\rho[(-g)^{1/2}\tilde J^\rho] + [\tilde\Gamma_\rho, \tilde J^\rho] = 0, \tag{3.3}$$

and in an expression of the components of the so(1,3)-valued current $\tilde J^\rho$ in terms of the matter properties described by means of the stress–energy–momentum of matter, T_{ab},[5]

$$J_{abc} = -(T_{ab;c} - T_{ac;b}) + (\eta_{ab} T_{,c} - \eta_{ac} T_{,b}). \tag{3.4}$$

This Lorentz current $\tilde J^a$ will be decomposed with respect to the chosen basis of so(1,3) into

$$\tilde J^a = -j^{(l)a} K_l + S^{(l)a} J_l; \tag{3.5}$$

the symmetries of V_4 and the meaning of σ require that

$$S^{(l)a} = 0, \quad \forall\, l, a \tag{3.6}$$

$$j^{(l)i} = j\delta^{li}, \quad i = 1,2,3, \quad j^{(l)0} = 0. \tag{3.7}$$

In terms of the total energy ρ and the pressure p, the current component j is locally given by (3.4):

$$j = j(\rho, p, \theta) = \tfrac{1}{2}(\dot{\rho} - 3\dot{p}) + \dot{p} + \theta(\rho + p). \qquad (3.8)$$

The current conservation (3.3), or equivalently $J^a_{bc;a} = 0$, which is the basic conservation law in the Yang–Mills dynamics, determines the divergence of the matter energy–momentum tensor

$$f_a = T_a{}^b{}_{;b}, \qquad (3.9)$$

in the form[1]

$$f_{a;b} - f_{b;a} = R_{bf} T^f{}_a - T_b{}^f R_{fa}. \qquad (3.10)$$

Since the Ricci tensor R_{ab} and the energy–momentum tensor T_{ab} are both diagonal on homogeneous and isotropic space–times, the right-hand side of (3.10) vanishes, and the f_a determine a closed 1-form α, $\alpha = f_a \theta^a$ with $d\alpha = 0$. The only nonvanishing component f_0 describes a "nongeometric" change of the energy of the massive particles

$$\dot{\rho} + 3\theta(\rho + p) = f_0; \qquad (3.11)$$

for this reason, $f_0 = f_0(t)$ is at least a measure for the change in the entropy per particle s, $f_0 = nT\dot{s}$. For cosmic matter in detailed balancing (this is the case, e.g., for the collision-dominated matter in the plasma era, 4×10^3 K $< T < 10^9$ K, and for the lepton era with 8×10^9 K $< T < 5\times10^{11}$ K) this change in the entropy per particle is negligible. Equation (3.10) does not signal a breakdown of the principle of equivalence; the arbitrariness in f_0 on spatially homogeneous space–times, however, allows us to include such processes as entropy production by means of pair creation out of a hot photon gas. [The interpretation of (3.10) on a general space–time has to be carefully analyzed; over nonhomogeneous space–time, the right-hand side of (3.10) does not, in general, vanish since gravity itself contributes to f_a via a back reaction generated by gravitational waves. For a discussion of this point see also Ref. 1.]

By including $f_0 = nT\dot{s}$ together with an equation of state $\rho = \rho(n, s)$, we may simplify the expression for the current j

$$j = -\tfrac{1}{2}(\rho + p)\theta(1 - 3v_s^2) + \tfrac{1}{2}nT\left(1 - \frac{\partial \ln T}{\partial \ln n}\bigg|_s\right)\dot{s}. \qquad (3.12)$$

$v_s^2 = n(\rho + p)^{-1}(\partial p/\partial n)|_s$ describes the sound velocity of the medium. If $\dot{s} = 0$ for all times, the solution of the Yang–Mills dynamics presents the time evolution of massive matter with constant entropy per particle. Relativistic matter, in particular photons and neutrinos ($p = \rho/3$), cannot generate tidal forces for massive matter.

The Yang–Mills equations (3.2) give the following expressions:

$$\rho = 0, \qquad 0 = \tilde{J}^0, \qquad (3.13)$$

$$\rho = 1, \quad -(S\ddot{S})^{\cdot} S^{-2} + 2\theta(\theta^2 + k/S^2) = -\kappa j^{(1)1} S, \qquad (3.14)$$

$$\rho = 2, \quad -(S\ddot{S})^{\cdot} S^{-2} + 2\theta(\theta^2 + k/S^2) = -\kappa j^{(2)2} S\Sigma, \qquad (3.15)$$

$$\rho = 3, \quad -(S\ddot{S})^{\cdot} S^{-2} + 2\theta(\theta^2 + k/S^2) = -\kappa j^{(3)3} S\Sigma \sin\vartheta. \qquad (3.16)$$

Because of the relation

$$(S\ddot{S})^{\cdot} S^{-2} = \ddot{\theta} + 4\theta\dot{\theta} + 2\theta^3 \qquad (3.17)$$

and $j = j^{(1)1}S = j^{(2)2}S\Sigma = j^{(3)3}S\Sigma\sin\vartheta$, there remains one fundamental equation of second order for the expansion parameter θ

$$\ddot{\theta} + 4\theta\dot{\theta} - 2k\theta S^{-2} = \kappa j. \qquad (3.18)$$

The time evolution of homogeneous and isotropic world models is therefore governed by a single component of the Lorentz current; (3.18) describes the simplest time-dependent model within the framework of the Lorentz dynamics.

The cosmological equation (3.18) and the Bianchi equation following from (3.1) define first order derivatives for the curvature function x and y, introduced in (2.14) and (2.15):

Bianchi equation

$$\dot{x} = -2\theta(x + y), \qquad (3.19)$$

Lorentz–Yang–Mills equation

$$\dot{y} = -2\theta(x + y) + \kappa j. \qquad (3.20)$$

Finally, the current conservation (3.3) is trivially satisfied as a consequence of (3.6) and (3.7).

4. THE CURVATURE PLANE $E_2(\Omega)$

The basic equations (3.19) and (3.20) define a *dynamical system* in the "curvature plane"[6] $E_2(\Omega)$

$$\frac{dx}{dt} = -2\theta(x + y), \qquad (4.1)$$

$$\frac{dy}{dt} = -2\theta(x + y + f); \qquad (4.2)$$

and, for matter with constant entropy per particle, f is given by

$$f = f(\rho, p) = \tfrac{1}{4}\kappa(\rho + p)(1 - 3v_S^2). \qquad (4.3)$$

This follows from the current j, defined in (3.12), and the energy conservation (3.11), $v_S^2 = dp/d\rho|_s$, and $\dot{s} = 0$.

As long as $\theta \neq 0$, we may use, instead of t, S itself as parameter for the curves $(x(S), y(S))$, given by their tangent field

$$\frac{dx}{dS} = -\frac{2}{S}(x + y), \qquad (4.4)$$

$$\frac{dy}{dS} = -\frac{2}{S}(x + y + f); \qquad (4.5)$$

the substitution $\xi = S_0/S$, $0 \le \xi < \infty$, gives finally the form of a nonautonomous dynamical system[6] $\mathcal{S}$ on $E_2(\Omega)$;

$$\frac{dx}{d\xi} = 2\xi^{-1}(x + y), \qquad (4.6)$$

$$\frac{dy}{d\xi} = 2\xi^{-1}(x + y + f). \qquad (4.7)$$

Here, S_0 is either max$S$, or min$S$ for regularly oscillating curvature, or $S_0 = \max S$ for singular but turning big-bang trajectories in $E_2(\Omega)$. Once the trajectories of the dynamical system (4.6) and (4.7) are known for a particular type of function f, the time t may be obtained as a function of S:

$$t = t(S) = t(\xi) = \int_{\xi_0}^{\xi} \frac{d\xi'}{\xi'\,\theta(\xi')}\,. \tag{4.8}$$

The trajectories of the free models, i.e., whenever $f\equiv 0$ (either $\rho\equiv 0$ or $v_S^2=\frac{1}{3}$), are straight lines $y=x+C$, $C\in\mathbb{R}$, in $E_2(\Omega)$, since

$$\frac{dy}{dx}=1,\ \forall x,\ \text{whenever } f\equiv 0, \tag{4.9}$$

and in this case the points on $y=-x$ are obviously critical points of (4.6), (4.7) (a point $P\in E_2$ is called a critical point if $P(t)=P$ for all $t\in\mathbb{R}$). By the way, the trajectories of (4.6), (4.7) are only semitrajectories, since ξ is positive; the negative semitrajectories would be generated by a negative S. The original dynamical system (4.1), (4.2) has a second type of critical points, the stationary, or equilibrium points defined by $\theta=0$ for all t; these equilibrium points lie on the axis $y=0$ and define the space–time geometries of constant 3-curvature, $x=-k/S^2$, $y=0$. In particular, the Minkowski space–time, $x=0=y$, appears as a critical point of the Yang–Mills dynamics. According to a general theorem for dynamical systems,[6] every neighborhood of a critical point contains a semitrajectory, and the set of all critical points in E_2 is closed; this means, that for any point on $y=-x$, we find trajectories approaching these points. A topology in $E_2(\Omega)$ may be defined, e.g., invariantly by means of the curvature invariant $I_1=x^2+y^2$, which is invariant under cross section transformations. In the following, we call $P\in E_2$ a *turning point* if $\theta(t_0)=0$ whenever $P=(x(t_0),\,y(t_0))$ and $\ddot{x}(t_0)\neq 0\neq\ddot{y}(t_0)$; this means that a curve $(x(t),\,y(t))$ in E_2 has a turning point for $t=t_0$, if the tangent vector in this point vanishes, and therefore

$$\dot{x}(t_0+\epsilon)=-\dot{x}(t_0-\epsilon),\quad \epsilon>0,$$
$$\dot{y}(t_0+\epsilon)=-\dot{y}(t_0-\epsilon). \tag{4.10}$$

Periodic orbits are, therefore, not given by closed loops in E_2, but by compact segments of corresponding trajectories.

The phase portrait for the dynamical system (4.4), (4.5) essentially depends on the properties of the external structure function $f(S)$. In general we shall distinguish between two types of positive semi-trajectories (defined for $S\in\mathbb{R}^+$):

TABLE I. The classification of the critical points on $y=-x$. Minkowski, de Sitter, and anti-de Sitter space–times are represented by the points in E_2 which are critical under the free Yang–Mills dynamics.

k	θ	S	C	remarks
$+1$	$\omega\tanh(\omega t)$	$\omega^{-1}\cosh(\omega t)$	$+\omega^2$	de Sitter
0	$\pm(\frac{1}{2}C)^{1/2}=\omega$	$S_0\exp(\omega t)$	>0	steady state
		$S_0=\text{const}$	$=0$	Minkowski
-1	$-\omega\tan(\omega t)$	$\omega^{-1}\cos(\omega t)$	$-2\omega^2$	anti-de Sitter
	$\omega\coth(\omega t)$	$\omega^{-1}\sinh(\omega t)$	$+2\omega^2$	

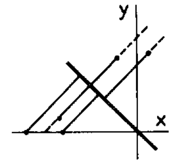


FIG. 1. The various types of trajectories for the free transition models, $k=\pm 1$, represented in the (x,y)-phase plane, $C>0$.

(i) the semitrajectory $\gamma^+(S)$ is called regular trajectory of the dynamical system if $\theta^2(S)<C$, $C>0$, for all $S\in\mathbb{R}^+$;

(ii) the semi-trajectory $\gamma^+(S)$ is called singular, if there does not exist for all $S\in\mathbb{R}$ a $C>0$ such that $\theta^2(S)<C$. In particular, a singular semitrajectory is called a big-bang trajectory if $\theta^2(S)$ is unbounded at $S_0=0$.

The usual Friedmann trajectory, $y=\frac{1}{2}x+\text{const}$, is a big-bang trajectory, where the metric also becomes singular at $S_0=0$, while for a non-big-bang, but singular trajectory a curvature (or matter) singularity occurs on a space–time of finite nonvanishing radius and volume. A corresponding example will be given in Sec. 7. Not any motion along a singular trajectory has to be singular itself; this depends on the initial conditions. In the following the phase portrait will be analyzed for the *restricted* Yang–Mills dynamics, where the current and therefore also the structure function f do not depend on the external function $f_0(t)$ defined by (3.9) and (3.10).

5. CRITICAL POINTS AND THE FREE MODELS ($j=0$)

The current j vanishes in particular whenever $dp/d\rho=\frac{1}{3}$, e.g., if $p=p(\rho)=\frac{1}{3}\rho$. Then the trajectories in $E_2(\Omega)$ are given by the half straight lines $y=x+C$, $C\in\mathbb{R}$. The free models do therefore not correspond to the matter-free space–times, but rather to the states of space–times where matter follows null curves (the so-called "radiation-filled" space–times); for this reason, the tidal action on the timelike congruence defined by the chosen observer cross section vanishes. The current $j(\rho,\,p)$ is a measure for this tidal action.

The motions of a particular model along the trajectories $y=x+C$ depend on the initial conditions; they are in general restricted to segments. Of special importance are the critical points on $y=-x$; they represent, by definition, constant motions in E_2 given by the following dynamical equations for θ and S

$$\dot{\theta}+2\theta^2+kS^{-2}=C,\quad C\in\mathbb{R}, \tag{5.1}$$

$$\dot{\theta}+\theta^2=\theta^2+k/S^2, \tag{5.2}$$

or in integrated form

$$\dot{S}^2=\tfrac{1}{2}CS^2-k. \tag{5.3}$$

The space–time geometries corresponding to (5.3) are summarized in the Table I.

The expansion parameter of the general free models satisfies Eq. (5.1); under the substitution $u=S^2$, we obtain finally

$$\ddot{u}-2Cu+2k=0. \tag{5.4}$$

The general solutions of (5.4) may be written as

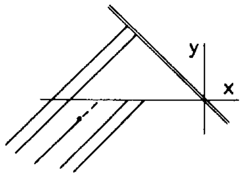


FIG. 2. The types of trajectories for the free big bang models, $k=\pm 1$ $(C>0)$.

$$u(t)=k/C+B_1\exp(\alpha t)+B_2\exp(-\alpha t), \quad \alpha^2=2C>0, \tag{5.5}$$

$$u(t)=-kt^2+B_1t+B, \quad C=0, \tag{5.6}$$

$$u(t)=k/C+B\cos(\omega t+\phi), \quad \omega^2=-2C>0. \tag{5.7}$$

They describe geometries of the transition type for $C>0$, "freely falling" models in the null case, $C=0$, and geometries of the oscillating type for $C<0$. The corresponding motions are shown in the phase space in Fig. 1–5 for the various combinations of the parameters B, B_1, B_2, and the constant C. There exist, in particular, regularly pulsating models for $k=-1$, whenever $|B|<2/\omega^2$; the models for $k=+1, 0$ and $y+x \geq 0$ are always regular.

The trajectories of the free solutions of the Lorentz Yang–Mills dynamics (4.6), (4.7) are identical with the trajectories spanned by the radiation-dominated solutions of Einstein's dynamics for cosmological models; Einstein's dynamics is namely a subsystem defined by the constraint equations

$$3x=-(\kappa\rho+\Lambda), \tag{5.8}$$

$$2y-x=-(\kappa p-\Lambda); \tag{5.9}$$

together with the Bianchi equation (4.1) and the energy conservation (3.11), these constraint equations reproduce the dynamical equation (4.2). For visualization of the constraint equations in the phase plane E_2 we introduce new coordinates in E_2

$$\tilde{x}=x+\tfrac{1}{3}\Lambda, \quad \tilde{y}=y-\tfrac{1}{3}\Lambda \tag{5.10}$$

to get the form of the constraint equations

$$3\tilde{x}=-\kappa\rho, \tag{5.11}$$

$$2\tilde{y}-\tilde{x}=-\kappa p. \tag{5.12}$$

Matter models dominated by an equation of state of the form $p=n\rho$, $0\leq n\leq 1$ and n constant, follow, therefore, the trajectories

$$\tilde{y}=\tfrac{1}{2}(3n+1)\tilde{x} \tag{5.13}$$

or

$$y=\tfrac{1}{2}(3n+1)x+B, \quad B=\tfrac{1}{2}(n+1)\Lambda; \tag{5.14}$$

for $n=\frac{1}{3}$ we regain the straight lines $y=x+C$ with

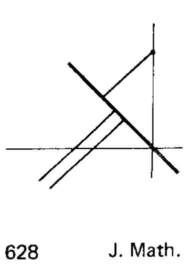


FIG. 3. The types of trajectories for the free transition and big bang models, $k=0$ $(C>0)$.

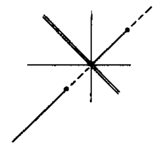


FIG. 4. The trajectory of the free null solution, $C=0$.

$C=\frac{2}{3}\Lambda$. The Einstein trajectories (5.14) span a cone in E_2, which we call in the following the "Einstein cone," of the form

$$2\tilde{x}\leq\tilde{y}\leq\tfrac{1}{2}\tilde{x}, \quad \text{for } 0\leq p\leq\rho. \tag{5.15}$$

The origin of this Einstein cone is located at $(-\frac{1}{3}\Lambda, \frac{1}{3}\Lambda)$ (see Fig. 6), which is identical with the location of the critical points of the free Yang–Mills dynamics; the axis of the cone, $y=x+C$ for $n=\frac{1}{3}$, coincides with the trajectory formed by the free Yang–Mills solutions. The cosmological constant Λ has, however, a completely different interpretation in the Yang–Mills dynamics; it always plays the role of a first integral of the dynamics.

6. THE PHASE PORTRAIT FOR PRESSURELESS MATTER AND $\dot{s}=0$

We consider the time evolution of matter models dominated by an equation of state $p=n\rho$, $0\leq n\leq 1$; then the current's structure function f, defined in (4.3), assumes a particularly simple form

$$f=\tfrac{1}{4}\kappa(n+1)(1-3n)\rho, \tag{6.1}$$

and the dependence of ρ from S is determined by (3.11)

$$\rho=\rho_0(S_0/S)^\delta, \quad \delta=3(n+1). \tag{6.2}$$

In this section, we want to discuss the phase portrait of the dynamical system (4.4,5) for $n=0$ and $\rho_0>0$; and afterwards we show that this phase portrait is structurally stable for $n<\frac{1}{3}$.

In the case $n=0$, we have to solve the following dynamical system:

$$\frac{dx}{dS}=-\frac{2}{S}(x+y), \tag{6.3}$$

$$\frac{dy}{dS}=-\frac{2}{S}(x+y)-\frac{b}{S^4}, \quad b\doteq\frac{\kappa}{2}\rho_0S_0^3. \tag{6.4}$$

By adding and substracting Eqs. (6.3,4) we get integrated expressions for x and y in terms of S,

$$y+x=AS^{-4}-bS^{-3}, \quad A\in\mathbb{R}, \tag{6.5}$$

$$y-x=\tfrac{1}{3}bS^{-3}+B, \quad B\in\mathbb{R} \tag{6.6}$$

or

$$x(S)=\tfrac{1}{2}(AS^{-4}-\tfrac{4}{3}bS^{-3}-B), \tag{6.7}$$

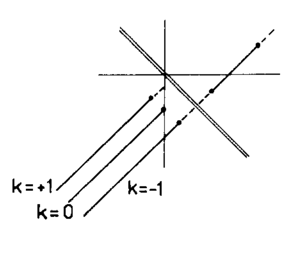


FIG. 5. The oscillatory types of trajectories, $k=0, \pm 1$ $(C<0)$.

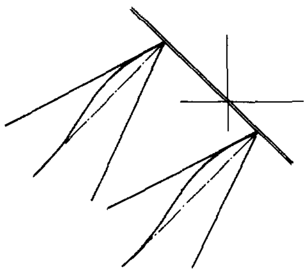


FIG. 6. The "Einstein cones" in the curvature phase plane. The Friedmann trajectory $y=\frac{1}{2}x+\frac{1}{2}\Lambda$ ($p=0$) and the extreme relativistic trajectory $y=2x+\Lambda$ ($p=\rho$) span the interior of the cone with origin at $y=-x$ (on the critical line). The value of the cosmological constant Λ gives the position of the origin of the cone. For a given Λ, any solution of Einstein's equations, with a realistic equation of state, $p\le\rho$, describes a trajectory which lies entirely in the Einstein cone. The origin of the cones are identical with the critical points of the free Yang–Mills dynamical system.

$$y(S)=\tfrac{1}{2}(AS^{-4}-\tfrac{2}{3}bS^{-3}+B). \qquad (6.8)$$

At the same time, an explicit expression is obtained for the expansion parameter θ from (6.7)

$$\theta^2=\tfrac{1}{2}(-AS^{-4}+\tfrac{4}{3}bS^{-3}-2kS^{-2}+B), \qquad (6.9)$$

which gives, implicitly over the integration in (4.8), the solution for $S=S(t)$. The same integration determines also the period of regularly oscillating models which exist, in general, for $A>0$, since then there exists a value of S, $S=S_R$, for which $dy/dx=0$; from Eq. (6.8) we obtain $S_R=2A/b$.

For the discussion of the phase portrait we introduce the translated coordinates

$$z=y-x-B, \quad z>0 \text{ for } b>0, \qquad (6.10)$$

$$\omega=y+x, \qquad (6.11)$$

and eliminate S from (6.5); then the trajectories in the (z,ω)-plane are given by

$$\omega(z)=\alpha z^{4/3}-3z, \quad \alpha\doteq A(3/b)^{4/3}, \qquad (6.12)$$

or in terms of the rescaled coordinates $\bar z\doteq z/B$ and $\bar\omega\doteq\omega/B$, $B\ne0$,

$$\bar\omega(\bar z)=\bar\alpha\bar z^{4/3}-3\bar z. \qquad (6.13)$$

Therefore, for a fixed B, $B\in\mathbb{R}$, the solution trajectories of (6.3), (6.4) form a one-parameter family of curves in $E_2(\Omega)$, only defined in the region $y-x-B>0$ and parametrized by the value of the constant α, which may be positive ($A>0$), zero ($A=0$), or negative ($A<0$); the points $(-\frac{1}{2}B, \frac{1}{2}B)$ on the line $y=-x$ remain critical points of the coupled dynamical system and act as attractors of all the trajectories defined by (6.13) for $z>0$, though a particular motion along a trajectory (defined by corresponding initial conditions) may have a turning point $\theta=0$ before reaching $z=0$. In the limit $z\to0$, all the trajectories approach the Friedmann trajectory $\omega=-3z$ (defined by $A=0$) (see Fig. 7), which in terms of x and y is given by

$$y=\tfrac{1}{2}x+\tfrac{3}{4}B, \quad \text{Friedmann trajectory.} \qquad (6.14)$$

Again, the Friedmann trajectory coincides with the trajectory formed by all solutions of Einstein's equations for $n=0$, which satisfy the constraint equation (5.14). Any trajectory in $E_2(\Omega)$ forms a closed dynamical subsystem of (4.4, 5) in particular, the Friedmann trajectory (for $\alpha=0$) represents the solutions of Einstein's dynamical system; however, the other trajectories for $A\ne0$ have no interpretation in terms of Einstein's equations since there is no direct relation between curvature and matter besides Eq. (3.18). For $\alpha>0$, $\bar\omega(\bar z)$ has always a zero at $\bar z=\bar z_1=(3/\alpha)^3$ with $\bar\omega'(\bar z_1)=1$ for all values of α; this condition is equivalent to $\dot x_1=0$, since for $\bar z_1$ we find $y_1+x_1=0$. A specific initial point $z=z_0$ and $\omega=\omega_0$ determines uniquely a motion along a trajectory (6.13). We discuss in the following the motions for $\alpha>0$; for this reason, we introduce the dimensionless variable $\xi=S_0/S$, where $S_0=S(t_0)$ denotes the initial value for the radius, together with the condition $\theta(t_0)=0$. From this condition, $\theta^2(\xi=1)=0$, we may express the parameters A, b, and S_0^2 in terms of the initial values $\bar z_0$ and $\bar\omega_0=\bar\omega(\bar z_0)$ (for a given α); then

$$\theta^2/B=\tfrac{1}{2}(1-\alpha_4\xi^4+\tfrac{4}{3}\alpha_3\xi^3-\alpha_2\xi^2), \qquad (6.15)$$

with the relations

$$\alpha_2=2k/(BS_0^2)=1+\bar z_0-\bar\omega_0, \qquad (6.16)$$

$$\alpha_3=b/(BS_0^3)=3\bar z_0, \qquad (6.17)$$

$$\alpha_4=A/(BS_0^4)=\bar\omega_0+3\bar z_0=\alpha\bar z_0^{4/3}. \qquad (6.18)$$

The values for the curvature coordinates $\bar x=x/B$ and $\bar y=y/B$ follow from the transformations

$$\bar x=\tfrac{1}{2}(\bar\omega-\bar z-1), \quad \bar\omega=\bar\omega(\bar z), \qquad (6.19)$$

$$\bar y=\tfrac{1}{2}(\bar\omega+\bar z+1). \qquad (6.20)$$

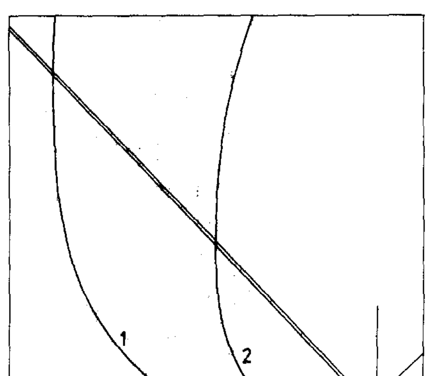


FIG. 7. The curvature phase plane portrait for the equation of state $p\ll\rho$. The double line corresponds to the z-axis ($y=-x$). The critical points lie on the z axis (de Sitter universes, e.g.,) and on the x axis (stable equilibrium points). The trajectories are given by $\omega=\alpha z^{4/3}-3z$, $z>0$; for curve 1 we have chosen $\alpha=2.321\times10^{-6}$; for 2, $\alpha=3\times10^{-6}$; for 3, $\alpha=0$; for 4, $\alpha=-0.5\times3$ coincides with the Friedmann trajectory.

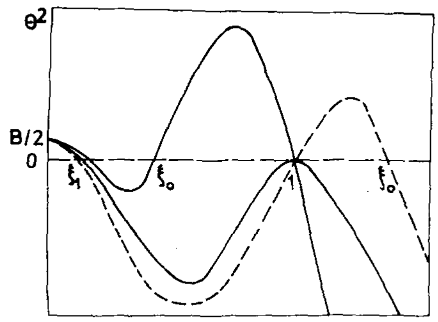


FIG. 8. The behavior of the function $\theta^2(\xi)$ decides about the existence of regularly oscillating solutions. $\theta^2(\xi)$ is determined by the initial conditions, either given by $(\bar{z}_0, \bar{\omega}_0)$ or $(\bar{z}_0, \bar{\alpha})$. For $\bar{\alpha} \leq 0$, all the solutions are of the big-bang type ($\theta^2 > 0$ for $\xi \to \infty$); for $\bar{\alpha} > 0$, regular oscillating models exist in the range $1 \leq \xi \leq \xi_0$ (initially contracting, curve 1) and for $\xi_0 \leq \xi \leq 1$ (initially expanding, curve 2). If $\theta^2(\xi)$ has a double root at $\xi = 1$, this corresponds to the equilibrium point in E_2 $(x_0, y_0 = 0)$. For $0 \leq \xi \leq \xi_1$ we get the motions directed towards the critical points on $y = -x$.

The expressions (6.16)–(6.18) show the following:

For any fixed value of B, $B \in \mathbb{R}$, a motion of (6.3) *and* (6.4) *is uniquely given by the initial point* $(x_0, y_0) \in E_2(\Omega)$; for any fixed value of B we generate in this way the phase portrait for an equation of state with $n = 0$. By changing the value of B we translate the whole phase portrait in $E_2(\Omega)$ by $x \to x + a$, $y \to y - a$, since the form of the trajectories determined by Eq. (6.12) is independent of B, while the meaning of a particular solution depends on the value of B [see Eq. (6.9)]. In this sense, even the coupled dynamical system (6.3), (6.4) is translation-invariant under $x \to x + a$, $y \to y - a$, $a \in \mathbb{R}$.

The behavior of the function $\theta^2(\xi)$ decides about the possible motions along the trajectories $\omega = \omega(z; \alpha)$ (see Fig. 8). For $\bar{\omega}_c = -\bar{z}_c - 1$ [i.e., in the points $(x_c, 0) \in E_2$] we find the equilibrium points of the Yang–Mills dynamical system; the trajectories $\bar{\omega} = \bar{\omega}(\bar{z}, \bar{\alpha})$ crosses, for $B > 0$, the x axis twice, namely for $\bar{z} = \bar{z}_1$ and $\bar{z} = \bar{z}_2$; thereby, $\bar{z}_c = \max(\bar{z}_1, \bar{z}_2)$ and $\bar{z}'_c = \min(\bar{z}_1, \bar{z}_2)$. The discussion of $\theta^2(\xi)$ implies the following classification of possible motions ($B > 0$) in terms of given initial conditions:

$\bar{\alpha} > 0$:

(i) If $\bar{z}_0 = \bar{z}_c$, then $\bar{x}_0 = \bar{x}_c$, $\bar{y}_0 = 0$: stable equilibrium point;

(ii) if $\bar{z}_c < \bar{z}_0 < \bar{z}_{sc}$, then $\bar{y}_0 > 0$ and $\xi_0 \leq \xi \leq 1$ $(\theta^2(\xi_0) = 0)$: initially expanding periodic model;

(iii) if $\bar{z}'_c < \bar{z}_0 < \bar{z}_c$, then $\bar{y}_0 < 0$ and $1 \leq \xi \leq \xi_0$ $(\theta^2(\xi_0) = 0)$: initially contracting periodic model;

(iv) if $\bar{z}_0 = \bar{z}'_c$, then $\bar{y}_0 = 0$: unstable equilibrium point;

(v) $$\text{if} \begin{Bmatrix} 0 < \bar{z}_0 < \bar{z}'_c \\ \bar{z}_0 > \bar{z}_{sc} \end{Bmatrix}, \text{ then } \bar{y}_0 > 0, \ \xi \leq \xi_1, \ \theta^2(\xi_1) = 0:$$

model running into the critical point $y = -x$.

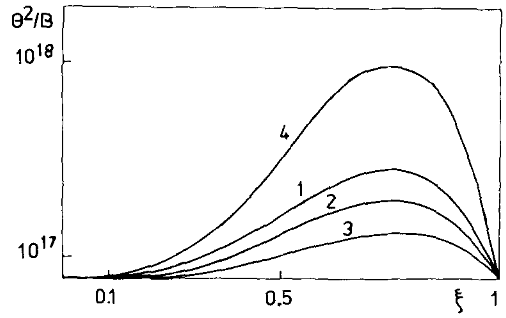


FIG. 9. Plots of $\theta^2(\xi)$ for initially expanding models ($\alpha > 0$, the expansion sets in for $\xi = 1$). The maximum of the expansion depends on the chosen initial conditions. 1: $\bar{\alpha} = 3\times10^{-6}$, $\bar{z}_0 = 2.370\times10^{18}$; 2: $\bar{\alpha} = 3\times10^{-6}$, $\bar{z}_0 = 2.0\times10^{18}$; 3: $\bar{\alpha} = 3\times10^{-6}$, $\bar{z}_0 = 1\times10^{18}$; 4: $\bar{\alpha} = 2\times10^{-6}$, $\bar{z}_0 = 6\times10^{18}$.

For all trajectories ($\bar{\alpha} > 0$, $B > 0$) there exists a $\bar{z} = \bar{z}_{sc}$, $\bar{\omega}(\bar{z}_{sc}) > 0$, such that for $\bar{z}_0 > \bar{z}_{sc}$ the motion is no longer periodic, but forever expanding into the critical point on $y = -x$. For the initial conditions specified under (ii) and (iii) we get periodic motions, i.e., the radius of closed models ($k = +1$) oscillates between a maximal, S_{max}, and a minimal value, S_{min}. The relative amplitude of oscillation, S_{max}/S_{min}, depends on the initial value $(\bar{z}_0, \bar{\omega}_0)$ as well as on the value of $\bar{\alpha}$: the smaller $\bar{\alpha}$, the greater the relative amplitude), for $\bar{z}_0 - \bar{z}'_c = \epsilon$, $\epsilon > 0$, the amplitude reaches its maximum. In the case $B \leq 0$ and $\bar{\alpha} > 0$, all the motions are *periodic* motions, since there exist no initial conditions for motions running into the critical points on $y = -x$. While for $B \geq 0$ the underlying space–times of the periodic modes of evolution are always of positive curvature ($k = +1$), the curvature type for $B < 0$ depends now obviously on the initial data ($k = +1$, if $x_0 < 0$; $k = 0$, if $x_0 = 0$, and $k = -1$, if $x_0 > 0$). The motions for $\alpha = 0$ ($A = 0$) are identical with the motions given by Einstein's solutions; finally, we are not interested here in the trajectories for $\alpha < 0$ ($A < 0$), since they describe the time evolution of singular big-bang models analogous to the Einstein models with radiation.

Figure 9 shows how the maximum of θ^2 depends on the initial conditions (here we have chosen an initially expanding model); the expansion always reaches its maximum soon after the "initial explosion." In Fig. 10 we compare this behavior with that of initially contracting

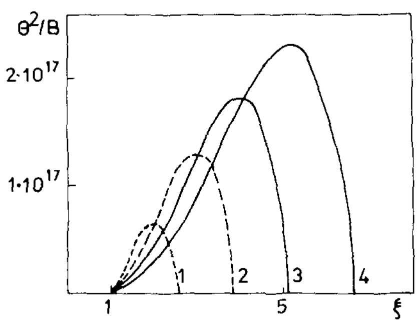


FIG. 10. Plots of $\theta^2(\xi)$ for initially contracting models of small relative amplitudes ($\alpha > 0$), $\bar{\alpha} = 3\times10^{-6}$; 1: $\bar{z}_0 = 5\times10^{16}$; 2: $\bar{z}_0 = 2\times10^{16}$; 3: $\bar{z}_0 = 1\times10^{16}$; 4: $\bar{z}_0 = 5\times10^{15}$.

models; for $\bar{z}_c - \bar{z}_0 = \epsilon$, $\epsilon > 0$ and a suitable α, we can generate periodic motions of arbitrarily high relative amplitudes. This indicates at the same time that the equilibrium points $\bar{z}_0 = \bar{z}_c$ at $\bar{x}_0 = -(1+\bar{z}_c)$ are stable, since for $|\bar{z}_0 - \bar{z}_c| < \epsilon$, $\epsilon > 0$, the corresponding motions are periodic around $(x_c, 0)$. The relative amplitude, $S_{\max}/S_{\min}$, is given by the third zero of θ^2, ξ_0, and $\theta^2(\xi_0) = 0$. The value of $\bar{\alpha}$ together with the initial value $\bar{z} = \bar{z}_0$ determine uniquely the coefficients α_2, α_3, α_4 of $\theta^2(\xi)$; from this, ξ_0 is determined numerically.

The existence of regular periodic motions in the Yang–Mills framework marks the difference to Einstein's dynamics, where initial conditions for recontracting models always give rise to singular motions. Let us relate the geometric initial condition $\bar{z} = \bar{z}_0$ to physical conditions in the case of *initially contracting* motions. From Eq. (6.17) we obtain together with $b = \frac{1}{2}\kappa\rho_0 S_0^3$ an expression for B

$$B = \kappa\rho_0(6\bar{z}_0)^{-1}, \tag{6.21}$$

and from (6.16) for $k = +1$ an expression for S_0

$$S_0^{-2} = \kappa\rho_0(1 + \bar{z}_0 - \bar{\omega}_0)(12\bar{z}_0)^{-1}, \tag{6.22}$$

or

$$\begin{aligned} \bar{\omega}_0 &= \bar{\omega}_0(\bar{z}_0, \kappa\rho_0 S_0^2), \\ \bar{\omega}_0 &= 1 + (1 - (\tfrac{1}{12}\kappa\rho_0 S_0^2)^{-1})\,\bar{z}_0. \end{aligned} \tag{6.23}$$

On the other hand, $\bar{\omega}_0$ is determined by α and $\bar{z}_0$

$$\bar{\omega}_0 = \alpha\bar{z}_0^{4/3} - 3\bar{z}_0; \tag{6.24}$$

Equations (6.23) and (6.24) determine therefore $\bar{z}_0 = \bar{z}_0(\alpha, \kappa\rho_0 S_0^2)$

$$\bar{z}_0 = \frac{1 - \alpha\bar{z}_0^{4/3}}{(\frac{1}{12}\kappa\rho_0 S_0^2)^{-1} - 4}. \tag{6.25}$$

Since we are interested in the motions of "deeply falling" models, i.e., $\xi_0 > 10^3$, $\alpha\bar{z}_0^{4/3} \ll 1$; this term will be neglected in (6.25). In this case, the initial matter density ρ_0 and the initial radius S_0 have to satisfy the condition

$$\tfrac{1}{12}\kappa\rho_0 S_0^2 < \tfrac{1}{4}. \tag{6.26}$$

On the other hand, $\bar{y}_0 < 0$, and therefore $\bar{z}_0 > \bar{z}_c'$. Since $\alpha \ll 1$, the motion follows initially the Friedmann trajectory $\bar{\omega} = -3\bar{z}$, or $\bar{y} = \frac{1}{2}\bar{x} + \frac{3}{4}$; $\bar{y}_0 < 0$ is therefore equivalent to $\bar{x}_0 < -\frac{3}{2}$, or $\bar{z}_0 > \frac{1}{2}$. In summarizing, we have found the following:

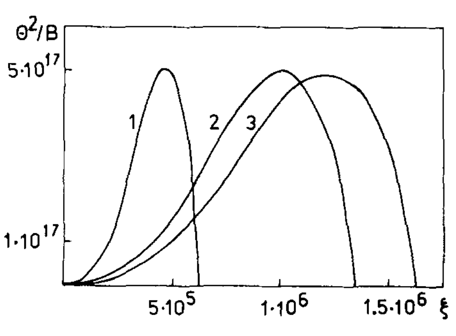


FIG. 11. Plots of $\theta^2(\xi)$ for initially contracting models of high relative amplitude, $\bar{\alpha} = 3\times10^{-6}$, generated by initial values $\bar{z}_0 \simeq 1$. For a given $\bar{\alpha} \ll 1$, the relative amplitude $S_{\max}/S_{\min} = \xi_0$ depends on $\bar{z}_0$ (i.e., on $\kappa\rho_0 S_{\max}^2/2$), while the maximum of the expansion remains constant for $\bar{z}_0 \simeq 1$. 1: $\bar{z}_0 = 10$ ($\xi_0 = 6.18\times10^5$); 2: $\bar{z}_0 = 1$ ($\xi_0 = 1.33\times10^6$); 3: $\bar{z}_0 = 0.55$ ($\xi_0 = 1.62\times10^6$).

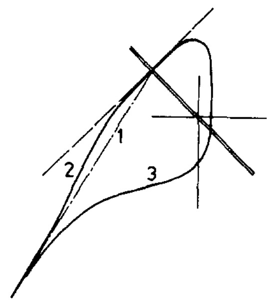


FIG. 12. The phase portrait for the superrelativistic equation of state $p = n\rho$, $n > \frac{1}{3}$. The double line is the z axis, the dotted line the ω axis. 1: Einstein trajectory $y = \frac{3}{2}x + \frac{5}{4}B$, defined for $z < 0$; 2: $\alpha > 0$; 3: $\alpha < 0$. In a realistic case we should have, however, $n \to 0$ for $y \to -x$. The equation of state separates the (z, ω)-plane into two half-planes, $z > 0$ for $n < \frac{1}{3}$, $z = 0$ for $n = \frac{1}{3}$, and $z < 0$ for $n > \frac{1}{3}$.

The matter density ρ_0 and the radius S_0 at the maximum state of regular *periodic motions* with high relative amplitude ($S_{\max}/S_{\min} \gtrsim 10^3$) satisfy the constraint equation

$$1 < \tfrac{1}{2}\kappa\rho_0 S_0^2 < \tfrac{3}{2}. \tag{6.27}$$

The behavior of these periodic motions is illustrated in Fig. 11 for various values of $\frac{1}{2}\kappa\rho_0 S_0^2$ and a fixed α; a rescaling of α only changes the relative amplitude $S_{\max}/S_{\min}$ and the maximum of the expansion.

7. THE PHASE PORTRAIT FOR THE EQUATION OF STATE $p = n\rho$, $0 \leqslant n \leqslant 1$, and $\dot{s} = 0$

For the equation of state $p = n\rho$ the Yang–Mills system has the form

$$\frac{dx}{dS} = -\frac{2}{S}(x + y), \tag{7.1}$$

$$\frac{dy}{dS} = -\frac{2}{S}(x + y) - bS^{-\delta-1}, \tag{7.2}$$

$$b = \tfrac{1}{2}\kappa(n+1)(1-3n)\rho_0 S_0^\delta, \quad \delta = 3(1+n). \tag{7.3}$$

The corresponding trajectories are then given in parametrized form

$$z(S) = b\delta^{-1}S^{-\delta}, \quad z = y - x - B, \; B \in \mathbf{R}, \tag{7.4}$$

$$\omega(S) = AS^{-4} - b(4-\delta)^{-1}S^{-\delta}, \quad A \in \mathbf{R}, \quad \delta \neq 4, \tag{7.5}$$

or in closed form

$$\omega = \omega(z) = A\delta^{4/\delta}(z/b)^{4/\delta} - \delta(4-\delta)^{-1}z. \tag{7.6}$$

Thereby the domain of definition depends on n:

$$\text{(i)} \;\; z > 0, \;\; b > 0 \;\; \text{if} \;\; 0 \leqslant n < \tfrac{1}{3}; \tag{7.7}$$

$$\text{(ii)} \;\; z < 0, \;\; b < 0 \;\; \text{if} \;\; \tfrac{1}{3} < n \leqslant 1. \tag{7.8}$$

For $A = 0$, we again obtain the Einstein trajectory

$$y = \tfrac{1}{2}(\delta - 2)x + \tfrac{1}{4}\delta B. \tag{7.9}$$

For $0 < n < \frac{1}{3}$, the phase portrait given by (7.6) is structurally equivalent with the phase portrait for $n = 0$, with the exception that for $z \to 0$ ($S \to \infty$) the generalized Einstein trajectory (7.9) acts as attractor. In this sense, the phase portrait is structurally stable in the domain $0 \leqslant n < \frac{1}{3}$.

Under the transition from a nonrelativistic to the superrelativistic equation of state ($n > \frac{1}{3}$), the phase portrait changes its features in the neighborhood of the end points at $y = -x$ (see Fig. 12), while for high values of z the trajectories are now attracted by the Einstein trajectories (7.9). The application of the equation of state $p = n\rho$, $n > \frac{1}{3}$, in the asymptotic domain $S \to \infty$ is, however, physically not quite reasonable; for this domain we would expect $p \ll \rho$, which means that the Friedman picture is correct in the asymptotic limit $S \to \infty$ for all physically reasonable equations of state.

The detailed structure of the phase portrait may vary with the chosen equation of state; as an example, we consider $p = p(\rho) = a\rho^m$, $m > 0$ ($\rho < 10^{16}$ gcm^{-3}). In this case the energy density is found to be

$$\rho(\xi) = \rho_1 \xi^3 (1 - a\rho_1^{m-1} \xi^{3(m-1)})^{-1/(m-1)}, \quad \xi = S_0/S. \qquad (7.10)$$

The phase portrait is essentially dictated by the structure function f, defined in (4.3) for $\dot{s} = 0$,

$$f(\xi) = \tfrac{1}{4}\kappa\xi^3\rho_1[1 - (3m+1)a\rho_1^{m-1}\xi^{3(m-1)}] \times [1 - a\rho_1^{m-1}\xi^{3(m-1)}]^{-(2m-1)/(m-1)}. \qquad (7.11)$$

In the asymptotic limit $S \to \infty$, it has the same form as (6.1). Without a cutoff at $\rho \sim 10^{16}$ gcm^{-3}, f becomes singular at ξ_∞:

$$\xi_\infty = (a\rho_1^{m-1})^{-1/3(m-1)} = (3m+1)^{1/3(m-1)}\xi_2, \qquad (7.12)$$

with the introduction of $f(\xi_2) = 0$. Therefore, we have for $m \neq 1$ always a $S = S_\infty$ with $0 < S_\infty < S_2$; for $m = \frac{4}{3}$, $S_\infty = S_2/5$, and for $m = \frac{5}{3}$, $S_\infty = S_2/6^{1/2}$. Already in Einstein's theory, this equation of state, $p = a\rho^m$, would generate for the early epoch in the time evolution a matter, or curvature, singularity at a finite radius $S = S_\infty > 0$, the metric being completely regular under this equation of state in contrast to the usual Friedmann singularity. The behavior of the curvature follows immediately from (5.11) and (5.12):

$$\tilde{x}_E(\xi) = -\tfrac{1}{3}\kappa\rho(\xi) < 0, \qquad (7.13)$$

$$\tilde{y}_E(\xi) = \tfrac{1}{2}\tilde{x}_E(\xi) - \tfrac{1}{2}\kappa a\rho_1^m \xi^{3m}(1 - a\rho_1^{m-1}\xi^{3(m-1)})^{-m/(m-1)}. \qquad (7.14)$$

We calculate the Yang–Mills trajectories given by the structure functions (7.11) only in the first approximation

$$f(\xi) = \tfrac{1}{4}\kappa\xi^3\{1 - [m/(m-1)](3m-4)a\rho_1^{m-1}\xi^{3(m-1)}\}\rho_1. \qquad (7.15)$$

For not too high densities this is a good approximation to (7.11). For $m = 4/3$, this first correction vanishes, while for $m > 4/3$ it always gives a negative contribution to the lowest order (6.1). The phase portrait generated by (7.15) with $m = \frac{5}{3}$ is shown in Fig. 13; the trajectories are parametrized by

$$z(\xi) = \tfrac{2}{3}\alpha_3\xi^3 - \tfrac{2}{5}\alpha_5\xi^5, \quad \alpha_5 > 0, \qquad (7.16)$$

$$\omega(\xi) = -2\alpha_3\xi^3 + \alpha_4\xi^4 - 2\alpha_5\xi^5; \qquad (7.17)$$

they are considered as a superposition of (6.5), (6.6) and (7.4), (7.5); the coefficient α_5 is determined by $\kappa a\rho_1^{5/3}$. If $\alpha_4 > 0$, regularly oscillating motions exist, their

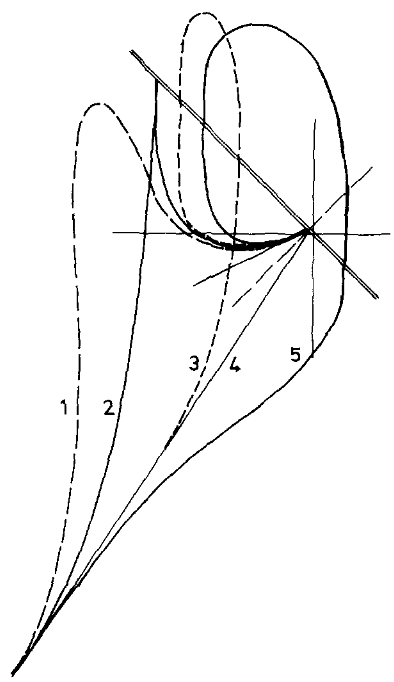


FIG. 13. An example of a phase portrait for an equation of state which is stiffer at high densities. $f = a_3S^{-3} - a_5S^{-5}$, $a_3 > 0$, $a_5 > 0$, i.e., $n \to 0$ for $S \to \infty$ and $n \to \frac{2}{3}$ for $S \to 0$. The dotted straight lines are the z and ω axes. The trajectories are given in parametrized form by $\bar{z}(\xi) = \frac{2}{3}\alpha_3\xi^3 - \frac{2}{5}\alpha_5\xi^5$, $\bar{\omega}(\xi) = -2\alpha_3\xi^3 + \alpha_4\xi^4 - 2\alpha_5\xi^5$, $\alpha_4 > 0$. The behavior of the trajectories depends critically on the contribution from the equation of state at the various epochs; this can be characterized by means of the parameter $D = \alpha_4^2 - 16\alpha_3\alpha_5$. 1: $D < 0$ ($\bar{\omega} < 0$); 2: $D = 0$ ($\bar{\omega} \leq 0$); 3: $D > 0$; 4: Einstein trajectory $y = \frac{3}{2}x + \frac{5}{4}B$ ($B > 0$, $\alpha_3 = 0 = \alpha_4$),; 5: $D > 0$, α_5 sufficiently small. The motions along 1–3 are of the big-bang type, while a motion along the upper segment of 5 describes the time evolution of a regularly oscillating closed space–time model ($k = +1$). For all the trajectories we have $\alpha_5 \ll \alpha_4 \ll \alpha_3 \simeq 1$, $\alpha_4 = 3\times10^{-6}$, and $\alpha_5 \simeq 10^{-13}$, e.g.

periods and relative amplitudes being determined by the coefficients α_4 and α_5.

This investigation of the phase portrait over the range $0 \leq n \leq 1$ for the equation of state $p(\rho) = n\rho$ shows that reliable equations of state cannot guarantee the existence of regular motions within the Yang–Mills dynamics characterized by a vanishing exterior form α defined in (3.9) and (3.10). This additional freedom may be used to generate a structure function $f(S)$ having a positive influence on the evolution along regular trajectories; with the interpretation of α and the existence of "regular big-bang cosmologies" we shall be concerned in a forthcoming paper. That under well posed conditions regular trajectories exist in the Yang–Mills dynamics will be shown in the next section.

8. ON THE EXISTENCE OF REGULAR TRAJECTORIES IN THE YANG-MILLS DYNAMICS

For any realistic equation of state the structure function $f(S)$ has the asymptotic limit $f \sim \frac{1}{4}\kappa\rho_0\xi^3$. This means:

Lemma 1: In a neighborhood of the critical points on $y = -x$ the phase portrait of the dynamical system

(4.1)–(4.3) is structurally stable under the transition from Einstein's dynamics to the Yang–Mills dynamics, i.e., the Friedmann trajectory always acts as an attractor in the asymptotic limit.

For the limit $S \to 0$, the phase portrait depends on the equation of state in the case $\alpha \equiv 0$:

(i) for dust matter, this limit depends on the sign of A defined in (6.5);

(ii) for $p = n\rho$, $n < \frac{1}{3}$, this limit is qualitatively the same as in (i);

(iii) for $p = n\rho$, $n > \frac{1}{3}$, the Yang–Mills trajectories are attracted by the corresponding Einstein trajectories; they all are confined to the Einstein cone in the limit $S \to 0$.

In the following we want to give a type of a structure function $f(S)$ for generation of regular trajectories in $E_2(\Omega)$:

Lemma 2: If the structure function $f(S)$ satisfies $f \simeq a_m S^{-m}$ for $S \to 0$, i.e., for $\rho \to \infty$, with $m > 4$, then

(i) no singular big-bang trajectory exists for $a_m > 0$; any trajectory is attracted by $\omega(z) = [m/(m-4)]z$, $z > 0$;

(ii) any trajectory is of the singular big-bang type for $a_m < 0$; they are attracted by $\omega(z) = [m/(m-4)]z$, $z < 0$.

Thereby the coordinates z and ω are defined in Sec. 6. This statement gives a criterion for the existence of singular and regular trajectories in the Yang–Mills dynamics. It is beyond of the aim of this paper to generate this type of structure function by introducing a nonvanishing 1-form α; however, note that $a_m < 0$ for $\alpha = 0$ and $p = n\rho$, $n > \frac{1}{3}$.

The attractors are calculated by solving the dynamical system

$$\frac{dx}{dS} = -\frac{2}{S}(x+y). \tag{8.1}$$

$$\frac{dy}{dS} = -\frac{2}{S}(x + y + a_3 S^{-3} + a_m S^{-m}), \quad m > 4. \tag{8.2}$$

In parametrized form they are given by

$$z(S) = \tfrac{2}{3} a_3 S^{-3} + (2/m) a_m S^{-m}, \tag{8.3}$$

$$\omega(S) = -2a_3 S^{-3} + AS^{-4} + [2/(m-4)] a_m S^{-m}, \tag{8.4}$$

with the following domain of definitions:

(i) $z > 0$ and $\omega > 0$, if $a_m > 0$,

(ii) $z < 0$ and $\omega < 0$, if $a_m < 0$.

The corresponding expansion function,

$$2\theta^2(S) = B - 2kS^{-2} + \tfrac{8}{3} a_3 S^{-3} - AS^{-4} - [8/m(m-4)] a_m S^{-m}, \tag{8.5}$$

proves to give regular motions for $a_m > 0$ (no matter or curvature singularity can occur) while for $a_m < 0$ (8.5) characterizes singular big-bang trajectories. The influence of the term AS^{-4} is no longer important for $m > 4$.

9. THE PRESENT STATE OF THE UNIVERSE AND THE RELATIONSHIP BETWEEN THE OBSERVATIONAL PARAMETERS

Any point in the phase plane $E_2(\Omega)$ determines for a fixed value of $S = S_1$ a dynamical state of a world model at some particular instant of time, $t = t_1$ with $S(t_1) = S_1$. The Hubble constant H_0, which is equal to the present value of the expansion θ_0, the deceleration parameter q_t, $q_t = -y_t/\theta_t^2$, and the matter parameter σ_t, $\sigma_t = \frac{1}{6}\kappa\rho_t/\theta_t^2$ are the observational parameters of every dynamical system based on the Robertson–Walker space–time geometry. In general relativity, two relations are given between the state (x_t, y_t) and the energy content of the universe; see, e.g., Eqs. (5.8) and (5.9); the integration of the Yang–Mills dynamics also provides two corresponding relations, in general given by (for the late epoch)

$$y_t - x_t = B + \tfrac{1}{3} b S_t^{-3} + \tilde{F}(S_t), \quad B \in \mathbf{R}, \tag{9.1}$$

$$y_t + x_t = AS_t^{-4} - bS_t^{-3} - \tilde{G}(S_t), \quad A \in \mathbf{R}. \tag{9.2}$$

Here, $b = \frac{1}{2}\kappa\rho_0 S_0^3$ describes the contribution from the structure function f generated by the asymptotic form $(S \to \infty)$ of the equation of state, $p \to 0$ for $\rho \to 0$; the two functions $\tilde{F}(S)$ and $\tilde{G}(S)$ are responsible for the contributions from the form of the equation of state at high densities (in the lepton era). Their influence on (9.1) and (9.2), however, vanishes sufficiently rapidly in the asymptotic limit.

The A term, as well as the $\tilde{F}$- and $\tilde{G}$-terms determine how much the motion of the present universe deviates from the Friedmann trajectory

$$y_t - \tfrac{1}{2}x_t - \tfrac{3}{4}B = \tfrac{1}{4}(AS_t^{-4} + 3\tilde{F}_t - \tilde{G}_t). \tag{9.3}$$

By solving (9.1) and (9.2) for x and y

$$x_t = \tfrac{1}{2}(AS_t^{-4} - \tfrac{4}{3}bS_t^{-3} - \tilde{F}(S_t) - \tilde{G}(S_t) - B), \tag{9.4}$$

$$y_t = \tfrac{1}{2}(AS_t^{-4} - \tfrac{2}{3}bS_t^{-3} + \tilde{F}(S_t) - \tilde{G}(S_t) + B), \tag{9.5}$$

we obtain the two observational relations

$$k/S_t^2\theta_t^2 = 2\sigma_t - 1 - A/(2\theta_t^2 S_t^4) + B/2\theta_t^2 + \tilde{F}_t/2\theta_t^2 + \tilde{G}_t/2\theta_t^2, \tag{9.6}$$

$$q_t = \sigma_t - A/(2\theta_t^2 S_t^4) - B/2\theta_t^2 - \tilde{F}_t/2\theta_t^2 + \tilde{G}_t/2\theta_t^2. \tag{9.7}$$

In the case $A = 0 = \tilde{F} = \tilde{G}$, they coincide with the general relativistic relations (if we identify $B = \frac{2}{3}\Lambda$); for $t = t_0$, the residual contributions from the early universe are negligible, i.e., $|\tilde{F}_0|/H_0^2 \ll 1$ and $|\tilde{G}_0|/H_0^2 \ll 1$, while the A term, which defines the initial state of the model, might have a nonnegligible influence on q_0 and $k/S_0^2 H_0^2$. If, furthermore, the present state of the universe is long away from the turning point $\theta(S_{\max}) = 0$, $|B|/H_0^2 \ll 1$. The present data $q_0 \simeq 1$ and $\sigma_0 \gtrsim 0.02$ would require a negative A term, $A < 0$, in order to account for the relation (9.7) with

$$-A/2H_0^2 S_0^4 = q_0 - \sigma_0, \quad \text{if } |B|/H_0^2 \ll 1. \tag{9.8}$$

Despite the small value of σ_0, the geometry might be closed, since from (9.6) we find in general, by eliminating the A term,

$$k/S_0^2 H_0^2 = q_0 + \sigma_0 - 1 + B/H_0^2; \qquad (9.9)$$

the geometry remains closed for

$$q_0 > 1 - \sigma_0 - B/H_0^2. \qquad (9.10)$$

In the same way we obtain an upper limit for q_0 determined by $A/S_0^4 H_0^2$; as a consequence, the geometry is closed if and only if

$$1 - \sigma_0 - B/H_0^2 < q_0 < 3\sigma_0 - 1 - A/S_0^4 H_0^2. \qquad (9.11)$$

In Einstein's theory, A is determined by the present radiation-energy content of the universe, $A/S_0^4 = -\frac{2}{3}\kappa\rho_\gamma^0 \simeq -10^{-61}$ cm^{-2}, and B by the cosmological constant Λ, $B = \frac{2}{3}\Lambda$. In the "Yang–Mills" cosmology, B is an expression for the total energy integral and A follows from the physics of the early universe whenever the energy transfer between the massive constituents of the cosmic fluid and the photon gas is no longer negligible. The functional dependence of the late epoch is therefore exactly the same as in Einstein's theory, while the physical interpretation of the corresponding terms is essentially different.

ACKNOWLEDGMENTS

The author is grateful to W. Kundt and A. King for critical discussions and comments on this subject.

[1]M. Camenzind, "Weak and Strong Sources of Gravity; I. An SO(1,3)-Gauge Theory of Gravity," Hamburg 1976 (to be published).

[2]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[3]M. Camenzind; J. Math. Phys. **16**, 1023 (1975).

[4]Latin indices denote the tetrad components of geometric quantities, Greek indices refer to the coordinate components.

[5]We use a two-component description of the cosmic matter; T_{ab} only contains the massive constituents of the cosmic fluid (protons, electrons, myons, ..., and their antiparticles). The photon gas and the neutrino background will be described in the transfer approach.

[6]N. P. Bhatia and G. P. Szegö, *Stability Theory of Dynamical Systems*, Die Grundlehren der mathematischen Wissenschaften **161** (Springer-Verlag, Berlin, 1970).

Integral representations of particular integrals of a class of inhomogeneous linear ordinary differential equations[a)]

George L. Johnston[b)]

Lawrence Berkeley Laboratory, Berkeley, California 94720
and Plasma Fusion Center and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 5 August 1977)

A method of obtaining integral representations of particular integrals of a class of inhomogeneous second-order linear ordinary differential equations is presented. The integrands of the representations are $\exp[F(z;t)]$, where $F(z;t) = z^2a(t) + zb(t) + c(t)$, z is the independent variable of the differential equation, and a, b, and c are initially unspecified functions of the variable of integration, t. The lower limit of the contours of integration is zero. The upper limits of integration and the contours along which the integral is taken are initially unspecified. In the general class of inhomogeneous differential equations considered, the coefficients of the dependent variable and its derivatives are polynomials in z with complex constants and the homogeneous term is $\exp(k_2z^2 + k_1z + k_0)$, where the k_n are complex constants. One imposes the conditions that the application of the homogeneous operator to the assumed form of integral representation give the integral of $-\partial F/\partial t$, that $\exp[F(z;0)]$ be equal to the inhomogeneous term of the differential equation, and that the limit of $\exp F$ as t approaches the upper limit along the contour of integration be zero. By equating coefficients of different powers of z separately to zero, one obtains a set of coupled equations for a, b, and c. The basic class of inhomogeneous differential equations to which the method is applicable is determined by requiring that a, b, and c be algebraic or elementary transcendental functions of t. The class of equations to which the method is applicable is extended to include inhomogeneous terms of the form $z^n\exp(k_2z^2 + k_1z + k_0)$, where n is a positive integer, by repeated differentiation of integral representations of members of the basic class with respect to k_1, treated as a parameter. More general inhomogeneous terms may be treated by superposition and, in appropriate cases, by approximation in terms of sets of orthogonal functions.

1. INTRODUCTION

The theory of integral representations of solutions of homogeneous linear ordinary differential equations is well established. Such integral representations are a significant element of the theory of ordinary differential equations and are extremely important in mathematical physics.[1] In the case of inhomogeneous equations, integral representations do not appear to have received substantial attention. This is somewhat surprising, considering the importance of inhomogeneous equations in mathematical physics. Integral representations of particular integrals of inhomogeneous equations whose integrands contain only algebraic or elementary transcendental functions of the variable of integration may have substantial advantages over solutions obtained by the usual methods of variation of parameters or Green's functions. The latter solutions contain products of solutions of the corresponding homogeneous equation and integrals whose integrands contain solutions of the homogeneous equation as a factor. If the solutions of the homogeneous equation are higher transcendental functions, the relative advantage of integral representations of the sort described may be considerable. Furthermore, integral representations may be particularly convenient for approximate analytic or numerical evaluation.

[a)]This work was supported in part by the U.S.E.R.D.A. under the auspices of the Division of Magnetic Fusion Energy under contract W-7405-ENG-48, and by the N.S.F. under contract ENG 75-06242.

[b)]Present address: California Energy Resources Conservation and Development Commission, Sacramento, California 95825.

In this paper a method of obtaining integral representations of particular integrals of a class of inhomogeneous second-order linear ordinary differential equations is presented. In Sec. 2 the motivation for the method is presented. It is based upon a consideration of two particular inhomogeneous equations. The inhomogeneous term of both equations is a constant. The corresponding homogeneous equation is, in one case, Airy's equation and, in the other case, the parabolic cylinder equation. These equations appear in the warm fluid theory of linear mode conversion of a time-harmonic uniform external electrostatic field in the direction of linear[2] and parabolic plasma density profiles, respectively.

In the case of the inhomogeneous Airy's equation, there exists an integral representation of a particular integral.[3] In the case of the inhomogeneous parabolic cylinder equation, an integral representation apparently has not been published.[4] An integral representation is easily obtained by transformations of the dependent and independent variables and the method of Fourier transforms. The derivation of this result is presented in the Appendix.

An examination of the manner in which the two integral representations satisfy the corresponding equations suggests a method for obtaining integral representations of a broader class of differential equations. The elements of the method are presented in Sec. 3. The basic class of equations to which the method may be applied consists of two subclasses, depending upon the set of restrictions imposed upon the form of solution assumed. Subclass 1, which is treated in Sec. 4, consists of equations of the form

 0022-2488/78/1903-635$1.00

$$(f_1 z + f_0)y'' + (g_1 z + g_0)y' + (h_1 z + h_0)y(z) = \exp(k_1 z + k_0) \quad (1)$$

and related equations obtained by setting various of the constant coefficients f_n, g_n, h_n, and k_n equal to zero. Subclass 2, which is treated in Sec. 5, consists of equations of the form

$$f_0 y'' + g_0 y' + (h_2 z^2 + h_1 z + h_0)y(z) = \exp(k_2 z^2 + k_1 z + k_0) \quad (2)$$

and related equations obtained by setting various of the constant coefficients equal to zero.

The class of equations to which the method can be applied can be extended by differentiating integral representations of the basic class with respect to k_1, treated as a parameter, and by superposition. Under some conditions it is possible to treat, in an approximate manner, equations with more general inhomogeneous terms. In the case of Subclass 1, it may be possible to approximate an inhomogeneous term on the range $0 \leq z < \infty$ by a reasonable number of terms of an expansion in weighted Laguerre polynomials. In the case of Subclass 2, it may be possible to approximate an inhomogeneous term on the range $-\infty < z < \infty$ by a reasonable number of terms of an expansion in weighted Hermite polynomials.

In Sec. 6, various aspects of the method are discussed. In particular, the relation of the method to the theory of integral representations of solutions of homogeneous equations is examined. The extension of the latter theory to inhomogeneous equations apparently has never been undertaken. It yields integral representations of a narrower class of inhomogeneous equations than the method presented here—it cannot be applied to some equations of Subclass 2.

2. TWO EQUATIONS OF INTEREST

The two inhomogeneous equations which motivate the method presented in this paper are

$$v'' - zv(z) = \pi^{-1} \quad (3)$$

and

$$y'' + (z^2 - \alpha)y(z) = 1. \quad (4)$$

Equation (3) appears in the warm fluid theory of linear mode conversion of a time-harmonic uniform external electrostatic field in the direction of a linear plasma density profile.[2] The dependent variable is the complex amplitude of the self-consistent electric field in the plasma. If the plasma density profile is parabolic instead of linear, the corresponding equation is (4).

A particular integral of (3) is given by the integral representation[3]

$$Hi(z) = \frac{1}{\pi}\int_0^\infty \exp(zt - t^3/3)dt. \quad (5)$$

A set of linearly independent solutions of the corresponding homogeneous equation, denoted by $Ai(z)$ and $Bi(z)$, are expressible as integral representations in which, except for a constant multiplier, the integrand is the same as that of (5) but the contours of integration are different from that of (5). The integral representation $Hi(z)$ can be obtained by applying the method of Fourier transforms. The integral representations $Ai(z)$ and $Bi(z)$ can be obtained by applying the theory of integral representations of solutions of homogeneous linear ordinary differential equations, using the Laplace kernel $K(z,t) = \exp(zt)$.[1]

Integral representations of solutions of the homogeneous equation corresponding to (4) have been obtained in various forms by the theory of integral representations. For example, the asymptotic behavior $y(z) \sim \exp(\frac{1}{2}iz^2)$ can be removed by assuming a solution of the form

$$y(z) = \exp(iz^2/2)\Phi(w), \quad (6)$$

where $w = -iz^2$. The transformed equation for $\Phi(w)$ is the canonical form of the confluent hypergeometric equation with parameters $a = \frac{1}{4}(1 + i\alpha)$ and $c = \frac{1}{2}$. Integral representations of solutions of the confluent hypergeometric equation can be obtained by using the Laplace kernel $K(w,t) = \exp(wt)$. Integral representations of particular integrals of (4) apparently have not been published.[4] Comparison of the Eq. (3) and (4) and the integral representations of the corresponding homogeneous equations suggests that it should be possible to obtain integral representations of particular integrals of (4).

Integral representations can in fact be obtained by straightforward application of the method of Fourier transforms to the differential equation for $\Psi(v) = \Phi(w)$, where $v = iw = z^2$. Introducing a convenient transformation of the variable of integration and subtracting a solution of the homogeneous equation, one obtains the integral representations

$$y(z) = \int_{C_\pm} \exp\{-\tfrac{1}{2}z^2 \tan\sigma + \tfrac{1}{2}\alpha\sigma + \log[\tfrac{1}{2}(\cos\sigma)^{-1/2}]\}d\sigma, \quad (7)$$

where the contours of integration C_+ and C_- apply to the cases $\mathrm{Im}(\alpha) > -1$ and $\mathrm{Im}(\alpha) < 1$, respectively. Note that both contours of integration are available in the range $-1 < \mathrm{Im}(\alpha) < 1$. The contours C_+ and C_- proceed from the origin to $r \pm i\infty$, respectively, where r is a finite real number. The procedure described above for obtaining the integral representation (7) is presented in the Appendix.

Examination of the manner in which the integral representations (5) and (7) satisfy the Eq. (3) and (4), respectively, suggests a method for obtaining integral representations of a broader class of differential equations.

3. ELEMENTS OF THE METHOD

Operation of the linear differential operators of (3) and (4) on the integral representations (5) and (7), respectively, produces an integral of a derivative of a function of the variable of integration. At the upper limit of integration the function vanishes. At the lower limit it reproduces the inhomogeneous term. If a particular inhomogeneous equation is represented as

$$Ly(z) = R(z), \quad (8)$$

integral representations of particular integrals are of the form

$$y(z)=\int_C \exp F(z;t)dt. \quad (9)$$

The function F satisfies the conditions

$$L(\exp F)+\frac{\partial}{\partial t}\exp F=0 \quad (10)$$

and

$$\exp F(z;0)=R(z). \quad (11)$$

The contour of integration, C, proceeds from the origin to an upper limit $t=u$ along a path such that the limit of $\exp F$ as t approaches u on the contour is zero.

Generalizing the form of F in a manner such that the forms of its dependence on z in (5) and (7) are comprehended, we have

$$F(z;t)=z^2a(t)+zb(t)+c(t), \quad (12)$$

where a, b, and c are initially unspecified functions of t. Subject to additional limitations that may appear in the further development of the method, the form of F given in (12) permits the treatment of inhomogeneous equations in which the inhomogeneous term has the form

$$R(z)=\exp(k_2z^2+k_1z+k_0), \quad (13)$$

where k_2, k_1, and k_0 are (possibly complex) constants. The consistency of (12) and (13) imposes the conditions $a(0)=k_2$, $b(0)=k_1$, and $c(0)=k_0$.

The basic class of differential equations for which we shall endeavor to obtain integral representation of the form which we have described is

$$\begin{aligned}(f_2z^2+f_1z+f_0)y''+(g_2z^2+g_1z+g_0)y'\\ +(h_2z^2+h_1z+h_0)y(z)=\exp(k_2z^2+k_1z+k_0),\end{aligned} \quad (14)$$

where the quantities f_n, g_n, and h_n, $n=0,1,2$, are (possibly complex) constants, some of which may vanish. Since the functions z^n, $n=0,1,2,\ldots$, are linearly independent, the imposition of condition (10), where L is given by (14), yields a set of coupled equations for a, b, c, and their first derivatives. The process of obtaining an integral representation of the assumed form consists of the determination of a set of algebraic or elementary transcendental functions which satisfies this set of equations subject to the conditions $a(0)=k_2$, $b(0)=k_1$, $c(0)=k_0$, and a contour of integration C which satisfies the conditions states above. We shall find that the class of Eq. (14) is broader than the class of equations for which integral representations of the assumed form can be obtained. The exploration of the limitations on the class of equations for which integral representations can be found is facilitated by initial inclusion of inadmissible terms and an examination of the considerations which necessitate their exclusion.

Proceeding in the manner which we described above, we obtain the following set of five coupled equations for a, b, c, and their first derivatives:

$$4f_2a^2=0, \quad (15)$$

$$4f_2ab+4f_1a^2+2g_2a=0, \quad (16)$$

$$f_2(b^2+2a)+4f_1ab+4f_0a^2+g_2b+2g_1a+h_2=-a', \quad (17)$$

$$f_1(b^2+2a)+4f_0ab+g_1b+2g_0a+h_1=-b', \quad (18)$$

$$f_0(b^2+2a)+g_0b+h_0=-c'. \quad (19)$$

If $f_2\neq 0$, it is possible to obtain an integral representation of the assumed form only for members of the class of Eq. (14) in which certain of the constant coefficients are so interrelated that the corresponding equations form a subclass of (14) which is negligible interest. If $f_2\neq 0$, (15) imposes the condition that $a=0$. The condition $a(0)=k_2$ cannot be satisfied unless $k_2=0$. Relation (16) is then identically satisfied. Relations (17) and (18) are a pair of quadratic equations for b. The condition $b(0)=k_1$ must also be satisfied. These three relations can be satisfied simultaneously only if $f_2/f_1=g_2/g_1=h_2/h_1$ and if $f_2k_1^2+g_2k_1+h_2=0$.

Accordingly, we shall henceforth assume that $f_2=0$. With this assumption, (15) is identically satisfied. There are then three exclusive alternative sets of assumptions that result in the satisfaction of (16): $a=-g_2/2f_1$, $g_2\neq 0$, $f_1\neq 0$; $a\equiv 0$; and $f_1=g_2=0$.

The first set of assumptions is that $a=-g_2/2f_1$, $f_1\neq 0$, and $g_2\neq 0$. It is possible to obtain an integral representation of the assumed form only for a subclass of (14) which is so restricted as to be of negligible interest. Condition $a(0)=k_2$ can be satisfied only if $k_2=-g_2/2f_1$. Relation (17) determines a constant value of b that is consistent with relations (18) and $b(0)=k_1$ only in extremely limited circumstances. Accordingly we exclude henceforth the set of assumptions $a=-g_2/2f_1$, $f_1\neq 0$, and $g_2\neq 0$.

The second set of assumptions for satisfying (16) is that $a\equiv 0$. This assumption yields Subclass 1, which is treated in the next section. As we shall see there, it is characterized by the restrictions $f_2=g_2=h_2=k_2=0$. Additional coefficients may, of course, vanish.

The third set of assumptions is that a is not identically equal to zero but that $f_1=g_2=0$. This leads to Subclass 2, which is treated in Sec. 5. It is characterized by the restrictions $f_2=f_1=g_2=g_1=0$. Additional coefficients may, of course, vanish.

The class of equations to which the method is applicable can be extended beyond the basic class of equations composed of Subclasses 1 and 2. This is accomplished in the following manner. First, suppose that $y(z)$ is an integral representation of a particular integral of an inhomogeneous equation which is a member of the basic class. The constant k_1 is replaced by the parameter λ. The nth derivative of $y(z)$ with respect to λ, evaluated at $\lambda=k_1$, is an integral representation of a particular integral of an inhomogeneous equation which differs from the original equation in that the inhomogeneous term is z^n times that of the inhomogeneous term in the original equation. Second, integral representations of particular integrals of equations in which the inhomogeneous term consists of a sum of such terms can be obtained by superposition.

Under some conditions it is possible to treat, in an approximate manner, equations with more general inhomogeneous terms. This is accomplished by expressing the inhomogeneous term as an expansion in an appropriate set of orthogonal polynomials.

4. SUBCLASS 1

If (16) is satisfied by choosing $a\equiv 0$, the restriction

$k_2=0$ is imposed. If (17) is satisfied by choosing $b=-h_2/g_2$ and imposing the restriction $g_2\neq 0$, the simultaneous additional satisfaction of (18) and the relation $b(0)=k_1$ imposes restrictions on the coefficients which deprive the resulting subclass of equations of substantial interest.

Accordingly, we choose the imposition of the restrictions $g_2=h_2=0$ as the means of satisfying (17). The two remaining equations of the set, (18) and (19), determine functions $b(t)$ and $c(t)$ which are algebraic or elementary transcendental functions of t. Contours of integration which satisfy the requirements of the method can be obtained. Thus Subclass 1 consists of the equations

$$(f_1z+f_0)y''+(g_1z+g_0)y'+(h_1z+h_0)y(z) = \exp(k_1z+k_0), \tag{20}$$

in which additional coefficients may vanish.

The determination of the complete set of possible contours of integration for integral representations is simplified considerably by choosing b instead of t as the variable of integration. We thereby obtain representations in the form

$$y(z)=-\int_{C'}\exp[bz+c(b)+j(b)]db, \tag{21}$$

where

$$c(b)=\int_{C'(b)}\frac{f_0s^2+g_0s+h_0}{f_1s^2+g_1s+h_1}ds+k_0 \tag{22}$$

and $j(b)=-\log(f_1b^2+g_1b+h_1)$. The contours of integration C' proceed from k_1 to a finite or infinite upper limit along a path such that the limit of $\exp[zb+c(b)]$ as b approaches the upper limit on the contour is zero. The contours of integration $C'(b)$ proceed from k_1 to b, a point on C'. Note that $c(b)$ is a multivalued function of b depending on the character of C'. Contours of integration C' are of two types: those for which the upper limit is a root of the equation $P_1=0$, where $P_1=f_1b^2+g_1b+h_1$, and the limit of $\exp c(b)$ as b approaches the upper limit on the contour is zero; and those for which the upper limit is at infinity and the limit of $\exp[zb+c(b)]$ as b approaches the upper limit on the contour is zero.

We consider the former case first. If the polynomial P_1 is of second degree and has equal roots or if P_1 is of first degree, the values of the coefficients in the integrand of $c(b)$ determine whether or not it is possible to satisfy the condition that the limit of $\exp c(b)$ as b approaches the upper limit on the contour is zero. If P_1 is a perfect square, it is always possible to satisfy this condition by choosing the contour of integration so that its directed tangent at the upper limit lies within a suitable range of directions. In this case it is convenient to introduce the transformation of variable of integration $p=(b-b_s)^{-1}$, where b_s is the (repeated) root of the equation $P_1=0$.

The possibility that the equation $P_0=0$, where $P_0=f_0b^2+g_0b+h_0$, has a root (or roots) in common with the equation $P_1=0$ must be investigated in specific cases.

Consider now contours of integration C' for which the upper limit is at infinity and the limit of $\exp[zb+c(b)]$ as b approaches the upper limit on the contour is zero. The values of the coefficients in the integrand of $c(b)$ determine the dominant asymptotic behavior of $c(b)$ as b approaches infinity. Since $f_0\neq 0$, it can be proportional to b^3, b^2, or b. If it is proportional to b^3 or b^2, the contour of integration can always be chosen to approach infinity in a range of directions such that the limit of $\exp[zb+c(b)]$ is zero. If the dominant asymptotic behavior of $c(b)$ is proportional to b, the contour of integration can be chosen to approach infinity in a range of directions such that the limit of $\exp[zb+c(b)]$ is zero, except for a single (possibly complex) value of z.

The extension of Subclass 1 to equations in which the inhomogeneous term is $z^n\exp(k_1z+k_0)$, where n is a positive integer, does not introduce integral representations distinct from the basic integral representation. To see this, it is sufficient to treat the case $n=1$. An integral representation for $n=1$ is given by $y_1=(\partial y_0/\partial\lambda)|_{\lambda=k_1}$, where y_0 is the representation given in (21). The result is

$$y_1=\exp[zk_1+k_0+j(k_1)]-\frac{f_0k_1^2+g_0k_1+h_0}{f_1k_1^2+g_1k+h_1}y_0. \tag{23}$$

It may be possible to treat, in an approximate manner, equations with more general inhomogeneous terms. On the range $0\leq z<\infty$, it may be possible to approximate an inhomogeneous term by a reasonable number of terms of an expansion in weighted Laguerre polynomials. We have shown that the extension of Subclass 1 to equations in which the inhomogeneous term is $z^n\exp(k_1z+k_0)$ does not introduce integral representations distinct from the basic integral representation. Thus the approximate representation of a particular integral of an equation with a more general inhomogeneous term may be relatively convenient.

5. SUBCLASS 2

The remaining alternative for satisfying (16) is that a is not identically equal to zero but that $f_1=g_2=0$. Then a can be determined in terms of elementary transcendental functions of t by integration of (17) subject to the condition $a(0)=k_2$.

If $g_1\neq 0$ and $h_2\neq 0$, it is impossible to determine a, b, and c in terms of algebraic and elementary transcendental functions. One can see this by inspection of the explicit expression for $a(t)$ obtained from (17) and the expressions for b as a functional of a and for c as a functional of a and b obtained from (18) and (19), respectively. In terms of the quantities $\gamma=4f_0a$, $\gamma_\pm=-g_1\pm(g_1^2-4f_0h_2)^{1/2}$, and $\kappa_2=4f_0k_2$, $a(t)$ is given by the relation

$$\gamma(t)=\frac{\gamma_+\exp[\frac{1}{2}(\gamma_+-\gamma_-)t-\frac{1}{2}\ln P]-\gamma_-\exp[-\frac{1}{2}(\gamma_+-\gamma_-)t+\frac{1}{2}\ln P]}{\exp[\frac{1}{2}(\gamma_+-\gamma_-)t-\frac{1}{2}\ln P]-\exp[-\frac{1}{2}(\gamma_+-\gamma_-)t+\frac{1}{2}\ln P]}, \tag{24}$$

where $P=[(\kappa_2-\gamma_+)/(\kappa_2-\gamma_-)]$. The value of γ is independent of the branch of $\ln P$ chosen. Note that if $g_1=0$ and $h_2\neq 0$, γ is given by the relation

$$\gamma=\frac{d}{dt}\ln\{\sinh[\delta(t-t_0)]\}, \tag{25}$$

where $\delta = 2(-f_0h_2)^{1/2}$ and $t_0 = (2\delta)^{-1}\ln[(\kappa_2 - \delta)/(\kappa_2+\delta)]$. The value of γ is independent of the branch of the square root which is chosen for δ and the branch of the logarithm which is chosen in (25). The expressions for b and c are

$$b(t) = \exp\{-\int_0^t [4f_0a(t') + g_1]\,dt'\} \times [k_1 - \int_0^t \exp\{\int_0^{t'} [4f_0a(t'') + g_1]dt''\} \times [2g_0a(t') + h_1]dt'] \quad (26)$$

and

$$c(t) = k_0 - \int_0^t [f_0\{[b(t')]^2 + 2a(t')\} + g_0b(t') + h_0]dt'. \quad (27)$$

If, on the other hand, $g_1 = g_0 = h_1 = 0$ and $h_2 \neq 0$, the determination of a, b, and c in terms of algebraic and elementary transcendental functions is a relatively simple matter. If we impose the condition that $g_1 = 0$, it is unnecessary to limit further the class of differential equations for which we can obtain ingegral representations by imposing the conditions $g_0 = h_1 = 0$. Instead, we take advantage of the perceived simplification by introducing certain transformations of dependent and independent variables. These transformations permit us to apply our method to a transformed differential equation, instead of the original equation. In it, the coefficients of powers of the transformed independent variable in the linear differential operator which vanish are those that correspond to g_0 and h_1, in addition to those that correspond to f_2, f_1, g_2, and g_1. Thus Subclass 2 consists of the equations

$$f_0y'' + g_0y' + (h_2z^2 + h_1z + h_0)y(z) = \exp(k_2z^2 + k_1z + k_0), \quad (28)$$

in which additional coefficients may vanish.

The transformations of dependent and independent variables are the following. The solution of (28) is given by $y(z) = u(z)v(z)$, where $u(z) = \exp\{-(g_0/2f_0)[z + (h_1/2h_2)]\}$, $v(z) = w(x)$, $x = (h_2/f_0)^{1/4}[z + (h_1/2h_2)]$, and $w(x)$ is a particular integral of the equation

$$\frac{d^2w}{dx^2} + (x^2 - \beta)w(x) = \exp(l_2x^2 + l_1x + l_0). \quad (29)$$

Note that, even if z is real, x may be complex. In this equation

$$\beta = \frac{1}{4}\frac{g_0^2}{f_0^{3/2}h_2^{1/2}} + \frac{1}{4}\frac{h_1^2}{f_0^{1/2}h_2^{3/2}} - \frac{h_0}{f_0^{1/2}h_2^{1/2}}, \quad (30)$$

$$l_2 = k_2(f_0/h_2)^{1/2}, \quad (31)$$

$$l_1 = -\frac{k_2f_0^{1/4}h_1}{h_2^{5/4}} + \frac{k_1f_0^{1/4}}{h_2^{1/4}} + \frac{1}{2}\frac{g_0}{f_0^{3/4}h_2^{1/4}}, \quad (32)$$

$$l_0 = \frac{1}{4}\frac{k_2h_1^2}{h_2^2} - \frac{1}{2}\frac{k_1h_1}{h_2} + k_0 - \log(f_0h_2)^{1/2}. \quad (33)$$

Henceforth, for the purpose of determining integral representations of the transformed equation (29), we make the replacements $w \to y$, $x \to z$, $l_2 \to k_2$, $l_1 \to k_1$, and $l_0 \to k_0$. We assume integral representations of the form described in Sec. 3.

The determination of the complete set of possible contours of integration for integral representations is simplified considerably by choosing as the variable of integration a instead of t. We thereby obtain representations in the form

$$y(z) = -\int_{C''} \exp[z^2a + zb(a) + c(a) + m(a)]da, \quad (34)$$

where $m(a) = -\log[4(a^2 + \frac{1}{4})]$ and C'' are contours of integration which proceed from the point $a = k_2$ and satisfy the conditions stated in Sec. 3. The functions $b(a)$ and $c(a)$ are determined in the following manner. Combining (17) and (18) and integrating the resulting differential equation subject to the conditions $a(t=0) = k_2$ and $b(t=0) = k_1$, we obtain the result

$$b(a) = B(a^2 + \tfrac{1}{4})^{1/2}, \quad (35)$$

where $B = k_1(k_2^2 + \frac{1}{4})^{-1/2}$. Proceeding in similar fashion with (17) and (19), eliminating b^2 from (19) by the use of (35), and imposing the conditions $a(t=0) = k_2$ and $c(t=0) = k_0$, we obtain the result

$$c(a) = \tfrac{1}{4}B^2(a - k_2) + \tfrac{1}{4}\log\left(\frac{a^2 + \frac{1}{4}}{k_2^2 + \frac{1}{4}}\right) - \tfrac{1}{2}\beta\tan^{-1}(2a) + \tfrac{1}{2}\beta\tan^{-1}(2k_2) + k_0. \quad (36)$$

It is convenient for the purpose of determining contours of integration to express the inverse tangent in terms of logarithmic functions by the relation

$$\tan^{-1}(2a) = \tfrac{1}{2}i\log\left[-\left(\frac{a + \frac{1}{2}i}{a - \frac{1}{2}i}\right)\right]. \quad (37)$$

Contours of integration C'' are of two types: those for which the upper limit is $a = \pm\frac{1}{2}i$ and the limit of $\exp c(a)$ as a approaches the upper limit on the contour is zero; and those for which the upper limit is at infinity and the limit of $\exp[z^2a + zb(a) + c(a)]$ as a approaches the upper limit on the contour is zero. With regard to the former type of contour, note from (36) and (37) that if $\mathrm{Im}\beta > -1$, the upper limit $a = -\frac{1}{2}i$ must be chosen. If $\mathrm{Im}\beta < 1$, the upper limit $a = \frac{1}{2}i$ must be chosen. Note that either contour of integration may be used in the region $-1 < \mathrm{Im}\beta < 1$. The use of the variable of integration $\sigma = 2t$, where $\sigma = -\tan^{-1}(2a)$, may be convenient for purposes of approximation and computation because the upper limit of integration is at infinity. Compare the Appendix. With regard to the latter type of contour, note that the dominant asymptotic behavior of the quantity $[z^2a + zb(a) + c(a)]$ as a increases without limit is $(z^2 + Bz + \frac{1}{4}B^2)a$. Thus one requires that the contour of integration approach infinity within the range of directions for which $\mathrm{Re}[(z^2 + Bz + \frac{1}{4}B^2)a] < 0$. At the value of z for which the quantity $(z^2 + Bz + \frac{1}{4}B^2)$ vanishes, the condition cannot be satisfied.

In contrast to the case of Subclass 1, the extension of Subclass 2 to equations in which the inhomogeneous term is $z^n\exp(k_2z^2 + k_1z + k_0)$, where n is a positive integer, introduces integral representations distinct from the basic integral representation. To see this, it is sufficient to treat the case $n = 1$. An integral representation for $n = 1$ is given by $y_1 = (\partial y_0/\partial\lambda)|_{\lambda=k_1}$, where y_0 is the representation given in (34). The integrand of y_1 differs from that of y_0 by the factor $[z(k_2^2 + \frac{1}{4})^{-1/2} \times (a^2 + \frac{1}{4})^{1/2} + \frac{1}{2}k_1(k_2^2 + \frac{1}{4})^{-1}(a - k_2)]$. The contours of integration for the additional integral representations are the same as the contours for y_0.

It may be possible to treat, in an approximate manner, equations with more general inhomogeneous terms. On the range $-\infty < z < \infty$, it may be possible to approximate an inhomogeneous term by a reasonable number of terms of an expansion in weighted Hermite polynomials. We have shown that the extension of Subclass 2 to equations in which the inhomogeneous term is $z^n \exp(k_2 z^2 + k_1 z + k_0)$ introduces integral representations distinct from the basic integral representation. Thus the approximate representation of a particular integral of an equation with a more general inhomogeneous term may be relatively less convenient than in the case of Subclass 1.

6. DISCUSSION

We now discuss various aspects of the method which we have presented.

An obvious question arises concerning the relation of the method to the theory of integral representations of homogeneous differential equations.[1] The integrands of integral representations of equations of Subclass 1 are the same as those of the corresponding homogeneous equations based upon the Laplace kernel $\exp(zb)$. The contours of integration of the integral representations of inhomogeneous equations differ from those of representations of homogeneous equations. In the former case, the limits of integration and the contour along which the integral is taken are such that at the lower limit of integration the bilinear concomitant is equal to minus the inhomogeneous term and, as the upper limit of integration is approached, the bilinear concomitant approaches zero. In the latter case, of course, the limits of integration and the contour along which the integral is taken is such that the bilinear concomitant returns to its initial value at the end of the contour. The integral representations of equations of Subclass 2 are similarly related to those of the corresponding homogeneous equations based upon the Laplace kernel $\exp(z^2 a)$ only if l_1 vanishes. Otherwise, a kernel canot be found. As a practical matter, this requires that $k_1 = g_0 = 0$ and that $k_2 = 0$ or $h_1 = 0$. Thus, for example, in consequence of the requirement that $k_1 = 0$, representations of equations with inhomogeneous terms $z^{(2n+1)} \exp(k_2 z^2 + k_1 z + k_0)$, where $n = 0, 1, 2, \ldots$, cannot be related to integral representations of the corresponding homogeneous equations.

Various transformations of the variable of integration may be used to cast an integral representation in a form which is particularly convenient for approximate analytic or numerical evaluation. Although b and a have been used in presenting the development of the method for Subclasses 1 and 2, respectively, in particular cases it may be desirable to use the original variable of integration, t.

The method presented here should be applicable to a wide range of problems in theoretical physics. One particular area of application is wave propagation in inhomogeneous media, such as plasma. Another is resistive boundary layer problems of magnetohydrodynamics, such as those of tearing modes and resistive internal kink modes.[5]

ACKNOWLEDGMENTS

I wish to thank A. Baños, Jr., A. Bers, H. Feshbach, R. Galvão, A. N. Kaufman, G. J. Morales, and A. Sen for helpful discussions.

APPENDIX: DERIVATION OF (7)

In the case of (4), integral representations can be obtained in a straightforward manner by the use of Fourier transforms. A solution $y(z) = \exp(\frac{1}{2} i z^2)\, \Psi(v)$, where $v = z^2$, is assumed. The function $\Psi(v)$ is a solution of the inhomogeneous differential equation

$$iv\Psi'' + (ic - v)\Psi' - a\Psi(v) = \tfrac{1}{4} i \exp(-\tfrac{1}{2} iv), \tag{A1}$$

where $a = \frac{1}{4}(1 + i\alpha)$ and $c = \frac{1}{2}$. Introducing the Fourier transform

$$\tilde{\Psi}(q) = \int_{-\infty}^{\infty} \Psi(v) e^{-iqv} dv, \tag{A2}$$

we obtain the following first-order differential equation in the transform variable,

$$\frac{d}{dq}[q(q+1)\tilde{\Psi}] - (cq + a)\tilde{\Psi} = \tfrac{1}{2}\pi i \delta(q + \tfrac{1}{2}). \tag{A3}$$

In this equation, δ is the Dirac delta function. Defining the quantities $\tilde{\Theta}(q) = q(q+1)\tilde{\Psi}(q)$ and $S(q) = (cq + a)/q(q+1)$, we obtain a differential equation for $\tilde{\Theta}$, which may be integrated from the constant value q_0 to q to give

$$\tilde{\Theta}(q) = \exp\left(\int_{q_0}^{q} S dq'\right)\tilde{\Theta}(q_0) + \tfrac{1}{2}\pi i \int_{q_0}^{q} \exp\left(\int_{q'}^{q} S dq''\right)\delta(q' + \tfrac{1}{2}) dq'. \tag{A4}$$

We may choose $q_0 \to \pm\infty$ and $\tilde{\Theta}(\pm\infty) = 0$. We thereby obtain, respectively,

$$\tilde{\Theta}(q) = \mp \tfrac{1}{2}\pi i \exp\left(\int_{-1/2}^{q} S dq'\right) u[\mp(q + \tfrac{1}{2})], \tag{A5}$$

where u is the unit step function. Expressing $\tilde{\Psi}(q)$ in terms of $\tilde{\Theta}$, determining the inverse transform,

$$\Psi(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{iqv}\tilde{\Psi}(q) dq, \tag{A6}$$

and expressing $y(z)$ in terms of $\Psi(v)$, we finally obtain the integral representations

$$y(z) = \tfrac{1}{4} i \int_{-1/2}^{\mp\infty} \exp\left[i(z^2 + \tfrac{1}{2})q + \int_{-1/2}^{q} S dq'\right] \frac{1}{q(q+1)} dq. \tag{A7}$$

In order to determine the conditions under which these two solutions are valid, it is necessary to examine the quantity

$$\int_{-1/2}^{q} S dq' - \log[q(q+1)] = \tfrac{1}{4}(-3 + i\alpha)\log q - \tfrac{1}{4}(3 + i\alpha)\log(q+1) - \tfrac{1}{4}(1 + i\alpha)\log(-\tfrac{1}{2}) + \tfrac{1}{4}(-1 + i\alpha)\log(\tfrac{1}{2}). \tag{A8}$$

If the upper limit of integration in (A7) is $-\infty$, the path of integration contains the point $q = -1$. Accordingly, the existence of the integral requires that the condition $\mathrm{Im}\alpha > -1$ be satisfied. If the upper limit is $+\infty$, the path of integration contains the point $q = 0$ and the condition $\mathrm{Im}\alpha < 1$ must be satisfied. Note that either representation may be used in the region $-1 < \mathrm{Im}\alpha < 1$.

The latter choice is appropriate in the case of electrostatic linear mode conversion in a parabolic density profile because $\mathrm{Im}(\alpha)$ is proportional to minus the collision frequency.

Operation of the linear differential operator of (3) on the integral representation (5) produces an integral of a derivative of a function of the variable of integration. At the upper limit of integration the function vanishes. At the lower limit it produces the inhomogeneous term. Such is not the case when the linear differential operator of (4) operates on the integral representations (A7). There is an important difference between the two sets of integral representations (5) and (A7). Representing a particular inhomogeneous equation as

$$Ly(z)=R(z), \tag{A9}$$

we see that the representation (5) is of the form

$$y(z)=\int_C \exp F(z;t)dt, \tag{A10}$$

where $t=0$ is the lower limit of integration and $\exp f(z;0)=R(z)$. In contrast, the representations (A7) are of the form

$$y(z)=\int_C \exp[F(z;t)]G(t)dt, \tag{A11}$$

where, as before, $\exp F(z;0)=R(z)$, and $G(t)$ is a nonconstant function of the variable of integration. Integral representations of (4) which are of the form (A10) can be obtained by the introduction of a transformation of the variable of integration. The desired transformation of variable of integration is defined by the relation $d\sigma=\frac{1}{2}i[q(q+1)]^{-1}dq$ and the correspondence of $q=-\frac{1}{2}$ with $\sigma=0$. The direct and inverse transformations are, respectively, $\sigma=\frac{1}{2}i\{\log[q/(q+1)]-\log(-1)\}$, where the same branch of the logarithmic function is choosen in both terms, and $q=-(1+\exp 2i\sigma)^{-1}$. Introducing the transformation into (A7), we obtain the integral representations

$$y(z)=\int_{C_{0\pm}} \exp\{-\tfrac{1}{2}z^2\tan\sigma+\tfrac{1}{2}\alpha\sigma + \log[\tfrac{1}{2}(\cos\sigma)^{-1/2}]\}d\sigma. \tag{A12}$$

The contours of integration $C_{0\pm}$ consist of two segments. The first, which we denote by $C_\pm$, proceeds from the origin to upper limits at $r\pm i\infty$, where r is a finite real number. The second segment proceeds from $r\pm i\infty$ to $\frac{1}{2}\pi(2n+1)$, where $n=0, \pm1, \pm2, \cdots$. Since in this case $F(z;r\pm i\infty)=F[z;\frac{1}{2}\pi(2n+1)]=0$, the parts of the integral representations whose contours of integration are the second segments are solutions of the homogeneous equation corresponding to (4). We choose to omit them and thereby obtain the integral representations

$$y(z)=\int_{C_\pm} \exp\{-\tfrac{1}{2}z^2\tan\sigma+\tfrac{1}{2}\alpha\sigma + \log[\tfrac{1}{2}(\cos\sigma)^{-1/2}]\}d\sigma, \tag{A13}$$

which appear in (7).

[1]P.M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part I. Chapter 5, Sec. 3.

[2]G.J. Morales and Y.C. Lee, Phys. Rev. Lett. 33, 1016 (1974).

[3]M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), Secs. 10.4.44 and 10.4.56.

[4]Ref. 3, pp. 1031, 1040. In the Subject Index, cf. the headings "Airy functions: differential equations" and "Parabolic cylinder functions: differential equation."

[5]R. Galvão, private communication.

Quantization of spinor fields

Piotr Garbaczewski

Institute of Theoretical Physics, University of Wrocław, Cybulskiego 36, 50-205 Wrocław, Poland
(Received 16 February 1977)

Influenced by Klauder's investigations on the same subject, we study the question of correspondence principle for Dirac fields, looking for its formulation without use of Grassman algebras. We prove that with each Fermi operator (the series with respect to asymptotic free fields): $\Omega(\psi,\bar{\psi})$: one can associate the functional $\Omega^C(\psi^C, \bar{\psi}^C)$ with respect to classical spinor fields. Here the projector 1_F and the Hilbert (Fock) space $\mathcal{J}_F = 1_F \mathcal{J}_{F_B}$ are given such that the identity 1_F: $\Omega^C(\psi^B, \bar{\psi}^B)$: $1_F \mathcal{J}_{F_F} = :\Omega(\psi, \bar{\psi}):\mathcal{J}_F$ defines the mediating boson level, where coherent state expectation values of operator expressions are in order: $\langle :\Omega^C(\psi^B, \bar{\psi}^B): \rangle = \Omega^C(\psi^C, \bar{\psi}^C)$. For proofs we employ functional differentiation (resp. integration) methods, especially in connection with the use of functional representations of the CCR and CAR algebras.

1. THE CORRESPONDENCE PRINCIPLE FOR SCALAR FIELDS

In the present paper we shall not go beyond the framework of the conventional quantum field theory, and all considerations are essentially based on its LSZ formulation.[1] The basic assumption here is that any operator quantity characterizing a given quantum system (scalar field) admits a decomposition into power series expansions with respect to normal ordered products of free asymptotic fields. With a given scalar quantum field

$$\phi(x) \underset{t\pm\infty}{\to} \varphi_{\substack{\mathrm{in}\\ \mathrm{out}}}(x) = \varphi(x),$$

we associate an algebra of all operators,

$$:F(\varphi): = \sum_n (f_n, :\varphi^n:) \tag{1.1}$$

where $(\cdot,\cdot)$ is a bilinear form, and the Schwartz nuclear theorem allows us to consider $(f_n, :\varphi^n:)$ in the form

$$(f_n, :\varphi^n:) = \int dx_1 \cdots \int dx_n f_n(\mathbf{x}_n) : \varphi(x_1) \cdots \varphi(x_n):$$
$$\mathbf{x}_n = (x_1, \ldots, x_n), \quad x_k \in M_4. \tag{1.2}$$

In general there appears the highly nontrivial task of recovering conditions, necessary to impose on coefficient functions (distributions) $\{f_n\}$, to get proper algebraic properties on a suitably chosen domain. We do not bother with this question in the course of the paper. With the Fock representation of the CCR algebra (asymptotic condition) in mind, $\{a^*, a, \Omega_B\}_K$, $K = \mathcal{L}^2(\mathbb{R}^3)$, we introduce a coherent state domain for our operator algebra according to

$$\mathcal{L}^2(\mathbb{R}^3) \ni \alpha, \quad (\alpha, \bar{\alpha}) = \int_{\mathbb{R}^3} dk\, \alpha(k)\bar{\alpha}(k) = \|\alpha\|^2,$$
$$|\alpha) = \exp(-\|\alpha\|^2/2)\exp(\alpha, a^*)\Omega_B, \tag{1.3}$$
$$(\alpha|a(k)|\alpha) = \langle a(k)\rangle = \alpha(k), \quad k \in \mathbb{R}^3.$$

If $\alpha, \bar{\alpha}$ appear as classical (complex) Fourier amplitudes of $\overset{c}{\varphi}(x) : \alpha, \bar{\alpha} \to a, a^* \Rightarrow \overset{c}{\varphi}(x) \to \varphi(x)$, we get

$$(\alpha|\varphi(x)|\alpha) = \langle\varphi(x)\rangle = \overset{c}{\varphi}(x),$$
$$(\alpha| :F(\varphi): |\alpha) = F(\overset{c}{\varphi}) = \sum_n (f_n, \overset{c}{\varphi}{}^n). \tag{1.4}$$

The formula (1.4) establishes a *correspondence between the quantum and classical level of a given (scalar) field and the associated algebra.* All the algebraic manipulations appearing on the quantum level induce corresponding relations on the classical level, and therefore many essential questions as, e.g., estimates (concerning the convergence of operator series, criterions for joint multiplication) are transferred onto the classical level, where powerful analytic methods allowing to solve them are known (compare Ref. 2). In connection with those problems, it is extremely useful to employ so-called *functional representations* of the CCR algebra, arising in the theory of the functional power series.[2] Namely, let us assume that we have given the Hilbert space [Bargmann space $\mathcal{B}(\mathcal{L}^2(\mathbb{R}^3))$] of all functional power series $V(\bar{\alpha})$:

$$V(\bar{\alpha}) = \sum_n (1/\sqrt{n!})(v_n, \bar{\alpha}^n)$$
$$= \sum_n (1/\sqrt{n!}) \int d\mathbf{k}_n\, v_n(\mathbf{k}_n)\, \bar{\alpha}(k_1)\cdots\bar{\alpha}(k_n),$$
$$d\mathbf{k}_n = dk_1 \cdots dk_n,$$
$$\|V\|^2 = (\bar{V}, V) = \bar{V}\left(\frac{d}{d\bar{\alpha}}\right) V(\bar{\alpha})_{|\bar{\alpha}=0} = \sum_n (\bar{v}_n, v_n)$$
$$= \sum_n \|v_n\|^2 = \int \bar{V}(\gamma)\, V(\bar{\gamma}) \exp[-(\bar{\gamma}, \gamma)]\, d\,\frac{\gamma}{\sqrt{\pi}}\,. \tag{1.5}$$

where $d/d\bar{\alpha}$ symbolizes the Gateaux derivative with respect to $\bar{\alpha} \in \mathcal{L}^2(\mathbb{R}^3)$, while $d(\gamma/\sqrt{\pi})$ the functional (Gaussian path) integration measure, compare Ref. 2 and J. Rzewuski's monograph.[3]

In $\mathcal{B}(\mathcal{L}^2(\mathbb{R}^3))$ we assume (Ref. 2) to have defined an algebra of double power series:

$$A(\bar{\alpha}, \alpha) = \sum_{nm} \frac{1}{\sqrt{n!m!}}(a_{nm}, \bar{\alpha}^n\alpha^m) = \sum_{nm}\frac{1}{\sqrt{n!m!}} \int d\mathbf{k}_n \int d\mathbf{p}_m$$
$$\times a_{nm}(\mathbf{k}_n, \mathbf{p}_m)\,\bar{\alpha}(k_1)\cdots\bar{\alpha}(k_n)\,\alpha(p_1)\cdots\alpha(p_m),$$
$$(AB)(\bar{\alpha}, \alpha) = \sum_{nm}\frac{1}{\sqrt{n!m!}}\left(\sum_k (a_{nk}, b_{km}), \bar{\alpha}^n\alpha^m\right)$$
$$= A\left(\bar{\alpha}, \frac{d}{d\bar{\gamma}}\right) B(\bar{\gamma}, \alpha)_{|\bar{\gamma}=0}$$
$$= \int A(\bar{\alpha}, \gamma)\, B(\bar{\gamma}, \alpha) \exp[-(\bar{\gamma}, \gamma)]\, d\left(\frac{\gamma}{\sqrt{\pi}}\right),$$

$$(AV)(\bar{\alpha}) = V'(\bar{\alpha}) = \sum_n \frac{1}{\sqrt{n!}}\left(\sum_k (a_{nk}, v_k), \bar{\alpha}^n\right)$$
$$= A\left(\bar{\alpha}, \frac{d}{d\bar{\gamma}}\right) V(\bar{\gamma})_{|\bar{\gamma}=0} = \int A(\bar{\alpha}, \gamma)\, V(\bar{\gamma})$$
$$\times \exp[-(\bar{\gamma}, \gamma)]\, d\left(\frac{\gamma}{\sqrt{\pi}}\right). \tag{1.6}$$

 0022-2488/78/1903-0642$1.00

The underlying Hilbert space and the algebra can be derived from much worse defined objects by applying suitable analytic restrictions (their study in the framework of double functional power series is given in Ref. 2).

In the course of the paper, we do not pretend to give highly correct meaning to the notion of functional (path) integrals (see, e.g., Ref. 4); all the definitions establishing a sufficient axiomatization of the formalism can be found in Rzewuski's monograph.[3]

Theorem 1 (*functional representation of the CCR*): Double power series $(\bar{\alpha},f)\exp(\bar{\alpha},\alpha)=a(f)^*(\bar{\alpha},\alpha)$, $(\bar{f},\alpha)\exp(\bar{\alpha},\alpha)=a(f)(\bar{\alpha},\alpha)$, $f\in\mathcal{L}^2(\mathbb{R}^3)$, play in $\mathcal{J}_B=\mathcal{B}(\mathcal{L}^2(\mathbb{R}^3))$ the role of generators $a(f)^*$, $a(f)$ respectively of the Fock representation of the CCR algebra with the vacuum vector $\Omega_B=1$ (the whole set of complex numbers $\mathbb{C}$ spans in fact the vacuum sector).

Proof: Given in Refs. 2, 5; for further convenience we shall only quote

$$[a(f),a(g)^*]_-(\bar{\alpha},\alpha)=(\bar{f},g)\exp(\bar{\alpha},\alpha)=(\bar{f},g)\,\mathbb{1}_B(\bar{\alpha},\alpha),$$
$$[a(f),a(g)]_-(\bar{\alpha},\alpha)=0,\quad (a(f)\Omega_B)(\bar{\alpha})=0. \tag{1.7}$$

As a corollary to Theorem 1, one can easily prove:

Lemma 1: For any $F(\mathring{\varphi})=A(\bar{\alpha},\alpha)$ (after suitable reordering of summations and integrations), the double power series $F(\mathring{\varphi})\exp(\bar{\alpha},\alpha)$ play in $\mathcal{J}_B=\mathcal{B}(\mathcal{L}^2(\mathbb{R}^3))$ the role of the operator $:F(\varphi):$

$$:F(\varphi):(\bar{\alpha},\alpha)=F(\mathring{\varphi})\exp(\bar{\alpha},\alpha). \tag{1.8}$$

Proof: Immediate, by applying (1.6); see also Ref. 2. In consequence, in addition to the correspondence rule (1.4) we can formulate the quantization principle (1.8) allowing to reconstruct immediately the quantum object from a given classical object. Here (see Rzewuski's monograph) the algebraic structure on the quantum level induces a corresponding structure on the classical level:

$$\begin{aligned}:F_1(\varphi)::F_2(\varphi):&\Rightarrow(:F_1(\varphi)::F_2(\varphi):)(\bar{\alpha},\alpha)\\&=\exp(\bar{\alpha},\alpha)\cdot\{F_1(\mathring{\varphi})(*)F_2(\mathring{\varphi})\}=\exp(\bar{\alpha},\alpha)F_{12}(\mathring{\varphi})\\&=:F_{12}(\varphi):(\bar{\alpha},\alpha)\Rightarrow:F_{12}(\varphi):,\end{aligned} \tag{1.9}$$

where

$$(*)=\exp\left(\frac{\overleftarrow{d}}{d\mathring{\varphi}}\Delta\frac{\overrightarrow{d}}{d\mathring{\varphi}}\right),$$
$$\frac{\overleftarrow{d}}{d\mathring{\varphi}}\Delta\frac{\overrightarrow{d}}{d\mathring{\varphi}}:=\int\frac{\overleftarrow{d}}{d\mathring{\varphi}(x)}\Delta(x-y)\frac{\overrightarrow{d}}{d\mathring{\varphi}(y)}\,dx\,dy. \tag{1.10}$$

Arrows indicate the direction in which operators act, and $\Delta(x-y)$ is the Pauli–Jordan distribution.

The identity (1.9) recovers what is the relation between the quantum and (implied) classical multiplication rules. The situation appearing can be summarized in the following:

Correspondence principle: (*i*) *Correspondence rule*: $\{:F(\varphi):\}\to\{F(\mathring{\varphi})\}$:

$$(\alpha\,|:F(\varphi):|\,\alpha)=F(\varphi),$$
$$(\alpha\,|:F_1(\varphi)::F_2(\varphi):|\,\alpha)=F_1(\mathring{\varphi})\exp\left(\frac{\overleftarrow{d}}{d\mathring{\varphi}}\Delta\frac{\overrightarrow{d}}{d\mathring{\varphi}}\right)F_2(\mathring{\varphi})=F_{12}(\varphi). \tag{1.11}$$

(*ii*) *Quantization rule*: $\{F(\mathring{\varphi})\}\to\{:F(\varphi):\}$:

$$F(\mathring{\varphi})\exp(\bar{\alpha},\alpha)=:F(\varphi):(\bar{\alpha},\alpha)\Rightarrow:F(\varphi):,$$
$$F_{12}(\mathring{\varphi})\exp(\bar{\alpha},\alpha)=(:F_1(\varphi)::F_2(\varphi):)(\bar{\alpha},\alpha)\Rightarrow:F_{12}(\varphi):. \tag{1.12}$$

Commonly, the quantization is believed to be performed, if the Green's functions are given. For this purpose, one needs, however, the knowledge of the generating functional:

$$Z(\eta)=\frac{\int\exp\{i[S+\int dx\,\eta(x)\,\mathring{\varphi}(x)]\}\rho d(M\mathring{\varphi}/\sqrt{i\pi})}{\int\exp(iS)\,d(M\mathring{\varphi}/\sqrt{i\pi})} \tag{1.13}$$

where S is the classical action, M is an arbitrary linear operator, and $\mathring{\varphi}$ a quite arbitrary scalar field. The integration measure $d(M\mathring{\varphi}/\sqrt{i\pi})$ is defined according to Rzewuski's monograph[2] (Fresnel integral).

The two-point Green's function is then given by

$$G(x,y)=i\,\frac{d}{d\eta(x)}\,\frac{d}{d\eta(y)}\,Z(\eta)_{|\eta=0}. \tag{1.14}$$

In the above, η is a suitable classical source function. It is useful to know that, in the free field case, $Z(\eta)$ reduces to

$$Z(\eta)=\exp[-(i/2)\int\eta(x)\,G(x,y)\,\eta(y)\,dx\,dy], \tag{1.15}$$

where Δ is one of the Green's functions of the *KG* operator (the arbitrariness exists), usually chosen to be the causal function.

2. INTRODUCTION TO THE PROBLEM: FERMIONS

Pragmatists working in the domain of quantum field theory are strongly convinced (see, e.g., Coleman's opinion expressed in Ref. 6) that quite satisfactory (though even not fully correct) classical level for the algebra associated with any Fermi (Dirac, say) field is given in the framework of Grassman algebras, which are built of *c-number-like, but anticommuting objects*. This last property manifestly exhibits the Pauli exclusion principle, influencing the starting Fermion level. Investigations[3,7] have been going in this direction (especially because of the similarity of the formal scheme, allowing us to reproduce all the results in the manner analogous to this of Bose case). There was even founded a complete mathematical theory (Berezin's[3] monograph) of anticommuting numbers in functional-like differentiation and integration procedures.

Let us add that if in Theorem 1 we formally put elements of the anticommuting ring in place of square integrable functions, a Fock representation of the CAR would be obtained (see, e.g., Garbaczewski,[8] where a complete construction is given).

If we follow the Grassmanian way, the generating functional (the notion used here in rather ambiguous meaning) for the Green's functions of the Dirac field reads

$$Z(\eta,\bar{\eta})=\frac{\int \exp(i\{S+\int[\bar{\eta}(x)\psi(x)+\eta(x)\bar{\psi}(x)]dx\}\cdot d(M\psi/\sqrt{i\pi})}{\int \exp(iS)\,d(M\psi/\sqrt{i\pi})} \tag{2.1}$$

(M is an arbitrary linear operator). If electromagnetic interactions are taken into account (with the Faddeev–Popov measure $\delta\mu_A$; see Popov's monograph[2]), then

$$Z(\eta,\bar{\eta},\eta_\mu)=\frac{\int \exp\{i[S+\int(\bar{\eta}\psi+\bar{\psi}\eta+\eta_\mu A^\mu)\,dx]\}\,\delta\mu_A\,d(M/\sqrt{i\pi})}{\int \exp(iS)\,\delta\mu_A\,d(M\psi/\sqrt{i\pi})}, \tag{2.2}$$

where $\eta,\bar{\eta},\eta_\mu$ are "sources" of fields $\psi,\bar{\psi},A_\mu$ respectively. Notice that $\eta,\bar{\eta},\psi,\bar{\psi}$ belong to the Grassman algebra, and $\delta\mu_A$ integrates over classes (orbits with respect to the gauge group).

On the other hand it is perfectly well known that one can always construct the set of (c-valued!) functional power series with respect to free Dirac fields and equip this set with a suitable topology and algebraic structure. So, it is rather surprising that no reasonable correspondence with the (prospective) quantum level was found. Really, *the Pauli exclusion principle does not govern the considered classical level, in contrast to the Grassman approach.*

At this point we do not wish to tilt at windmills and advocate this pure c-number point of view, against the just-described Grassman tools (especially because these last are widely spread and quite convenient in explicit calculations). We wish, however, to prove that the reasonable *correspondence principle can be established between functional power series of Dirac spinors and operator series with respect to normal products of Dirac fermions.* This correspondence will be established in a correct and unambiguous way with no reference to Grassman methods.

Let us mention the two isolated attempts in this direction which are known to the author; see, e.g., Ref. 8. It was proved that c-number images of Fermion functionals do exist. Another possibility[8] was to construct a pure operator theory where functional-like differentiations and integrations would be carried out with respect to operators. In the free field case, the generating functional (2.2) reads

$$Z(\eta,\bar{\eta},\eta_\mu)=\exp[-i\int\bar{\eta}(x)G(x-y)\eta(y)\,dx\,dy - \tfrac{1}{2}\int\eta_\mu(x)D^{\rm tr}_{\mu\nu}(x-y)\eta_\nu(y)\,dx\,dy], \tag{2.3}$$

where

$$G(x-y)=(2\pi)^{-4}\int dp\,\exp[ip(x-y)]\cdot\frac{\hat{p}+m}{p^2-m^2+i0},$$
$$D^{\rm tr}_{\mu\nu}(x-y)=(2\pi)^{-4}\int dk\,\exp[ik(x-y)]\cdot\frac{-k^2\delta_{\mu\nu}+k_\mu k_\nu}{(k^2+i0)^2}. \tag{2.4}$$

The choice of the causal function is justified by the need for uniqueness of the expressions in exponents; the arbitrariness mentioned in connection with scalar case is thus removed. The integral received gives the photon part in the Lorentz gauge. For more details, see Ref. 3.

Let us add that the functional (Grassman level) definition of the two-point Green's function corresponding to the spinor field, by the use of (anticommuting) derivatives with respect to sources, reads

$$G_{\alpha\beta}(x,y)=-i\,\frac{d^2Z(\eta,\bar{\eta},\eta_\mu)}{d\bar{\eta}_\alpha(x)\,d\eta_\beta(y)}\bigg|_{\eta=\bar{\eta}=0}. \tag{2.5}$$

3. INTERLUDE: BOSON EXPANSION METHOD IN THE QUANTUM THEORY OF FERMIONS

Realizing the program sketched in Sec. 2, we intend to close, by the present paper, the series,[5] developing the method of Boson expansions in application to Fermi systems. The first two papers of Ref. 8, of these series, include in fact an attempt to apply a c-number language in the functional formulation of the quantum theory of Fermi systems: So-called *functional representations* of the CAR algebra were invented there. The third paper of Ref. 5, of these series, generalizing results of the previous two onto the algebraic level, began a systematic study of the "*bosonization*" question (the term used by us as the shorthand version of the title of this section) from both mathematical and physical points of view.

Theorem 2 (*representation of the CAR*): Let us denote $\sigma_n(\mathbf{k}_n)=\sigma(k_1,\ldots,k_n)$, $k\in\mathbb{R}^3$, the Friedrichs–Klauder sign function[8,9] being a continuous generalization of the n-point Levi-Civitta tensor. Let $\{a^*,a,\Omega_B\}_{L^2(\mathbb{R}^3)}$ generate a Fock representation of the CCR algebra over the Hilbert space $L^2(\mathbb{R}^3)$. The underlying Fock space we denote $\mathcal{F}_B$. Then, the triple $\{b^*,b,\Omega_B\}_{L^2(\mathbb{R}^3)}$ with

$$(a^*,a)=\int dk\,a^*(k)\,a(k),$$
$$b(f)=:\exp[-(a^*,a)]\cdot\sum_{nm}(1/\sqrt{n!\,m!})\times\int d\mathbf{k}_n\int d\mathbf{p}_m\,f_{nm}(\mathbf{k}_n,\mathbf{p}_m)\times a^*(k_1)\cdots a^*(k_n)\,a(p_1)\cdots a(p_m):,$$
$$f_{nm}(\mathbf{k}_n,\mathbf{p}_m)=\sqrt{n+1}\,\delta_{m,1+n}\,\sigma_n(\mathbf{k}_n)\bar{f}(p_1)\,\sigma_{1+n}(\mathbf{p}_{1+n})\times\delta(k_1-p_2)\,\delta(k_2-p_3)\cdots\delta(k_n-p_{1+n}) \tag{3.1}$$

generates a Fock representation of the CAR algebra over $L^2(\mathbb{R}^3)$, whose (Fock) representation space $\mathcal{F}_F$ is selected from $\mathcal{F}_B$ due to projection properties of the operator unit $\mathbb{1}_F$:

$$[b(f),b(g)^*]_+=(\bar{f},g)\,\mathbb{1}_F, \tag{3.2}$$
$$\mathbb{1}_F=:\exp[-(a^*,a)]\cdot\sum_n\frac{1}{n!}\int d\mathbf{k}_n\,a^*(k_1)\cdots a^*(k_n)\,\sigma_n^2(\mathbf{k}_n)\times a(k_1)\cdots a(k_n):,$$
$$\mathcal{F}_F=\mathbb{1}_F\mathcal{F}_B,$$

which implies the coincidence of the vacuum and one-particle sectors for the representations (CCR and CAR respectively).

Proof: Details are given in Ref. 8 and in the first paper of Ref. 5. The only difference lies in that we use an explicit form $E_n(\mathbf{k}_n,\mathbf{p}_n)=\sigma_n(\mathbf{k}_n)\,\delta(k_1-p_1)\cdots\delta(k_n-p_n)$ of the integral kernel of the square root of the abstract projector E_n^2 appearing in the original derivation.

Comments: (i) The extension of Theorem 2 to an arbitrary number of internal degrees of freedom is nearly immediate, and, by the substitutions

$$\bar{f} \to \bar{f}_s, \quad a, a^* \to a_s, a_s^*,$$
$$\delta(k_l - p_{l+1}) \to \delta_{s_l, t_{l+1}} \cdot \delta(k_l - p_{l+1})$$
$$\int d\mathbf{k}_n \to \sum_{\{s\}} \int d\mathbf{k}_n, \quad \sigma_n(K_n) \to \sigma_n(S_{n1}K_n),$$
$$(a^*, a) \to (a^*, a) = \sum_s \int dk\, a_s^*(k)\, a_s(k), \tag{3.3}$$

we get the pair of Fock representations (CCR and CAR) spanned over $\oplus_1^N \mathcal{L}^2(\mathbb{R}^3) \ni f_s$, and hence with the number N of internal degrees of freedom.

(ii) By virtue of this result and the Haag–LSZ conjecture (compare Sec. 1), we can associate with each quantum field theory (QFT) of the boson system (asymptotic free bosons) the corresponding QFT of the fermion system (asymptotic free fermions). In the relativistic theory when the number of space–time dimensions is equal to two, the above conclusion can be proved in many ways; compare, e.g., Ref. 10. If Minkowski space is taken into account, then because both fermions and bosons have same number of internal degrees of freedom, one of those systems should violate assumptions of the spin-statistics theorem. Hence if the former field is the physical one, the latter can appear as a *subsidiary* (ghost) entity, or conversely.

(iii) On the other hand, if relativistic restrictions can be abandoned, the whole variety of interesting correspondences can be studied. For example, if we consider the low temperature limit of the Heisenberg ferromagnet, it is well known that the free magnon gas (bosons) behaves like the Heisenberg crystal itself. And really we have proved[5] that if the ferromagnet Hamiltonian is H, then there exists the boson (magnons) lattice Hamiltonian H_B and a projection P_0 in the boson Fock space $\mathcal{J}_B$ such that $H = P_0 H_B P_0$ and $P_0 \mathcal{J}_B = \mathcal{J}_0$ is the Hilbert space of spin states of the Heisenberg ferromagnet.

An analogous effect was observed in the macroscopic model of the atomic nuclei (four-fermion interaction) where atomic spectra in weak excitation limit look like those of quadrupole bosons. Here the underlying boson Hamiltonian H_B includes a two-boson interaction term, where each Boson corresponds to the Cooper pair of (starting) fermions.

Suitable modification of Theorem 2 was also used by us to make a transition from boson to fermion variables in the ultralocal quantization attempt for sine–Gordon 1-solitons. (This was the model study of the quantization procedure, where *by starting from the classical level, through the subsidiary boson level, the final physical Fermion level is achieved*).

Theorem 3 (*functional representation of the CAR*): Double power series

$$b(f)(\bar{\alpha}, \alpha) = \sum_{nm} \frac{1}{\sqrt{n!m!}} \int d\mathbf{k}_n \int d\mathbf{p}_m \{\sqrt{n+1}\, \delta_{m,1+n} \sigma_n(\mathbf{k}_n)$$
$$\times \bar{f}(p_1)\, \sigma_{1+n}(\mathbf{p}_{1+n})\, \delta(k_1 - p_2)$$
$$\times \delta(k_2 - p_3) \cdots \delta(k_n - p_{1+n})\}$$
$$\times \bar{\alpha}(k_1) \cdots \bar{\alpha}(k_n)\, \alpha(p_1) \cdots \alpha(p_{1+n})$$
$$= \sum_{nm} \frac{1}{\sqrt{n!m!}} (f_{nm}, \bar{\alpha}^n \alpha^m)\, b(f)^*(\bar{\alpha}, \alpha) = b(\bar{f})(\alpha, \bar{\alpha}) \tag{3.4}$$

play in $\mathbb{1}_F \mathcal{J}_B = \mathcal{J}_F$ the role of generators $b(f)^*$, $b(f)$ respectively of the Fock representation of the CAR algebra:

$$[b(f), b(g)^*]_+(\bar{\alpha}, \alpha) = (\bar{f}, g)\, \mathbb{1}_F(\bar{\alpha}, \alpha),$$
$$\Omega_B = 1, \quad \mathbb{1}_F(\bar{\alpha}, \alpha) = \sum_n \frac{1}{n!} (\bar{\alpha}^n, \sigma_n^2 \alpha^n). \tag{3.5}$$

Proof: The above theorem is a corollary to Theorem 2, and can be proved by making use of Theorem 1 and calculating the functional representation of objects appearing in (3.1), (3.2). It is useful to recall the formula (1.8): $F(\bar{\alpha}, \alpha) \exp(\bar{\alpha}, \alpha) = :F(a^*, a): (\bar{\alpha}, \alpha)$.

The fermion subspace of the Bargmann space is here given by

$$V \in B(\mathcal{L}^2(\mathbb{R}^3)),$$
$$\mathbb{1}_F \left(\bar{\alpha}, \frac{d}{d\bar{\gamma}}\right) V(\bar{\gamma})_{|\bar{\gamma}=0}$$
$$= \int \mathbb{1}_F(\bar{\alpha}, \gamma)\, V(\bar{\gamma})\, d\left(\frac{\gamma}{\sqrt{\pi}}\right)$$
$$= \sum_n \frac{1}{n!} \left(\bar{\alpha}^n, \sigma_n^2 \frac{d^n}{d\bar{\gamma}^n}\right) \sum_n \frac{1}{\sqrt{m!}} (v_n, \bar{\gamma}^n)_{|\bar{\gamma}=0}$$
$$= \sum_n \frac{1}{\sqrt{n!}} (v_n \sigma_n^2, \bar{\alpha}^n), \tag{3.6}$$

and includes vectors received by the Fock construction from symmetric functions $(v_n\sigma_n^2)(\mathbf{k}_n)$, which vanish if any two of variables coincide.

In the Fock construction *there is no difference* between such functions and the antisymmetric functions:

$$\sigma_n(v_n\sigma_n^2)(\mathbf{k}_n) = (v_n\sigma_n)(\mathbf{k}_n),$$
$$(v_n\sigma_n, \sigma_n\bar{\alpha}^n) = (v_n\sigma_n^2, \bar{\alpha}^n). \tag{3.7}$$

Both kinds of them appear in the theory on an equal footing. In this connection compare also Refs. 5, 8, where the study of Hilbert spaces of symmetric and antisymmetric functions is given (together with suitable isometries between them).

4. PROJECTION THEOREMS

Let us consider an arbitrary operator:

$$:F(b^*, b): = \sum_{nm} \frac{1}{\sqrt{n!m!}} (f_{nm}, b^{*n} b^m), \tag{4.1}$$

whose generating triple $\{b^*, b, \Omega_B\}$ is associated with the starting Bose triple $\{a^*, a, \Omega_B\}$, f_{nm} is a totally antisymmetric $(n+m)$-point function (distribution in general). We have:

Lemma 2(*boson expansions*):

$$:F(b^*, b): = :\exp[-(a^*, a)] \circ \sum_{nm} \frac{1}{\sqrt{n!m!}} \sum_k \frac{1}{k!}$$
$$\times (\sigma_{n+k} f_{n\underset{\sim}{m}} \sigma_{m+k}, a^{*\,k+n} a^{k+m}):, \tag{4.2}$$

where $\underset{\sim}{m}$ denotes the reversed order of variables:

$f_{nm}(\mathbf{k}_n, \mathbf{p}_m) = f_{nm}(k_1, \ldots, k_n, p_m, p_{m-1}, \ldots, p_1).$

Proof: Immediate by applying the functional tools. Here, the fermion analog of (1.8) can be easily derived (see Ref. 8):

$$:F(b^*, b):(\bar{\alpha}, \alpha) = \sum_{nm} \frac{1}{\sqrt{n!m!}} \sum_k \frac{1}{k!} \times (\sigma_{n+k} f_{nm} \sigma_{m+k}, \bar{\alpha}^{k+n} \alpha^{k+m}). \quad (4.3)$$

The only difference if compared with the original[8] formula lies in the use of the explicit form σ_{n+k} of the operators E_{n+k} (σ_{n+k} is the *alternating* function).

One can also easily check the following identity:

$$:F(b^*, b): = :\exp[-(a^*, a)] \cdot \overset{c}{F}(a^*, a): = \sum_n \frac{(-1)^n}{n!} (a^{*n}, :\overset{c}{F}(a^*, a): a^n),$$

$$:F(b^*, b): \Omega_B = \sum_{nm} \frac{1}{\sqrt{n!m!}} (\sigma_n f_{nm} \rho \sigma_m, a^{*n} a^m) \Omega_B: = :\overset{c}{F}(a^*, a): \Omega_B, \quad (4.4)$$

suggesting the equivalence relation between $:F(b^*, b):$ and $:\overset{c}{F}(a^*, a):$, where $\overset{c}{f}_{nm} = \sigma_n f_{nm} \sigma_m$ is a symmetric function in groups of variables (m) and (n) respectively, but antisymmetric with respect to permutations from (m) into (n), and converselv.

In connection with (4.4) we have the following:

Theorem 4 (*projection theorem*): Let $\mathbb{1}_F$ be given by (3.2), $\mathcal{J}_F = \mathbb{1}_F \mathcal{J}_B$. The following identity

$$\mathbb{1}_F : \overset{c}{F}(a^*, a): \mathbb{1}_F \mathcal{J}_F = :F(b^*, b): \mathcal{J}_F, \quad (4.5)$$

holds for all operators $:F(b^*, b):$ and $:\overset{c}{F}(a^*, a):$ related by (4.4).

Proof: The study of isometries between Hilbert spaces of symmetric and antisymmetric functions, performed in Ref. 8, results in the basic projection formula:

$$\mathbb{1}_F \mathcal{J}_B = \mathcal{J}_F \ni V = \sum_n \frac{1}{\sqrt{n!}} (\overset{s}{v}_n \sigma_n^2, a^{*n}) \Omega_B = \sum_n \frac{1}{\sqrt{n!}} (\overset{s}{v}_n \sigma_n, b^{*n}) \Omega_B \quad (4.6)$$

so that $\mathbb{1}_F V = V \Rightarrow (\mathbb{1}_F V)(\bar{\alpha}) = V(\bar{\alpha})$. We denote $v_n = \overset{s}{v}_n \sigma_n$, where $\overset{s}{v}_n$ is the n-point, symmetric function and thus v_n is antisymmetric. Here, for all $V \in \mathcal{J}_F$, (4.5) reduces to $\mathbb{1}_F : \overset{c}{F}(a^*, a) : V = :F(b^*, b): V$. [Note that (4.5) is an identity on the whole of $\mathcal{J}_B$]. By (1.8)

$$\mathbb{1}_F(\bar{\alpha}, \alpha) = \sum_n \frac{1}{n!} (\bar{\alpha}^n, \sigma_n^2 \alpha^n),$$

$$:\overset{c}{F}(a^*, a):(\bar{\alpha}, \alpha) = \exp(\bar{\alpha}, \alpha) \cdot \overset{c}{F}(\bar{\alpha}, \alpha) = \exp(\bar{\alpha}, \alpha) \sum_{nm} \frac{1}{\sqrt{n!m!}} (\overset{c}{f}_{nm}, \bar{\alpha}^n \alpha^m) = \sum_{nm} \frac{1}{\sqrt{n!m!}} \sum_k \frac{1}{k!} (\bar{\alpha}^{k+n} \overset{c}{f}_{nm}, \alpha^{k+m}) \quad (4.7)$$

with

$$(\bar{\alpha}^{k+n}, \overset{c}{f}_{nm} \alpha^{k+m}) = \int d\mathbf{q}_k \int d\mathbf{p}_n \int d\mathbf{r}_m \overset{c}{f}_{nm}(\mathbf{p}_n, \mathbf{r}_m) \bar{\alpha}(p_1) \cdots \bar{\alpha}(p_n) \times \alpha(r_1) \cdots \alpha(r_m) \bar{\alpha}(q_1) \alpha(q_1) \cdots \bar{\alpha}(q_k) \alpha(q_k) \quad (4.8)$$

Applying (1.6), we get at once

$$\mathbb{1}_F : \overset{c}{F}(a^*, a) : (\bar{\alpha}, \alpha) = \sum_{nm} \frac{1}{\sqrt{n!m!}} \sum_k \frac{1}{k!} (\bar{\alpha}^{k+n} \sigma_{k+n}^2 \overset{c}{f}_{nm}, \alpha^{k+m}), \quad (4.9)$$

while (4.3) can be written in complete analogy with (4.7):

$$:F(b^*, b):(\bar{\alpha}, \alpha) = \sum_{nm} \frac{1}{\sqrt{n!m!}} \sum_k \frac{1}{k!} (\bar{\alpha}^{k+n} \sigma_{k+n} f_{nm} \sigma_{k+m}, \alpha^{k+m}). \quad (4.10)$$

In consequence (with the use of the identity $v_n \sigma_n^2 = v_n$) we get

$$(:F(b^*, b): V)(\bar{\alpha}) = \sum_{nm} \frac{1}{\sqrt{n!m!}} \sum_k \frac{\sqrt{(k+m)!}}{k!} \times (\bar{\alpha}^{k+n}, \sigma_{k+n} f_{nm} v_{k+m}) \quad (4.11)$$

and

$$(\mathbb{1}_F : \overset{c}{F}(a^*, a) : V)(\bar{\alpha}) = \sum_{nm} \frac{1}{\sqrt{n!m!}} \sum_k \frac{\sqrt{(k+m)!}}{k!} \times (\bar{\alpha}^{k+n}, \sigma_{k+n}^2 \sigma_n f_{nm} \sigma_m \sigma_{k+m} v_{k+m}). \quad (4.12)$$

To make the comparison between (4.11) and (4.12) there is enough to restrict considerations to respective bilinear forms. The integrations symbolized by the sign $(\cdot, \cdot)$ induce a nonzero counterpart only from these functions which are *totally symmetric both in the group of* $(n+k)$ *and* $(m+k)$ *variables and vanish if any two of variables coincide.*

(i) $(\bar{\alpha}^{k+n}, \sigma_{k+n}^2 \sigma_n f_{nm} \sigma_m \sigma_{k+m} v_{k+m})$. The coefficient function integrated with $\bar{\alpha}^{k+n}$, due to the $(n) \overset{a}{\sim} (m)$ symmetry [the change of sign if the variable from the group (n) is permuted with any from (m)], can be decomposed into a sum of irreducible parts with respect to the symmetry group. Denoting $\mathcal{S}(n, m)$ as the symmetrization operator, we indicate the term of interest in explicit fashion:

$$\mathcal{S}(n, m)[\sigma_n \sigma_m f_{nm}] = f_{nm}[\mathcal{A}(n, m) \sigma_n \sigma_m],$$
$$\sigma_n f_{nm} \sigma_m = \mathcal{S}(n, m)[\sigma_n \sigma_m \cdot f_{nm}] + \text{other decomposition terms.} \quad (4.13)$$

Here we have clearly emphasized $[\mathcal{A}(n, m)]$ the fact that symmetrization of the expression is achieved by the antisymmetrization of the product $\sigma_n \sigma_m$.

In this way we have explicitly disclosed the totally symmetric in $(k+n)$ and $(k+m)$ variables function, whose decomposition terms possessing another symmetry are annihilated by the bilinear form:

$$f_{nm} v_{k+m} = \{\sigma_{k+n}^2 \mathcal{A}(n, m)[\sigma_n \sigma_m] \cdot \sigma_{k+m}\} \cdot f_{nm} v_{k+m} + \text{other decomposition terms.} \quad (4.14)$$

(ii) $(\bar{\alpha}^{k+n}, \sigma_{k+n} f_{nm} v_{k+m})$. Repeating arguments of (i) we must select a totally symmetric in variables $(k+n)$ and $(k+m)$ decomposition term of the function $\sigma_{k+n} f_{nm} v_{k+m}$. This can be obviously done by making use of (4.14):

$$\sigma_{k+n} f_{nm} v_{k+m} = \{\sigma_{k+n} \mathcal{A}(n, m)[\sigma_n \sigma_m] \cdot \sigma_{k+m}\} \sigma_{k+n} f_{nm} v_{k+m} + \text{other decomposition terms.} \quad (4.15)$$

The above symmetry analysis clearly shows that though visually the forms (i) and (ii) are different, they clearly coincide by virtue of performed integrations. Hence (4.11), (4.12) coincide also. The theorem is proved.

To complete the above analysis, let us prove one more theorem, concerning the relations

$$\mathbb{1}_F a(f)\mathbb{1}_F \mathcal{J}_F = b(f)\mathcal{J}_F, \quad \mathbb{1}_F a(f)^* \mathbb{1}_F \mathcal{J}_F = b(f)^* \mathcal{J}_F, \tag{4.16}$$

which is the special example satisfying Theorem 4.

Theorem 5 (*projected representation*): Given the Bose triple $\{a^*, a, \Omega_B\}$ and the associated Fermi triple $\{b^*, b, \Omega_B\}$. The CAR hold on $\mathcal{J}_F$ for operators $\mathbb{1}_F a(f)\mathbb{1}_F$ and $\mathbb{1}_F a(f)^*\mathbb{1}_F$. The corresponding representation of the CAR is called the projected representation [notice that formal operator expressions received after normal ordering of $\mathbb{1}_F a(f)\mathbb{1}_F$, $\mathbb{1}_F a(f)^*\mathbb{1}_F$, respectively, are quite different from these for $b(f)$, $b(f)^*$].

Proof: We make use of (4.11), (4.12).

(i) $n=0$, $m=1$ implies

$$[b(f)V](\bar\alpha) = \sum_k \frac{\sqrt{k+1}}{k!}(\bar\alpha^k, \sigma_k \bar f v_{k+1}), \quad [\mathbb{1}_F a(f)V](\bar\alpha) = \sum_k \frac{\sqrt{k+1}}{k!}(\bar\alpha^k, \sigma_k^2 \bar f \sigma_{k+1} v_{k+1}). \tag{4.17}$$

Let us notice that $\sigma_k^2\sigma_{k+1} = \sigma_{k+1}$, so that the second of our bilinear forms reads $(\bar\alpha^k \bar f, \sigma_{k+1} v_{k+1})$.

In the case of $(\bar\alpha^k \bar f, \sigma_k v_{k+1})$ we discover the antisymmetry (change of sign) for permutations $(k)^{\mathfrak{L}}(1)$ so that the only part of the symmetry group decomposition of the product $\bar\alpha^k \bar f$ which does not vanish while integrated with the former function reads $\sigma_{k+1}\sigma_k \bar\alpha^k \bar f$:

$$\bar\alpha^k f = \sigma_{k+1}\sigma_k \bar\alpha^k f + \text{other decomposition terms.}$$

But it means that

$$(\bar\alpha^k \bar f, \sigma_k v_{k+1}) = (\sigma_{k+1}\sigma_k \bar\alpha^k \bar f, \sigma_k v_{k+1}) = (\bar\alpha^k \bar f, \sigma_{k+1}\sigma_k^2 v_{k+1}) = (\bar\alpha^k \bar f, \sigma_{k+1} v_{k+1}),$$

which proves the coincidence of both expressions (4.17).

(ii) $n=1$, $m=0$ implies

$$[b(f)^* V](\bar\alpha) = \sum_k \frac{1}{\sqrt{k!}}(\bar\alpha^{k+1}, \sigma_{k+1} f v_k), \quad [\mathbb{1}_F a(f)^* V](\bar\alpha) = \sum_k \frac{1}{\sqrt{k!}}(\bar\alpha^{k+1}, \sigma_{k+1}^2 f \sigma_k v_k). \tag{4.18}$$

Here the function $\sigma_{k+1}^2 f\sigma_k v_k$ is $(k+1)$-symmetric and appears as a suitable symmetry group decomposition term of the function

$$\sigma_{k+1} f v_k = \{\sigma_{k+1}\sigma_k\}\sigma_{k+1} f v_k + \text{other decomposition terms,} \tag{4.19}$$

which is the only term not annihilated by the bilinear form. The coincidence of both expressions (4.18) is thus immediate. Because identities (4.17), (4.18) hold for all vectors $V \in \mathcal{J}_F$ there is obvious that denoting

$$V'(\bar\alpha) = [b(g)V](\bar\alpha) = (\mathbb{1}_F a(g) V)(\bar\alpha),$$

we get at once

$$[b(f)^* V'](\bar\alpha) = [b(f)^* b(g) V](\bar\alpha) = [\mathbb{1}_F a(f)^* \mathbb{1}_F a(g) V](\bar\alpha) \tag{4.20}$$

and, in an analogous way, with $V''(\bar\alpha) = [b(f)^* V](\bar\alpha) = (\mathbb{1}_F a(f)^* V)(\bar\alpha)$, we get

$$[b(g) V''](\bar\alpha) = [b(g) b(f)^* V](\bar\alpha) = [\mathbb{1}_F a(g)\mathbb{1}_F a(f)^* V](\bar\alpha), \tag{4.21}$$

which by virtue of

$$[b(f)^*, b(g)]_+ = (\bar f, g)\mathbb{1}_F \tag{4.22}$$

trivially implies

$$[b(f)^*, b(g)]_+ \mathcal{J}_F = [\mathbb{1}_F a(f)^* \mathbb{1}_F, \mathbb{1}_F a(g)\mathbb{1}_F]_+ \mathcal{J}_F = (f, \bar g)\mathcal{J}_F, \tag{4.23}$$

proving Theorem 5.

5. DIRAC FIELD: THE CORRESPONDENCE RULE

To get Fock representation of the CAR, suitable for the description of a free Dirac field, we must start from the triples $\{a^*, a, \Omega_B\}_{\oplus_1^4 L^2(\mathbb{R}^3)}$ and $\{b^*, b, \Omega_B\}_{\oplus_1^4 L^2(\mathbb{R}^3)}$ exhibiting the number four of the internal degrees (two charge and two spin degrees) of freedom in the theory. All previous results hold without any change for these representations (see, e.g., Theorem 2 and comments following it).

In the fourth paper of Ref. 8 we have analyzed the standard construction

$$b^+ = \frac{1}{\sqrt2}\begin{bmatrix} b_1^* + ib_3^* \\ b_2^* + ib_4^* \end{bmatrix}, \quad b^- = \frac{1}{\sqrt2}\begin{bmatrix} b_1 + ib_3 \\ b_2 + ib_4 \end{bmatrix}, \quad b^{*+} = \frac{1}{\sqrt2}\begin{bmatrix} b_1^* - ib_3^* \\ b_2^* - ib_4^* \end{bmatrix}, \quad b^{*-} = \frac{1}{\sqrt2}\begin{bmatrix} b_1 - ib_3 \\ b_2 - ib_4 \end{bmatrix} \tag{5.1}$$

(the analogous formulas for boson operators), allowing us to get the quintets: $\{b^\pm, b^{*\pm}, \Omega_B\}_{\oplus_1^2 L^2(\mathbb{R}^3)}$, $\{a^\pm, a^{*\pm}, \Omega_B\}_{\oplus_1^2 L^2(\mathbb{R}^3)}$ with

$$[b^+(f), b^{*-}(g)]_+ = (\bar f, g)\mathbb{1}_F = [b^-(f), b^{*+}(g)]_+; \tag{5.2}$$

the other anticommutators vanish.

On the level of functional representations in the place of $\alpha, \bar\alpha \in \oplus_1^4 L^2(\mathbb{R}^3)$, we introduce the new Fourier amplitudes $\alpha, \alpha^*, \beta, \beta^* \oplus_1^2 L^2(\mathbb{R}^3)$, so that

$$\begin{aligned}
&(\bar\alpha, \alpha) = (\alpha, \alpha^*) + (\beta, \beta^*) \quad \text{and} \quad f \in \oplus_1^2 L^2(\mathbb{R}^3),\\
&a^+(f)(\bar\alpha, \alpha) = (\alpha, \bar f)\exp[(\alpha, \alpha^*) + (\beta, \beta^*)],\\
&a^{*+}(f)(\bar\alpha, \alpha) = (\alpha^*, f)\exp[(\alpha, \alpha^*) + (\beta, \beta^*)],\\
&a^-(f)(\bar\alpha, \alpha) = (\beta, \bar f)\exp[(\alpha, \alpha^*) + (\beta, \beta^*)],\\
&a^{*-}(f)(\bar\alpha, \alpha) = (\beta^*, f)\exp[(\alpha, \alpha^*) + (\beta, \beta^*)].
\end{aligned} \tag{5.3}$$

Functional differentiations with respect to $\alpha, \bar\alpha$ can be apparently translated to the language of $\alpha, \alpha^*, \beta, \beta^*$ according to : $k=1,2$, $(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}) := \alpha$,

$$\frac{d}{d\alpha_{1k}} = \frac{1}{\sqrt2}\left(\frac{d}{d\alpha_k} + \frac{d}{d\beta_k^*}\right), \quad \frac{d}{d\bar\alpha_{1k}} = \frac{1}{\sqrt2}\left(\frac{d}{d\beta_k} + \frac{d}{d\alpha_k^*}\right),$$

$$\frac{d}{d\alpha_{2k}}=\frac{i}{\sqrt{2}}\quad\frac{d}{d\alpha_k}-\frac{d}{d\beta_k^*}\quad,\quad\frac{d}{d\overline{\alpha}_{2k}}=\frac{i}{\sqrt{2}}\quad\frac{d}{d\beta_k}-\frac{d}{d\alpha^*}\,.\tag{5.4}$$

Again, in close analogy to (1.8), any normal ordered in the a^+, a^{*+}, a^-, a^{*-} operator expression,

$$:F(a^+, a^{*+}, a^-, a^{*-}):=\sum_{nmkl}\frac{1}{\sqrt{n!m!k!l!}}\times(f_{nmkl}, a^{+n}a^{*+m}a^{-k}a^{*-l}),\tag{5.5}$$

admits a straightforward functional representation:

$$:F(a^+, a^{*+}, a^-, a^{*-}):(\overline{\alpha}, \alpha)=F(\alpha, \alpha^*, \beta, \beta^*)\exp(\alpha, \alpha),\tag{5.6}$$

where classical Fourier amplitudes $\alpha, \alpha^*, \beta, \beta^*$ appear in the place of boson operators.

Let us extend the *Haag–LSZ expansion conjecture* to the case of the Dirac field algebra ($\psi, \overline{\psi}$ are asymptotic free Dirac fields):

$$\begin{aligned}:\Omega(\psi, \overline{\psi}):&=\sum_{nm}\frac{1}{n!m!}(\omega_{nm}, :\psi^n\overline{\psi}^m:)\\&=\sum_{nm}\frac{1}{n!m!}\sum_{\sigma\tau}\quad d\mathbf{x}_n\quad\omega_{nm}^{\sigma\tau}(\mathbf{x}_n, \mathbf{y}_m):\psi_{\sigma_1}(x_1)\\&\times\cdots\psi_{\sigma_n}(x_n)\overline{\psi}_{\tau_n}(y_1)\cdots\overline{\psi}_{\tau_m}(y_m):.\end{aligned}\tag{5.7}$$

σ, τ are bispinor indices and the overbar denotes Dirac conjugation of bispinors.

It was proved in Ref. 8 that by the use of functional representations of the CCR and CAR the operator $:\Omega(\psi, \overline{\psi}):$ admits a straightforward c-number image:

$$:\Omega(\psi, \overline{\psi}):(\overline{\alpha}, \alpha)=\sum_{nm}\frac{1}{\sqrt{n!m!}}(s_{nm}, \sigma_n\overline{\alpha}^n\sigma_m\alpha^m)=\overset{c}{S}(\overline{\alpha}, \alpha),\tag{5.8}$$

with a suitable (rather involved function of ω_{kl}) coefficient function $s_{nm}^{\mu\nu}(\mathbf{k}_n, \mathbf{p}_m)$, $\mu, \nu=1, 2, 3, 4$, denoting vector indices in $\oplus_1^4 \mathcal{L}^2(\mathbb{R}^3)$.

Unfortunately, this c-number image of $:\Omega(\psi, \overline{\psi}):$ cannot be related so simply as in scalar case, with functional power series of *classical fermion fields*.

This seems to be a disadvantage of (5.8) if we compare it to a canonical classical-like image being based on the use of Grassman algebras (see also the fourth paper of Ref. 8). In this last case, one can satisfactorily reproduce operator identities on the functional-like level (though not in the language of ordinary c-number functionals). We wish now to remove this difficulty, and to find the functional power series of classical spinor (Dirac) fields, being in the correspondence relation with the starting operator series $:\Omega(\psi, \overline{\psi}):$.

Theorem 6 (*the correspondence rule*): For each operator series $:\Omega(\psi, \overline{\psi})$ one can find the functional power series $\overset{c}{\Omega}(\overset{c}{\psi}, \overset{c}{\overline{\psi}})$ with respect to classical free Dirac fields $\overset{c}{\psi}, \overset{c}{\overline{\psi}}$, such that:

$$\text{(i)}\ :\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):(\overline{\alpha}, \alpha)=\overset{c}{\Omega}(\overset{c}{\psi}, \overset{c}{\overline{\psi}})\exp(\overline{\alpha}, \alpha),\tag{5.9}$$

where $\overset{B}{\psi}, \overset{B}{\overline{\psi}}$ are the subsidiary Dirac fields obeying (the thus improper) bose statistics, and

$$\text{(ii)}\ :\Omega(\psi, \overline{\psi}):\mathcal{J}_F=\mathbb{1}_F:\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\mathbb{1}_F\mathcal{J}_F.\tag{5.10}$$

The set of all functionals $\overset{c}{\Omega}(\overset{c}{\psi}, \overset{c}{\overline{\psi}})$ may stand for an exact classical image of the former set of operators $:\Omega(\psi, \overline{\psi}):$ realized via the mediation of the subsidiary boson level.

Proof: $:\Omega(\psi, \overline{\psi}):$ can be written in the following form, manifestly exhibiting the normal ordering of operators (below, the total antisymmetry of ω_{nm} in $n+m$ variables is essential):

$$\begin{aligned}:\Omega(\psi, \overline{\psi}):&=\sum_{nm}\frac{1}{n!m!}(\omega_{nm}, :(\psi^++\psi^-)^n(\overline{\psi}^++\overline{\psi}^-)^m:)\\&=\sum_{nm}\frac{1}{n!m!}\left(\omega_{nm}, :\sum_k^n\binom{n}{k}\psi^{+k}(\psi^-)^{n-k}\sum_l^m\right.\\&\quad\left.\times\binom{m}{l}\overline{\psi}^{+l}(\overline{\psi}^-)^{m-l}:\right)\\&=\sum_{nmkl}\frac{1}{n!m!k!l!}\sum_{\mu\nu\sigma\rho}\int d\mathbf{x}_n\int d\mathbf{y}_m\int d\mathbf{z}_k\int d\mathbf{u}_l\\&\quad\times\omega_{nmkl}^{\mu\nu\rho\sigma}(\mathbf{x}_n, \mathbf{y}_m, \mathbf{z}_k, \mathbf{u}_l)\psi_{\mu_1}^+(x_1)\cdots\psi_{\mu_m}^+(x_m)\\&\quad\times\overline{\psi}_{\nu_1}^+(y_1)\cdots\overline{\psi}_{\nu_m}^+(y_m)\psi_{\rho_1}^-(z_1)\cdots\psi_{\rho_k}^-(z_k)\\&\quad\times\overline{\psi}_{\sigma_1}^-(u_1)\cdots\overline{\psi}_{\sigma_l}^-(u_l)\\&=\sum_{nmkl}\frac{1}{n!m!k!l!}(\omega_{n+k, m+l}, \psi^{+n}(\overline{\psi}^+)^m(\psi^-)^k(\overline{\psi}^-)^l),\end{aligned}\tag{5.11}$$

where the operators $\psi^{\pm}, \overline{\psi}^{\pm}$ depend linearly (through Fourier transformations) on the Fermi operators $b^{\pm}, b^{*\pm}$ defined by (5.1):

$$[b_i^+(k), b_j^{*-}(p)]_+=\delta_{ij}\delta(k-p)\,\mathbb{1}_F=[b_i^-(k), b_j^{*+}(p)]_+.\tag{5.12}$$

The other anticommutators vanish. Indices i, j denote here helicity states $i, j=1, 2$ in contrast to bispinor indices μ, ν. In (5.11) we have clearly distinguished two groups of operators: $\psi^{+n}(\overline{\psi}^+)^m$ and $\psi^{-k}(\overline{\psi}^-)^l$, which involve, by (5.1) the $(n+m)$-point product of b^*'s and $(k+l)$-point product of b's respectively.

The validity of Theorem 5 is here immediate (compare also Comments to Theorem 2) so that

$$\begin{aligned}&b_{i_1}^+(k_1)\cdots b_{i_n}^+(k_n)b_{j_1}^{*+}(p_1)\cdots b_{j_m}^{*+}(p_m)b_{s_1}^-(q_1)\cdots b_{s_k}^-(q_k)\\&\quad\times b_{t_1}^{*-}(r_1)\cdots b_{t_l}^{*-}(r_l)\mathcal{J}_F\\&\quad\triangleq\sigma_{n+m}(\mathbf{k}_n, \mathbf{p}_m)\,\underset{\sim}{\sigma_{k+l}}(\mathbf{q}_k, \mathbf{r}_l)\mathbb{1}_F a_{i_1}^*(k_1)\cdots a_{i_n}^*(k_n)\\&\quad\times a_{j_1}^{*+}(p_1)\cdots a_{j_m}^{*+}(p_m)a_{s_1}^-(q_1)\cdots a_{s_k}^-(q_k)\\&\quad\times a_{t_1}^{*-}(r_1)\cdots a_{t_l}^{*-}(r_l)\mathbb{1}_F\mathcal{J}_F,\end{aligned}\tag{5.13}$$

where $\triangleq$ means that the identity holds true only if integrated from both sides over all variables, with the suitable (antisymmetric) $(n+m+k+l)$-point function, $\underset{\sim}{\sigma_{k+l}}(\mathbf{q}_k, \mathbf{r}_l)=\sigma_{k+l}(r_l, \ldots, r_1, q_k, \ldots, q_1)$, i.e., the tilde reverses the order of variables.

The operators $a_j^{\pm}(k)$, $a_j^{*\pm}(k)$ stand here for operators of the *ideal*, *fictitious*, *subsidiary bosons*, constituting the mediating level in the transition from (5.9) to (5.10).

Here obviously the fermion Fock space $\mathcal{J}_F$ appears as a subspace $\mathbb{1}_F \mathcal{J}_B$ of the boson Fock space $\mathcal{J}_B$. They are representation spaces for triples $\{b^*, b, \Omega_B\}_{\oplus_1^4 \,\mathcal{L}^2(\mathbb{R}^3)}$, $\{a^*, a, \Omega_B\}_{\oplus_1^4 \,\mathcal{L}^2(\mathbb{R}^3)}$ respectively.

Let us now restrict consideration to the two-point product $\psi_\mu^+(x)\psi_\nu^+(y)$, where we immediately get

$$\psi_\mu^+(x)\psi_\nu^+(y)\,\mathcal{J}_F$$
$$\overset{a}{=} (1/2\pi)^3 \int dk(\sqrt{2\omega_k})^{-1} \int dp(\sqrt{2\omega_p})^{-1} \sum_{ij} v_\mu^{+i}(k) v_\nu^{+j}(p)$$
$$\times \exp[i(kx+py)]\cdot \sigma_2(k,p)\,\mathbb{1}_F\, a_i^+(k)\, a_j^+(p)\,\mathbb{1}_F \mathcal{J}_F. \qquad (5.14)$$

Here again $\overset{a}{=}$ means the validity of (5.14) only after smearing with an antisymmetric two-point test function. Here, by the use of four-dimensional Fourier transformations we can introduce the sign operator $\mathcal{E}_2$, with the integral kernel:

$$\mathcal{E}_2(x'-x, y'-y) = \frac{1}{(2\pi)^4} \int dq \int dr\, \sigma_2(q,r) \exp(-iqx - iry)$$
$$\times \exp[i(qx' + ry')], \qquad (5.15)$$

where

$$x, y \in M^4, \quad q = (\mathbf{q}, q_0), \quad \sigma_2(q,r)\big|_{\substack{q_0=\omega_k\\ p_0=\omega_p}} = \sigma_2(\mathbf{q}, \mathbf{r}), \quad \mathbf{q}, \mathbf{r} \in \mathbb{R}^3.$$

Now, (5.14) reads

$$\psi_\mu^+(x')\psi_\nu^+(y')\,\mathcal{J}_F$$
$$\overset{a}{=} \frac{1}{(2\pi)^4} \int dx \int dy\, \mathcal{E}_2(x'-x, y'-y)\, \frac{1}{(2\pi)^3} \int dk(\sqrt{2\omega_k})^{-1}$$
$$\times \int dp(\sqrt{2\omega_p})^{-1} \sum_{ij} v_\mu^{+i}(k)\, v_\nu^{+j}(p)$$
$$\times \exp[i(kx+py)]\,\mathbb{1}_F\, a_i^+(k)\, a_j^+(p)\,\mathbb{1}_F \mathcal{J}_F$$
$$\overset{a}{=} \frac{1}{(2\pi)^4} \int dx \int dy\, \mathcal{E}_2(x'-x, y'-y)$$
$$\times \mathbb{1}_F \overset{B}{\psi}_\mu(x) \overset{B}{\psi}_\nu(y)\,\mathbb{1}_F \mathcal{J}_F$$
$$\overset{a}{=} \mathbb{1}_F(\mathcal{E}_2 \overset{B}{\psi}_\mu \overset{B}{\psi}_\nu)(x', y')\,\mathbb{1}_F \mathcal{J}_F, \qquad (5.16)$$

where the superscript B means that $\overset{B}{\psi^\pm}, \overset{B}{\overline{\psi}^\pm}$ appear as positive and negative frequency parts of fictitious (as violating the spin-statistics theorem) spinor fields in which Fermi operators $b^\pm, b^{*\pm}$ are replaced by boson operators $a^\pm, a^{*\pm}$ of the associated boson representation.

The generalization of (5.16) is obvious, leading thus to the identity:

$$\psi_{\mu_1}^+(x_1)\cdots\psi_{\mu_m}^+(x_m)\,\overline{\psi}_{\nu_1}^+(y_1)\cdots\overline{\psi}_{\nu_m}^+(y_m)$$
$$\times \psi_{\rho_1}^-(z_1)\cdots\psi_{\rho_k}^-(z_k)\,\overline{\psi}_{\sigma_1}^-(u_1)\cdots\psi_{\sigma_l}^-(u_l)\,\mathcal{J}_F$$
$$\overset{a}{=} \mathbb{1}_F(\mathcal{E}_{n+m} \overset{B}{\psi}{}_{\mu_1}^+ \cdots \overset{B}{\psi}{}_{\mu_m}^+ \overset{B}{\overline{\psi}}{}_{\nu_1}^+ \cdots \overset{B}{\overline{\psi}}{}_{\nu_m}^+ \overset{B}{\psi}{}_{\rho_1}^- \cdots \overset{B}{\psi}{}_{\rho_k}^- \overset{B}{\overline{\psi}}{}_{\sigma_1}^- \cdots$$
$$\times \overset{B}{\overline{\psi}}{}_{\sigma_l}^- \underset{\sim}{\mathcal{E}_{k+l}})(\mathbf{x}_n, \mathbf{y}_m, \mathbf{z}_k, \mathbf{u}_l)\,\mathbb{1}_F \mathcal{J}_F, \qquad (5.17)$$

where the undertilde means that the order of the $k+l$ variables is reversed, $(z_1, \ldots, z_k, u_1, \ldots, u_l) \to (u_l, \ldots, u_1, z_k, \ldots, z_1)$. By virtue of (5.17) we get at once the required equivalence formula (5.9):

$$:\Omega(\psi, \overline{\psi}):\mathcal{J}_F = \mathbb{1}_F :\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\mathbb{1}_F \mathcal{J}_F$$
$$= \sum_{nm} \frac{1}{n!\,m!}\,(\omega_{nm}\mathcal{E}_n \underset{\sim}{\mathcal{E}_m},\ \mathbb{1}_F :\overset{B}{\psi}{}^n \overset{B}{\overline{\psi}}{}^m: \mathbb{1}_F)\,\mathcal{J}_F. \qquad (5.18)$$

Here the notation $\overset{c}{\omega}_{nm} = \omega_{nm}\mathcal{E}_n \underset{\sim}{\mathcal{E}_m}$ is used. We have proved that, with each Fermi field algebra, one can associate a projection of the subsidiary (mediating) Bose field algebra, so that on $\mathcal{J}_F$ both algebras coincide. On the (not projected) boson level, we have trivially realized (5.10) as a consequence of (5.3)–(5.6), so that with each operator $:\Omega(\psi, \overline{\psi}):$ we have finally associated the functional

$$\overset{c}{\Omega}(\overset{c}{\psi}, \overset{c}{\overline{\psi}}) = :\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):(\overline{\alpha}, \alpha)\cdot\exp[-(\overline{\alpha}, \alpha)], \qquad (5.19)$$

depending on classical spinor fields $\overset{c}{\psi}, \overset{c}{\overline{\psi}}$ differing from $\psi, \overline{\psi}$ by the replacement of operators $b^\pm, b^{*\pm}$ by classical amplitudes [see (5.3)] $\alpha, \beta, \alpha^*, \beta^*$ respectively. The theorem is proved.

Comment: (i) As a consequence of Theorem 6, there is enough to start from the set of functionals $\{\overset{c}{\Omega}(\overset{c}{\psi}, \overset{c}{\overline{\psi}})\}$ to get a functional representation $\{:\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):(\overline{\alpha}, \alpha)\}$ of the set of boson operators $\{:\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\}$, whose projection $\{\mathbb{1}_F :\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\mathbb{1}_F\}$ on the Fock space $\mathcal{J}_F$ is equivalent to the pure Fermi set $\{:\Omega(\psi, \overline{\psi}):\}$. This sequence of steps allows to state the question of *quantization* of classical spinor fields.

(ii) Note that operators $:\Omega(\psi, \overline{\psi}):$ and $\mathbb{1}_F :\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\mathbb{1}_F$ have the same matrix elements if calculated between arbitrary states from $\mathcal{J}_B$:

$$(m\,|:\Omega(\psi, \overline{\psi}):|\,n) = (m\,|\,\mathbb{1}_F :\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\mathbb{1}_F|\,n),$$

where $|m), |n) \in \mathcal{J}_B$.

6. THE QUESTION OF ALGEBRAIC STRUCTURE

As was emphasized in the discussion of scalar fields, the operator multiplication on the quantum level, via the correspondence rule, results in the multiplication (*) on the classical level; see, e.g., (1.9).

In the case of Dirac fields, the situation is not so obvious, because classical spinors by *no reason can account for the Pauli exclusion principle*. The appearance of it on the quantum level should involve serious restrictions on the classical level.

(i) Let us recall that the set of operators $\{:\Omega(\psi, \overline{\psi}):\}$ is in the equivalence relation on $\mathcal{J}_F$ with the reduction $\{\mathbb{1}_F :\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\mathbb{1}_F\}$ of the set $\{:\overset{c}{\Omega}(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):\}$ of operators belonging to the (subsidiary) boson field algebra. By the use of functional representation of the CCR, one can easily reproduce a corresponding (*) operation [see, e.g., (3.7)]:

$$[:\overset{c}{\Omega}_1(\overset{B}{\psi}, \overset{B}{\overline{\psi}}): \ :\overset{c}{\Omega}_2(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):](\overline{\alpha}, \alpha)$$
$$= \{\overset{c}{\Omega}_1(\overset{c}{\psi}, \overset{c}{\psi})(*)\,\overset{c}{\Omega}_2(\psi_c, \overline{\psi}_c)\}\exp(\overline{\alpha}, \alpha), \qquad (6.1)$$

$$(*) = \exp\left(\frac{\overleftarrow{d}}{d\alpha}, \frac{\overrightarrow{d}}{d\overline{\alpha}}\right)$$
$$= \exp\left[\left(\frac{\overleftarrow{d}}{d\alpha^*}, \frac{\overrightarrow{d}}{d\alpha}\right) + \left(\frac{\overleftarrow{d}}{d\beta}, \frac{\overrightarrow{d}}{d\beta^*}\right)\right]$$
$$= \exp\left[-i\left(\frac{\overleftarrow{d}}{d\overset{c}{\overline{\psi}}} G^+ \frac{\overrightarrow{d}}{d\overset{c}{\psi}}\right) - i\left(\frac{\overleftarrow{d}}{d\overset{c}{\psi}} G^- \frac{\overrightarrow{d}}{d\overset{c}{\overline{\psi}}}\right)\right]. \qquad (6.2)$$

We have exploited here the fact that

$$\overset{c}{\psi}_\sigma(x) = \overset{c}{\psi}_\sigma(x, \alpha, \beta)$$
$$= \int dk \sum_j \{v^+_{j\sigma}(k, x)\,\alpha_j(x) + v^-_{j\sigma}(k, x)\,\beta_j(k)\} \qquad (6.3)$$

and integrals over products of v's allow us to get Green's functions of the Dirac equation, $G^\pm$, respectively:

$$\left(\frac{\overleftarrow{d}}{d\overset{c}{\psi}} G^+ \frac{\overrightarrow{d}}{d\overset{c}{\overline{\psi}}}\right) = \sum_{\sigma\tau} \int dx \int dy \frac{\overleftarrow{d}}{d\overset{c}{\psi}_\sigma(x)} G^+_{\sigma\tau}(x-y) \frac{\overrightarrow{d}}{d\overset{c}{\overline{\psi}}_\tau(y)}\,; \qquad (6.4)$$

arrows indicate the direction in which differential operators act. In formulas above $\overset{c}{\psi}, \overset{c}{\overline{\psi}}$ are classical functions (the commuting ring).

(ii) If we follow the Grassman methods[13] (anticommuting ring of spinors), formulas, nearly identical with (6.1)–(6.4) appear:

$$[:\Omega_1(\psi,\overline{\psi}): \,:\Omega_2(\psi,\overline{\psi}):](\overline{\alpha},\alpha) = \exp(\overline{\alpha},\alpha)\{\Omega_1(\psi,\overline{\psi})(*)\,\Omega_2(\psi,\overline{\psi})\}, \qquad (6.5)$$

with (*) given by (6.2). However, here $\overline{\alpha}, \alpha$ belong to the Grassman algebra, so that we deal with the functional-like representation (see, e.g., Ref. 8) of the CAR, formally coinciding (*in form*) with the functional representation of the CCR. On the lhs of (6.5) $\psi, \overline{\psi}$ are Fermi fields, while on the rhs there are functions from the anticommuting ring. Obviously functional expansion coefficients ω_{nm} in (6.5) are totally antisymmetric, while in (6.1) we have dealt with $\mathcal{E}_n\mathcal{E}_m\omega_{nm}$. Obviously, (6.5) can be rewritten with the use of functional-like (measures on Grassman algebras) integrals. We prefer however the differential way, as significantly simpler and easier to work with (notice an analogy of functional power series with power series of complex variables).

(iii) The relations between expansion coefficients of operators $:\Omega_1(\psi,\overline{\psi}):$ and $:\Omega_2(\psi,\overline{\psi}):$ following from their multiplications are well reproduced by (6.5). One may, however, proceed along less formal, though unfortunately not so elegant here, c-number way of previous sections. Here

$$:\Omega_1(\psi,\overline{\psi}): \,:\Omega_2(\psi,\overline{\psi}): \mathcal{J}_F$$
$$= :\Omega_{12}(\psi,\overline{\psi}): \mathcal{J}_F = \mathbb{1}_F :\Omega_{12}(\psi,\overline{\psi}): \mathbb{1}_F\,\mathcal{J}_F$$
$$= \mathbb{1}_F : \overset{c}{\Omega}_1(\overset{B}{\psi}, \overset{B}{\overline{\psi}}): \mathbb{1}_F : \overset{c}{\Omega}_2(\overset{B}{\psi}, \overset{B}{\overline{\psi}}): \mathbb{1}_F\,\mathcal{J}_F \qquad (6.6)$$

so that, by Theorem 6,

$$\overset{c}{\Omega}_{12}(\overset{c}{\psi}, \overset{c}{\overline{\psi}}) = [: \overset{c}{\Omega}_1(\overset{B}{\psi}, \overset{B}{\overline{\psi}}): \mathbb{1}_F : \overset{c}{\Omega}_2(\overset{B}{\psi}, \overset{B}{\overline{\psi}}):](\overline{\alpha}, \alpha)\cdot\exp[-(\overline{\alpha},\alpha)] \qquad (6.7)$$

would establish the required translation of quantum mutliplication rule (fermions) into the classical language. Because the representations of the CCR were defined with respect to primary Fourier amplitudes $\alpha, \overline{\alpha} \in \oplus_1^4 \mathcal{L}^2(\mathbb{R}^3)$, we indicate the possibility of suitable reordering of summations and integrations, writing

$$\overset{c}{\Omega}(\overset{c}{\psi}, \overset{c}{\overline{\psi}}) = F(\overline{\alpha}, \alpha) = \sum_{nm} \frac{1}{\sqrt{n!\,m!}}(f_{nm}, \overline{\alpha}^n\alpha^m). \qquad (6.8)$$

Now

$$\overset{c}{\Omega}_{12}(\overset{c}{\psi}, \overset{c}{\overline{\psi}}) = F_{12}(\overline{\alpha}, \alpha)$$
$$= [:F_1(a^*, a): \mathbb{1}_F : F_2(a^*, a):](\overline{\alpha}, \alpha)\exp[-(\overline{\alpha}, \alpha)]$$
$$= F_1(\overline{\alpha}, \alpha)(*)\,F_2(\overline{\alpha}, \alpha), \qquad (6.9)$$

where

$$(*) = \exp\left(\frac{\overleftarrow{d}}{d\alpha}, \frac{\overrightarrow{d}}{d\overline{\gamma}}\right) 1_F(\overline{\gamma}, \gamma)\exp\left(\frac{\overleftarrow{d}}{d\gamma}, \frac{\overrightarrow{d}}{d\overline{\alpha}}\right)\Big|_{\substack{\gamma=\alpha\\ \overline{\gamma}=\overline{\alpha}}} \qquad (6.10)$$

(after performing all differentiations one puts $\alpha = \gamma$, $\overline{\alpha} = \overline{\gamma}$) and

$$1_F(\overline{\gamma}, \gamma) = \exp[-(\overline{\gamma}, \gamma)]\cdot \mathbb{1}_F(\overline{\gamma}, \gamma)$$
$$= \sum_n \frac{1}{n!}\sum_k \frac{(-1)^k}{k!}(\overline{\gamma}^{k+n}, \sigma_n^2\gamma^{k+n}). \qquad (6.11)$$

We have been not able to find any sensible representation of (6.9) in terms of pure c-number functions $\overset{c}{\psi}, \overset{c}{\overline{\psi}}$, and thus not in terms of amplitudes $\alpha, \beta \in \oplus_1^4 \mathcal{L}^2(\mathbb{R}^3)$. However, the formal rules (6.5) can be used as a *complementary tool*, satisfactorily reflecting relations between expansion coefficients, which follow from (6.9), and then allow us to define a c-number functional $\overset{c}{\Omega}_{12}(\overset{c}{\psi}, \overset{c}{\overline{\psi}})$ while starting from the Grassman functional $\Omega_{12}(\psi, \overline{\psi})$: $\omega^{12}_{nm} \to \mathcal{E}_n\mathcal{E}_m\omega^{12}_{nm}$.

7. QUANTIZATION OF DIRAC FIELD

In the case of the scalar field, having given an asymptotic free field $\overset{c}{\varphi}$, we could define sets of operators (functionals respectively) $:\Omega(\varphi):$, $\overset{c}{\Omega}(\overset{c}{\varphi})$.

In the case of the Dirac fields $\psi, \overline{\psi}$ we map $:\Omega(\psi, \overline{\psi}):$ onto a classical level through the mediation of the subsidiary boson level. However, this boson level itself allows us to consider its own classical map consisting from the set $\mathcal{S}$ of all functionals with respect to $\overset{c}{\psi}, \overset{c}{\overline{\psi}}$ whose expansion coefficients ω_{nm} are totally $(n+m)$-symmetric:

$$\overset{s}{\Omega}(\overset{c}{\psi}, \overset{c}{\overline{\psi}}) = \sum_{nm}\frac{1}{n!\,m!}(\overset{s}{\omega}_{nm}, \overset{c}{\psi}{}^n\overset{c}{\overline{\psi}}{}^m). \qquad (7.1)$$

In the quantization attempts of any classical spinor field theory one starts from functionals $\overset{s}{\Omega}$ rather than from $\overset{c}{\Omega}$. At first we must have a reduction tool allowing us to transform $\mathcal{S}$ into the set $\mathcal{S}_0$ of functionals $\overset{c}{\Omega}$, which are the only ones of interest if it is required that the Fermi level be achieved.

Lemma 3: There exists the reduction operator P_0 on $\mathcal{S}$, such that

$$P_0\mathcal{S} = \mathcal{S}_0.$$

Proof: We shall introduce into our considerations the following functional:

$$P_0(\overset{c}{\psi},\overset{c}{\bar\psi})=\sum_{nm}\frac{1}{n!m!}(\tilde{\mathcal{E}}_n\tilde{\mathcal{E}}_{\underset{\sim}{m}}\tilde{\mathcal{E}}_{n+m}\overset{c}{\psi}{}^n\overset{c}{\bar\psi}{}^m,\overset{c}{\psi}{}^n\overset{c}{\bar\psi}{}^m$$
$$=\sum_{nm}\frac{1}{n!m!}(\overset{c}{\psi}{}^n\overset{c}{\bar\psi}{}^m,\mathcal{E}_{n+m}\mathcal{E}_{\underset{\sim}{m}}\mathcal{E}_n\overset{c}{\psi}{}^n\overset{c}{\bar\psi}{}^m). \tag{7.2}$$

If we consider P_0 as an operator in $\mathcal{S}$, acting according to the following (functional) rule,

$$(P_0\overset{s}{\Omega})(\overset{c}{\psi},\overset{c}{\bar\psi})=\sum_{nm}\frac{1}{n!m!}\left(\overset{c}{\psi}{}^n\overset{c}{\bar\psi}{}^m,\mathcal{E}_{n+m}\mathcal{E}_{\underset{\sim}{m}}\mathcal{E}_n\frac{d^n}{d\overset{c}{\bar\varphi}{}^n}\frac{d^m}{d\overset{c}{\varphi}{}^m}\right)$$
$$\times\overset{s}{\Omega}(\overset{c}{\varphi},\overset{c}{\bar\varphi})_{|\overset{c}{\varphi}=0=\overset{c}{\bar\varphi}}$$
$$=\sum_{nm}\frac{1}{n!m!}(\mathcal{E}_n\mathcal{E}_{\underset{\sim}{m}}\mathcal{E}_{n+m}\overset{s}{\omega}_{nm},\overset{c}{\psi}{}^n\overset{c}{\bar\psi}{}^m)=\overset{c}{\Omega}(\overset{c}{\psi},\overset{c}{\bar\psi}), \tag{7.3}$$

where $\overset{c}{\omega}_{nm}=\mathcal{E}_{\underset{\sim}{m}}\mathcal{E}_n\mathcal{E}_{n+m}\overset{s}{\omega}_{nm}=\mathcal{E}_{\underset{\sim}{m}}\mathcal{E}_n\omega_{nm}$, and as possessing the expected symmetry properties, then ω_{nm} is totally $(n+m)$-antisymmetric. The lemma is proved. With this selection tool, we can formulate:

Theorem 7 (*quantization rule*): Given the set $\mathcal{S}$ of functionals $\overset{s}{\Omega}(\overset{c}{\psi},\overset{c}{\bar\psi})$, then $P_0\mathcal{S}=\mathcal{S}_0$, if equipped with the algebraic structure (6.9), allows us the quantization map

$$\overset{c}{\Omega}(\overset{c}{\psi},\overset{c}{\bar\psi})\to:\overset{c}{\Omega}(\overset{B}{\psi},\overset{B}{\bar\psi}):\Rightarrow \mathbb{1}_F:\overset{c}{\Omega}(\overset{B}{\psi},\overset{B}{\bar\psi}):\mathbb{1}_F\mathcal{J}_F=:\Omega(\psi,\bar\psi):\mathcal{J}_F, \tag{7.4}$$

connecting with each element $\overset{c}{\Omega}$ of $\mathcal{S}_0$ the corresponding element $:\Omega(\psi,\bar\psi):$ of the Fermi field algebra. The converse map is realized by the correspondence rule of Theorem 6.

Proof: Repeats in fact arguments of Theorem 6.

Theorems 6 and 7, combined together, form a *correspondence principle* for Dirac fields.

8. ON GENERATING FUNCTIONALS FOR THE GREEN'S FUNCTIONS

The commonly used functionals (2.1) are based on Grassman concepts. Let us consider the functional of the same form:

$$Z(\eta^c,\overset{c}{\bar\eta})=\frac{\int\exp\{i[\overset{c}{S}+\int(\overset{c}{\bar\eta}\overset{c}{\psi}+\overset{c}{\bar\eta}\overset{c}{\psi})dx]\}\cdot d(M\overset{c}{\psi}/\sqrt{i\pi})}{\int\exp(i\overset{c}{S})\,d(M\overset{c}{\psi}/\sqrt{i\pi})}, \tag{8.1}$$

with the only difference lying in the replacement of Grassman objects by corresponding c-numbers (commuting ring) $\overset{c}{\eta},\overset{c}{\bar\eta},\overset{c}{\psi},\overset{c}{\bar\psi}$, $d(M\overset{c}{\psi}/\sqrt{i\pi})$. Functionals of the form $Z(\overset{c}{\eta},\overset{c}{\bar\eta})$ play the role played in the previous section by $\overset{s}{\Omega}$.

Let us introduce the following *reduction of* $Z(\overset{c}{\eta},\overset{c}{\bar\eta})$: $Z_0(\overset{c}{\eta},\overset{c}{\bar\eta})=(P_0Z)(\overset{c}{\eta},\overset{c}{\bar\eta})$

$$=\sum_{nm}\frac{1}{n!m!}\left(\tilde{\mathcal{E}}_n\tilde{\mathcal{E}}_{\underset{\sim}{m}}\tilde{\mathcal{E}}_{n+m}\overset{c}{\eta}{}^n\overset{c}{\bar\eta}{}^m,\frac{d^n}{d\overset{c}{\varphi}{}^n}\frac{d^m}{d\overset{c}{\bar\varphi}{}^m}\right)$$
$$\times Z(\overset{c}{\varphi},\overset{c}{\bar\varphi})_{|\overset{c}{\varphi}=\overset{c}{\bar\varphi}=0}. \tag{8.2}$$

For the general case, the reduction formula (8.2) does not look too attractive. Let us see, however, what happens in the free field case, when

$$Z(\overset{c}{\eta},\overset{c}{\bar\eta})=\exp[-i(\overset{c}{\bar\eta},G\overset{c}{\eta})], \tag{8.3}$$

where $G_{\sigma\tau}(x-y)$ is the Green's function of the Dirac equation. We have defined at once the two-point Green's function by

$$\left(\tilde{\mathcal{E}}_2\frac{d}{d\overset{c}{\eta}}\frac{d}{d\overset{c}{\bar\eta}}\right)(x,y)\cdot Z_0(\overset{c}{\eta},\overset{c}{\bar\eta})_{|\overset{c}{\eta}=\overset{c}{\bar\eta}=0}$$
$$=\left(\tilde{\mathcal{E}}_2\frac{d}{d\overset{c}{\eta}}\frac{d}{d\overset{c}{\bar\eta}}\right)(x,y)\cdot\int dx'\,dy'$$
$$\times(\tilde{\mathcal{E}}_2\overset{c}{\bar\eta}_\sigma\overset{c}{\eta}_\tau(x',y')\,G_{\sigma\tau}(x-y)=G^\rho_{\sigma\tau}(x-y) \tag{8.4}$$

which allows us to consider the reduced (boson) generating functional (8.1) as a (c-number) generating functional for the Green's functions of the Dirac field.

Note added in proof: In the course of the paper the words "classical" and "quantum" concern the c-number and q-number levels respectively of the given theory, and have nothing to do with any $\hbar\to 0$ limit. The natural system of units $\hbar=c=1$ is employed.

A complete operator formulation of steps (5.14)—(5.16), which should be more convincing for an unfamiliar reader, can be found in the Phys. Rep. C (1978) paper of Ref. 5.

Let us emphasize that by virtue of the projection theorems each Bose field, which obeys the Haag—LSZ expansion conjecture, has its corresponding fermion contents. It happens independently of whether the spin-statistics theorem holds or not, and makes less surprising the fact that in some Bose field theory models (as, e.g., the sine-Gordon one) fermions are allowed to appear.

[1]The general structure of the quantum field theory (QFT) including functional formulation of the LSZ approach with respect to scalar fields is studied in: H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento I, 206 (1955); VI, 319 (1957); R. Haag, Dansk. Mat. Fyz. Medd. 29, 13 (1955); V. Glaser, H. Lehmann, and W. Zimmermann, Nuovo Cimento VI, 1122 (1957); N.N. Bogolubov, A.A. Logunov, and I.I. Todorov, *Axiomatic Quantum Field Theory*, in Russian (Nauka, Moscow, 1963): K. Hepp, in *Brandeis Summer Institute* 1965, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1965).

[2]For an introduction to the Hilbert spaces of functional power series and functional representations of the CCR, study of estimates, and convergence criteria, see: J. Rzewuski, Rep. Math. Phys. 1, 1 (1970); 1, 195 (1971); see also S. Schweber, J. Math. Phys. 3, 831 (1962); V. Bargmann, Comm. Pure Appl. Math. 14, 187 (1961).

[3]Most systematic studies of functional integration and differentiation methods in application to QFT have been performed in monographs (spinors treated in Grassman language): F.A. Berezin, *The Method of Second Quantization*, in Russian (Nauka, Moscow, 1965); J. Rzewuski, *Field Theory, Vol. II. Functional Formulation of the S-Matrix Theory* (Iliffe, London, PWN, Warsaw, 1969); V.N. Popov, *Path Integrals in Quantum Field Theory and Statistical Physics*, in Russian (Atomizdat, Moscow, 1976).

[4]The so-called mathematical theory of Feynman path integrals (with a few limitations) is covered by: C. DeWitt-Morette, Comm. Math. Phys. 28, 47 (1972); 37, 63 (1974); S. Albeverio and R. Høegh-Krohn, *Mathematical Theory of*

Feynman Path Integrals, Lecture Notes in Mathematics (Springer-Verlag, Berlin, 1976).

[5]For boson expansion methods, as developed by the present author, see: Comm. Math. Phys. **43**, 131 (1975), "Representations of the CAR generated by representations of the CCR. Fock case"; Bull. Acad. Polon. Sci. Ser. Astr. Phys. Math. **23**, 113 (1975), "Spin-1 vector Boson structure of free spin-1/2 quantum field"; *ibid.* 24, 201 (1976), "Representations of the CAR generated by representations of the CCR. II /isometries/"; *ibid*, "Boson expansions of Jordan-Wigner representation," Bull. Acad. Pol. Sci. **25**, 711 (1977); with Z. Popowicz, Rep. Math. Phys. **11**, 57 (1977), "Representations of the CAR generated by representations of the CCR. III, Non Fock extension"; Proc. of the 13th Karpacz School, Acta Univ. Wrat. 1976, "Bosonization of Fermions in QFT", "Bosonization of Fermions in Heisenberg ferromagnet," in *Theoretical Physics, Memorial Book on J. Rzewuski's 60th Birthday* (Wroclaw, 1976); "The method of Boson expansions in the q.t. of Fermions," Phys. Rept. C in print; with Z. Popowicz, "Ultralocal quantizational of Sine—Gordon 1-solitons," submitted for publication; Int. Journ. Theor. Phys. **115**, 809 (1977), "Remark on Kalnay theory of Fermions constructed from Bosons."

[6]For the voice of a pragmatist, see: S. Coleman, "Secret symmetry," in *Laws of Subnuclear Matter*, Erice Summer Institute, edited by A. Zichichi (Academic, New York, 1975).

[7]Extension of the LSZ methods in terms of functional integrals onto Dirac spinors by making use of Grassman algebra tools, is considered in P.T. Matthews and A. Salam, Nuovo Cimento X, 120 (1955); Yu. Novoshilov and A.V. Tulub, Usp. Fiz. Nauk 61, 53 (1957).

[8]For attempts to find a c-number formulation of the QT of fermions, without use of Grassman algebras, see: J.R. Klauder, Ann. Phys. (N.Y.) **11**, 123 (1960), "The action option and Feynman quantization of spinor fields in terms of ordinary c-numbers"; F. Rohrlich, in *Analytic Methods in Mathematical Physics*, edited by R.P. Gilbert and R.G. Newton (Gordon and Breach, New York, 1970)," The coherent state representation and q.f.t."; P. Garbaczewski and J. Rzewuski, Rep. Math. Phys. **6**, 431 (1974), "On generating functionals for antisymmetric functions and their application in q.f.t."; P. Garbaczewski, Rep. Math. Phys. **7**, 321 (1975), "Functional representations of the CAR."

[9]K.O. Friedrichs, *Mathematical Aspects of the Quantum Field Theory Fields* (Interscience, New York, 1953).

[10]For strongly selected number of investigations on fermion—boson correspondence, especially in connection with Thiring and sine—Gordon systems, see: S. Coleman, Phys. Rev. D **11**, 2088 (1975); S. Mandelstam, Phys. Rev. D **11**, 3026 (1975); A.K. Pogrebkov and V.N. Sushko, Teor. Mat. Fiz. 24, 425 (1975); B. Schroer and J.A. Swieca, "Spin and statistics of quantum kinks," CERN preprint 1976); B. Schroer, Q.f.t. of kinks in two-dimensional space-time," Cargese lecture notes (1976); H. Neuberger, "Bosonization in field theory in two space-time dimensions," Tel-Aviv preprint (1976) (Grassman algebras involved in the study of correspondence between Hamiltonians).

[11]The influence of Grassman methods: for works on pseudomechanics, as an example, we suggest: F. Casalbuoni, Nuovo Cimento A **33**, 115 (1976), "On the quantization of systems with anticommuting variables."

The local von Neumann algebras for the massless scalar free field and the free electromagnetic field

G. Benfatto[a)]

Istituto di Matematica dell'Università, L'Aquila, Italy

F. Nicolò[a)]

Istituto di Fisica dell'Università, Roma, Italy
(Received 3 March 1977)

The properties of the von Neumann algebras of local observables for the free scalar field of zero mass are studied. The local algebras possess a lattice structure, and the duality condition is satisfied. The problem of duality for the free electromagnetic field is discussed.

1. INTRODUCTION

In the algebraic approach to quantum field theory one is involved with the problem of giving the general properties (axioms) of the local algebras of the observables of the theory. The program of determining the general algebraic structure which must constitute the natural framework of the quantum field theory has been initiated by Haag and Kastler[1] and developed in great detail in a series of papers by Doplicher, Haag, and Roberts.[2] Between their very general axioms, the locality postulate, in the more stringent form of the *duality condition*, has a central role.

To exploit in detail the meaning and the power of these axioms, it is interesting to examine if they are satisfied in some simple models, as the free field ones. It turns out that, in the free field models, all the axioms are very easy to verify, except the duality condition, which requires a detailed investigation of the structure of the local algebras. Some years ago Araki[3,4] has proven, for the scalar free field of mass $m>0$, that the duality condition is satisfied and that the local algebras have a well defined lattice structure.

In this paper we extend Araki's proof to the scalar free field of mass $m=0$. Then we give a clear definition of the local algebras for the electromagnetic free field and show that the proof of the duality condition rests, in this case, on a well defined technical problem. We make the ansatz that the answer to this problem should imply that the duality condition is still satisfied.

The plan of the paper is the following. In Sec. 2 we define the formalism needed to study the free scalar field and we recall its well known properties. In Sec. 3 we define the Von Neumann algebras of local observables and state, in Theorems 3.3 and 3.4, their properties, already proven[4] for $m>0$. Section 4 is devoted to the technical lemmas and theorems which we need to prove our results. In Sec. 5 we prove Theorem 3.4 for $m=0$. Finally, in Sec. 6, we study the Von Neumann local algebras for the electromagnetic free field.

a)Research partially supported by the Istituto Nazionale di Fisica Nucleare, Roma.

2. THE FREE SCALAR FIELD

Let us consider the complex Hilbert space $L^2 = L^2(\mathbb{R}^3, d^3p/(\mathbf{p}^2+m^2)^{1/2})$, $m \geq 0$. From L^2 we construct the symmetric Fock space[5] $\mathcal{F}_s = \oplus_{n=0}^{\infty} \mathcal{F}_s^{(n)}$, where $\mathcal{F}_s^{(n)}$ is the symmetrized tensor product of n copies of L^2. For any $\tilde{f} \in L^2$ there is a self-adjoint operator on $\mathcal{F}_s$,[6] the Segal field operator, which is defined by

$$\Phi_s(\tilde{f}) = \frac{1}{\sqrt{2}} [a(\tilde{f}) + a^*(\tilde{f})], \tag{2.1}$$

where

$$a^*(\tilde{f})\, \Psi^{(n)} = \sqrt{n+1}\, (\tilde{f} \otimes \Psi^{(n)})_s, \quad \Psi^{(n)} \in \mathcal{F}_s^{(n)}, \tag{2.2}$$

and $a(\tilde{f})$ is the adjoint of $a^*(\tilde{f})$.

Let us consider the following closed subsets of L^2,

$$K = \{\tilde{f} \in L^2 \,|\, \overline{\tilde{f}(-\mathbf{p})} = \tilde{f}(\mathbf{p})\},$$
$$K' = \{\tilde{f} \in L^2 \,|\, \overline{\tilde{f}(-\mathbf{p})} = -\tilde{f}(\mathbf{p})\}, \tag{2.3}$$

where $\overline{\tilde{f}(\mathbf{p})}$ is the complex conjugate of $\tilde{f}(\mathbf{p})$. K and K' are not (complex) subspaces of L^2, but if we define the real Hilbert space $\mathcal{L}$ whose elements are those of L^2, considered as a real linear space with the real scalar product

$$(\tilde{f}_1, \tilde{f}_2)_{\mathcal{L}} = \mathrm{Re}(\tilde{f}_1, \tilde{f}_2)_{L^2}, \tag{2.4}$$

K and K' are orthogonal closed subspaces of $\mathcal{L}$ and they satisfy the relation

$$K' = \beta K = K^{\perp}, \tag{2.5}$$

where β is the operator of multiplication by the imaginary unit. Then, if $\tilde{f} \in L^2$, there is a unique decomposition

$$\tilde{f} = \tilde{g} + i\tilde{h}, \quad \tilde{g} \in K, \quad \tilde{h} \in K \tag{2.6}$$

which implies

$$\Phi_s(\tilde{f}) = \Phi_s(\tilde{g}) + \Phi_s(i\tilde{h}) = \varphi(\tilde{g}) + \pi(\tilde{h}) \tag{2.7}$$

where

$$\varphi(\tilde{g}) = \frac{1}{\sqrt{2}} [a^*(\tilde{g}) + a(\tilde{g})] \tag{2.8}$$

and

$$\pi(\tilde{h}) = \frac{i}{\sqrt{2}} [a^*(\tilde{h}) - a(\tilde{h})] \tag{2.9}$$

 0022-2488/78/1903-0653$1.00

are the canonical free field and the canonical conjugate momentum.

We now want to recall the correspondence between the Segal field operators and a class of solutions of the Klein–Gordon wave equation,

$$(\Box + m^2) F(x) = 0. \tag{2.10}$$

Let $\mathcal{S}_r(\mathbb{R}^4)$ be the space of the C^∞ real functions of rapid decrease, in the four-dimensional coordinate space, with the real inner product given by

$$\begin{aligned}(f_1, f_2)_H &= \int f_1(x)\,\Delta^{(1)}(x-y) f_2(y)\, d^4x\, d^4y \\ &= \int \overline{\hat{f}_1(p)}\hat{f}_2(p)\,\delta(p^2 - m^2)\, d^4p,\end{aligned} \tag{2.11}$$

where

$$p = (p_0, \mathbf{p}), \quad p^2 = p_0^2 - \mathbf{p}^2,$$

$$x = (x_0, \mathbf{x}), \quad px = p_0 x_0 - \mathbf{p}\cdot\mathbf{x},$$

$$\hat{f}(p) = (2\pi)^{-3/2} \int d^4x \exp(ipx) f(x), \tag{2.12}$$

$$\Delta^{(1)}(x) = (2\pi)^{-3} \int d^4p \exp(-ipx)\,\delta(p^2 - m^2). \tag{2.13}$$

Let us define

$$\mathcal{S}_r^0(\mathbb{R}^4) = \{h \in \mathcal{S}_r(\mathbb{R}^4) \mid (h, h)_H = 0\}. \tag{2.14}$$

The quotient space $\mathcal{S}_r(\mathbb{R}^4)/\mathcal{S}_r^0(\mathbb{R}^4)$ is a real pre-Hilbert space with respect to the scalar product

$$(\{f\}, \{g\})_H = (f, g)_H, \tag{2.15}$$

where $\{f\}$, $f \in \mathcal{S}_r(\mathbb{R}^4)$, is the equivalence class of f. We call H the completion of $\mathcal{S}_r(\mathbb{R}^4)/\mathcal{S}_r^0(\mathbb{R}^4)$. It is immediate to recognize that H can be identified with $\mathcal{L}$ by means of the correspondence

$$\{f\} \in \mathcal{S}_r(\mathbb{R}^4)/\mathcal{S}_r^0(\mathbb{R}^4) \to \tilde{f}(\mathbf{p}) \equiv \hat{f}((\mathbf{p}^2 + m^2)^{1/2}, \mathbf{p}) \in \mathcal{L} \tag{2.16}$$

extended to all H by continuity. Moreover, to each $\{f\} \in \mathcal{S}_r(\mathbb{R}^4)/\mathcal{S}_r^0(\mathbb{R}^4)$ corresponds a smooth solution of the wave equation (2.10),

$$F(x) = \int d^4y\,\Delta(x-y) f(y), \tag{2.17}$$

where

$$\Delta(x) = \frac{-i}{(2\pi)^3} \int d^4p \exp(-ipx)\,\delta(p^2 - m^2)\,\epsilon(p_0). \tag{2.18}$$

The function $F(x)$ is uniquely determined by the knowledge of its initial conditions $a(\mathbf{x}) = \dot{F}(0, \mathbf{x})$ and $b(\mathbf{x}) = F(0, \mathbf{x})$, which have the following expressions in terms of f [see Eq. (2.17)].

$$\begin{aligned} a(\mathbf{x}) &= -(2\pi)^{-3/2} \int d^3p \exp(i\mathbf{p}\cdot\mathbf{x})\,\tilde{g}(\mathbf{p}), \\ b(\mathbf{x}) &= (2\pi)^{-3/2} \int d^3p \exp(i\mathbf{p}\cdot\mathbf{x})(\mathbf{p}^2 + m^2)^{-1/2}\,\tilde{h}(\mathbf{p}), \end{aligned} \tag{2.19}$$

where [see Eq. (2.16)]

$$\tilde{g}(\mathbf{p}) + i\tilde{h}(\mathbf{p}) = \tilde{f}(\mathbf{p}).$$

Equations (2.19) imply that

$$\tilde{a}(\mathbf{p}) = -\tilde{g}(\mathbf{p}), \quad \tilde{b}(\mathbf{p}) = (\mathbf{p}^2 + m^2)^{-1/2}\,\tilde{h}(\mathbf{p}), \tag{2.20}$$

where $\tilde{a}(\mathbf{p})$ and $\tilde{b}(\mathbf{p})$ are the Fourier transforms of $a(\mathbf{x})$ and $b(\mathbf{x})$. From Eqs. (2.5) and (2.20) it easily follows that $\mathcal{L}$ can be identified with the real Hilbert space $\mathcal{H} = \mathcal{H}_\varphi^{(m)} \oplus \mathcal{H}_\pi^{(m)}$, where $\mathcal{H}_\varphi^{(m)}$ and $\mathcal{H}_\pi^{(m)}$ are the real Hilbert spaces obtained by completing the space $\mathcal{S}_r(\mathbb{R}^3)$ of the real C^∞ functions of rapid decrease with respect to the scalar products,

$$(f, g)_\varphi = \int \overline{\tilde{f}(\mathbf{p})}\,\tilde{g}(\mathbf{p})\,\frac{d^3p}{\omega}, \tag{2.21}$$

$$(f, g)_\pi = \int \tilde{f}(\mathbf{p})\,\tilde{g}(\mathbf{p})\,\omega\, d^3p, \tag{2.22}$$

where

$$\omega = (\mathbf{p}^2 + m^2)^{1/2}. \tag{2.23}$$

The correspondence is the following:

$$\begin{aligned}\tilde{f} &= \tilde{g} + i\tilde{h} \in \mathcal{L} = K \oplus \beta K \\ &\to \langle -\tilde{g}, \omega^{-1}\tilde{h}\rangle \in \mathcal{H} = \mathcal{H}_\varphi^{(m)} \oplus \mathcal{H}_\pi^{(m)}.\end{aligned} \tag{2.24}$$

Remark 1: The multiplication by ω^{-1} is a unitary operator from $\mathcal{H}_\varphi^{(m)}$ onto $\mathcal{H}_\pi^{(m)}$; then we can write

$$\mathcal{H} = \mathcal{H}_\varphi^{(m)} \oplus \omega^{-1}\mathcal{H}_\varphi^{(m)}. \tag{2.25}$$

Remark 2: The identification of H with $\mathcal{H}$ allows us to express the locality properties of the Segal field operators $\Phi_s(\{f\})$, associated with supp$f(x)$, in terms of the supports of $a(\mathbf{x})$ and $b(\mathbf{x})$. In fact, if $f \in \mathcal{S}_r(\mathbb{R}^4)$ and supp$f \subset \mathcal{O} \subset \mathbb{R}^4$; the function $F(x) = (\Delta * f)(x)$ has initial conditions $a(\mathbf{x}) = \dot{F}(0, \mathbf{x})$ and $b(\mathbf{x}) = F(0, \mathbf{x})$ which satisfy

$$\text{supp}\,a, \ \text{supp}\,b \subset (\mathcal{O}')' \cap S_{t=0}, \tag{2.26}$$

where S_t is the three-dimensional hypersurface at t fixed and $\mathcal{O}'$ is the casual complement of $\mathcal{O}$,

$$\mathcal{O}' = \{x \in \mathbb{R}^4 \mid (x-y)^2 < 0, \ \text{for } y \in \mathcal{O}\}. \tag{2.27}$$

Remark 3: For $m > 0$, $\mathcal{H}_\varphi^{(m)}$ and $\mathcal{H}_\pi^{(m)}$ are essentially the well-known Sobolev spaces $H^{-1/2}(\mathbb{R}^3)$ and $H^{1/2}(\mathbb{R}^3)$, which have been studied extensively in the literature.

3. THE LOCAL ALGEBRAS FOR THE FREE SCALAR FIELD

If L is a linear subspace of H, we associate with L the Von Neumann algebra $R(L)$ defined in the following way[7]:

$$R(L) = \{\exp[i\Phi_s(h)] \mid h \in L\}''. \tag{3.1}$$

Analogously, if K_1 and K_2 are two linear subspaces of K, we define

$$R(K_1, K_2) = \{\exp[i\varphi(g)\exp(i\pi(h)] \mid g \in K_1, h \in K_2\}''. \tag{3.2}$$

If L is a linear subspace of H and K_1 and K_2 are two linear subspaces of K such that $L \sim K_1 \oplus \beta K_2$, (2.6) and (2.16) imply that

$$R(L) = R(K_1, K_2). \tag{3.3}$$

Furthermore, since $\exp[i\Phi_s(h)]$ is a strongly continuous function of h,[3] we have

$$R(L) = R(\bar{L}) = R(\bar{K}_1, \bar{K}_2). \tag{3.4}$$

The lattice properties of the algebras $R(L)$ have been investigated in great detail by Araki.[3] Let us consider the lattice of all closed subspaces of H, $\{H_\alpha\}$, and the lattice of the corresponding Von Neumann algebras

$\{R(H_\alpha)\}$, the lattice operations being,

$$\bigwedge_\alpha H_\alpha = \bigcap_\alpha H_\alpha, \quad \bigwedge_\alpha R(H_\alpha) = \bigcap_\alpha R(H_\alpha),$$
$$\bigvee_\alpha H_\alpha = \bigoplus_\alpha H_\alpha, \quad \bigvee_\alpha R(H_\alpha) = [\bigcup_\alpha R(H_\alpha)]''. \tag{3.5}$$

We define also a complementation,

$$H_\alpha^c = \beta H_\alpha^\perp, \quad R(H_\alpha)^c = R(H_\alpha)'. \tag{3.6}$$

Araki has shown[3] that:

Theorem 3.1: The complemented lattices of all closed subspaces of H and of the corresponding Von Neumann algebras are isomorphic:

$$(1)\ R(H_1) \supset R(H_2) \text{ iff } H_1 \supset H_2, \tag{3.7}$$

$$(2)\ R(H_1) = R(H_2) \text{ iff } H_1 = H_2, \tag{3.8}$$

$$(3)\ R(\bigvee_\alpha H_\alpha) = \bigvee_\alpha R(H_\alpha), \tag{3.9}$$

$$(4)\ R(\bigwedge_\alpha H_\alpha) = \bigwedge_\alpha R(H_\alpha), \tag{3.10}$$

$$(5)\ R(H_\alpha)' = R(\beta H_\alpha^\perp). \tag{3.11}$$

This theorem has an obvious translation in terms of the algebras $R(K_1, K_2)$. In particular the duality relation (3.11) becomes

$$R(K_1, K_2)' = R(K_2^\perp, K_1^\perp). \tag{3.12}$$

If $\mathcal{O} \subset \mathbb{R}^4$ is an open set and $L = D_r(\mathcal{O})/\mathcal{S}_r^0(\mathbb{R}^4)$ where $D_r(\mathcal{O}) = \{\varphi \in C_r^\infty(\mathbb{R}^4) \mid \operatorname{supp}\varphi$ is a compact set contained in $\mathcal{O}\}$, the algebra $R(L)$, which we will denote also by $R(\mathcal{O})$, has the physical interpretation of the algebra of the observables, which can be measured in the space—time region $\mathcal{O}$. It is well known that

Theorem 3.2: Let $\mathcal{O}$ and $\mathcal{O}_\alpha$ be open sets in $\mathbb{R}^4$. Then:

$$\text{If } \mathcal{O} = \bigcup_\alpha \mathcal{O}_\alpha,\ R(\mathcal{O}) = \bigvee_\alpha R(\mathcal{O}_\alpha) \quad \text{(field property)}, \tag{3.13}$$

$$\text{If } \mathcal{O}_1' \supset \mathcal{O}_2,\ \ R(\mathcal{O}_1)' \supset R(\mathcal{O}_2) \quad \text{(local commutativity)}, \tag{3.14}$$

$$U(\alpha, \Lambda) R(\mathcal{O}) U(\alpha, \Lambda)^{-1} = R((\alpha, \Lambda)\mathcal{O}) \quad \text{(covariance)}. \tag{3.15}$$

The algebras $R(\mathcal{O})$ have some other interesting properties when $\mathcal{O} = C(B)$ where $B \subset \mathbb{R}^3$ is an open set and

$$C(B) = \{x \in \mathbb{R}^4 \mid (x-y)^2 < 0,\ y \in S_0 \cap B^c\}. \tag{3.16}$$

If we put $L = D_r(C(B))/\mathcal{S}_r^0(\mathbb{R}^4)$, it easily follows from (2.20), (2.24), and (2.26) that $\bar{L} = F_\varphi^{(m)}(B) \oplus F_\pi^{(m)}(B)$, where

$$F_{\varphi,\pi}^{(m)}(B) = \overline{\{\varphi \in D_r(B)\}}^{\mathcal{H}_{\varphi,\pi}^{(m)}}. \tag{3.17}$$

Equation (3.3) then implies that

$$R(C(B)) = R(F_\varphi^{(m)}(B), \omega F_\pi^{(m)}(B)). \tag{3.18}$$

For any $\Delta \subset \mathbb{R}^4$ and any $E \subset \mathbb{R}^3$, we define

$$\bar{R}(\Delta) = \bigwedge_{\substack{\mathcal{O} \supset \bar{\Delta} \\ \mathcal{O} \text{ open}}} R(\mathcal{O}), \tag{3.19}$$

$$\bar{F}_{\varphi,\pi}^{(m)}(E) = \bigwedge_{\substack{B \supset \bar{E} \\ B \text{ open}}} F_{\varphi,\pi}^{(m)}(B). \tag{3.20}$$

We can now formulate the main theorem. Let $\mathcal{B}$ be the family of the sets $B \subset \mathbb{R}^3$ such that:

(a) $B = \bigcup_{i=1}^n B_i$, where n is a finite integer depending on B and $\{B_i\}_{i=1}^n$ is a family of mutually disjoint open connected sets.

(b) $B_i = \operatorname{int}\bar{B}_i$, $i = 1, \ldots, n$, and $B = \operatorname{int}\bar{B}$.

(c) The boundary of B_i, ∂B_i, is, for $i = 1, \ldots, n$, a surface made up of a finite number of many-times differentiable surfaces joined together along many-times differentiable curves.

$\mathcal{B}$ is a complemented lattice with respect to the operations:

$$B_1 \wedge B_2 = B_1 \cap B_2,$$
$$B_1 \vee B_2 = \operatorname{int}\overline{B_1 \cup B_2}, \tag{3.21}$$
$$B^\perp = \bar{B}^c.$$

Define now by $\tilde{\mathcal{B}}$ the subset of $\mathcal{B}$ made up of the sets $B = \bigcup_{i=1}^n B_i$, such that:

(d) If p is a singular point of ∂B_i $(i = 1, \ldots, n)$, there is a neighborhood U_p of p and a vector $\boldsymbol{\lambda}$, such that

$$\{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x} = \mathbf{y} + \boldsymbol{\lambda},\ \mathbf{y} \in U_p \cap \bar{B}_i\} \subset B_i.$$

(e) $\partial B_i \cap \partial B_j = \emptyset$ if $i \neq j$.

Observe that $\tilde{\mathcal{B}}$ is not closed with respect to the operations (3.21). We have

Theorem 3.3:

$$(1)\ R(C(B_1)) \wedge R(C(B_2)) = R(C(B_1 \wedge B_2)),$$
$$B_1, B_2, B_1^\perp \vee B_2^\perp \in \tilde{\mathcal{B}}, \tag{3.22}$$

$$(2)\ R(C(B_1)) \vee R(C(B_2)) = R(C(B_1 \vee B_2)),$$
$$B_1, B_2 \in \tilde{\mathcal{B}}, \tag{3.23}$$

$$(3)\ R(\mathbb{R}^4) = R(C(\mathbb{R}^3)) = \mathcal{B}(\mathcal{J}_s) \quad \text{(completeness)}, \tag{3.24}$$

$$(4)\ \bar{R}(\partial B) = \{\lambda I\}, \quad B \in \tilde{\mathcal{B}}, \tag{3.25}$$

$$(5)\ R(C(B))' = R(C(B^\perp)), \quad B, B^\perp \in \tilde{\mathcal{B}}. \tag{3.26}$$

By using Theorem 3.1 and Eq. (3.18), it is easy to see that Theorem 3.3 is equivalent to the following one:

Theorem 3.4:

$$(1)\ F_{\varphi,\pi}^{(m)}(B_1) \wedge F_{\varphi,\pi}^{(m)}(B_2) = F_{\varphi,\pi}^{(m)}(B_1 \wedge B_2),$$
$$B_1, B_2, B_1^\perp \vee B_2^\perp \in \tilde{\mathcal{B}}, \tag{3.27}$$

$$(2)\ F_{\varphi,\pi}^{(m)}(B_1) \vee F_{\varphi,\pi}^{(m)}(B_2) = F_{\varphi,\pi}^{(m)}(B_1 \vee B_2),$$
$$B_1, B_2 \in \tilde{\mathcal{B}}, \tag{3.28}$$

$$(3)\ F_{\varphi,\pi}^{(m)}(\mathbb{R}^3) = \mathcal{H}_{\varphi,\pi}^{(m)}, \tag{3.29}$$

$$(4)\ \bar{F}_{\varphi,\pi}^{(m)}(\partial B) = \{0\}, \quad B \in \tilde{\mathcal{B}}, \tag{3.30}$$

$$(5)\ F_\varphi^{(m)}(B)^\perp = \omega F_\pi^{(m)}(B^\perp), \quad B^\perp \in \tilde{\mathcal{B}}. \tag{3.31}$$

Theorem 3.4 has been proven by Araki[4] for $m > 0$ many years ago, but his proof does not work for $m = 0$, because the spaces $\mathcal{H}_\varphi^{(0)}$ and $\mathcal{H}_\pi^{(0)}$ are essentially different from $\mathcal{H}_\varphi^{(m)}$ and $\mathcal{H}_\pi^{(m)}$, $m > 0$. In the next section we will prove some properties of the spaces $\mathcal{H}_\varphi^{(0)}$ and $\mathcal{H}_\pi^{(0)}$, that are relevant, in the case $m = 0$, to the proof of Theorem 3.4, which is postponed to Sec. 5.

We observe now that (3.22) and (3.26) imply

$$R(C(B))' \cap R(C(B)) = \{\lambda I\}, \quad B, B^\perp \in \tilde{\mathcal{B}}, \tag{3.32}$$

that is, $R(C(B))$ is a factor. It is also possible to show[4,8] that $R(C(B))$ is of type III. The proof is essentially based on algebraic arguments, but there are some technical points (see Sec. 8 in Ref. 4) which require a different approach in the cases $m>0$ and $m=0$. The result is true also for $m=0$, however we will not give the details here.

4. THE ZERO MASS HILBERT SPACES $H_\varphi^{(0)}$ AND $H_\pi^{(0)}$

We recall that $H_\varphi^{(0)}$ and $H_\pi^{(0)}$ were defined as the real Hilbert spaces obtained by completing $S_r(\mathbb{R}^3)$ with respect to the scalar products

$$(f,g)_{\substack{\varphi\\ \pi}} = \int d^3p\,|\mathbf{p}|^{\mp 1}\,\overline{\widetilde{f}(\mathbf{p})}\,\widetilde{g}(\mathbf{p}). \tag{4.1}$$

We want now to characterize $H_\varphi^{(0)}$ and $H_\pi^{(0)}$ as spaces of distributions.[9] In the following we will denote simply by S the space of the real C^∞ functions of rapid decrease and by $D\subset S$ the set of the functions of compact support; moreover S will be sometimes considered as a linear topological space with the topology given by the seminorms

$$\|\varphi\|_{\alpha,\beta} = \sup_{x\in\mathbb{R}^3}|x^\alpha D^\beta\varphi(x)|. \tag{4.2}$$

S' will be the dual of S and its topology will be the weak topology.

If F and G are two topological spaces, the expression $F\subset_{\text{cont}} G$ will mean that F can be injected in G and that the injection is continuous.

Definition 4.1: $L^2_{\varphi,\pi}$ is the real Hilbert space of the measurable functions $\widetilde{f}(\mathbf{p})=\overline{\widetilde{f}(-\mathbf{p})}$ with scalar product

$$(\widetilde{f},\widetilde{g})_{\widetilde{\varphi},\widetilde{\pi}} = \int d^3p\,|\mathbf{p}|^{\mp1}\overline{\widetilde{f}(\mathbf{p})}\,\widetilde{g}(\mathbf{p}). \tag{4.3}$$

Lemma 4.2:

$$(1)\ S \underset{\text{cont}}{\subset} L^2_{\varphi,\pi} \underset{\text{cont}}{\subset} S', \tag{4.4}$$

$$(2)\ \overline{S}^{\widetilde{\varphi},\widetilde{\pi}} = L^2_{\varphi,\pi}. \tag{4.5}$$

Proof: Let $\widetilde{f}\in L^2_\varphi$ and $\epsilon>0$. There exists $R>0$ such that $\|\widetilde{f}-\widetilde{g}\|_{\widetilde{\varphi}}<\epsilon/2$, where $\widetilde{g}(\mathbf{p})=\chi_{[1/R,R]}(|\mathbf{p}|)\widetilde{f}(\mathbf{p})$ and $\chi_{[a,b]}(\lambda)$ denotes the characteristic function of $[a,b]$. Furthermore there exists $\widetilde{\varphi}\in D$ such that $\operatorname{supp}\widetilde{\varphi}\subseteq\{\mathbf{p}\,|\,(1/2R)\le|\mathbf{p}|\le 2R\}$ and $\int d^3\mathbf{p}\,|\widetilde{g}(\mathbf{p})-\widetilde{\varphi}(\mathbf{p})|^2<(2R)^{-1}(\epsilon/2)^2$. Therefore, $\|\widetilde{g}-\widetilde{\varphi}\|_{\widetilde{\varphi}}<\epsilon/2$ and $\|\widetilde{f}-\widetilde{\varphi}\|_{\widetilde{\varphi}}<\epsilon$ which implies (2) for L^2_φ. Analogously we can prove (2) for L^2_π. Let us now suppose that $\widetilde{\varphi}_n\in S$, and $\widetilde{\varphi}_n\to\widetilde{\varphi}$ in the topology of S. We have

$$\begin{aligned}\|\widetilde{\varphi}_n-\widetilde{\varphi}\|^2_{\widetilde{\varphi}} &= \int_{|\mathbf{p}|\le1}|\mathbf{p}|^{-1}|\widetilde{\varphi}_n(\mathbf{p})-\widetilde{\varphi}(\mathbf{p})|^2d^3\mathbf{p}\\ &\quad+\int_{|\mathbf{p}|\ge1}|\mathbf{p}|^{-1}|\widetilde{\varphi}_n(\mathbf{p})-\widetilde{\varphi}(\mathbf{p})|^2d^3\mathbf{p}\\ &\le[\sup_{\mathbf{p}}|\widetilde{\varphi}_n(\mathbf{p})-\widetilde{\varphi}(\mathbf{p})|]\int_{|\mathbf{p}|\le1}d^3\mathbf{p}\,|\mathbf{p}|^{-1}\\ &\quad+[\sup_{\mathbf{p}}|\mathbf{p}|^3|\widetilde{\varphi}_n(\mathbf{p})-\widetilde{\varphi}(\mathbf{P})|^2]\int_{|\mathbf{p}|\ge1}|\mathbf{p}|^{-4/3}d^3\mathbf{p}\\ &\xrightarrow[n\to\infty]{}0\end{aligned}$$

analogously for $\|\widetilde{\varphi}_n-\widetilde{\varphi}\|^2_{\widetilde{\pi}}$. Therefore, S is injected continuously in L^2_φ and L^2_π by the identity application. Let us finally show that $L^2_{\varphi,\pi}$ can be injected continuously in S' by the natural injection

$$\widetilde{i}(\widetilde{f})(\widetilde{\varphi})=\int d^3\mathbf{p}\,\widetilde{f}(\mathbf{p})\,\widetilde{\varphi}(\mathbf{p}),\quad \widetilde{f}\in L^2_{\varphi,\pi},\quad \widetilde{\varphi}\in S. \tag{4.6}$$

If $\widetilde{f}\in L^2_\varphi$, we have

$$\begin{aligned}|\widetilde{i}(\widetilde{f})(\widetilde{\varphi})|^2 &\le \|\widetilde{f}\|^2_{\widetilde{\varphi}}\int d^3\mathbf{p}\,|\mathbf{p}|\,|\widetilde{\varphi}(\mathbf{p})|^2\\ &\le\|\widetilde{f}\|^2_{\widetilde{\varphi}}\{[\sup|\widetilde{\varphi}(\mathbf{p})|^2]\int_{|\mathbf{p}|\le1}|\mathbf{p}|\,d^3\mathbf{p}\\ &\quad+[\sup|\mathbf{p}|^5\,|\widetilde{\varphi}(\mathbf{p})|^2]\int_{|\mathbf{p}|\ge1}|\mathbf{p}|^{-4}d^3\mathbf{p}\}.\end{aligned}$$

This shows that $\widetilde{i}(\widetilde{f})\in S'$ and that the injection $\widetilde{i}$ is continuous. The same result is true for L^2_π. ■

Proposition 4.3:

$$(1)\ S\underset{\text{cont}}{\subset} H^{(0)}_{\varphi,\pi}\underset{\text{cont}}{\subset}S', \tag{4.7}$$

$$(2)\ \overline{D}^{\varphi,\pi}=H^{(0)}_{\varphi,\pi}. \tag{4.8}$$

Proof: (1) immediately follows from the definition of $H^{(0)}_{\varphi,\pi}$ and Lemma 4.2, if we recall that the Fourier transform is a continuous bijection[10] of S onto S and of S' onto S' and if we define the injection of $H^{(0)}_{\varphi,\pi}$ in S' by

$$i(f)(\varphi)=\widetilde{i}(\widetilde{f})(\widetilde{\varphi}),\quad f\in H^{(0)}_{\varphi,\pi},\quad \varphi\in S, \tag{4.9}$$

where now $\widetilde{f}$ and $\widetilde{\varphi}$ are the Fourier transforms of f and φ—(2) follows from the fact that D is dense in S in the topology of S, $\overline{S}^{\varphi,\pi}=H^{(0)}_{\varphi,\pi}$ and $S\subset_{\text{cont}}H^{(0)}_{\varphi,\pi}$, which implies $\overline{D}^{\varphi,\pi}\subset\overline{S}^{\varphi,\pi}$. ■

In the following we will identify $f\in H^{(0)}_{\varphi,\pi}$ with $i(f)\in S'$.

Lemma 4.4: If

$$\widetilde{f}\in L^2_\pi,\quad \widetilde{f}_1(\mathbf{p})=\chi_{[0,1]}(|\mathbf{p}|)\widetilde{f}(\mathbf{p}),\quad \widetilde{f}_2(\mathbf{p})=\chi_{[1,\infty)}(|\mathbf{p}|)\widetilde{f}(\mathbf{p})$$

we have

$$\widetilde{f}_1\in L^p(\mathbb{R}^3),\quad 1\le p<\tfrac32,\ \|\widetilde{f}_1\|_p\le C_p\|\widetilde{f}\|_{\widetilde{\pi}}, \tag{4.10}$$

$$\widetilde{f}_2\in L^p(\mathbb{R}^3),\quad \tfrac32<p\le2,\quad \|\widetilde{f}_2\|_p\le C'_p\|\widetilde{f}\|_{\widetilde{\pi}}. \tag{4.11}$$

Moreover in (4.10) and (4.11) the value $p=\frac32$ is excluded.

Proof: By using Hölder's inequality we have

$$\begin{aligned}\|\widetilde{f}_1\|_p &= [\int_{|\mathbf{p}|\le1}(|\mathbf{p}|^{1/2}|\widetilde{f}(\mathbf{p})|)^p|\mathbf{p}|^{-p/2}d^3\mathbf{p}]^{1/p}\\ &\le[\int_{|\mathbf{p}|\le1}|\mathbf{p}|^{-q/2}d^3\mathbf{p}]^{1/q}\|\widetilde{f}\|_{\widetilde{\pi}},\end{aligned} \tag{4.12}$$

where $1/p=\frac12+1/q$. Since $C_p=[\int_{|\mathbf{p}|\le1}|\mathbf{p}|^{-1/2}d^3\mathbf{p}]^{1/q}<\infty$ only if $q<6$, (4.12) implies (4.10). In order to show that the value $p=\frac32$ is excluded, we give a counterexample. Let us consider the function $\widetilde{f}(\mathbf{p})=|\mathbf{p}|^{-2}[\ln|\mathbf{p}|]^{-\alpha}\chi_{[0,1]}(|\mathbf{p}|)$. It is easy to see that $\widetilde{f}=\widetilde{f}_1\in L^2_\pi$ iff $2\alpha>1$, but $\widetilde{f}_1\in L^{3/2}$ iff $3\alpha>2$; then, if $\frac12<\alpha\le\frac23$, $\widetilde{f}\in L^2_\pi$ but $\widetilde{f}_1\notin L^{3/2}$. The proof of (4.11) is analogous. ■

Lemma 4.5: If $f\in H^{(0)}_\pi$, we can write

$$f=f_1+f_2,\quad f_1\in L^{q_1}(\mathbb{R}^3)\cap C^\infty(\mathbb{R}^3),\quad f_2\in L^{q_2}(\mathbb{R}^3),\quad q_1>3,\quad 2\le q_2<3, \tag{4.13}$$

and

$$\|f_1\|_{q_1}\le d_{q_1}\|f\|_\pi,\quad \|f_2\|_{q_2}\le d'_{q_2}\|f\|_\pi. \tag{4.14}$$

Proof: Let $\widetilde{f}$ be the Fourier transform of f and $\widetilde{f}_{1,2}$ be defined as in Lemma 4.4. By the Hausdorff–Young inequality[11]:

$$\begin{aligned}\|f_1\|_{q_1}&\le(2\pi)^{3/2-3/p_1}\|\widetilde{f}_1\|_{p_1}\\ &\le(2\pi)^{3/2-3/p_1}C_{p_1}\|\widetilde{f}\|_{\widetilde{\pi}}=d_{q_1}\|f\|_\pi\end{aligned}$$

where $1/q_1+1/p_1=1$; analogously for f_2. Furthermore, since $\tilde{f}_1$ has compact support, by the Paley–Wiener theorem[11] f_1 is an entire function, then $f_1\in C^\infty(\mathbb{R}^3)$. ∎

Theorem 4.6: If $\varphi\in D$ the multiplication by φ is a continuous operator of $\mathcal{H}_\tau^{(0)}$ in itself.

Proof: Let $f\in\mathcal{H}_\tau^{(0)}$ and f_1, f_2 be defined as in Lemma 4.5. We have

$$\|\varphi f\|_\tau=\|\widetilde{\varphi f}\|_{\tilde\tau}=(2\pi)^{-3/2}\|\tilde\varphi*\tilde f\|_{\tilde\tau}$$
$$\leq(2\pi)^{-3/2}[\|\tilde\varphi*\tilde f_1\|_{\tilde\tau}+\|\tilde\varphi*\tilde f_2\|_{\tilde\tau}].$$

Furthermore, if we denote by $\omega^{1/2}$ the operator of multiplication by $|\mathbf{p}|^{1/2}$, and by $\|\cdot\|_p$ the L^p norm, we have:

$$\|\tilde\varphi*\tilde f_{1,2}\|_{\tilde\tau}=\|\omega^{1/2}|\tilde\varphi*\tilde f_{1,2}|\;\|_2$$
$$\leq\|(\omega^{1/2}|\tilde\varphi|)*|\tilde f_{1,2}|\;\|_2$$
$$+\|\;|\tilde\varphi|*(\omega^{1/2}|\tilde f_{1,2}|)\|_2,$$
$$\|\tilde\varphi*\tilde f_1\|_{\tilde\tau}\leq\|\omega^{1/2}\tilde\varphi\|_2\|\tilde f_1\|_1+\|\tilde\varphi\|_1\|\tilde f_1\|_{\tilde\tau}$$
$$\leq(C_1\|\omega^{1/2}\tilde\varphi\|_2+\|\tilde\varphi\|_1)\|f\|_\tau,$$
$$\|\tilde\varphi*\tilde f_2\|_{\tilde\tau}\leq\|\omega^{1/2}\tilde\varphi\|_1\|\tilde f_2\|_2+\|\tilde\varphi\|_1\|\tilde f_2\|_{\tilde\tau}$$
$$\leq(C_2'\|\omega^{1/2}\tilde\varphi\|_1+\|\tilde\varphi\|_1)\|f\|_\tau,$$

where we used the Young inequality[11] and Lemma 4.4. Therefore, there exists a constant K_φ, depending only on φ, such that $\|\varphi f\|_\tau\leq K_\varphi\|f\|_\tau$. ∎

Let $\Psi\in D$ be a function such that $\Psi(\mathbf{x})=1$ if $|\mathbf{x}|\leq 1$ and $0\leq\Psi(\mathbf{x})\leq 1$; we put $\Psi_\epsilon(\mathbf{x})=\Psi(\epsilon\mathbf{x})$.

Theorem 4.7: Let $f\in\mathcal{H}_\tau^{(0)}$, then

$$(1)\;(\varphi,\Psi_\epsilon f)_\tau\xrightarrow[\epsilon\to 0]{}(\varphi,f)_\tau,\quad\forall\varphi\in\mathcal{S},\tag{4.15}$$

$$(2)\;\|\Psi_\epsilon f\|_\tau\leq C\|f\|_\tau,\quad C>0,\quad 0\leq\epsilon\leq 1.\tag{4.16}$$

Proof: If $f\in\mathcal{H}_\tau^{(0)}$ and $\varphi\in\mathcal{S}$, we have

$$(\varphi,f)_\tau=\int\chi(\mathbf{x})f(\mathbf{x})\,d^3\mathbf{x},\quad\tilde\chi(\mathbf{p})=|\mathbf{p}|\tilde\varphi(\mathbf{p}).\tag{4.17}$$

$\tilde\chi(\mathbf{p})$ is not a C^∞ function; however it is easy to show that $D^\alpha\tilde\chi\in L^2$ if $|\alpha|\leq 2$. This implies that

$$\mathbf{x}^\alpha\chi(\mathbf{x})\in L^2,\quad|\alpha|\leq 2.\tag{4.18}$$

Then, if we define f_1 and f_2 as in Lemma 4.5, we have

$$|\int\chi(\mathbf{x})f(\mathbf{x})\,d^3\mathbf{x}|\leq|\int\chi(\mathbf{x})f_1(\mathbf{x})\,d^3\mathbf{x}|+|(\chi,f_2)_2|$$
$$\leq\|(1+|\mathbf{x}|^2)\chi(\mathbf{x})\|_2\,\|(1+|\mathbf{x}|^2)^{-1}f_1(\mathbf{x})\|_2$$
$$+\|\chi\|_2\,\|f_2\|_2\tag{4.19}$$

because $(1+|\mathbf{x}|^2)^{-1}f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ belong to L^2. By Theorem 4.6 $\Psi_\epsilon f\in\mathcal{H}_\tau^{(0)}$, then we can apply (4.19) to the function $\Psi_\epsilon f-f$. Furthermore $\Psi_\epsilon(\mathbf{x})f_{1,2}(\mathbf{x})\xrightarrow[\epsilon\to 0]{}f_{1,2}(\mathbf{x})$ pointwise and $|\Psi_\epsilon(\mathbf{x})f_{1,2}(\mathbf{x})-f_{1,2}(\mathbf{x})|\leq|f_{1,2}(\mathbf{x})|$; then (4.15) follows from (4.19) and the Lebesgue theorem.

In order to prove (4.16) we will use a trick analogous to that used in the proof of Theorem (4.6), but we will make the decomposition of $\tilde f$:

$$\tilde f=\tilde f_{1,\epsilon}+\tilde f_{2,\epsilon},\quad\tilde f_{1,\epsilon}(\mathbf{p})=\chi_{[0,\epsilon]}(|\mathbf{p}|)\tilde f(\mathbf{p}),$$
$$\tilde f_{2,\epsilon}(\mathbf{p})=\chi_{[\epsilon,\infty]}(|\mathbf{p}|)\tilde f(\mathbf{p}).\tag{4.20}$$

Obviously Lemma 4.4 is true also for this decomposition; then we have

$$(2\pi)^{3/2}\|\Psi_\epsilon f\|_\tau\leq\|\tilde\Psi_\epsilon*\tilde f_{1,\epsilon}\|_{\tilde\tau}+\|\tilde\Psi_\epsilon*\tilde f_{2,\epsilon}\|_{\tilde\tau}$$
$$\leq\|\omega^{1/2}\tilde\Psi_\epsilon\|_2\|\tilde f_{1,\epsilon}\|_1+\|\tilde\Psi_\epsilon\|_1\|\omega^{1/2}\tilde f_{1,\epsilon}\|_2$$
$$+\|\omega^{1/2}\tilde\Psi_\epsilon\|_1\|\tilde f_{2,\epsilon}\|_2+\|\tilde\Psi_\epsilon\|_1\|\omega^{1/2}\tilde f_{2,\epsilon}\|_2.\tag{4.21}$$

On the other hand,

$$\|\tilde f_{1,\epsilon}\|_1\leq\|\tilde f\|_{\tilde\tau}[\int_{|\mathbf{p}|\leq\epsilon}|\mathbf{p}|^{-1}d^3\mathbf{p}]^{1/2}=(2\pi)^{1/2}\epsilon\|f\|_\tau,\tag{4.22}$$

$$\|\tilde f_{2,\epsilon}\|_2=[\int_{|\mathbf{p}|\geq\epsilon}|\mathbf{p}|\,|\tilde f(\mathbf{p})|^2|\mathbf{p}|^{-1}d^3\mathbf{p}]^{1/2}\leq\epsilon^{-1/2}\|f\|_\tau,\tag{4.23}$$

$$\|\omega^{1/2}\tilde\Psi_\epsilon\|_1=\epsilon^{1/2}\|\omega^{1/2}\tilde\Psi\|_1,\tag{4.24}$$

$$\|\omega^{1/2}\tilde\Psi_\epsilon\|_2=\epsilon^{-1}\|\omega^{1/2}\tilde\Psi\|_2,\tag{4.25}$$

$$\|\tilde\Psi_\epsilon\|_1=\|\tilde\Psi\|_1.\tag{4.26}$$

Equations (4.21)–(4.26) imply that (4.16) is satisfied, putting

$$C=(2\pi)^{-3/2}[(2\pi)^{1/2}\|\omega^{1/2}\tilde\Psi\|_2+2\|\tilde\Psi\|_1+\|\omega^{1/2}\tilde\Psi\|_1].$$ ∎

Corollary 4.8: If $f\in\mathcal{H}_\tau^{(0)}$, f is the limit of $\Psi_\epsilon f$ for $\epsilon\to 0$ in the weak topology.

Proof: We have to prove that $(g,\Psi_\epsilon f)_\tau\xrightarrow[\epsilon\to 0]{}(g,f)_\tau$ for any $g\in\mathcal{H}_\tau^{(0)}$, which is an easy consequence of (4.15) and (4.16), because $\bar{\mathcal{S}}^\tau=\mathcal{H}_\tau^{(0)}$. ∎

5. PROOF OF THEOREM 3.4 IN THE $m=0$ CASE

The essential part of the proof is the following lemma:

Lemma 5.1: If $B\in\tilde{\mathcal{B}}$ (see Sec. 3 for the definition),

$$F_\tau^{(0)}(B)=\overline{\{f\in\mathcal{H}_\tau^{(0)}\,|\,\mathrm{supp}f\subset\bar B\}}.\tag{5.1}$$

Proof: Let us define $\mathcal{J}_\tau(B)\equiv\{f\in\mathcal{H}_\tau^{(0)}\,|\,\mathrm{supp}f\subset\bar B\}$. Equation (4.7) implies that $\mathcal{J}_\tau(B)$ is a closed subspace of $\mathcal{H}_\tau^{(0)}$ and that $F_\tau^{(0)}(B)\subseteq\mathcal{J}_\tau(B)$.

Corollary 4.8 implies that the subset of $\mathcal{J}_\tau(B)$ made up of the compact support functions, $\mathcal{J}_\tau^c(B)$, is dense in $\mathcal{J}_\tau(B)$ in the weak topology; but $\mathcal{J}_\tau^c(B)$ is a convex set, then its weak closure and its norm closure coincide (see Ref. 12, Theorem V. 3.13), that is, $\mathcal{J}_\tau^c(B)$ is dense in $\mathcal{J}_\tau(B)$. In order to prove (5.1), that is $\mathcal{J}_\tau^{(0)}(B)=\mathcal{J}_\tau(B)$, it is then sufficient to show [see (3.17)] that any $f\in\mathcal{J}_\tau^c(B)$ can be approximated by a function $\varphi\in D$ with $\mathrm{supp}\varphi\subset B$, as well as we want. Due to Theorem 4.6, if $\{\alpha_i\}_{i=1}^N$ is a partition of the unity

$$\alpha_i\in D,\quad\sum_1^N{}_i\,\alpha_i(\mathbf{x})=1\quad\text{on supp}f\;,$$

we have

$$f=\sum_1^N{}_i f_i\quad\text{with } f_i=\alpha_i f\in\mathcal{J}_\tau^c(B).$$

Due to conditions (d) and (e) in the definition of $\tilde{\mathcal{B}}$ we can choose $\{\alpha_i\}$ in such a way that there exists, for any i, a vector $\mathbf{b}_i$ such that $\mathrm{supp}f_i(\mathbf{x}-\lambda\mathbf{b}_i)\subset B$ for small $\lambda>0$. If we put $f_{i\lambda}(\mathbf{x})=f_i(\mathbf{x}-\lambda\mathbf{b}_i)$ we have: $\|f_{i\lambda}-f_i\|_\tau\xrightarrow[\lambda\to 0]{}0$. This implies that f can be approximated by a function $g\in\mathcal{J}_\tau^c(B)$ with $\mathrm{supp}g\subset B$. On the other hand, if $J_\epsilon(\mathbf{x})=\epsilon^{-3}J(\mathbf{x}/\epsilon)$, where $J\in D$, $J(\mathbf{x})\geq 0$, $J(\mathbf{x})=0$ for $|\mathbf{x}|>1$

and $\int J(\mathbf{x})\,d^3\mathbf{x}=1$, $J_\epsilon * g$ has support contained in B for ϵ small enough, $J_\epsilon * g \in D$ and $\|J_\epsilon * g - g\|_\pi \xrightarrow[\epsilon\to 0]{} 0$. Then (5.1) is proven. ∎

Equation (3.31) (the duality relation) easily follows from Lemma 5.1. In fact, if $h \in F_\varphi^{(0)}(B)^\perp$, we have

$$(\varphi, h)_\varphi = (\omega^{-1/2}\tilde{\varphi}, \omega^{1/2}(\omega^{-1}\tilde{h}))_2 = 0, \ \forall\ \varphi \in D_r(B). \quad (5.2)$$

If we put $\tilde{f} = \omega^{-1}\tilde{h}$, $\tilde{f}$ is the Fourier transform of a function $f \in \mathcal{H}_\pi^{(0)}$ such that $\operatorname{supp} f \subset B^c = \overline{B^\perp}$. Therefore, if $B^\perp \in \mathcal{B}$

$$F_\varphi^{(0)}(B)^\perp = \omega\{f \in \mathcal{H}_\pi^{(0)} \mid \operatorname{supp} f \subset \overline{B^\perp}\} = \omega F_\pi^{(0)}(B^\perp). \quad (5.3)$$

It is worth noticing that (5.3) implies that, if $B^\perp \in \tilde{\mathcal{B}}$

$$F_\varphi^{(0)}(B) = [\omega F_\pi^{(0)}(B^\perp)]^\perp = \{f \in \mathcal{H}_\varphi^{(0)} \mid \operatorname{supp} f \subset \bar{B}\}. \quad (5.4)$$

We now prove two other lemmas.

Lemma 5.2:

$$\bar{F}_{\varphi,\pi}^{(0)}(B) = \bar{F}_{\varphi,\pi}^{(0)}(\bar{B}) = F_{\varphi,\pi}^{(0)}(B), \quad \text{if } B \in \tilde{\mathcal{B}}. \quad (5.5)$$

Moreover, if E is closed

$$\bar{F}_{\varphi,\pi}^{(0)}(E) = \{f \in \mathcal{H}_{\varphi,\pi}^{(0)} \mid \operatorname{supp} f \subset E\}. \quad (5.6)$$

Proof: It is an immediate consequence of (3.20), (5.1), and (5.4). ∎

Lemma 5.3. If $B_1, B_2 \in \tilde{\mathcal{B}}$

$$F_\pi^{(0)}(B_1) \wedge F_\pi^{(0)}(B_2) = F_\pi^{(0)}(B_1 \cap B_2), \quad (5.7)$$

$$F_\pi^{(0)}(B_1) \vee F_\pi^{(0)}(B_2) = F_\pi^{(0)}(\operatorname{int}(\overline{B_1 \cap B_2})). \quad (5.8)$$

Proof: By Eqs. (5.1) and (3.20) $F_\pi^{(0)}(B_1) \wedge F_\pi^{(0)}(B_2) = \bar{F}_\pi^{(0)}(\bar{B}_1 \cap \bar{B}_2)$. The set $\bar{B}_1 \cap \bar{B}_2$ differs from the set $\overline{B_1 \cap B_2}$ for a set of measure zero and the elements of $\mathcal{H}_\pi^{(0)}$ are locally L^2 functions, by Lemma 4.5. Then Eq. (5.6) implies that $\bar{F}_\pi^{(0)}(\bar{B}_1 \cap \bar{B}_2) = \bar{F}_\pi^{(0)}(\overline{B_1 \cap B_2}) = F_\pi^{(0)}(B_1 \cap B_2)$ and (5.7) is proven. To prove (5.8), it is sufficient to show that, if $\varphi \in D$, $\operatorname{supp}\varphi \subset \operatorname{int}(\overline{B_1 \cup B_2})$ and $\epsilon > 0$, there exist two functions f_1 and f_2, such that $f_1 \in F_\pi^{(0)}(B_1)$, $f_2 \in F_\pi^{(0)}(B_2)$ and $\|f_1 + f_2 - \varphi\|_\pi < \epsilon$. Araki has shown[4] that there exist f_1 and f_2, such that $f_1 \in F_\pi^{(m)}(B_1)$ and $f_2 \in F_\pi^{(m)}(B_2)$, $m > 0$, and $\|f_1 + f_2 - \varphi\|_{\mathcal{H}_\pi^{(m)}} < \epsilon$. However, if $f \in \mathcal{H}_\pi^{(m)}$, $m > 0$, $f \in \mathcal{H}_\pi^{(0)}$ and $\|f\|_\pi \leq \|f\|_{\mathcal{H}_\pi^{(m)}}$; therefore the desired result is an immediate consequence of Araki's. ∎

Observe now that, by the duality relation (3.31)

$$\omega[F_\pi^{(0)}(B_1) \wedge F_\pi^{(0)}(B_2)]^\perp = F_\varphi^{(0)}(B_1^\perp) \vee F_\varphi^{(0)}(B_2^\perp), \quad (5.9)$$

$$\omega[F_\pi^{(0)}(B_1) \vee F_\pi^{(0)}(B_2)]^\perp = F_\varphi^{(0)}(B_1^\perp) \wedge F_\varphi^{(0)}(B_2^\perp). \quad (5.10)$$

Equations (5.3), (5.7), and (5.9) immediately imply that, if $B_1, B_2 \in \tilde{\mathcal{B}}$

$$F_\varphi^{(0)}(B_1) \vee F_\varphi^{(0)}(B_2) = F_\varphi^{(0)}(\operatorname{int}(\overline{B_1 \cup B_2})). \quad (5.11)$$

If also $B_1^\perp \vee B_2^\perp \in \tilde{\mathcal{B}}$, Eqs. (5.3), (5.8), and (5.10) imply that

$$F_\varphi^{(0)}(B_1) \wedge F_\varphi^{(0)}(B_2) = F_\varphi^{(0)}(B_1 \cap B_2). \quad (5.12)$$

Then Eqs. (3.27) and (3.28) are proven. Equation (3.29) is an easy consequence of (4.8). Finally, by Lemma 5.2 and Eq. (3.27), if $B \in \tilde{\mathcal{B}}$,

$$\bar{F}_{\varphi,\pi}^{(0)}(\partial B) = \bar{F}_{\varphi,\pi}^{(0)}(\bar{B}) \wedge \bar{F}_{\varphi,\pi}^{(0)}(\overline{B^\perp}) = F_{\varphi,\pi}^{(0)}(B) \wedge F_{\varphi,\pi}^{(0)}(B^\perp) = \{0\}. \quad (5.13)$$

The proof is now complete.

6. THE LOCAL ALGEBRAS FOR THE ELECTROMAGNETIC FREE FIELD

As we saw in Sec. 2, there are many equivalent definitions of the one-particle Hilbert space for the free scalar field, but the more suitable definition for the study of the local algebras is the definition in terms of the initial conditions of the classical solutions of the Klein–Gordon equation. This is obviously true also for the electromagnetic free field, provided we start from the Maxwell equations:

$$\partial_\mu F_{\mu\nu} = 0, \quad \epsilon_{\mu\nu\rho\sigma}\partial_\nu F_{\rho\sigma} = 0, \quad F_{\mu\nu} = -F_{\nu\mu}. \quad (6.1)$$

Equations (6.1) imply, as it is well known, the wave equation

$$\Box F_{\mu\nu} = 0 \quad (6.2)$$

but they are not completely contained in it. Therefore, we cannot simply reduce the problems about the electromagnetic free field to problems about the scalar free field of zero mass.

Sometimes the tensor $F_{\mu\nu}$ is expressed in terms of the vector potential A_μ: $F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$, but this adds some well known gauge invariance problems.[13] Therefore, we choose to make all the construction in terms of the physical quantity $F_{\mu\nu}$. Let us now define the one-particle Hilbert space H. The antisymmetric tensor $F_{\mu\nu}$ can be expressed in terms of the six components

$$E_i = F_{0i}, \quad B_i = \tfrac{1}{2}\epsilon_{0i\mu\nu}F_{\mu\nu}, \quad i = 1, 2, 3 \quad (6.3)$$

which satisfy the wave equation: $\Box E_i = \Box B_i = 0$. Therefore, a solution of Eq. (6.2) is uniquely determined if we give the values at $t=0$ of $E_i(t,\mathbf{x})$, $B_i(t,\mathbf{x})$, $\dot{E}_i(t,\mathbf{x})$, $\dot{B}_i(t,\mathbf{x})$. However the Eqs. (6.1) allow us to express $\dot{E}_i(t,\mathbf{x})$ and $\dot{B}_i(t,\mathbf{x})$ in terms of $B_i(t,\mathbf{x})$ and $E_i(t,\mathbf{x})$ and impose the conditions

$$\partial_i E_i(t,\mathbf{x}) = \partial_i B_i(t,\mathbf{x}) = 0. \quad (6.4)$$

Therefore, a C^∞ solution $F_{\mu\nu}(t,\mathbf{x})$ of the Maxwell equations is uniquely determined by giving two C^∞ divergence free real fields on $\mathbb{R}^3$,

$$e_i(\mathbf{x}) = E_i(0,\mathbf{x}), \quad b_i(\mathbf{x}) = B_i(0,\mathbf{x}), \quad (6.5)$$

$$\partial_i e_i(\mathbf{x}) = \partial_i b_i(\mathbf{x}) = 0. \quad (6.6)$$

Let us now recall that the energy of the classical state described by the function $F_{\mu\nu}(t,\mathbf{x})$ is given by

$$\mathcal{E} = \tfrac{1}{2}\sum_1^3{}_i \int d^3\mathbf{x}[E_i(0,\mathbf{x})^2 + B_i(0,\mathbf{x})^2]$$

$$= \tfrac{1}{2}\sum_1^3{}_i \int d^3\mathbf{p}[|\tilde{e}_i(\mathbf{p})|^2 + |\tilde{b}_i(\mathbf{p})|^2]. \quad (6.7)$$

This implies that the average number of photons in the corresponding coherent quantum state[14] is

$$\bar{m} = \tfrac{1}{2}\sum_1^3{}_i \int \frac{d^3\mathbf{p}}{|\mathbf{p}|}[|\tilde{e}_i(\mathbf{p})|^2 + |\tilde{b}_i(\mathbf{p})|^2]. \quad (6.8)$$

This justifies the following definition:

Definition 6.1: Let $\mathcal{G}$ be the space of the C^∞ solutions $F_{\mu\nu}$ of the Maxwell equations such that $e_i \in D \equiv D_r(\mathbb{R}^3)$, $b_i \in D$, $i = 1, 2, 3$; then the one-particle real Hilbert space of the electromagnetic free field, H, is the completion of $\mathcal{G}$ with respect to the scalar product,

$$(F', F) = \sum_1^3{}_i \int \frac{d^3\mathbf{p}}{|\mathbf{p}|} [\tilde{e}_i'^*(\mathbf{p})\tilde{e}_i(\mathbf{p}) + b_i'^*(\mathbf{p})\tilde{b}_i(\mathbf{p})]. \qquad (6.9)$$

Lemma 6.2: Let K be the closure in H of the C^∞ solution of the Maxwell equations such that $e_i \in D$, $b_i = 0$. K can be identified with the real Hilbert space

$$\mathcal{H}_\varphi^{(\mathrm{e.m.})} = \{\mathbf{f}(\mathbf{x}) \,|\, f_i \in \mathcal{H}_\varphi^{(0)}, \textstyle\sum_i p_i \tilde{f}_i(\mathbf{p}) = 0\}, \qquad (6.10)$$

where $\mathcal{H}_\varphi^{(0)}$ is the space defined in Sec. 2, $\mathbf{f}(\mathbf{x})$ denotes a field of distributions on $\mathbf{R}^3$ with components $f_i(\mathbf{x})$, $i = 1, 2, 3$, and $(\mathbf{f}, \mathbf{g})_\varphi = \sum_{1i}^3 (f_i g_i)_\varphi$ is the scalar product in $\mathcal{H}_\varphi^{(\mathrm{e.m.})}$.

Proof: Equations (4.8), (6.6), and (6.9) imply that $K \subset_{\mathrm{cont}} \mathcal{H}_\varphi^{(\mathrm{e.m.})}$. Then, to prove the lemma, it is sufficient to show that, if $\mathbf{g} \in \mathcal{H}_\varphi^{(\mathrm{e.m.})}$ and

$$\sum_i \int \frac{d^3\mathbf{p}}{|\mathbf{p}|} \tilde{\varphi}_i(\mathbf{p})\tilde{g}_i(\mathbf{p}) = 0, \ \forall\ \phi_i \in D, \quad \partial_i\varphi_i = 0, \qquad (6.11)$$

then $g = 0$. If we define $\tilde{h}_i(\mathbf{p}) = |\mathbf{p}|^{-1}\tilde{g}_i(\mathbf{p})$ and we consider the fields $\varphi = \nabla \times \mathbf{a}$, $a_i \in D$, Eq. (6.11) gives

$$h_i(\varphi_i) = -(\nabla \times \mathbf{h})_i(a_i) = 0, \ \forall\ a_i \in D; \qquad (6.12)$$

that is $\nabla \times \mathbf{h} = 0$. Thererore, there exists[10] a distribution λ such that $h_i = \partial_i\lambda$. But $\partial_i h_i = 0$, then $\Delta\lambda = 0$, which implies $\lambda = 0$. ∎

The space H can be identified with the space of the couples $\langle e_i, b_i\rangle$, $\mathbf{e} \in \mathcal{H}_\varphi^{(\mathrm{e.m.})}$, $\mathbf{b} \in \mathcal{H}_\varphi^{(\mathrm{e.m.})}$. Define now on H the operator

$$\beta\langle \mathbf{e}, \mathbf{b}\rangle = \langle \mathbf{e}', \mathbf{b}'\rangle,$$
$$\tilde{\mathbf{e}}' = -\frac{i}{|\mathbf{p}|}\mathbf{p} \times \tilde{\mathbf{b}}, \quad \tilde{\mathbf{b}}' = \frac{i}{|\mathbf{p}|}\mathbf{p} \times \tilde{\mathbf{e}}. \qquad (6.13)$$

It is easy to show that β is an antiunitary operator such that $\beta^2 = -1$ and

$$H = K \oplus \beta K. \qquad (6.14)$$

If we define $if = \beta f$, $f \in H$, H becomes a complex Hilbert space and we can define the Segal field operator and the canonical field and momentum as in Sec. 2. Moreover, we can define $R(L)$ and $R(K_1, K_2)$ as in Sec. 3 and Theorem 3.1 is still valid.

Now let $B \subset \mathbf{R}^3$ be an open set which satisfies the following conditions:

(a) $B \in \tilde{\mathcal{B}}$ (see Sec. 3).

(b) If $B = \bigcup_1^n{}_i B_i$, where $\{B_i\}_{i=1}^n$ is defined as in the definition of $\tilde{\mathcal{B}}$, B_i is simply connected for $i = 1, \ldots, n$.

Condition (b) implies that the compact support second cohomology for B is zero, which can be expressed in the following less technical form:

(c) If φ is the field such that $\partial_i\varphi_i = 0$, $\varphi_i \in D$ and $\mathrm{supp}\varphi_i \subset B$, there exists a field $\mathbf{a}$, such that $a_i \subset D$, $\mathrm{supp}a_i \subset B$ and $\varphi = \nabla \times \mathbf{a}$.

This result easily follows from the de Rham duality (see Theorem 17′ in Ref. 15).

If $F_{\mu\nu}$ is an element of H, such that $e_i = 0$ and $b_i \in D$, by (6.6) and condition (c) we can find a field $\mathbf{a}(\mathbf{x})$, such that $\mathbf{b} = \nabla \times \mathbf{a}$, $a_i \in D$. Equation (6.13) then implies that

$$-\beta\langle 0, \mathbf{b}\rangle = \langle \omega P\mathbf{a}, 0\rangle, \qquad (6.15)$$

where ω is the operator of multiplication of the Fourier transform by $|\mathbf{p}|$ and

$$(\widetilde{P\mathbf{a}})_i = \left(\delta_{i}, -\frac{p_i p_j}{|\mathbf{p}|^2}\right)\tilde{a}_j(\mathbf{p}). \qquad (6.16)$$

It is worth noticing that, in general, $(P\mathbf{a})_i \notin D$.

Lemma 6.3: βK can be identified with the real Hilbert space,

$$\mathcal{H}_\pi^{(\mathrm{e.m.})} = \{\mathbf{f}(\mathbf{x}) \,|\, f_i \in \mathcal{H}_\pi^{(0)}, \textstyle\sum_i p_i \mathbf{f}_i(\mathbf{p}) = 0\}, \qquad (6.17)$$

with scalar product $(\mathbf{f}, \mathbf{g})_\pi = \sum_{1\,i}^3 (f_i, g_i)_\pi$.

Proof: Equation (6.15) implies that βK can be identified with the completion with respect to the scalar product $(\mathbf{f}, \mathbf{g})_\pi = \sum_{1\,i}^3 (f_i, g_i)_\pi$ of the fields $P\mathbf{a}$, $a_i \in D$. However, if $\mathbf{f} \in \mathcal{H}_\pi^{(\mathrm{e.m.})}$, by (4.8) there exists a sequence $\mathbf{a}^{(n)}$, such that $a_i^{(n)} \in D$ and $\|a_i^{(n)} - f_i\|_\pi \to 0$. Moreover (6.16) implies that $\|Pa_i^{(n)} - f_i\|_\pi = \|Pa_i^{(n)} - Pf_i\|_\pi \to 0$, then $f \in \beta K$. ∎

Definition 6.4: Let $B \subset \mathbb{R}^3$ be an open set satisfying conditions (a), (b). Define:

$$F_\varphi^{(\mathrm{e.m.})}(B) = \overline{\{\varphi(\mathbf{x}) \,|\, \varphi_i \in D,\ \partial_i\varphi_i = 0, \mathrm{supp}\varphi_i \subset B\}}^{\mathcal{H}_\varphi^{(\mathrm{e.m.})}}$$
$$F_\pi^{(\mathrm{e.m.})}(B) = \overline{\{(\overline{P\mathbf{a}})(\mathbf{x}) \,|\, a_i \in D, \mathrm{supp}a_i \subset B\}}^{\mathcal{H}_\pi^{(\mathrm{e.m.})}} \qquad (6.18)$$

Now let $R(C(B))$ be the Von Neumann algebra associated by Eq. (3.1) to the subspace of H, (6.19)

$$L_B = \overline{\{F_{\mu\nu} \in H \,|\, e_i \in D, b_i \in D, \mathrm{supp}e_i \subset B, \mathrm{supp}b_i \subset B\}}^H. \qquad (6.20)$$

By Eqs. (6.15), (6.18), and (6.19), we have [see Eq. (3.3)]

$$R(C(B)) = R(F_\varphi^{(\mathrm{e.m.})}(B), \omega F_\pi^{(\mathrm{e.m.})}(B)). \qquad (6.21)$$

At this point our aim should be the proof of Theorem 3.3, which again is equivalent to Theorem 3.4, provided we substitute $F_{\varphi,\pi}^{(m)}(B)$ by $F_{\varphi,\pi}^{(\mathrm{e.m.})}(B)$. However now the task is more difficult than in the scalar case. In fact, the analog of Lemma 5.1,

$$F_\pi^{(\mathrm{e.m.})}(B) = \{P\ \mathbf{f} \in \mathcal{H}_\pi^{(\mathrm{e.m.})} \,|\, \mathrm{supp}f_i \subset \bar{B}\}, \qquad (6.22)$$

is not an easy consequence of Eq. (5.1) and the definition (6.19), moreover (6.22) should not simply imply the duality relation

$$F_\varphi^{(\mathrm{e.m.})}(B)^\perp = \omega F_\pi^{(\mathrm{e.m.})}(B^\perp). \qquad (6.23)$$

In fact it is easy to verify, using condition (c), that

$$F_\varphi^{(\mathrm{e.m.})}(B)^\perp = \omega\{\mathbf{f} \in \mathcal{H}_\pi^{(\mathrm{e.m.})} \,|\, \mathrm{supp}(\nabla \times \mathbf{f})_i \subset \bar{B}^\perp\}. \qquad (6.24)$$

The proof that the rhs of (6.23) and (6.24) are equal can be reduced to the following very reasonable ansatz:

Ansatz: If $\mathbf{f} \in \mathcal{H}_\pi^{(\mathrm{e.m.})}$ and $\mathrm{supp}(\nabla \times \mathbf{f})_i \subset \bar{B}^\perp$, there is a distribution λ such that $\partial_i\lambda \in \mathcal{H}_\pi^{(0)}$ and $\nabla\lambda = \mathbf{f}$ in B.

Let us in fact suppose that this ansatz is true. Then, if $\omega\mathbf{f} \in F_\varphi^{(\mathrm{e.m.})\perp}$ and we define $\mathbf{g} = \mathbf{f} - \nabla\lambda$, $\mathrm{supp}g_i \subset \bar{B}^\perp$ and $P\mathbf{g} = \mathbf{f}$. It follows, by (5.1), that there exists a sequence $\{\mathbf{a}^{(n)}\}$ with $a_i^{(n)} \in D$, $\mathrm{supp}a_i^{(n)} \subset B^\perp$ and $\|a_i^{(n)} - g_i\|_\pi \underset{n\to\infty}{\to} 0$, which implies $\|Pa_i^{(n)} - f_i\|_{\tau} \underset{n\to\infty}{\to} 0$. The remaining part of the proof of the analog of Theorem 3.4 is now essentially the same as in the scalar case. Then all the problem is the proof of the ansatz, which is equivalent by (6.23), to the following equation:

$$F_{\varphi}^{(e.m.)}(B) = \{\mathbf{f} \in \mathcal{H}_{\varphi}^{(e.m.)} \mid \operatorname{supp} f_i \subset \bar{B}\}. \qquad (6.25)$$

The previous argument also implies that Eq. (6.22) should be true. Unfortunately we do not still have a complete proof of the ansatz or Eq. (6.25).

ACKNOWLEDGMENTS

We are indebted to Professor S. Doplicher and Dr. Wolf-Dieter Garber for many useful discussions. We also acknowledge the referee of this work for some useful suggestions.

[1]R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964).
[2]S. Doplicher, R. Haag, and J.E. Roberts, Commun. Math. Phys. 13, 1 (1969); 15, 173 (1969); 23, 199 (1971); 35, 49 (1974).
[3]H. Araki, J. Math. Phys. 4, 1343 (1963).
[4]H. Araki, J. Math. Phys. 5, 1 (1964).
[5]V. Fock, Z. Phys. 75, 622 (1932).
[6]J. Cook, Trans. Amer. Math. Soc. 74, 222 (1953); I.E. Segal, Trans. Amer. Math. Soc. 81, 106 (1956); Ann. Math. 63, 160 (1956).
[7]If $\mathcal{A}$ is a set of bounded operators on $\mathcal{F}_s$, we denote by $\{\mathcal{A}\}$ the linear subspace generated by $\mathcal{A}$ and by $\mathcal{A}'$ the commutant of $\mathcal{A}$.
[8]G.F. Dell'Antonio, Commun. Math. Phys. 9, 81 (1968).
[9]We use in general the definitions of Ref. 11.
[10]L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1957).
[11]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vols. I, II (Academic, New York, 1972, 1975).
[12]N. Dunford and J.T. Schwartz, *Linear Operators I* (Interscience, New York, 1957).
[13]F. Strocchi and A.S. Wightman, J. Math. Phys. 15, 2198 (1974).
[14]R.J. Glauber, Phys. Rev. 131, 2766 (1963).
[15]G. De Rham, *Variétés différentiables* (Hermann, Paris, 1960).

Type N gravitational field with twist. II[a]

I. Hauser

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 20 June 1977)

A full derivation is given of that one parameter family of type N twisting gravitational fields which was previously reported by the author. Explicit forms of this solution are obtained, and the coordinate ranges are specified for all possible cases. The general problem of the search for other type N twisting gravitational fields is also discussed; the differential equations required for that search are derived, and the extension of these equations to type (3,1) is given without derivation.

1. INTRODUCTION

Let TNT denote any type N [i.e., type (4)] vacuum solution of the Einstein field equations such that the principal null rays are twisting.

The only presently known TNT is a one-parameter family of metrics each of which has exactly one Killing vector.[1] The derivation of this solution has not yet been published, since the author's first paper (I) on the subject was necessarily brief.[1] The main purpose of the present paper is to fill this gap, both to disseminate the possibly useful ideas involved in the derivation and to furnish a basis for the discussion of the current search for more TNT's. A third paper (III) will cover what little the author has been able to determine about the geometry and physics of the known solution; some interesting work along these lines has already been done by Sommers and Walker.[2]

The TNT field equations used by the author differ significantly from those used by others[3] and are given in Sec. 2, together with a sketch of their derivation. We do not follow the common practice of first constructting a general formalism for all algebraically special vacuums with diverging null rays and then abstracting the TNT equations as a particular case. Therefore, the author's choice of a null tetrad is dictated only by the requirement that the corresponding affine connection forms have maximal simplicity for any TNT. The choice of coordinates and of dependent variables are prompted by what is already known from the completely solved problem of type N with diverging null rays and zero twist.[4] The other Petrov types play no role in guiding our thinking, though it is conceivable that the author's treatment may be advantageously extended to type (3, 1).

In Sec. 3, the particular case for which the author has found a TNT is defined, and the derivation of the solution is given. The results include specific forms for the solution in terms of hypergeometric functions and a clear-cut description of the range of the chart. The unorthodox notations for coordinates in (I) are replaced by the notations of Robinson and Trautman[4,5]; other notations are unchanged except for the use of A^* in place of A.

[a]Research supported in part by the National Science Foundation under Grant No. PHY75-08750.

In the discussion of Sec. 4, we show that the number of essential parameters in the solution is one, and we describe some aspects of our search for new TNT's.

Since a type (3, 1) solution with twisting null rays and with a radiative term in the conform tensor has yet to be discovered,[5] there may be some interest in the extension of the TNT field equations of Sec. 2 to type (3, 1). This extension is given (without derivation) in Sec. 4. Also, we point out the nonexistence of the simplest imaginable type (3, 1) analog[6] of the author's TNT.

2. THE TNT FIELDS EQUATIONS

First, we give a brief summary of our notations and conventions respecting the Einstein field equations in terms of differential forms and a null tetrad.

Consider any null tetrad[7] which consists of the 1-forms k, m, t, t^* such that k and m are real, t^* is the complex conjugate of t, and their inner products have values $k\cdot m = t\cdot t^* = 1$. The symbol $\wedge$ is omitted from exterior products of forms; e.g., km means $k\wedge m$ and dk means $d\wedge k$. The exterior derivatives of the null tetrad are given, for any spacetime, by

$$dk = \tfrac{1}{2}(v_0 + v_0^*)k + v_1^* t + v_1 t^*, \tag{1a}$$

$$dm = -\tfrac{1}{2}(v_0 + v_0^*)m + v_{-1}t + v_{-1}^* t^*, \tag{1b}$$

$$dt = -v_{-1}^* k - v_1 m + \tfrac{1}{2}(v_0 - v_0^*)t, \tag{1c}$$

where the connection forms v_1, v_0, v_{-1} are defined by

$$v_1 := dx^\alpha t^\beta \nabla_\alpha k_\beta, \quad v_{-1} := dx^\alpha t^{\beta *}\nabla_\alpha m_\beta,$$

$$v_0 := dx^\alpha (m^\beta \nabla_\alpha k_\beta + t^{\beta *}\nabla_\alpha t_\beta).$$

The vacuum field equations are given by

$$dv_1 + v_1 v_0 = C_0 kt + C_1(km + tt^*) + C_2 mt^*, \tag{2a}$$

$$-\tfrac{1}{2}dv_0 + v_1 v_{-1} = C_{-1}kt + C_0(km + tt^*) + C_1 mt^*, \tag{2b}$$

$$dv_{-1} + v_0 v_{-1} = C_{-2}kt + C_{-1}(km + tt^*) + C_0 mt^*. \tag{2c}$$

The scripts A and B on v_A and on the conform tensor components C_{A+B} are helicity scripts and have values 1, 0, −1.

As is customary, the null tetrad is chosen so that k is a principal null form. Then, for type N,

$$C_{A+B} = 0 \quad \text{except } C_{-2} \neq 0. \tag{3}$$

Also, the Goldberg–Sachs theorem tells us that k is a shear-free null geodesic, which means $k\cdot v_1 = t\cdot v_1 = 0$.

Therefore, for nonzero expansion,

$$v_1 = z(A^*k + t), \quad z \neq 0, \tag{4}$$

where $z := t^* \cdot v_1$ and $zA^* := m \cdot v_1$. The real part of z is the expansion and the imaginary part of z is the twist of the null rays.

When making comparisons with I, *note that the author formerly*[1] *used A to denote the complex conjugate of the present A;* the reason for the notational change can be seen later by inspecting Eqs. (14) and (17c).

We are now ready to consider the derivation of the TNT field equations. The first phase of the derivation is the selection of an appropriate null tetrad. The null tetrad which we have so far is still arbitrary up to any transformation which leaves k invariant except for multiplication by a scalar field. The corresponding transformation of the connection forms is given by

$$v_1 \to e^{\psi} v_1,$$

$$v_0 \to v_0 + 2\lambda v_1 + d\psi,$$

$$v_{-1} \to e^{-\psi}(v_{-1} + \lambda v_0 + \lambda^2 v_1 + d\lambda), \tag{5}$$

where ψ and λ are any complex scalar fields. We now take advantage of this group of transformations to choose simple connection forms. With the aid of Eqs. (2a)–(2b), (4), and the complex version of the theorem of Frobenius, we select a null tetrad so that

$$v_1 = d\zeta, \quad d\zeta d\zeta^* \neq 0,$$
$$v_{-1} = h d\zeta, \quad v_0 = 0. \tag{6}$$

ζ and h are complex scalar fields.

Though we have considerably narrowed our choice of a null tetrad, there is a residual arbitrariness which is given by that subgroup of the group of transformations (5) such that

$$\psi = F(\zeta), \quad \lambda = -\frac{1}{2}\frac{dF(\zeta)}{d\zeta}, \tag{7}$$

where $F(\zeta)$ is an analytic function of ζ. There is no obvious way of employing this remaining freedom of choice to achieve a useful simplification of the *general* TNT equations. Therefore, the group defined by Eqs. (5) and (7) is held in reserve to be used in particular cases. For example, when a Killing vector exists, the residual group can be used to select a null tetrad whose members commute with the given Killing vector and whose connection forms still satisfy equations of the form (6).

That completes the first phase of the derivation. The second phase is the choice of a coordinate system. Let

$$z^{-1} = \rho + i\tau, \quad \text{real } \rho \text{ and } \tau. \tag{8}$$

ρ, ζ, ζ^* will be three of the coordinates. The task is to find a suitable fourth.

It is convenient to start by showing that k, $d\rho$, $d\zeta$, $d\zeta^*$ are linearly independent. Substitution from Eqs. (4), (6), and (8) into Eq. (1a) yields

$$dk = k\theta - 2i\tau\, d\zeta\, d\zeta^*, \quad \theta := A d\zeta + A^* d\zeta^*. \tag{9}$$

From a comparison of the expression for dt in Eq. (1c) with dt as computed from Eqs. (4) and (9),

$$dt\, d\zeta^* k = k(dz^{-1}) d\zeta\, d\zeta^* = (zz^*) kmtt^*. \tag{10}$$

Therefore, k, $d\rho$, $d\zeta$, $d\zeta^*$ are linearly independent, and the differential of any scalar field f is expressible as

$$df = (d\rho X + kY + d\zeta D + d\zeta^* D^*) f. \tag{11}$$

The commutators of the directional derivations X, Y, D, D^* are required for later computations. From Eqs. (9) and (11) and from $d^2 f = 0$,

$$[X, Y] = [X, D] = 0, \quad [D, Y] = AY,$$
$$[D, D^*] = 2i\tau Y. \tag{12}$$

The above Eqs. (9)–(11) imply the existence of a scalar field σ such that ρ, σ, ζ, ζ^* is a chart, and

$$X = \frac{\partial}{\partial \rho}, \tag{13a}$$

$$Y = \frac{1}{p}\frac{\partial}{\partial \sigma}, \tag{13b}$$

$$D = \frac{\partial}{\partial \zeta} - \Omega \frac{\partial}{\partial \sigma}. \tag{13c}$$

p is a real and Ω is a complex scalar field. The domain of the chart is an open connected set in which $p \neq 0$. Since the above defining equations for p, Ω, and σ are invariant under the transformation

$$p \to -p, \quad \Omega \to -\Omega, \quad \sigma \to -\sigma,$$

we can always select σ so that $p > 0$. It will be understood that the positive p alternative holds from now on.

Equations (12) and (13) imply that p and Ω are independent of ρ; i.e., both fields are constant along each principal null geodesic. All other ρ-independent scalar fields in our equations are expressible in terms of p and Ω or their derivatives. In particular, from Eqs. (12) and (13),

$$A = -p^{-1}\mathcal{D} p, \tag{14}$$

$$\Delta := p^{-1}\tau = \tfrac{1}{2} i (D\Omega^* - D^*\Omega), \tag{15}$$

where

$$\mathcal{D} := D - \partial_\sigma \Omega, \quad \partial_\sigma := \partial/\partial\sigma. \tag{16}$$

All of the above Eqs. (13)–(16) are for a fixed null tetrad. For a fixed null tetrad, the coordinates may sill be subject to any mapping of the form

$$\rho \to \rho, \quad \sigma \to \sigma'(\sigma, \zeta, \zeta^*), \quad \partial_\sigma \sigma' > 0,$$

$$\zeta \to \zeta + \zeta_0, \quad \zeta_0 = \text{const}.$$

The general admissible coordinate transformation is the direct product of the above mapping with the one which is induced by the null tetrad change represented by Eqs. (5) and (7). The various scalar fields which we have defined undergo the following corresponding transformations:

$$\zeta \to \zeta' \text{ such that } d\zeta' = e^F d\zeta, \tag{17a}$$

$$z' = [\exp\tfrac{1}{2}(F+F^*)]z, \tag{17b}$$

$$A' = e^{-F}\left(A - \frac{1}{2}\frac{dF}{d\zeta}\right), \tag{17c}$$

$$h' = e^{-2F}\left[h + \frac{1}{4}\left(\frac{dF}{d\zeta}\right)^2 - \frac{1}{2}\frac{d^2F}{d\zeta^2}\right], \tag{17d}$$

$$p' = [\exp\tfrac{1}{2}(F+F^*)](\partial_\sigma \sigma')^{-1}p, \tag{17e}$$

$$\Delta' = [\exp(-F-F^*)](\partial_\sigma\sigma')\Delta, \tag{17f}$$

$$\Omega' = e^{-F}\left(\frac{\partial\sigma'}{\partial\sigma}\Omega - \frac{\partial\sigma'}{\partial\zeta}\right). \tag{17g}$$

It is clear that the coordinate σ can always be chosen so as to make either $p=1$ or $\Delta = \pm 1$, and this can be done without commitment to a particular member of the admissible null tetrad family. However, it is preferable to be guided in the choice of σ as well as of the null tetrad by auxiliary conditions of an invariantive nature, like the requirements that a Killing vector exist or that the twist be zero. For example, when the twist is zero, Eq. (9) becomes $dk = k\theta$ which implies the existence of σ and p such that $k = pd\sigma$; thus, we are automatically led to a choice of σ for which $\Omega = 0$, and it remains to solve for p. Other examples are discussed in Sec. 4.

That completes the second phase. The final act is the derivation of a viable form of the TNT field equations. We start by substituting from Eqs. (3), (4), and (6) into Eqs. (1) and (2c). It is advisable to proceed by expressing all 1-forms in terms of the basis k, $d\rho$, $d\zeta$, $d\zeta^*$ and to write all differentials of scalar fields as in Eq. (11).

Many of the scalar equations which derive from the 2-form components in Eqs. (1) and (2c) are redundant. We skip further details and simply list the nonredundant results which are obtained after judicious applications of Eqs. (12)—(16).

The final expressions for the null tetrad in terms of the coordinate differentials are

$$k = p(d\sigma + \Omega d\zeta + \Omega^* d\zeta^*), \tag{18a}$$

$$t = (\rho + i\tau)d\zeta - A^* k, \tag{18b}$$

$$m = d\rho - \tfrac{1}{2}[p^{-1}(\mathcal{D}\mathcal{D}^* + \mathcal{D}^*\mathcal{D})p]k$$
$$- i(D\tau - 2A\tau)d\zeta + i(D\tau - 2A\tau)^* d\zeta^*. \tag{18c}$$

Except for ρ itself, *all of the scalar fields in the above expressions are ρ-independent and are to be computed from p and Ω by using Eqs.* (13c)—(16).

After its choice is motivated by some invariantive condition or by a criterion of simplicity, Ω *is to be regarded as a given function which is to be fed into the equations.* For each choice of Ω, the field p is to be found by solving the pair of equations

$$\mathcal{D}^2 p = -hp, \tag{19}$$

$$[(DD^* + D^*D)\Delta + 2(D\Delta\mathcal{D}^* + D^*\Delta\mathcal{D})$$
$$+ \Delta(\mathcal{D}\mathcal{D}^* + \mathcal{D}^*\mathcal{D})]p = 0, \tag{20}$$

where h is subject to the constraint

$$D^* h = 0. \tag{21}$$

In Eq. (20), Δ is to be treated as an operator; e.g., $[(DD^* + D^*D)\Delta]p$ means $(DD^* + D^*D)(\Delta p)$. The complex scalar field h is defined by Eqs. (6). It also appears in the Riemann tensor as follows:

$$C_{-2} := m^\alpha t^{*\beta} m^\gamma t^{*\delta} R_{\alpha\beta\gamma\delta} = zN,$$

$$N = p^{-1}\partial_\sigma h. \tag{22}$$

The above TNT *field equations* (19)—(21) were derived subject only to the assumption that the null rays are diverging; i.e., they cover zero twist as a special case. If the twist is zero, Eq. (20) vanishes identically; σ can be chosen so that $\mathcal{D} = \partial/\partial\zeta$, Eq. (21) implies that $h = h(\sigma, \zeta)$, and Eq. (19) has a solution *for any given choice* of $h(\sigma,\zeta)$. Thus, in the zero twist case, any member h of the null space of D^* admits solutions.

In contrast, when there is twist, not every member h of the null space of D^* necessarily admits a solution of p of both Eqs. (19) and (20). The process of solving Eqs. (19)—(21) must include the calculation of those h which admit a solution. For example, in the solution found by the author, h is uniquely determined by the choice of $\Omega = i\Delta_0(\zeta + \zeta^*)$ where Δ_0 = real nonzero constant.

The author regards Eqs. (19) and (20) as *a pair of linear equations in p for which h plays a role roughly analogous to an eigenvalue.* h is to be determined by analyzing the successive integrability conditions for Eqs. (19) and (20), as computed with the aid of the relations

$$[D, D^*] = 2i\Delta\partial_\sigma, \tag{23a}$$

$$\partial_\sigma D = \mathcal{D}\partial_\sigma. \tag{23b}$$

Incidentally, it is interesting to note that Eqs. (19), (20), and (21) are not completely independent of each other for any given Ω corresponding to a nonzero Δ. Specifically, we now prove that *Eqs.* (19) *and* (21) *imply that the left side of Eq* (20) *is independent of σ, i.e., depends at most on ζ and ζ^*.* The proof proceeds by introducing a field ϕ such that

$$p = \partial_\sigma \phi. \tag{24}$$

Use of Eqs. (23) then shows that Eq. (20) *is equivalent to the equation*

$$(D^2 D^{*2} - D^{*2} D^2)\phi = 0. \tag{25}$$

However, from Eqs. (19) and (21), and Eq. (23b),

$$(\mathcal{D}^2\mathcal{D}^{*2} - \mathcal{D}^{*2}\mathcal{D}^2)p = \partial_\sigma(D^2D^{*2} - D^{*2}D^2)\phi = 0. \tag{26}$$

So, the theorem is proven.

It is time to give the derivation of the only presently known twisting solution[1] of Eqs. (19)—(21). *We will designate this solution as* TNT_1.

3. DERIVATION OF TNT_1

A "sophisticated" characterization of TNT_1 will be given in Sec. 4. At present, let us be satisfied with a mere criterion of simplicity and consider any TNT which admits a choice of null tetrad and coordinates such that

$$\Omega = i\Delta_0(\zeta + \zeta^*), \quad \Delta_0 = \text{real const} \neq 0. \tag{27}$$

Then, from Eq. (15),

$$\Delta = \Delta_0.$$

Consider the continuous group of null tetrad-coordinate transformations (including translations of ζ and σ) which are represented by Eqs. (17). The subgroup which leaves Δ constant and which retains the above form of Ω is given by

$$\begin{aligned} &\zeta' = e^F(\zeta + \xi_0 + i\eta_0), \quad F = \alpha + i\beta, \\ &\rho' = \exp(-\alpha)\rho, \\ &\sigma' = \gamma\{\tfrac{1}{2}(\sigma + \sigma_0) - i\Delta_0[(1 + \exp(2i\beta))\xi_0 - (1 - \exp(2i\beta))i\eta_0] \\ &\quad + \tfrac{1}{2} i\Delta_0[1 - \exp(2i\beta)]\zeta^2\} + \text{c. c.} \end{aligned} \tag{28}$$

Above, $\alpha, \beta, \gamma, \sigma_0, \xi_0, \eta_0$ are real parameters such that $\gamma > 0$. An important part of the derivation of TNT_1 is to use Eqs. (28) to select our null tetrad and coordinates so as to obtain a simple expression for h.

With our choice of Ω, the TNT field Eqs. (19) and (20) reduce to

$$D^2 p = -hp, \quad D = D, \tag{29}$$

$$(DD^* + D^*D)p = 0. \tag{30}$$

Equation (21) has the general solution

$$h = h(\sigma - i\kappa, \zeta) \quad \kappa = \tfrac{1}{2}\Delta_0(\zeta + \zeta^*)^2. \tag{31}$$

Now we solve Eqs. (29) and (30) in three phases. The first phase is to use the commutators of Eqs. (23) to obtain successive *integrability conditions* for the field equations.

By operating on Eqs. (30) with D^* and using Eq. (29), we obtain the first integrability condition

$$(h^* D + 3i\Delta_0 \partial_\sigma D^*)p = 0. \tag{32}$$

Operation on Eq. (32) with D and the use of Eq. (30) yields a second integrability condition

$$(3\Delta_0{}^2\partial_\sigma{}^2 + hh^*)p = 0. \tag{33}$$

Operation on Eq. (32) with D^* and the use of Eqs. (29) and (30) yields

$$[(Dh)^* D - 4i\Delta_0 h^* \partial_\sigma - 3i\Delta_0 \partial_\sigma h^*]p = 0. \tag{34}$$

Operate on Eq. (34) with D, and use Eq. (32) to get

$$[i\Delta_0(\partial_\sigma h^*)D + \tfrac{4}{3} hh^* D^* + h(Dh)^*]p = 0. \tag{35}$$

Operate on Eq. (35) with D, and use Eqs. (29), (30), (34), and the complex conjugate of Eq. (34), and we finally obtain a condition on h alone:

$$[(Dh)(Dh)^* + 4i\Delta_0(h\partial_\sigma h^* - h^*\partial_\sigma h)]p = 0. \tag{36}$$

Our round of integrability conditions is now completed.

The second phase is the determination of $h(\sigma - i\kappa, \zeta)$. Equations (34) and (35) and their complex conjugates form a system of four homogeneous linear equations in $Dp, D^*p, i\Delta_0\partial_\sigma p$, and p. For a nontrivial solution to exist, the 4×4 determinant formed from the coefficients in these equations must vanish. With the aid of Eq. (36), this condition reduces to the following neat and not so obvious form:

$$|(Dh^{-1})^*(i\Delta\partial_\sigma h^{-1}) - \tfrac{4}{3} Dh^{-1}|^2 = 0.$$

Therefore,

$$i\Delta_0\partial_\sigma h^{-1} = \tfrac{4}{3}\exp(ib), \tag{37a}$$

$$\exp(ib) = \frac{Dh^{-1}}{(Dh^{-1})^*}, \tag{37b}$$

where b is a real scalar field. From Eq. (21), $D^*h^{-1} = 0$. So, since $\partial_\sigma D = D\partial_\sigma$ in our particular case, $D^*b = Db = 0$ and $(DD^* - D^*D)\psi = 2i\Delta_0\partial_\sigma b = 0$. Hence b is a constant.

The constant b can be made zero by an appropriate choice of the null tetrad and coordinate system, as can be proven by using that transformation in Eqs. (28) for which $\alpha = \gamma = \sigma_0 = \xi_0 = \eta_0 = 0$ and $\beta = \tfrac{1}{2}b$. Integration of Eq. (37a) then gives us

$$h^{-1} = (4/3i\Delta_0)[\sigma - i\kappa + g(\zeta)],$$

where $g(\zeta)$ remains to be determined. From Eq. (37b), Dh^{-1} is real; this condition yields $g(\zeta) = a_0 + ia_1 + ia_2\zeta$ where a_0, a_1, a_2 are real constants. We then use that transformation in Eqs. (28) for which $\alpha = \beta = \gamma = \eta_0 = 0$ and $\xi_0 = -a_2/4\Delta_0$ while $\sigma_0 = a_0$. The new h^{-1} is as before except that $g(\zeta) = ia_3$ where a_3 is a real constant. Substitution of this latest h^{-1} into Eq. (35) shows that $a_3 = 0$, and we have

$$h^{-1} = (4/3i\Delta_0)(\sigma - i\kappa), \quad \kappa = \tfrac{1}{2}\Delta_0(\zeta + \zeta^*)^2. \tag{38}$$

That completes the calculation of h.

The final phase consists of substituting the above expression for h^{-1} into Eqs. (33) and (34), solving these equations for p, and then verifying that the solution also satisfies the TNT field equations (29) and (30). To proceed, introduce real coordinates ξ and η in place of the complex coordinates ζ and ζ^* as follows:

$$\xi + i\eta := \sqrt{2}\,\zeta. \tag{39}$$

This facilitates the breaking up of Eq. (34) into its real and imaginary parts, which are readily solved together with Eq. (33). When making comparisons with I,[1] *please note that the author's old notations (u, ξ, ρ, σ) for the real coordinate have now been replaced by the Robinson–Trautman notations $(\rho, \sigma, \xi, \eta)$.*

The solution for p was originally obtained in a simple form which excludes points at which $\xi = 0$ from the domain. It is easiest first to give this form and then to detail the circumstances under which p can be extended across the hypersurface $\xi = 0$. The solution which is applicable to the separate regions $\xi > 0$ and $\xi < 0$ is

$$\begin{aligned} &p(\sigma, \xi) = (\xi^2)^{3/4} f(y), \\ &y := \sigma/(\Delta\xi^2). \end{aligned} \tag{40}$$

$f(y)$ is any solution of the equation

$$\frac{d^2 f}{dy^2} + \frac{3f}{16(1 + y^2)} = 0, \tag{41}$$

subject only to the following constraints on the physical domain of p:

(1) the domain of p is open and connected, and $p > 0$ at all (σ, ξ) in this domain;

(2) p is at least C^∞ in the domain.

To avoid any ambiguities now or later, the positive root of $(\xi^2)^{3/4}$ in Eq. (40) is to be understood, and the same convention is adopted for all roots of positive numbers which occur in subsequent expressions.

f is expressible in terms of hypergeometric functions.[8] The even and odd solutions of Eq. (41) are given for $|y| \leq 1$ by

$$f_0(y) = {}_2F_1(-\tfrac{1}{8}, -\tfrac{3}{8}, \tfrac{1}{2}, -y^2),$$
$$f_1(y) = y\,{}_2F_1(\tfrac{3}{8}, \tfrac{1}{8}, \tfrac{3}{2}, -y^2). \tag{42}$$

An alternative fundamental pair of solutions $F_{1/4}$ and $F_{3/4}$ are given for $|y| \geq 1$ by[9]

$$F_{1/4} = (y^2)^{1/8}{}_2F_1(-\tfrac{1}{8}, \tfrac{3}{8}, \tfrac{5}{4}, -1/y^2),$$
$$F_{3/4} = (y^2)^{3/8}{}_2F_1(-\tfrac{3}{8}, \tfrac{1}{8}, \tfrac{3}{4}, -1/y^2). \tag{43}$$

The relations between these two pairs of solutions are

$$f_0 = B_1 F_{1/4} + B_2 F_{3/4},$$
$$f_1 = (y/|y|)(B_3 F_{1/4} + B_4 F_{3/4}), \tag{44}$$

where

$$B_1 = \Gamma(\tfrac{1}{2})\Gamma(-\tfrac{1}{4})/\Gamma(-\tfrac{3}{8})\Gamma(\tfrac{5}{8}) = 1.5832\ldots,$$
$$B_3 = \Gamma(\tfrac{3}{2})\Gamma(-\tfrac{1}{4})/\Gamma(\tfrac{1}{8})\Gamma(\tfrac{9}{8}) = -0.6123\ldots,$$
$$B_1B_4 = \sqrt{2} + 1, \quad B_3B_2 = \sqrt{2} - 1. \tag{45}$$

The general solution for f will be expressed in the alternative forms

$$f = af_0 + bf_1 = \alpha F_{3/4} + \beta F_{1/4}, \tag{46}$$

which are respectively useful for discussing the metric in the regions $|y| \leq 1$ and $|y| \geq 1$. The coefficients α, β depends on the sign of y as well as on a, b; from Eqs. (44) and (45),

$$\alpha = B_2 a + (\mathrm{sgn}y)B_4 b, \quad \beta = B_1 a + (\mathrm{sgn}y)B_3 b. \tag{47}$$

The first fundamental form

$$g_{\alpha\beta}\,\delta x^\alpha\,\delta x^\beta = 2(k_\alpha\,\delta x^\alpha)(m_\beta\,\delta x^\beta) + 2\left|t_\alpha\,\delta x^\alpha\right|^2$$

can be constructed from the following TNT_1 expressions for the null tetrad as computed with the aid of Eqs. (18), (14), (27), and (40):

$$k = p(d\sigma - 2\Delta_0 \xi\, d\eta), \tag{48a}$$
$$m = d\rho + 3i\Delta_0 p(A\,d\zeta - A^* d\zeta^*), \tag{48b}$$
$$t = (\rho + i\Delta_0 p)\, d\zeta - A^* k, \tag{48c}$$
$$A = \sqrt{2}\,\xi^{-1}\left[(y+i)\frac{1}{f}\frac{df}{dy} - \frac{3}{4}\right]. \tag{48d}$$

It is obvious that TNT_1 has a Killing vector $\mathbf{K}$ such that $K^\alpha d_\alpha = \partial/\partial\eta$. This was an extra bonus which the author did not expect. It is the only Killing vector, since Collinson[10] has proven that any TNT cannot have more than one Killing vector.[11]

Note that the first fundamental form is invariant in value under the *coordinate* transformation

$$\rho \to \rho, \quad \sigma \to \sigma, \quad \xi \to -\xi, \quad \eta \to -\eta.$$

However, *this does not imply an inversion symmetry of* TNT_1, because, if $p(\sigma, \xi)$ is given for positive ξ, then $p(\sigma, -\xi)$ need not represent the continuation (if any) of p into the region of negative ξ.

We next discuss the range of the chart and the circumstances under which p has an extension across the hypersurface $\xi = 0$ $(\sigma \neq 0)$ if it is given in the region $\xi > 0$. In all cases,

$$-\infty < \rho < \infty, \quad -\infty < \eta < \infty.$$

So, the only problem is to find the set of (σ, ξ) or, equivalently, the set of (y, ξ) which correspond to a given choice of f. There are three distinct cases.

(I) First, suppose f has two zeros y_1 and $y_2 > y_1$ such that $f(y) > 0$ in the interval $y_1 < y < y_2$. *Then there is no extension of the chart*, and its range is given by

$$y_1 < y < y_2, \quad \xi > 0.$$

An example is $f = f_0$, for which $y_1 \approx -5.5$ and $y_2 \approx 5.5$.

(II) Suppose f has a zero at $y = y_2$ and is positive in the open interval (y_2, ∞). *Then there is an analytic continuation of p across the hypersurface* $\xi = 0$ $(\sigma \neq 0)$ into the region of negative ξ.[12] The extension is uniquely determined by the C^1 matching conditions at $\xi = 0$ $(\sigma \neq 0)$. With the aid of Eqs. (43) to (47), the result of the matching procedure is given by

$$\alpha_- = \alpha, \quad \beta_- = -\beta,$$
$$a_- = -\sqrt{2}\, a - B_3B_4 b,$$
$$b_- = B_1B_2 a + \sqrt{2}\, b, \tag{49}$$

where

$$f_- = a_- f_0 + b_- f_1 = \alpha_- F_{3/4} + \beta_- F_{1/4}$$
$$p(\sigma, \xi) = (\xi^2)^{3/4} f_-(y) \quad \text{for } \xi \leq 0. \tag{50}$$

Let y_{2-} be the maximum zero of f_-. (There may be only one zero.) Then the range of the chart is given by

$$y > y_2 \quad \text{for } \xi \geq 0,$$
$$y > y_{2-} \quad \text{for } \xi \leq 0.$$

As an example, consider $f = -f_0$ restricted to the domain $(5.5, \infty)$; in this example, $y_{2-} \approx -1.2$.

(III) Suppose f has a zero at $y = y_1$ and is positive in the open interval $(-\infty, y_1)$. Then there is an analytic continuation of p across the hypersurface $\xi = 0$ $(\sigma \neq 0)$ into the region of negative ξ.[12] The same equations hold as in the preceding case except that B_3B_4 and B_1B_2 are to be replaced by their negatives in Eqs. (49), and the range of the chart is

$$y < y_1 \quad \text{for } \xi \geq 0,$$
$$y < y_{1-} \quad \text{for } \xi \leq 0,$$

where y_{1-} is the smallest zero of f_-. An example is $f = -f_0$ restricted to the domain $(-\infty, -5.5)$; in this example, $y_{1-} \approx 1.2$.

For some purposes, it is advantageous to introduce a positive parameter n_0 and a real parameter μ by the equations[13]

$$n_0 \cos\mu := a, \quad n_0 \sin\mu := b. \tag{51}$$

These parameters are used in the discussion.

4. DISCUSSION

A. Number of parameters in TNT_1

As can be verified by examining the derivation of h which followed Eqs. (28), the remaining arbitrariness in our choice of a null tetrad and coordinates is given by that subgroup of Eqs. (28) for which all parameters are identically zero except α, γ, η_0 and except for the possibility $\beta = \pi$. Also, there is the total inversion $k \to -k$, $m \to -m$, $t \to -t$ which we have not had to consider until now.

The key transformations which are induced by the residual group are as follows:

$$\begin{aligned}
&\rho' = \exp(-\alpha)\rho, \qquad \sigma' = \pm\gamma\sigma,\\
&\xi' = \exp(\alpha + i\beta)\xi, \quad \eta' = \exp(\alpha + i\beta)(\eta + \eta_0), \quad (\beta = 0, \pi),\\
&p' = \exp(\alpha)\gamma^{-1}p, \qquad \Delta_0' = \pm\exp(-2\alpha)\gamma\Delta_0,\\
&n_0' = \exp(-\alpha/2)\gamma^{-1}n_0, \quad y' = y, \quad \mu' = \mu.
\end{aligned} \tag{52}$$

n_0 and μ are the metrical parameters in Eq. (51). $\exp\alpha$ and γ are arbitrary positive scaling factors [where this α is not to be confused with the α which occurs in Eqs. (46) and (47)].

It is clear that $\exp\alpha$ and γ can be selected at will to make both $n_0 = 1$ and $\Delta_0 = 1$. Therefore, TNT_1 has only one essential parameter,[14] viz., the parameter μ which remains invariant under the residual group.

B. Another approach to TNT_1

In Sec. 3, we defined TNT_1 by the condition that there exist a null tetrad k, m, t, t^* consistent with Eqs. (6) and a coordinate system $\rho, \sigma, \zeta, \zeta^*$ consistent with Eqs. (13) such that Eq. (27) is true. An alternative definition is expressed by the condition that there exists a null tetrad k, m, t, t^* such that Eqs. (6) hold and such that

$$k\theta = k d\chi, \tag{53}$$

where[15]

$$d\chi = -d\tau/\tau \tag{54}$$

and where θ is the 1-form defined by Eq. (9).

We now sketch a proof of the equivalence of the two definitions. Since both definitions admit only those null tetrads which satisfy Eqs. (6), we will grant Eqs. (6) throughout the proof. First, assume that Eqs. (53) and (54) are true. Let

$$p := \exp(-\chi), \quad \Delta := p^{-1}\tau. \tag{55}$$

Note that Eq. (55) is equivalent to the statement that Δ is a uniform field. Furthermore, from Eq. (9), we see that Eq. (54) is equivalent to the condition

$$d(p^{-1}k) = -2i\Delta d\zeta\, d\zeta^*. \tag{56}$$

Since Δ is uniform, it follows that there exists a scalar field σ such that

$$d\sigma = p^{-1}k - i\Delta(\zeta + \zeta^*)(d\zeta - d\zeta^*). \tag{57}$$

Equation (57) is equivalent to Eqs. (13) with $\Omega = i\Delta(\zeta + \zeta^*)$. So, we have proven that Eqs. (53) and (54) imply the existence of a null tetrad k, m, t, t^* and a coordinate system $\rho, \sigma, \zeta, \zeta^*$ such that (13) and (27) hold. The converse presents no problem and is proven by showing that Eqs. (13) and (27) imply Eq. (57), which then implies Eq. (56), which is equivalent to Eqs. (53) and (54).

C. Search for more TNT

There are two possibly feasible approaches to the hunt for more TNT. One is an extension of the methods used in this paper. It starts with the assumption that χ exists such that Eq. (53) holds,[16] but we no longer assume that the scalar field χ is given by Eq. (53). As before, scalar fields p and Δ are defined by Eqs. (55), whereupon Eq. (56) is still true. Δ is not necessarily a constant in this generalization, but Eq. (56) does imply that Δ is expressible as a function of ζ and ζ^* alone; moreover, Eq. (56) implies that there exist scalar fields $\Omega(\zeta, \zeta^*)$ and σ such that

$$k = p(d\sigma + \Omega\, d\zeta + \Omega^* d\zeta^*),$$

which implies that Eqs. (13) hold.

In other words, we are considering the class of TNT for which there exist a null tetrad k, m, t, t^* consistent with Eqs. (6) and a coordinate system $\rho, \sigma, \zeta, \zeta^*$ consistent with Eqs. (13) such that Ω depends on ζ and ζ^* alone. Fred Ernst and the author have already worked out the complete set of field equations and integrability conditions which are appropriate generalizations of Eqs. (29)–(36). The integrability conditions are immensely difficult to handle except for the special case of TNT_1. However, we have not yet subjected these conditions to a thorough examination.

The second approach which we have in mind leads to TNT spacetimes which are not necessarily in the category covered by Eq. (53). Specifically, we assume that a Killing vector **K** exists. Then, we can always specialize the null tetrad and the coordinates defined by Eqs. (6) and (13) so that $K^\alpha d_\alpha = \partial/\partial\eta$. The differential equations for p and h then involve only the two coordinates σ and ξ.

Even then, the problem is tough. However, it is possible that a combination of the assumption that a Killing vector exists taken together with some ansatz concerning Ω may lead to less resistant equations. The author is working on that possibility.

D. Some comments on type (3,1) with twist

Finally, we give the extension of our TNT equations to type (3, 1) (also called type III). For type (3, 1), as well as for type N, *all of Eqs. (7)–(20) still hold.* The only one of Eqs. (6) which are changed is the expression for v_{-1}, which now becomes

$$v_{-1} = h d\zeta + zLk \tag{58}$$

where L is a scalar field which does not depend on ρ.

As regards Eq. (21), it is replaced by the pair of equations

$$D^*h = -L, \quad D^*L = 2A^*L. \tag{59}$$

As regards the conform tensor,

$$C_{-1} = z^2L, \quad C_{-2} = zN - z^2DL + 2iz^3L(D\tau - A\tau). \tag{60}$$

That is all.

Wild and the author[6] have examined the question as to whether there exists a type (3, 1) twisting gravitational field *such that* Ω *is given by* Eq. (27) *and such that* η *is an ignorable coordinate*. The answer is negative; i.e., the spacetime is type N if Ω is given by Eq. (27) and if a Killing vector **K** exists such that $K^\alpha d_\alpha = \partial/\partial\eta$.

On the other hand, there may exist type (3, 1) twisting gravitational fields such that $N \neq 0$, η is an ignorable coordinate, and there exists a scalar field χ such that Eq. (53) is true. This is an interesting possibility. [Ω is not given by Eq. (27) in this case.]

ACKNOWLEDGMENTS

I thank Fred Ernst and Bob Malhiot for many stimulating discussions, and I wish to thank the referee of this paper for helpful comments and corrections concerning the number of essential parameters in my solution.

[1] I. Hauser, Phys. Rev. Lett. 33, 1112 (1974). Equation (15) in this article has a typographical error; $1+y^2$ should be in the *denominator*. Also, omit the factor 4 in the right side of Eq. (16).

[2] P. Sommers and M. Walker, J. Phys. A 9, 357 (1976).

[3] For a review and bibliography on algebraically special gravitational fields with twisting rays, in the context of a review of exact solutions, see W. Kinnersley, in *General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975), pp. 109–35. Equation (49) of this reference contains relatively simple forms of the type N field equations which differ from those in the present paper and which are due to A. Exton (private communication).

[4] I. Robinson and A. Trautman, Phys. Rev. Lett. 4, 431 (1960); I. Robinson and A. Trautman, Proc. Roy. Soc. A 265, 463 (1962).

[5] Type (3, 1) twisting gravitational fields for which there is no radiative term in the conform tensor have been found by I. Robinson, Gen. Rel. Grav. 6, 423 (1975). Also, see I. Robinson and J.R. Robinson, Int. J. Theor. Phys. 2, 231 (1969).

[6] I. Hauser and W. Wild, Bull. Am. Phys. 21, 37 (1976). The term "analogue" which appears in this reference as well as in Sec. 1 ("analog") of the present paper is, of course, terribly ambiguous. A precise wording is given in Sec. 4.

[7] Our notation for the null tetrad is used, e.g., by R.K. Sachs, in *Relativity, Groups, and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964), p. 530. Our signature is +2.

[8] D.E. Novoseller, Phys. Rev. D 11, 2330 (1975). This reference gives our Eqs. (42)–(44).

[9] The notations $F_{1/4}$ and $F_{3/4}$ are taken from Ref. 2.

[10] C.D. Collinson, J. Phys. A: Gen. Phys. 2, 621 (1969).

[11] For some TNT_1, a proof of the uniqueness of the Killing vector is given in Ref. 2.

[12] There is only one exception to the rule in case (II), viz., when $\alpha = 0$, $\beta \neq 0$. In this anomalous case, $f(y) > 0$ when $y > y_2 \cong -1.85$. However, $p \to 0$ as $\xi \to 0$; so our chart has no continuation across $\xi = 0$. A like exception occurs for case (III) when $\alpha = 0$, $\beta \neq 0$; although $f(y) > 0$ when $y < y_1 \cong 1.85$, we have $p \to 0$ as $\xi \to 0$.

[13] A different parametrization is used in Ref. 2. Their μ is also a mixing parameter, but it is defined differently than ours.

[14] It was incorrectly stated that there are two essential parameters by I. Hauser, *8th International Conference on General Relativity and Gravitation*, Abstracts of Contributed Papers (University of Waterloo, Waterloo, Ontario, Canada, 1977), p. 61. A misconception which led to this error was corrected by the referee of the present paper.

[15] This condition (54) is not invariant under the group of transformations defined by Eqs. (17); i.e., Eq. (54) is true only for a subset of all of the null tetrads which are consistent with Eqs. (6).

[16] This ansatz is invariant in form under the transformations (17), since $\theta \to \theta - \frac{1}{2}d(F + F^*)$.

Birkhoff's theorem and magnetic monopole solutions for a system of generalized Einstein–Maxwell field equations

Gregory Walter Horndeski

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada
(Received 7 June 1977)

In this note I establish a "Birkhoff's type theorem" for the most general second-order vector-tensor theory of gravitation and electromagnetism, which is such that its field equations are (i) derivable from a variational principle, (ii) consistent with the notion of conservation of charge, (iii) compatible with Maxwell's equations in a flat space, and (iv) in agreement with Einstein's equations in the absence of electromagnetic fields. I also present solutions to these field equations which can be regarded as representing the gravitational and electromagnetic field outside of a magnetic monopole. It turns out that these magnetic monopole solutions admit event horizons when the mass of the source is sufficiently large.

1. INTRODUCTION

The results presented in Ref. 1 demonstrate that the Einstein–Maxwell field theory is *not* unique amongst all possible second-order vector–tensor field theories of gravitation and electromagnetism which satisfy the following four conditions.

(i) There exists a Lagrange scalar density L of the form[2]

$$L = L(g_{ab}; g_{ab,i_1}; \ldots; g_{ab,i_1\ldots i_\alpha}; \psi_a; \psi_{a,i_1}; \ldots; \psi_{a,i_1\ldots i_\beta})$$

($\alpha \geq 2$, $\beta \geq 1$), which is such that in the absence of sources the field equations are given by[3] $E^{ij}(L)=0$ and $E^i(L)=0$.

(ii) In the presence of sources the field equations assume the form $E^{ij}(L)=8\pi\sqrt{-g}T_M^{ij}$ and $E^i(L)=16\pi\sqrt{-g}J^i$, where T_M^{ij} and J^i denote the energy–momentum tensor and charge current vector of the sources.

(iii) $E^i(L)_{;i}=0$, in general, and $E^i(L)=4\sqrt{-g}F^{ij}{}_{;j}$ when evaluated for a flat metric tensor.[4]

(iv) If electromagnetic fields are not present, then the field equations $E^{ij}(L)=8\pi\sqrt{-g}T_M^{ij}$ reduce to Einstein's equations; viz., $G^{ij}=8\pi T_M^{ij}$.

In fact the field equations of any second-order vector–tensor field theory of gravitation and electromagnetism satisfying the above four conditions may be expressed as follows[5]:

$$G^{ij} = 8\pi(T^{ij} + kA^{ij}) + 8\pi T_M^{ij} \tag{1.1}$$

and

$$F^{ij}{}_{;j} + \tfrac{1}{2}kF_{bc;a}\,{}^*R^{*iabc} = 4\pi J^i, \tag{1.2}$$

where k is an arbitrary constant with units of (length)^2,

$$T^{ij} := \frac{1}{4\pi}(F^{ia}F^j{}_a - \tfrac{1}{4}g^{ij}F_{ab}F^{ab}) \tag{1.3}$$

and

$$A^{ij} := \frac{1}{8\pi}\{F_{al}F_b{}^l\,{}^*R^{*iajb} + {}^*F^{ia;b}\,{}^*F^j{}_{b;a}\}. \tag{1.4}$$

When the constant k appearing in Eqs. (1.1) and (1.2) is equated to zero, these equations reduce to the usual Einstein–Maxwell field equations. Due to this observation we shall refer to Eqs. (1.1) and (1.2) as the *generalized Einstein–Maxwell field equations.*

In Ref. 6 static, spherically symmetric, pure electric, source-free solutions to Eqs. (1.1) and (1.2) were presented. The purpose of this paper is to examine the spherically symmetric source-free solutions to Eqs. (1.1) and (1.2) without assuming the spacetime under consideration to be either static or pure electric. Our primary objective is to prove that under certain conditions any spherically symmetric source-free solution to the generalized Einstein–Maxwell field equations must be static, and hence these field equations satisfy a theorem similar to Birkhoff's theorem.[7] It turns out that these "certain conditions" are quite reasonable when one is concerned with asymptotically flat solutions to Eqs. (1.1) and (1.2). Consequently, we can safely say that the generalized Einstein–Maxwell field equations do *not* admit gravitational or electromagnetic monopole radiation fields. However, as we shall see, they do admit magnetic monopole and "Bertotti–Robinson type" solutions.[8,9] In addition these magnetic monopole solutions bear a strong resemblance to the (pure magnetic) Reissner–Nordstrom solution, and, like the Reissner–Nordstrom solution, they possess event horizons when the mass of the source is sufficiently large.

We shall now turn our attention to the Birkhoff's theorem satisfied by the generalized Einstein–Maxwell field equations.

2. A PRELIMINARY VERSION OF BIRKHOFF'S THEOREM

A spacetime containing an electromagnetic field will be said to be *spherically symmetric* if it admits the Lie group SO(3) as an effective Lie transformation group of isometries. We also require that:

(i) the orbits of SO(3) be diffeomorphic to the 2-sphere S^2;

(ii) the restriction of the metric tensor to each orbit be positive definite; and

(iii) the electromagnetic field tensor be invariant under the action of SO(3).

Under these assumptions it can be shown that locally it is always possible to introduce a chart $w := (t, r, \theta, \phi)$ with connected domain U which is such that on U the line element, ds^2, and electromagnetic field tensor, F,

 0022-2488/78/1903-0668$1.00

have the following form[10]

$$ds^2 = -e^{2\alpha}dt^2 + e^{2\beta}dr^2 + X^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{2.1}$$

and

$$F = E\, dt\wedge dr + b\sin\theta\, d\theta\wedge d\phi, \tag{2.2}$$

where α, β, X, and E are functions of r and t and $b\in\mathbb{R}$. In Eqs. (2.1) and (2.2) θ and ϕ have the same range as do the usual spherical polar coordinates on S^2. We shall refer to the chart w with the above properties as a *standard chart* for a spherically symmetric spacetime containing an electromagnetic field.

Throughout the remainder of this paper we shall confine our attention to a spherically symmetric spacetime containing an electromagnetic field and satisfying the source-free generalized Einstein–Maxwell field equations. Let $w := (t, r, \theta, \phi)$ be a standard chart for such a spacetime with domain U. In order to establish a "Birkhoff's type theorem" for the generalized Einstein–Maxwell field equations there are various cases for us to consider depending upon the behavior of the "hypersurfaces" $\{X = \text{const}\}$, where $X^2 := g(\partial/\partial\theta, \partial/\partial\theta)$. In this paper we shall *not* treat the case in which the normal vector to these hypersurfaces is null on a set of measure zero. However, in the next section we will examine the case in which X equals a constant on U. This case leads to Bertotti–Robinson type solutions to Eqs. (1.1) and (1.2). For the purposes of the present section we shall only be concerned with the case in which the normal vector to the hypersurfaces $\{X = \text{const}\}$ are either spacelike at each point of U, or timelike at each point of U, or null at each point of U. To begin with, we will show that the hypersurfaces $\{X = \text{const}\}$ *cannot* be null at each point of U.

When working on the domain U of a standard chart w which is such that $dX \neq 0$ at a single point of U, then it can be shown that the only functionally independent source-free field equations of the system (1.1), (1.2) are

$$\begin{aligned}
&e^{-2\beta}[-(X^{-1}X')^2 - 2X^{-1}X'' + 2X^{-1}X'\beta'] + X^{-2}\\
&\quad + e^{-2\alpha}[(X^{-1}\dot{X})^2 + 2X^{-1}\dot{X}\dot{\beta}]\\
&\qquad = E^2e^{-2\alpha}e^{-2\beta} + b^2X^{-4} + 2kb^2e^{-2\beta}X^{-5}[-X'' + X'\beta'\\
&\qquad\quad + \dot{X}\dot{\beta}e^{-2\alpha}e^{2\beta}] - kE^2e^{-2\alpha}e^{-2\beta}X^{-2}[1 - (X')^2e^{-2\beta}\\
&\qquad\quad + (\dot{X})^2e^{-2\alpha}] + ke^{-2\beta}X^{-4}[6b^2(X^{-1}X')^2 - 2E^2(X\dot{X})^2e^{-4\alpha}],
\end{aligned} \tag{2.3}$$

$$\begin{aligned}
&e^{-2\alpha}X^{-1}(\dot{X}' - X'\dot{\beta} - \dot{X}\alpha')\\
&\quad = kb^2e^{-2\alpha}X^{-5}(\dot{X}' - X'\dot{\beta} - \dot{X}\alpha')\\
&\qquad + ke^{-2\alpha}e^{-2\beta}X^{-6}\dot{X}X'(E^2e^{-2\alpha}X^4 - 3b^2e^{2\beta}),
\end{aligned} \tag{2.4}$$

$$\begin{aligned}
&e^{-2\beta}[(X^{-1}X')^2 + 2X^{-1}X'\alpha'] - X^{-2}\\
&+ e^{-2\alpha}[2X^{-1}\dot{X}\dot{\alpha} - 2X^{-1}\ddot{X} - (X^{-1}\dot{X})^2]\\
&\quad = -E^2e^{-2\alpha}e^{-2\beta} - b^2X^{-4} + kE^2e^{-2\alpha}e^{-2\beta}X^{-2}[1 - (X')^2e^{-2\beta}\\
&\qquad + (\dot{X})^2e^{-2\alpha}] + 2kb^2e^{-2\alpha}X^{-5}[-\ddot{X} + \dot{X}\dot{\alpha} + X'\alpha'e^{2\alpha}e^{-2\beta}]\\
&\qquad + ke^{-2\alpha}X^{-4}[6b^2(X^{-1}\dot{X})^2 - 2E^2(XX')^2e^{-4\beta}],
\end{aligned} \tag{2.5}$$

$$\begin{aligned}
&-E' + E(\alpha' + \beta') - 2EX^{-1}X' + kX^{-2}[E' - E(\alpha' + \beta')]\\
&\quad \times[1 - (X')^2e^{-2\beta} + (\dot{X})^2e^{-2\alpha}] + 2kEX^{-2}\dot{X}e^{-2\alpha}\\
&\quad \times(\dot{X}' - X'\dot{\beta} - \dot{X}\alpha')\\
&\quad + 2kEX^{-2}X'e^{-2\beta}(-X'' + X'\beta'\\
&\quad + \dot{X}\dot{\beta}e^{-2\alpha}e^{2\beta}) = 0,
\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}
&\dot{E} - E(\dot{\alpha} + \dot{\beta}) + 2EX^{-1}\dot{X} - kX^{-2}[\dot{E} - E(\dot{\alpha} + \dot{\beta})]\\
&\quad \times[1 - (X')^2e^{-2\beta} + (\dot{X})^2e^{-2\alpha}] + 2kEX^{-2}X'e^{-2\beta}(\dot{X}' - X'\dot{\beta} - \dot{X}\alpha')\\
&\quad + 2kEX^{-2}\dot{X}e^{-2\alpha}(-\ddot{X} + \dot{X}\dot{\alpha} + X'\alpha'e^{2\alpha}e^{-2\beta}) = 0,
\end{aligned} \tag{2.7}$$

where the dot and the prime denote partial differentiation with respect to t and r respectively.[11] Equations (2.3)–(2.7) represent the following equations of the system (1.1), (1.2):

$$-G^0_0 = -8\pi\mathcal{T}^0_0,\quad \tfrac{1}{2}G^0_1 = 4\pi\mathcal{T}^0_1,\quad G^1_1 = 8\pi\mathcal{T}^1_1,$$

$$e^{2\alpha}e^{2\beta}(F^{0j}{}_{;j} + \tfrac{1}{2}kF_{bc;a}{}^*R^{*0abc}) = 0$$

and

$$e^{2\alpha}e^{2\beta}(F^{1j}{}_{;j} + \tfrac{1}{2}kF_{bc;a}{}^*R^{*1abc}) = 0,$$

respectively, where $\mathcal{T}^i_j := T^i_j + kA^i_j$. The only other nontrivial equations of the system (1.1), (1.2) are $G^2_2 = 8\pi\mathcal{T}^2_2$ and $G^3_3 = 8\pi\mathcal{T}^3_3$, where $G^2_2 = G^3_3$ and $\mathcal{T}^2_2 = \mathcal{T}^3_3$. However, under our present assumptions, these equations are automatically satisfied whenever Eqs. (2.3)–(2.7) hold. This is so since $G^i_{j;i} = 0$ and $\mathcal{T}^i_{j;i} = 0$ (in general) when the source-free version of equation (1.2) is satisfied.

Now suppose that $\tilde{w} := (\tilde{t}, \tilde{r}, \theta, \phi)$ is a standard chart with domain $\tilde{U}$ which is such that the hypersurfaces $\{\tilde{X} = \text{const}\}$ are null, and $d\tilde{X} \neq 0$ at a single point of $\tilde{U}$, where $\tilde{X}^2 := g(\partial/\partial\theta, \partial/\partial\theta)$. Then it is easily seen that given any point $P\in\tilde{U}$ there exists a standard chart $w := (t, r, \theta, \phi)$ at P with domain $U\subset\tilde{U}$ which is such that on U

$$ds^2 = e^{2\alpha}(-dt^2 + dr^2) + (t + r)^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

while F has the form presented in Eq. (2.2) with α and E being functions of r and t. Upon subtracting Eq. (2.5) from (2.3) we find that on U

$$1 = E^2e^{-4\alpha}[(t + r)^2 - k] + b^2/(t + r)^2, \tag{2.8}$$

while Eq. (2.6) tells us that

$$(E' - 2\alpha'E)[-1 + k/(t + r)^2] - 2E/(t + r) = 0. \tag{2.9}$$

When Eq. (2.8) is differentiated with respect to r, we discover that Eq. (2.9) can be used to rewrite the resultant equation as follows:

$$0 = (Ee^{-2\alpha})^2 + b^2/(t + r)^4,$$

and hence $E = b = 0$. However, this fact is incompatible with Eq. (2.8) and thus our original assumption that the hypersurfaces $\{\tilde{X} = \text{const}\}$ are null must be incorrect.

We shall now examine the case in which all of the hypersurfaces $\{\tilde{X} = \text{const}\}$ are timelike. In this case it can be shown that given any point $P\in\tilde{U}$ there exists a standard chart $w := (t, r, \theta, \phi)$ at P with domain $U\subset\tilde{U}$ which is such that on U

$$ds^2 = -e^{2\alpha}dt^2 + e^{2\beta}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{2.10}$$

while F has the form presented in Eq. (2.2) with α, β, and E being functions of r and t. For this form of line element equations (2.3), (2.4), and (2.5) imply that

$$\frac{2\beta'}{r} - \frac{1}{r^2} + \frac{e^{2\beta}}{r^2} = \frac{E^2 e^{-2\alpha}}{r^2}[r^2 + k(e^{-2\beta} - 1)] + \frac{b^2}{r^6}(r^2 e^{2\beta} + 2kr\beta' + 6k), \tag{2.11}$$

$$\dot{\beta} = 0 \tag{2.12}$$

and

$$\frac{2\alpha'}{r} + \frac{1}{r^2} - \frac{e^{2\beta}}{r^2} = -\frac{E^2 e^{-2\alpha}}{r^2}[r^2 + k(3e^{-2\beta} - 1)] + \frac{b^2}{r^5}(2k\alpha' - re^{2\beta}). \tag{2.13}$$

Upon differentiating Eq. (2.11) with respect to t, and noting Eq. (2.12), we find that

$$0 = [r^2 + k(e^{-2\beta} - 1)]\frac{\partial(E^2 e^{-2\alpha})}{\partial t}. \tag{2.14}$$

Now suppose that there exists a point $Q \in U$ which is such that $\partial(E^2 e^{-2\alpha})/\partial t \neq 0$ at Q. Then Eq. (2.14) would imply that $r^2 + k(e^{-2\beta} - 1) = 0$ on a neighborhood V of Q and hence $e^{-2\beta} = 1 - r^2/k$ on V. However, this expression for $e^{-2\beta}$ does not satisfy Eq. (2.11), and thus we may conclude that

$$\frac{\partial(E^2 e^{-2\alpha})}{\partial t} = 0 \tag{2.15}$$

on U.

If we now differentiate Eq. (2.13) with respect to t, we see that we can use Eq. (2.15) to deduce that $\dot{\alpha}' = 0$ and hence

$$\alpha = A(r) + B(t), \tag{2.16}$$

where A and B are differentiable functions of r and t respectively.

Due to Eqs. (2.15) and (2.16) we see that there must exist a differentiable function $\xi = \xi(r)$ on U which is such that

$$E = \xi e^{B(t)}. \tag{2.17}$$

Equations (2.2), (2.10), (2.16), and (2.17) suggest that we define a new time coordinate T on U by means of the equation $dT = e^B dt$. In terms of the chart (T, r, θ, ϕ) with domain U we have

$$ds^2 = -e^{2A} dT^2 + e^{2\beta} dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)$$

and

$$F = \xi\, dT \wedge dr + b \sin\theta\, d\theta \wedge d\phi,$$

where A, β, and ξ are functions only of r. As a result of this observation we can conclude that our spacetime must be static on U.

Employing an argument similar to the one presented above, we can show that if the hypersurfaces $\{\tilde{X} = \text{const}\}$ are all spacelike in $\tilde{U}$, then given any point $P \in \tilde{U}$ there exists a standard chart (t, R, θ, ϕ) at P with domain $U \subset \tilde{U}$ which is such that on U

$$ds^2 = -e^{2\alpha} dt^2 + e^{2B} dR^2 + t^2(d\theta^2 + \sin^2\theta\, d\phi^2)$$

and

$$F = \epsilon\, dt \wedge dR + b \sin\theta\, d\theta \wedge d\phi$$

where α, B, and ϵ are functions only of t.

To recapitulate the above work, we have the following preliminary version of Birkhoff's theorem for the generalized Einstein–Maxwell field equations.

Theorem 1: Let (M, g, F) be a spherically symmetric spacetime containing an electromagnetic field and let $\tilde{w} := (\tilde{t}, \tilde{r}, \theta, \phi)$ be a standard chart with domain $\tilde{U}$. Suppose in addition that

(i) g and F satisfy the generalized Einstein–Maxwell field equations without sources; and

(ii) the function $\tilde{X}^2 := g(\partial/\partial\theta, \partial/\partial\theta)$ has no critical points in $\tilde{U}$.

Then the gradient of $\tilde{X}$ cannot be a null vector field on $\tilde{U}$.

If the hypersurfaces $\{\tilde{X} = \text{const}\}$ are timelike (resp. spacelike), then, given any point $P \in \tilde{U}$, there exists a standard chart $w := (t, r, \theta, \phi)$ at P with domain $U \subset \tilde{U}$ which is such that on U

$$ds^2 = -e^{2\alpha} dt^2 + e^{2\beta} dr^2 + X^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{2.18}$$

and

$$F = E\, dt \wedge dr + b \sin\theta\, d\theta \wedge d\phi, \tag{2.19}$$

where α, β, and E are functions only of r (resp. t), and $X^2 = r^2$ (resp. t^2).

When seeking spherically symmetric, asymptotically flat solutions to the generalized Einstein–Maxwell field equations, one requires that the surfaces $\{\tilde{X} = \text{const}\}$ must be timelike hypersurfaces in the asymptotic domain. Thus due to the above theorem we see that *in the asymptotic domain any spherically symmetric, asymptotically flat solution to the generalized Einstein–Maxwell field equations must be static*. Consequently, the generalized Einstein–Maxwell field equations do not admit gravitational or electromagnetic monopole radiation fields.

We shall now turn our attention to the case in which $\tilde{X}$ equals a constant on the entire domain of a standard chart.

3. BERTOTTI–ROBINSON TYPE SOLUTIONS

Suppose that $\tilde{w} := (\tilde{t}, \tilde{r}, \theta, \phi)$ is a standard chart with domain $\tilde{U}$ which is such that $\tilde{X}^2 = Q^2$ (a positive constant) on $\tilde{U}$, where $\tilde{X}^2 := g(\partial/\partial\theta, \partial/\partial\theta)$. In this case it can be shown that given any point $P \in \tilde{U}$ there exists a standard chart $w := (t, r, \theta, \phi)$ at P with domain $U \subset \tilde{U}$ which is such that on U

$$ds^2 = Q^2(-dt^2 + e^{2\lambda} dr^2 + d\theta^2 + \sin^2\theta\, d\phi^2) \tag{3.1}$$

and

$$F = E\, dt \wedge dr + b \sin\theta\, d\theta \wedge d\phi, \tag{3.2}$$

where λ and E are functions of t and r. Without loss of generality we may suppose that λ is such that

$$\lambda(0, r) = 0 \quad \text{and} \quad \dot{\lambda}(0, r) = 0. \tag{3.3}$$

The chart w with the above properties is sometimes referred to as a Novikov-type coordinate system.[12]

Under the present assumptions equations (2.3)—(2.7) reduce to

$$Q^2 = E^2 e^{-2\lambda}\left(1 - \frac{k}{Q^2}\right) + b^2, \qquad (3.4)$$

$$\left(1 - \frac{k}{Q^2}\right)\frac{\partial (Ee^{-\lambda})}{\partial r} = 0, \qquad (3.5)$$

and

$$\left(1 - \frac{k}{Q^2}\right)\frac{\partial (Ee^{-\lambda})}{\partial t} = 0. \qquad (3.6)$$

One solution to the above equations is $k = Q^2 = b^2$. If $k \neq Q^2$, then Eq. (3.4) implies that $E^2 e^{-2\lambda}$ must be a constant and hence Eqs. (3.5) and (3.6) are a consequence of Eq. (3.4). We shall presently demonstrate that the case $k = Q^2$ cannot occur.

At this time it should be recalled that since $d\tilde{X} = 0$ on $\tilde{U}$ Eqs. (2.3)—(2.7) do *not* represent all of the functionally independent equations of the system (1.1), (1.2). The equation of this system which is not represented by Eqs. (2.3)—(2.7) is the equation $G_2^2 = 8\pi(T_2^2 + kA_2^2)$. In terms of the chart w this equation can be written as follows:

$$\left(\frac{kb^2}{Q^4} - 1\right)[\ddot{\lambda} + (\dot{\lambda})^2] = \frac{1}{Q^2}(b^2 + E^2 e^{-2\lambda}). \qquad (3.7)$$

Thus we see that if $kb^2 = Q^4$, then we must also have $b^2 = 0$ which is impossible in view of the fact that $Q^2 \neq 0$. Consequently, we must require that $kb^2 \neq Q^4$ and hence the case $k = Q^2$ mentioned above cannot arise.

Since $kb^2 \neq Q^4$, we can use Eq. (3.4) to rewrite Eq. (3.7) in the following manner:

$$\frac{\partial^2 e^\lambda}{\partial t^2} = \frac{Q^2 e^\lambda}{k - Q^2}. \qquad (3.8)$$

In order to solve this equation, there are two cases for us to consider: viz., case (i) $k - Q^2 < 0$ and case (ii) $k - Q^2 > 0$. We shall now examine each of these cases in turn.

Case (i): $k - Q^2 < 0$.

Upon setting

$$\mu^2 := Q^2/|k - Q^2| \qquad (3.9)$$

we find that Eq. (3.8) becomes

$$\frac{\partial^2 e^\lambda}{\partial t^2} + \mu^2 e^\lambda = 0.$$

Thus we may use Eqs. (3.3) and (3.4) to conclude that

$$e^\lambda = \cos\mu t, \quad E = C_1 \cos\mu t, \quad \text{and} \quad Q^2 \geq b^2,$$

where

$$(C_1)^2 := Q^2(Q^2 - b^2)/(Q^2 - k). \qquad (3.10)$$

Case (ii): $k - Q^2 > 0$.

In this case Eq. (3.8) becomes

$$\frac{\partial^2 e^\lambda}{\partial t^2} - \mu^2 e^\lambda = 0,$$

where μ^2 is defined by Eq. (3.9). Using Eqs. (3.3) and (3.4), we find that

$$e^\lambda = \cosh\mu t, \quad E = C_2 \cosh\mu T, \text{ and } b^2 \geq Q^2,$$

where

$$(C_2)^2 := Q^2(b^2 - Q^2)/(k - Q^2). \qquad (3.11)$$

Thus we see that if the magnetic charge b were set equal to zero, Case (ii) could not occur.

To summarize the above results concerning the Bertotti—Robinson type solutions to the generalized Einstein—Maxwell field equations, we have the following:

Theorem 2: Let (M, g, F) be a spherically symmetric spacetime containing an electromagnetic field, and let $\tilde{w} := (\tilde{t}, \tilde{r}, \theta, \phi)$ be a standard chart with domain $\tilde{U}$. Suppose in addition that

(i) g and F satisfy the generalized Einstein—Maxwell field equations without sources; and

(ii) the function $\tilde{X}^2 := g(\partial/\partial\theta, \partial/\partial\theta)$ is equal to a positive constant Q^2 on $\tilde{U}$.

Then Q^2 cannot equal k. If $k < Q^2$ ($k > Q^2$, respectively), then, given any point $P \in \tilde{U}$, there exists a standard chart $w := (t, r, \theta, \phi)$ at P with domain $U \subset \tilde{U}$ which is such that on U

$$ds^2 = Q^2(-dt^2 + e^{2\lambda} dr^2 + d\theta^2 + \sin^2\theta\, d\phi^2)$$

and

$$F = E\, dt \wedge dr + b \sin\theta\, d\theta \wedge d\phi,$$

where $e^\lambda = \cos(\mu t)$ [$\cosh(\mu t)$, respectively] and $E = C_1 \cos(\mu t)$ [$C_2 \cosh(\mu t)$, respectively]. In these formulas the constants μ and C_1 (C_2, respectively) are defined by Eqs. (3.9) and (3.10) [(3.11), respectively], and $Q^2 \geq b^2$ ($Q^2 \leq b^2$, respectively).

At present I regard the combination of Theorems 1 and 2 as representing Birkhoff's theorem for the generalized Einstein—Maxwell field equations. The only case which the combination of these two theorems fails to treat (and which is treated by Birkhoff's theorem for the Einstein—Maxwell field equations) is the case in which the gradient vector of the function $\tilde{X}^2$ is null on a set of measure zero. It is not clear to me how this case can be handled without first determining (in closed form) the general solution to Eqs. (1.1) and (1.2) for the spacetime whose metric and electromagnetic field have the form presented in Eqs. (2.18) and (2.19). The determination (in closed form) of such a general solution appears to be impossible.

4. ASYMPTOTICALLY FLAT MAGNETIC MONOPOLE SOLUTIONS

In Ref. 6 static, spherically symmetric, pure electric, source-free solutions to the generalized Einstein—Maxwell field equations were presented. The unfortunate aspect of these solutions was that they had to be expressed in series form. The purpose of this section is to show that it is possible to determine spherically symmetric, asymptotically flat, pure magnetic solutions to the source-free generalized Einstein—Maxwell field equations. As we shall see, the only obstacle to expressing these pure magnetic solutions in an elementary manner is a single quadrature.

To begin with, let (t, r, θ, ϕ) denote the standard spherical polar coordinate chart of[13] $\mathbf{R}^4 \approx \mathbf{R}\times\mathbf{R}^3$. We now seek a spherically symmetric, asymptotically flat, pure magnetic solution to the source-free generalized Einstein—Maxwell field equations whose underlying manifold M_ρ is of the form

$$M_\rho = \{P \in \mathbf{R}^4 \mid r(P) > \rho\}, \tag{4.1}$$

where ρ is some nonnegative real number. SO(3) acts on M_ρ in the obvious manner and we assume that the line element, ds^2, and electromagnetic field tensor, F, have the following form on M_ρ

$$ds^2 = -e^{2\alpha}dt^2 + e^{2\beta}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{4.2}$$

and

$$F = b\sin\theta\, d\theta \wedge d\phi, \tag{4.3}$$

where (due to the proof of Theorem 1) α and β are functions only of r and $b \in \mathbf{R}$. Using Eqs. (2.3)—(2.7), we find that under the present assumptions the only functionally independent equations of the system (1.1), (1.2) are

$$e^{-2\beta}\left(\frac{2\beta'}{r} - \frac{1}{r^2}\right) + \frac{1}{r^2} = \frac{b^2}{r^4} + \frac{2kb^2\beta' e^{-2\beta}}{r^5} + \frac{6kb^2e^{-2\beta}}{r^6} \tag{4.4}$$

and

$$e^{-2\beta}\left(\frac{2\alpha'}{r} + \frac{1}{r^2}\right) - \frac{1}{r^2} = \frac{-b^2}{r^4} + \frac{2kb^2\alpha' e^{-2\beta}}{r^5}. \tag{4.5}$$

Upon adding Eqs. (4.4) and (4.5) together, we find that

$$\alpha' = -\beta' + 3kb^2/r(r^4 - kb^2). \tag{4.6}$$

In what follows we shall assume that the constant ρ appearing in Eq. (4.1) has been chosen so that when $k \le 0$, $\rho = 0$ and when $k > 0$, $\rho^4 = kb^2$. Due to this assumption we can use Eq. (4.6) to conclude that

$$e^{2\alpha} = (Ae^{-2\beta}/r^6)(r^4 - kb^2)^{3/2}, \tag{4.7}$$

where A is a positive real constant. We shall now proceed to determine as expression for $e^{-2\beta}$.

Upon multiplying Eq. (4.4) by r we find that the resulting equation can be written as follows:

$$\frac{d}{dr}(e^{-2\beta}) + \frac{(r^4 + 6kb^2)}{r(r^4 - kb^2)}e^{-2\beta} = \frac{r(r^2 - b^2)}{r^4 - kb^2}. \tag{4.8}$$

The general solution to this equation is

$$e^{-2\beta} = [r^6/(r^4 - kb^2)^{7/4}](r - 2M + I), \tag{4.9}$$

where M is an arbitrary real constant and $I = I(r)$ is defined by

$$I(r) := \int_\infty^r \left(\frac{(x^2 - b^2)(x^4 - kb^2)^{3/4}}{x^5} - 1\right) dx. \tag{4.10}$$

Thus we may now employ Eqs. (4.7) and (4.9) to conclude that

$$e^{2\alpha} = A(r^4 - kb^2)^{-1/4}(r - 2M + I). \tag{4.11}$$

Using the binomial series it can be shown that $e^{2\alpha}$ admits the following series expansion:

$$e^{2\alpha} = A\left(1 - \frac{2M}{r} + \frac{b^2}{r^2} + \frac{kb^2}{2r^4} - \frac{Mkb^2}{2r^5} + \frac{kb^4}{10r^6} + \frac{13k^2b^4}{56r^8} - \frac{5Mk^2b^4}{16r^9} + \frac{13k^2b^6}{120r^{10}} + \frac{201k^3b^6}{1232r^{12}} - \frac{15Mk^3b^6}{64r^{13}} + \frac{55k^3b^8}{624r^{14}} + O(r^{-16})\right) \tag{4.12}$$

provided $r^4 > |kb^2|$.

The demand that our spacetime be asymptotically flat requires that for large r

$$e^{2\alpha} = 1 - 2m/r + O(r^{-2}),$$

where m represents the mass of the source which is responsible for the gravitational field. Consequently, we can use Eq. (4.12) to conclude that $A = 1$ and $M = m$. As a result of this fact, we see that the expression for $e^{2\alpha}$ presented in Eq. (4.12) agrees with the corresponding quantity in the Reissner—Nordstrom solution out through terms involving r^{-3}.

In summary, we have shown that the line element and electromagnetic field tensor of our spherically symmetric, asymptotically flat, pure magnetic spacetime, (M_ρ, g, F), must assume the following form if they satisfy the source-free generalized Einstein—Maxwell field equations:

$$ds^2 = -\frac{(r - 2M + I)}{(r^4 - kb^2)^{1/4}}dt^2 + \frac{(r^4 - kb^2)^{7/4}}{r^6(r - 2M + I)}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{4.13}$$

and

$$F = b\sin\theta\, d\theta \wedge d\phi, \tag{4.14}$$

where I is defined by Eq. (4.10) and $\rho = 0$ when $k \le 0$ while $\rho^4 = kb^2$ when $k > 0$. The constants M and b appearing in Eqs. (4.13) and (4.14) represent the mass and magnetic charge of the source respectively.

In passing it should be noted that the metric presented in Eq. (4.13) reduces to the Reissner—Nordstrom (pure magnetic) metric when k is set equal to zero.

The natural question to ask at this time is: Does the metric presented in Eq. (4.13) have any[14] "singularities"? The answer to this question is evidently in the affirmative if there exist points in M_ρ which are such that $r - 2M + I = 0$. We shall now demonstrate that solutions to this equation can exist when M is chosen to be sufficiently large.

Employing the binomial series, we find that when $r^4 > |k|b^2$,

$$I(r) = \frac{b^2}{r} + \frac{kb^2}{4r^3} - \frac{3kb^4}{20r^5} + 3\sum_{l=2}^{\infty}\frac{1\cdot 5\cdot 9\cdot \ldots \cdot(4l-7)k^lb^{2l}}{4^l l!(4l-1)r^{4l-1}} - 3b^2\sum_{l=2}^{\infty}\frac{1\cdot 5\cdot 9\cdot \ldots \cdot(4l-7)k^lb^{2l}}{4^l l!(4l+1)r^{4l+1}},$$

and hence $\lim_{r\to\infty} I(r) = 0$. Due to this fact it is clear that the image of M_ρ under the real valued function $f := r + I(r)$, must be an interval of the form (a, ∞). Consequently, if $2M > a$, then there will exist points in M_ρ at which $r - 2M + I = 0$. Due to the form of the line element presented in Eq. (4.13) these points will lie on a "null hypersurface."

When $k < 0$, I am not sure how small the real number a defined above will be. However, if $k \ge 0$, then $a = (kb^2)^{1/4}$. This is so since in this case the function $I(r)$ defined by Eq. (4.10) is strictly positive and monotonically decreasing as r increases. As a result $f(r) > r$ on M_ρ and hence $f(r) > (kb^2)^{1/4}$, since $r^4 > kb^2$ on M_ρ. Consequently, when $k \ge 0$ the constant M cannot be chosen arbitrarily close to zero if there is to be a singularity in M_ρ. I suspect that a similar remark pertains to the case $k < 0$; however, I have no proof of this claim.

Further examination of the function f shows that

(i) if $b^2 > k$ then f is critical at each point on the hypersurface $r = |b|$ and attains its absolute minimum on this hypersurface, while

(ii) if $b^2 \leq k$, then f is monotonically increasing as r increases.

Thus if M_ρ experiences a singularity and $b^2 > k$, then the value of r at the "outermost" singularity must be greater than or equal to $|b|$.

Now suppose that the constant M has been chosen so that the spacetime (M_ρ, g, F) experiences a singularity. If τ denotes the value of r at the outermost singularity, we let $M_{\rho,\tau} := \{P \in M_\rho | r(P) > \tau\}$. The spacetime $(M_{\rho,\tau}, g, F)$ can be extended to a larger spacetime containing an event horizon in the following manner.

Let v denote the function defined on $M_{\rho,\tau}$ by

$$v := t + \int_{2\tau}^{r} \frac{(x^4 - kb^2)\,dx}{x^3[x - 2m + I(x)]} .$$

It is easily seen that the collection of functions (v, r, θ, ϕ) defines a global chart for $M_{\rho,\tau}$. In terms of this chart the metric and electromagnetic field assume the following form:

$$ds^2 = -\frac{(1 - 2M/r + I/r)}{(1 - kb^2/r^4)^{1/4}}\,dv^2 + 2\left(1 - \frac{kb^2}{r^4}\right)^{3/4} dv\,dr + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{4.15}$$

and

$$F = b \sin\theta\, d\theta \wedge d\phi. \tag{4.16}$$

The manifold $M_{\rho,\tau}$ corresponds to the region $\tau < r < \infty$, but the metric (4.15) is nonsingular on the larger manifold N for which $\rho < r < \infty$. We now let (N, h, F) denote the spacetime containing an electromagnetic field whose metric, h, and electromagnetic field, F, are given by Eqs. (4.15) and (4.16). This spacetime is evidently an extension[15] of $(M_{\rho,\tau}, g, F)$ and it has an event horizon[16] at $r = \tau$. Moreover, (N, h, F) satisfies the source-free generalized Einstein–Maxwell field equations.

The spacetime (N, h, F) experiences a "genuine singularity at $r = \rho$." This is so since a straightforward calculation using the metric presented in Eq. (4.15) shows that

$$C_{abcd}C^{abcd} = \frac{4}{3}\left(\frac{(-2r^6 + 3b^2r^4 + kb^2r^2 - 2kb^4)}{(r^4 - kb^2)^2} + \frac{r^4(r - 2M + I)(3r^8 + 3kb^2r^4 - 2k^2b^4)}{(r^4 - kb^2)^{15/4}} - \frac{1}{r^2}\right)^2,$$

where C_{abcd} denotes the Weyl tensor. Thus we see that as $r \to \rho^+$, $C_{abcd}C^{abcd} \to \infty$, and hence there exists a curvature singularity "at $r = \rho$." Consequently, the spacetime (N, h, F) cannot be extended beyond $r = \rho$. This implies that if one were to require the constant k to be positive, then the "radius" of any magnetic monopole would have to be greater than or equal to $(kb^2)^{1/4}$. For if this were not the case, then it would be impossible to find a spherically symmetric, asymptotically flat, source-free solution to the generalized Einstein–Maxwell field equations which is valid at all points outside of the magnetic monopole. In other words, *when k is positive, there are no asymptotically flat magnetic monopole solutions to the generalized Einstein–Maxwell field equations corresponding to a point source.*

Due to the above work, we see that the source-free generalized Einstein–Maxwell field equations admit static, spherically symmetric, asymptotically flat, pure magnetic solutions containing an event horizon. Since the generalized Einstein–Maxwell field equations are *not* invariant under a duality transformation,[17] we cannot employ such a transformation to turn these pure magnetic solutions into pure electric solutions. Thus it is still an open question as to whether there exist spherically symmetric, asymptotically flat, pure electric solutions to the source-free generalized Einstein–Maxwell field equations which contain an event horizon.

In Ref. 18 it was argued that a possible alternative to the usual energy–momentum tensor T_{ij} of the electromagnetic field used in general relativity is provided by $\mathcal{T}_{ij} := T_{ij} + kA_{ij}$, where T_{ij} and A_{ij} are defined by Eqs. (1.3) and (1.4) respectively. We shall now conclude this paper by examining the behavior of T_{ij} and $\mathcal{T}_{ij}$ for the magnetic monopole spacetime $(M_{\rho,\tau}, g, F)$ considered above. For the purposes of this discussion we shall assume that $(M_{\rho,\tau}, g, F)$ is embedded in the spacetime (N, h, F).

Let O be an observer in $M_{\rho,\tau}$ whose world line is an integral curve of the vector field $u := e^{-\alpha}\,\partial/\partial t$, where e^α is defined by Eq. (4.11). Thus O is a Killing observer. Using Eq. (2.3), it can be shown that under our present assumptions[19]

$$T(u, u) = b^2/8\pi r^4 \tag{4.17}$$

and

$$\mathcal{T}(u, u) = \frac{1}{8\pi}\left(\frac{b^2}{r^4} - \frac{kb^2}{r^5}\frac{d}{dr}e^{-2\beta} + \frac{6kb^2e^{-2\beta}}{r^6}\right). \tag{4.18}$$

Due to Eq. (4.8) we may rewrite Eq. (4.18) as follows:

$$\mathcal{T}(u, u) = \frac{(r^2 - k)b^2 + 7kb^2e^{-2\beta}}{8\pi r^2(r^4 - kb^2)}, \tag{4.19}$$

where $e^{-2\beta}$ is given by Eq. (4.9).

Equation (4.17) clearly indicates that $T(u, u)$ is positive on $M_{\rho,\tau}$. However, in view of Eq. (4.19), it is by no means obvious whether a similar remark applies to $\mathcal{T}(u, u)$, although it is apparent that $\mathcal{T}(u, u)$ is positive once r gets sufficiently large since in that case $\mathcal{T}(u, u) = T(u, u) + O(r^{-6})$. Employing Eq. (4.19) in conjunction with our earlier remarks concerning the size of τ, it is not difficult to show that:

(i) if $k < 0$, then $\mathcal{T}(u, u)$ is positive for all observers near the event horizon;

(ii) if $b \neq 0$ and $b^2 \geq k \geq 0$, then $\mathcal{T}(u, u)$ is positive on $M_{\rho,\tau}$; and

(iii) if $\tau^2 < k$, then $\mathcal{T}(u, u)$ is negative for all observers in a neighborhood of the event horizon.[20]

I presently suspect that $\mathcal{T}(u, u)$ is positive on $M_{\rho,\tau}$

when $k < 0$, since it is positive near the event horizon and near infinity. Unfortunately, I have been unable to establish this claim. However, if this conjecture could be proved, then we would be able to say that $\mathcal{T}(u,u)$ is positive on $M_{\rho,\tau}$ except when $\tau^2 < k$ (and hence $b^2 < k$).

ACKNOWLEDGMENTS

I would like to thank J. Wainwright for many interesting and informative discussions on the topics dealt with in this note. I also wish to thank the National Research Council of Canada for financial assistance.

[1]G.W. Horndeski, J. Math. Phys. **17**, 1980 (1976).

[2]The notational conventions used in this paper are the same as those employed in C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), with the following exceptions: (i) tensor indices will be denoted by small Latin letters, and (ii) the vector potential of the electromagnetic field will be denoted by ψ_a, and hence the electromagnetic field tensor F_{ij} is given by $F_{ij} := \psi_{j,i} - \psi_{i,j}$.

[3]The Euler–Lagrange tensors $E^{ij}(L)$ and $E^i(L)$ are defined by

$$E^{ij}(L) := \sum_{\mu=0}^{\alpha} (-1)^\mu \frac{d^\mu}{dx^{i_1}\cdots dx^{i_\mu}}\left(\frac{\partial L}{\partial g_{ij,i_1\cdots i_\mu}}\right)$$

and

$$E^{i}(L) := \sum_{\mu=0}^{\beta} (-1)^\mu \frac{d^\mu}{dx^{i_1}\cdots dx^{i_\mu}}\left(\frac{\partial L}{\partial \psi_{i,i_1\cdots i_\mu}}\right).$$

[4]It should be noted that these constraints upon $E^i(L)$ imply that: (i) $J^i{}_{;i} = 0$ when the field equations are satisfied, and hence charge is conserved; and (ii) the field equation $E^i(L) = 16\pi\sqrt{-g}\, J^i$ reduces to Maxwell's equation in a flat space.

[5]A Lagrangian L which yields the source-free part of Eqs. (1.1) and (1.2) is provided by

$$L = -\sqrt{-g}R + \sqrt{-g}F_{ab}F^{ab} + \tfrac{1}{2}k\sqrt{-g}F_{ab}F^{cd}\,{}^*R^{*ab}{}_{cd}.$$

[6]G.W. Horndeski, "Static Spherically Symmetric Solutions to a System of Generalized Einstein-Maxwell Field Equations," to appear in Phys. Rev. D, 1978.

[7]More exactly I should say that these field equations satisfy a theorem which is similar to the "major lemma" needed to prove Birkhoff's theorem, which, in effect, says that if a spacetime satisfying the source-free Einstein–Maxwell field equations admits a G_3 [which is isomorphic to the Lie algebra of SO(3)], then it admits a G_4. For a precise statement of Birkhoff's theorem for the Einstein–Maxwell field equations see exercise 32.1 on pp. 844–46 of Ref. 2.

[8]B. Bertotti, Phys. Rev. **116**, 1331 (1959).

[9]I. Robinson, Bull. Acad. Polon. Sci. 7, 351 (1959).

[10]For a classical proof that the metric must have the form presented in equation (2.1) see J. Eiesland, Trans. Amer. Math. Soc. **27**, 213 (1925). A more modern proof of this fact may be found in B. Schmidt, Z. Naturforsch. **22a**, 1351 (1967). In order to prove that F has the form presented in Eq. (2.2), one needs to use assumption (iii) along with the fact that $dF = 0$.

[11]In deriving these equations I have made use of the formulas presented on pp. 270–72 of J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1971). One should note that Synge's convention for the Einstein tensor is the negative of ours.

[12]See p. 845 of Ref. 2 for Novikov coordinates.

[13]t denotes the standard chart of $\mathbb{R}$, while (r, θ, ϕ) denotes the spherical polar coordinate chart of $\mathbb{R}^3$.

[14]The type of "singularity" which we are presently concerned with is a "coordinate singularity."

[15]In fact each of the singularity free regions of (M_ρ, g, F) can be isometrically imbedded in (N, h, F).

[16]For a detailed discussion of event horizons see part two of B. Carter's article in *Black Holes*, Proceedings of 1972 session of Ecole d'été de physique theorique, edited by C. DeWitt and B.S. DeWitt (Gordon and Breach, New York, 1973).

[17]A duality transformation is the transformation which replaces F_{ij} by $\cos(\alpha)F_{ij} + \sin(\alpha)\,{}^*F_{ij}$, where α is an arbitrary real number.

[18]G.W. Horndeski and J. Wainwright, Phys. Rev. D **16**, 1691 (1977).

[19]Recall that Eq. (2.3) represents the equation $-G^0_0 = -8\pi\mathcal{T}^0_0$.

[20]It should be noted that, in order to have $\tau^2 < k$, we require $b^2 < k$.

Rigorous results on ferromagnetic lattice spin systems. I

Yukihiko Karaki

Computer Centre, University of Tokyo, Tokyo, Japan
(Received 18 April 1977)

The Lee–Yang theorem was extended to the case of the correlation functions of the Ising ferromagnets ($s = 1/2$). Each adjacent pair of zeros of the $(n+1)$th correlation function in the complex fugacity plane is separated by one and only one zero of the nth correlation function ($n = 0,1,2,\ldots,N-1$), and none of zeros of each function degenerate except for the infinite temperature in the completely connected system. The first Griffiths' inequality for the correlation function was elaborated such as $\langle\sigma_1\sigma_2\cdots\sigma_n\rangle \geq \tanh^n(mh/kT)$. The inequality for the free energy in the presence of the external field was obtained as $-mh \leq (h,T) - \mathfrak{F}(0,T) \leq -kT\log[\cosh(mh/kT)]$.

1. INTRODUCTION

Lee and Yang first proved the theorem that all fugacity zeros of the partition function of the Ising ferromagnet of spin $\frac{1}{2}$ lie on the unit circle of the complex plane.[1] This theorem gives useful information about the cooperative phenomena.[2] A few years ago, on examining computer experiments,[3,4] this theorem was extended to Ising ferromagnets of arbitrary spin[5–7] and also proved for the anisotropic Heisenberg ferromagnets, the general Ising models, etc.[8–10]

After these developments, we know that the Lee—Yang theorem is universal for any ferromagnetic lattice spin system, independent of the lattice size, the dimension, the strength and the range of the spin—spin interaction, the boundary condition, the component spin, and the commutativity of the spin variables of the system.

On the other hand, there exists the well-known fact that the physical properties seem to depend on the dimension, the interaction range, etc., of the system. For example, in the one-dimensional Ising chane there occurs no phase transition,[11] but in the system of more than two dimensions the ferromagnetic phase transition does occur,[12,13] where the critical exponents of the physical quantities really depend on the structure of the system.[14]

The aim of our work is to narrow the gap between the universal property and the structure-dependent one for the ferromagnetic lattice spin systems from the side of the former.[15] We tried to extend Lee—Yang one-circle theorem to other properties, and obtained several rigorous theorems. Using these theorems, we also obtained the new inequality for the correlation functions and that for the free energy in the presence of a uniform external field, of which the former contains Griffiths' first inequality.[13,16,8]

This work was suggested from the results of computer experiments on finite spin systems.[17,18] Particularly, our motive comes from the question why numerical results for small systems, much smaller than statistical mechanical magnitude, give more suitable information than expected.

The obtained results show that the spin functions, which we call the partition function and the unnormalized correlation functions, have some topological invariant properties for various values of coupling parameters of spins. This fact seems not only to assure the scale-invariant properties of the free energy or correlation functions,[19–22] but also to indicate some dimensional-invariant one.

In this paper we propose the results for the Ising ferromagnets of spin $\frac{1}{2}$. In paper II we will extend the results to the case of the Ising ferromagnets of arbitrary spin. In paper III the results will be extended to the case of the Heisenberg ferromagnets. We note that another approach, with results similar to ours, has been reported for the case of the Ising ferromagnets of spin $\frac{1}{2}$.[23]

First we introduce the definitions. Secondly, the one-circle theorem for the partition function is extended to the case of the correlation functions. Next the rigorous relations between the fugacity zeros of the spin functions are proved. Then the fundamental theorem on the distribution of the fugacity zeros of the spin functions is presented. Then several new inequalities for the correlation functions and the free energy are proved in the presence of the external field. Finally we give a discussion.

2. DEFINITIONS

The Hamiltonian of the Ising model of spin $\frac{1}{2}$ is written as

$$\mathcal{H}_N = \mathcal{H}_0 + \sum_{i=1}^{N} m h_i \sigma_i \tag{2.1a}$$

and

$$\mathcal{H}_0 = -\sum_{(ij)} J_{ij}\sigma_i\sigma_j, \tag{2.1b}$$

where σ_i is the spin variable at the ith site and takes the values $\sigma_i = \pm 1$, h_i is the external field at the ith site, m is the magnetic moment, J_{ij} is the coupling constant between the ith and jth spins, N is the total number of spins, and $\sum_{(ij)}$ denotes the summation over all spin pairs.

We define the generalized partition function, that is, the partition function under the nonuniform external field, as the following function of N variables:

$$F_N(z_1, z_2, \ldots, z_N, \beta) = \mathrm{tr}\exp(-\beta\mathcal{H}_N) = \mathrm{tr}\left\{\exp(-\beta\mathcal{H}_0)\prod_{i=1}^{N} z_i^{\sigma_i}\right\}, \tag{2.2}$$

 0022-2488/78/1903-0675$1.00

where $z_i=\exp(\beta m h_i)$, $\beta=1/kT$, T is the temperature, k is the Boltzmann constant, and tr denotes the trace over all Ising spin states, that is, the summation over all values of the spin variables.

The spin function of the nth order in the presence of the uniform external field is defined as

$$f_n(z,\beta;\, i_1,\ldots,i_n) = \mathrm{tr}\left\{\left(\prod_{k=1}^{n}\sigma_{i_k}\right)\exp(-\beta H_0)\prod_{i=1}^{N} z^{\sigma_i}\right\}, \tag{2.3}$$

where $n=0,1,2,\ldots,N-1,N$, $z=\exp(\beta m h)$, h is the uniform external field, and z^2 is usually called "fugacity." The partition function in the presence of the uniform external field is the spin function of the 0th order, $f_0(z,\beta)$.

3. LEE-YANG THEOREM FOR CORRELATION FUNCTIONS

In this section the Lee–Yang theorem on the fugacity zeros of the partition function of the Ising ferromagnet ($s=\frac{1}{2}$) is extended to the case of correlation functions.

Lemma 1: The spin function of the nth order is generated from the generalized partition function, such as

$$f_n(z,\beta;\, i_1,\ldots,i_n)=(-i)^n F_N(z_1,\ldots,z_N,\beta) \quad (i=\sqrt{-1}), \tag{3.1}$$

where $z_k=z$ for $k\notin\{i_1,\ldots,i_n\}$, and $z_{k'}=iz$ for $k'\in\{i_1,\ldots,i_n\}$. (The notation $\{\ \}$ denotes the set of sites. $k\in\{i_1,\ldots,i_n\}$ shows that the site k does not belong to the set of sites $i_1,\ldots,i_n$, and so on.)

Proof: Using the fundamental identities of the following,

$$\sigma_k=(-i)^{\sigma_k}, \tag{3.2}$$

we can write the spin function of the nth order as

$$\begin{aligned} f_n(z,\beta;\, i_1,\ldots,i_n) &= \mathrm{tr}\left(\prod_{k=1}^{n}\sigma_{i_k}\right)\exp(-\beta H_0)\prod_{i=1}^{N} z^{\sigma_i} \\ &= (-i)^n\,\mathrm{tr}\exp(-\beta H_0)\prod_{i=1}^{N} z_i^{\sigma_i}, \end{aligned} \tag{3.3}$$

where z_i is the variable defined in the above statement. Then the lemma is immediately proved. (Q.E.D.)

Theorem 1: Every fugacity zero of a spin function of any order lies on the unit circle of the complex plane.

Proof: As is well known, the Lee–Yang lemma ($\forall J_{ij}\geq 0$) states that if $F_N(z_1,\ldots,z_N,\beta)$ vanishes and $|z_i|\geq 1$ for all i, then we have $|z_i|=1$ for all i. Using the Lee–Yang lemma and Lemma 1, we easily obtain that if $f_n(z,\beta;i_1,\ldots,i_n)=0$ and $|z|\geq 1$, then we have $|z|=1$, because $|iz|=|z|$. (Q.E.D.)

This theorem is the extension of the Lee–Yang theorem for the partition function (f_0) to the correlation functions (f_n; $0\leq n\leq N$).

4. BASIC THEOREMS ON THE DISTRIBUTION OF THE ZEROS OF THE CORRELATION FUNCTIONS

In this section we investigate the problem of how the fugacity zeros of the correlation functions relate to each other on the unit circle of the complex plane. For the sake of this aim, let us introduce the connected spin function of the nth order,

$$\Phi_n(z,\beta;\,\gamma;\, i_1,\ldots,i_n,j) = f_n(z,\beta;\, i_1,\ldots,i_n)+\gamma f_{n+1}(z,\beta;\, i_1,\ldots,i_n,j), \tag{4.1}$$

where γ is a complex parameter.

Lemma 2: The connected spin function of the nth order is generated from the generalized partition function,

$$\Phi_n(z,\beta;\,\gamma;\, i_1,\ldots,i_n,j) = C(\gamma)(-i)^n F_N(z_1,\ldots,\tilde{z}_j,\ldots,z_N,\beta) \quad (\gamma\neq\pm 1) \tag{4.2}$$

or

$$\Phi_n(z,\beta;\,\gamma;\, i_1,\ldots,i_n,j) = 2\,(-i)^n F_{N-1}(z_1',\ldots,z_N',\beta)z^{\gamma} \quad (\gamma=\pm 1), \tag{4.3}$$

where $z_k=z$ for $k\notin\{i_1,\ldots,i_n\}$, $z_{k'}=iz$ for $k'\in\{i_1,\ldots,i_n\}$, $\tilde{z}_j=z(1+\gamma)/C(\gamma)$, $z_k'=z_k\exp(\gamma\beta J_{kj})$ $(k\neq j)$, and $C(\gamma)=(1-\gamma^2)^{1/2}$.

Proof: Using Lemma 1, we can write the connected function as

$$\begin{aligned}\Phi_n(z,\beta;\,\gamma;\, i_1,\ldots,i_n,j) &= (-i)^n\{F_N(z_1,\ldots,z_j,\ldots,z_N,\beta) \\ &\quad - i\gamma F_N(z_1,\ldots,iz_j,\ldots,z_N,\beta)\} \quad (z_j=z).\end{aligned} \tag{4.4}$$

By the reduction formula, the generalized partition function can be written as

$$\begin{aligned}F_N(z_1,\ldots,z_j,\ldots,z_N,\beta) &= z_jF_{N-1}(z_1^+,\ldots,z_N^+,\beta)+z_j^{-1}F_{N-1}(z_1^-,\ldots,z_N^-,\beta),\end{aligned} \tag{4.5}$$

where $z_k^+=z_k\exp(\beta J_{kj})$ and $z_k^-=z_k\exp(-\beta J_{kj})$ $(k\neq j)$. From Eqs. (4.4) and (4.5), the following relation is obtained:

$$\begin{aligned}\Phi_n(z,\beta;\gamma;\, i_1,\ldots,i_n,j) &= (-i)^n\{(1+\gamma)F_{N-1}(z_1^+,\ldots,z_N^+,\beta)z \\ &\quad +(1-\gamma)F_{N-1}(z_1^-,\ldots,z_N^-,\beta)z^{-1}\}.\end{aligned} \tag{4.6}$$

For $\gamma\neq\pm 1$, Eq. (4.6) can be written as

$$\begin{aligned}\Phi_n(z,\beta;\,\gamma;\, i_1,\ldots,i_n,j) &= (-i)^nC(\gamma)[\{z(1+\gamma)/C(\gamma)\}F_{N-1}(z_1^+,\ldots,z_N^+,\beta) \\ &\quad +\{z(1+\gamma)/C(\gamma)\}^{-1}F_{N-1}(z_1^-,\ldots,z_N^-,\beta)],\end{aligned} \tag{4.7}$$

where $C(\gamma)=(1-\gamma^2)^{1/2}$. Applying again the reduction formula (4.5) to Eq. (4.7) and introducing a variable $\tilde{z}_j=z(1+\gamma)/C(\gamma)$, we obtain Eq. (4.2).

For $\gamma=\pm 1$ we immediately obtain Eq. (4.3) from Eq. (4.6). (Q.E.D.)

Theorem 2: The two spin functions, $f_n(z,\beta;\, i_1,\ldots,i_n)$ and $f_{n+1}(z,\beta;\, i_1,\ldots,i_n,j)$, have no common fugacity zero in a completely connected system, except for the infinite temperature. Here the completely connected system is the system in which every lattice site is connected with one or more than one site through the nonzero coupling parameters.

Proof: Taking the value of γ in Eqs. (4.1) and (4.3) as -1, we have

$$f_n(z,\beta;\, i_1,\ldots,i_n) - f_{n+1}(z,\beta;\, i_1,\ldots,i_n,j)$$
$$= 2\,(-i)^n F_{N-1}(z_1^-,\ldots,z_N^-,\beta)z^{-1}, \tag{4.8}$$

where $z_k^- = z_k\exp(-\beta J_{kj})$ $(z_k = z$ or $z_k = iz)$ $(k\neq j)$.

In the completely connected system, there always exists the nonzero coupling parameter for any site. Then suppose k'' be the nonzero coupling site for j, that is, $J_{k''j}\neq 0$ $(k''\neq j)$.

If $|z|=1$, we have $|z_k^-| = \exp(-\beta J_{kj})$ and we obtain that $|z_k^-|\leq 1$ for any k $(\neq k'', j)$ and $|z_{k''}^-|<1$ except for $\beta=0$. Thus using the Lee–Yang lemma, we immediately obtain

$$F_{N-1}(z_1^-,\ldots,z_{k''}^-,\ldots,z_N^-,\beta)z\neq 0$$
$$\text{for } |z_k^-|\leq 1 \ (k\neq k'', j) \text{ and } |z_{k''}|<1, \tag{4.9a}$$

that is,

$$f_n(z,\beta;\, i_1,\ldots,i_n)\neq f_{n+1}(z,\beta;\, i_1,\ldots,i_n,j)$$
$$\text{for } |z|=1 \text{ and } \beta\neq 0. \tag{4.9b}$$

On the other hand, according to Theorem 1, both f_n and f_{n+1} vanish only when $|z|=1$. From the above fact we find that f_n and f_{n+1} have no common fugacity zero on the unit circle of the complex plane. (Q.E.D.)

Extending Theorem 2, we obtain the following theorem.

Theorem 3: The connected spin function, $\Phi_n(z,\beta;\gamma;i_1,\ldots,i_n,j)$, has the following property.

If $|(1+\gamma)/C(\gamma)|=1$, every fugacity zero of Φ_n lies on the unit circle of the complex plane.

If $|(1+\gamma)/C(\gamma)|\neq 1$ or $\gamma=\pm 1$, Φ_n cannot be zero for $|z|=1$.

Proof: First we consider the case of $\gamma\neq\pm 1$. In this case we know from Lemma 2 that the connected function is written as Eq. (4.2). If $|(1+\gamma)/C(\gamma)|=1$, we have $|z_j|=|z|$ in Eq. (4.2). Besides we know $|z_k|=|z|$ for any k $(\neq j)$ in Eq. (4.2). Then using the Lee–Yang lemma, we obtain that it holds for $|z|=1$ if $F_N=0$ and $|z|\geq 1$ in Eq. (4.2). As $\Phi_n = C(\gamma)(-i)^n F_N$, thus the proof of the first statement is found. If $|(1+\gamma)/C(\gamma)|\neq 1$, we have $|z_j|\neq|z|$ in Eq. (4.2). It follows from the Lee–Yang lemma that F_N never vanishes for $|z_k|=1$ $(k\neq j)$ and $|z_j|\neq 1$ in Eq. (4.2). This gives the second statement for Φ_n.

For $\gamma=\pm 1$ we obtain Eq. (4.3) and a similar proof to that of Theorem 2 is found. (Q.E.D.)

Applying Theorem 3 to the cases of real γ and pure imaginary γ, we obtain the following theorem.

Theorem 4: The function, $\Gamma_n(z,\beta;\,\delta;\, i_1,\ldots,i_n,j)$, which is defined as

$$\Gamma_n(z,\beta;\,\delta;\, i_1,\ldots,i_n,j)$$
$$= f_n^2(z,\beta;\, i_1,\ldots,i_n) + \delta f_{n+1}^2(z,\beta;\, i_1,\ldots,i_n,j), \tag{4.10}$$

has the following property.

If $\delta\geq 0$, every fugacity zero of Γ_n lies on the unit circle of the complex plane.

If $\delta<0$, Γ_n cannot be zero on the unit circle, that is, for $|z|=1$.

Proof: When $\delta\geq 0$, we may write δ as $\delta=\gamma^2$ where γ is real. Then the function Γ_n can be written as follows:

$$\Gamma_n(z,\beta;\,\delta;\, i_1,\ldots,i_n,j)$$
$$=\{f_n(z,\beta;\, i_1,\ldots,i_n) + i\gamma f_{n+1}(z,\beta;\, i_1,\ldots,i_n,j)\}$$
$$\times\{f_n(z,\beta;\, i_1,\ldots,i_n) - i\gamma f_{n+1}(z,\beta;\, i_1,\ldots,i_n,j)\}$$
$$=\Phi_n(z,\beta;\, i\gamma;\, i_1,\ldots,j)\Phi_n(z,\beta;\, -i\gamma;\, i_1,\ldots,j), \tag{4.11}$$

where Φ_n is the connected spin function of the nth order. Since $|(1\pm i\gamma)/C(\pm i\gamma)|=1$, we find with the use of Theorem 3 that every fugacity zero of Γ_n lies on the unit circle of the complex plane.

When $\delta<0$, we may write δ as $\delta=-\gamma^2$ where γ is positive real. Similarly to the previous case, Γ_n can be expressed as

$$\Gamma_n(z,\beta;\,\delta;\, i_1,\ldots,i_n,j)$$
$$=\Phi_n(z,\beta;\,\gamma;\, i_1,\ldots,j)\Phi_n(z,\beta;\, -\gamma;\, i_1,\ldots,j). \tag{4.12}$$

Since $|(1\pm\gamma)/C(\pm\gamma)|\neq 1$ or $\gamma=\pm 1$, we find with the use of Theorem 3 that $\Gamma_n\neq 0$ for $|z|=1$. (Q.E.D.)

Theorem 4 will be used for discriminating the degeneracy, etc., of the fugacity zeros of the spin functions later.

5. MAIN THEOREM

In the previous sections several basic theorems and lemmas on the spin functions were proved. In this section some lemmas and the main theorem are demonstrated.

Lemma 3: The spin function, f_n $(0\leq n\leq N)$, has the following property.

(a) $f_n(z,\beta;\, i_1,\ldots,i_n)f_{n+1}(z,\beta;\, i_1,\ldots,i_n,j)$
$$=(z^2-z^{-2})\varphi(y)\psi(y), \tag{5.1}$$

where $y=z^2+z^{-2}$, and both $\varphi(y)$ and $\psi(y)$ are the polynomials of y.

(b) For the even system, we have

$$f_{2n}(z,\beta;\, i_1,\ldots,i_{2n}) = \varphi(y) \tag{5.2a}$$

and

$$f_{2n+1}(z,\beta;\, i_1,\ldots,i_{2n+1}) = (z^2-z^{-2})\psi(y). \tag{5.2b}$$

For the odd system, we have

$$f_{2n}(z,\beta;\, i_1,\ldots,i_{2n}) = (z+z^{-1})\varphi(y) \tag{5.2c}$$

and

$$f_{2n+1}(z,\beta;\, i_1,\ldots,i_{2n+1}) = (z-z^{-1})\psi(y) \quad (y=z^2+z^{-2}). \tag{5.2d}$$

Here we call the system even or odd, according to whether the total number of sites is even or odd.

Proof: Statement (a) can be proved from (b). Then we have only to prove statement (b).

From the definition of the spin function, we find

$$f_n(z,\beta;\, i_1,\ldots,i_n) = \sum_{s=-N,N,2} P_s^{(n)}(\beta;\, i_1,\ldots,i_n)z^s \tag{5.3}$$

and

$$P_s^{(n)}(\beta; i_1, \ldots, i_n) = \sum^{(\Sigma_i \sigma_i = s)n} \left(\prod_{k=1}^{n} \sigma_{i_k} \right) \exp(-\beta \mathcal{H}_0), \qquad (5.4)$$

where $\sum_{s=-N,N,2}$ denotes the summation over the values $s=-N, -N+2, \ldots, N-2, N$, and $\sum^{(\Sigma_i \sigma_i = s)}$ denotes the summation over the spin states of $\sum_{i=1}^{N} \sigma_i = s$ (s: fixed). As $\mathcal{H}_0$ has the spin reciprocity property, we can easily find

$$P_{-s}^{(n)}(\beta; i_1, \ldots, i_n) = (-1)^n P_s^{(n)}(\beta; i_1, \ldots, i_n). \qquad (5.5)$$

Then the spin function can be written as

$$f_n(z, \beta; i_1, \ldots, i_n) = \sum_{s \geq 0} P_s^{(n)}(\beta; i_1, \ldots, i_n)(z^s + (-1)^n z^{-s}), \qquad (5.6)$$

where $\sum_{s \geq 0}$ denotes the summation over the nonnegative values of s.

In the even system, s takes the values $0, 2, 4, \ldots, N-2, N$ $(=2N')$ in Eq. (5.6). If we write s as $s=2s'$ (s': nonnegative integer), we obtain

$$z^s + z^{-s} = z^{2s'} + z^{-2s'} \qquad (5.7a)$$

and

$$z^s - z^{-s} = (z^2 - z^{-2})(z^{2(s'-1)} + z^{2(s'-3)} + z^{2(s'-5)} + \cdots + z^{2(-s'+1)}). \qquad (5.7b)$$

By the well-known algebraic theorem, we know that the reciprocal polynomial in z^2 can be expressed as a polynomial in $z^2 + z^{-2}$. Thus we obtain

$$z^s + z^{-s} = \varphi_s(y) \qquad (5.8a)$$

and

$$z^s - z^{-s} = (z^2 - z^{-2}) \psi_s(y), \qquad (5.8b)$$

where both $\varphi_s(y)$ and $\psi_s(y)$ are the polynomials in $y = z^2 + z^{-2}$.

In the odd system, s takes the values $1, 3, 5, \ldots, N-2, N$ $(=2n'+1)$ in Eq. (5.6). If we write s as $s=2s'+1$ (s': nonnegative integer), we obtain

$$z^s + z^{-s} = (z + z^{-1})(z^{2s'} - z^{2(s'-1)} + z^{2(s'-2)} - \cdots + z^{-2s'}) = (z + z^{-1}) \varphi_s'(y) \qquad (5.9a)$$

and

$$z^s - z^{-s} = (z - z^{-1})(z^{2s'} + z^{2(s'-1)} + z^{2(s'-2)} + \cdots + z^{-2s'}) = (z - z^{-1}) \psi_s'(y), \qquad (5.9b)$$

where both $\varphi_s'(y)$ and $\psi_s'(y)$ are the polynomials in y.

A linear combination of the polynomials in y is also the polynomial in y. Then from Eqs. (5.6), (5.8), and (5.9) we immediately obtain the relations (5.2a)—(5.2d). (Q.E.D.)

Lemma 4: Let us take $y = z^2 + z^{-2}$.

(a) If $|z| = 1$, then y is real and $|y| \leq 2$.

(b) If y is real and $|y| \leq 2$, then we have $|z| = 1$ in the complex z plane.

Proof: (a) If $|z| = 1$, we may write

$$z^2 = \exp(i\theta) \quad \text{and} \quad z^{-2} = \exp(-i\theta), \qquad (5.10)$$

θ being real. Then we have

$$y = 2\cos\theta, \qquad (5.11)$$

and we find that y is real and $|y| \leq 2$.

(b) Let us express z^2 in the form

$$z^2 = r \exp(i\theta), \qquad (5.12)$$

where r is real and nonnegative, and θ is real. Then we have

$$y = r\exp(i\theta) + r^{-1}\exp(-i\theta) = (r + r^{-1})\cos\theta + i(r - r^{-1})\sin\theta. \qquad (5.13)$$

If y is real, we obtain

$$(r - r^{-1})\sin\theta = 0, \qquad (5.14)$$

which implies that $r = 1$ or $\theta = n\pi$ (n: integer). If $r \neq 1$, we have $\theta = n\pi$, and so $y = \pm(r + r^{-1})$ and $|y| > 2$. Hence, if y is real and also $|y| \leq 2$, then $r = 1$, that is, $|z| = 1$. (Q.E.D.)

Lemma 5: The square of the spin function can be factorized only by the variable y $(=z^2 + z^{-2})$.

In the even system $(N = 2N')$,

$$f_{2n}^2(z, \beta; i_1, \ldots, i_{2n}) = A^2 \prod_{s=1}^{N'} (y - \alpha_s)^2 \qquad (5.15a)$$

and

$$f_{2n\pm1}^2(z, \beta; i_1, \ldots, i_{2n\pm1}) = A^2(y^2 - 4) \prod_{s=1}^{N'-1} (y - \beta_s)^2. \qquad (5.15b)$$

In the odd system $(N = 2N' + 1)$,

$$f_{2n}^2(z, \beta; i_1, \ldots, i_{2n}) = A^2(y + 2) \prod_{s=1}^{N'} (y - \alpha_s)^2 \qquad (5.15c)$$

and

$$f_{2n\pm1}^2(z, \beta; i_1, \ldots, i_{2n\pm1}) = A^2(y - 2) \prod_{s=1}^{N'} (y - \beta_s)^2. \qquad (5.15d)$$

In Eqs. (5.15a)—(5.15d), both α_s and β_s are real, $|\alpha_s| \leq 2$ and $|\beta_s| \leq 2$ for every s.

Proof: The factorization expression follows easily from Lemmas 3 and 4, and Theorem 1. Here we have only to show the fact that the constant factor (A^2) is common in the consecutive spin functions.

As easily seen from Eq. (5.6), the highest term of the polynomial in y only comes from the term of $z^N + (-1)^n z^{-N}$. Then the coefficient of the highest order of the polynomial is common for z or y, and we have

$$A = P_N^{(n)}(\beta; i_1, \ldots, i_n). \qquad (5.16)$$

By the definition, $P_N^{(n)}$ contains only the states in which every spin variable takes the same value, that is, $\sigma_i = +1$ (or -1) for every i, and we obtain

$$P_N^{(n)}(\beta; i_1, \ldots, i_n) = \exp\left(\beta \sum_{(i,j)} J_{ij}\right) \quad \text{for any } n. \qquad (5.17)$$

Therefore, we obtain the expressions (5.15a)—(5.15d). (Q.E.D.)

Now we have been ready. Let us enter the main statement, in which a certain "universal" property of the ferromagnetic lattice spin system is found.

Theorem 5: Let us consider the consecutive two spin functions, $f_n(z,\beta;i_1,\ldots,i_n)$ and $f_{n+1}(z,\beta;i_1,\ldots,i_n,j)$, in a completely connected system $(0\le n\le N-1)$. The fugacity zeros of both functions separate each other on the unit circle of the complex plane, that is, each adjacent pair of zeros of f_{n+1} in z^2 is separated by one and only one zero of f_n, except for the infinite temperature. None of the fugacity zeros of the spin function, f_n $(0\le n\le N)$, degenerate for finite temperature.

Proof: First we consider the even system $(N=2N')$. As defined in Theorem 4, let us consider the following function,

$$\Gamma_{2n}(z,\beta;\delta;i_1,\ldots,i_{2n},j)$$
$$=f_{2n}^2(z,\beta;i_1,\ldots,i_{2n})+\delta f_{2n+1}^2(z,\beta;i_1,\ldots,i_{2n},j), \tag{5.18}$$

where δ is real. Using Lemma 5, Eq. (5.18) can be written as

$$\Gamma_{2n}(z,\beta;\delta;i_1,\ldots,i_{2n},j)$$
$$=A^2\left\{\prod_{s=1}^{N'}(y-\alpha_s)^2+\delta(y^2-4)\prod_{s=1}^{N'-1}(y-\beta_s)^2\right\}, \tag{5.19}$$

where α_s and β_s are real, $|\alpha_s|\le 2$ and $|\beta_s|\le 2$ for every s. Let the subscript be defined as $\alpha_1\ge\alpha_2\ge\alpha_3\ge\cdots\ge\alpha_{N'}$ and $2\ge\beta_1\ge\beta_2\ge\beta_3\ge\cdots\ge\beta_{N'-1}\ge-2$. (Let us denote 2 and -2 as β_0 and $\beta_{N'}$.)

Now we introduce the following rational polynomial:

$$\Xi(y;\{\alpha_s\},\{\beta_{s'}\})=\frac{\prod_{s=1}^{N'}(y-\alpha_s)^2}{(4-y^2)\prod_{s'=1}^{N'-1}(y-\beta_{s'})^2}. \tag{5.20}$$

We know from Theorem 2 that the sets of zeros, $\{\alpha_1,\alpha_2,\ldots,\alpha_{N'}\}$ and $\{2,\beta_1,\beta_2,\ldots,\beta_{N'-1},-2\}$, have no common element. Then the algebraic equation

$$\Gamma_{2n}(z,\beta;\delta;i_1,\ldots,i_{2n},j)=0, \tag{5.21}$$

is identical with the equation

$$\Xi(y;\{\alpha_s\},\{\beta_{s'}\})=\delta. \tag{5.22}$$

The number of roots of Eq. (5.21) is $2N$ with respect to z^2, and that of Eq. (5.22) is N with respect to y.

Suppose that there exist two or more than two α's between some interval of β's, for example, β_t and β_{t+1} $(0\le t\le N')$, given by

$$\beta_t>\alpha_{s_1}>\alpha_{s_2}>\cdots>\alpha_{s_m}>\beta_{t+1}\quad(m\ge 2). \tag{5.23}$$

In this case the rational polynomial, Ξ, has the following part, as seen in Fig. 1, so that its value is positive in the region $\beta_{t'}>y>\beta_{t'+1}$, and the minimum value is δ_0 (>0) there. Then if we take δ as $\delta_0>\delta\ge 0$, Eq. (5.22) must have the nonreal roots. That is, Eq. (5.21) must have the roots of $|z|\ne 1$ from Lemma 4. This is contrary to Theorem 4 for $\delta\ge 0$. The discussion is similarly performed for changing α and β. Therefore, we find that the set of α's, $\{\alpha_1,\alpha_2,\ldots,\alpha_{N'}\}$, must separate the set of β's, $\{\beta_0,\beta_1,\beta_2,\ldots,\beta_{N'}\}$, except for the degeneracy of each zero.

Next suppose that there exist some degenerate α's, $\alpha_s=\alpha_{s'}$ $(s\ne s')$. In this case, the number of real roots of Eq. (5.22) becomes less than N, and so the complex roots in y must appear. This again contradicts with

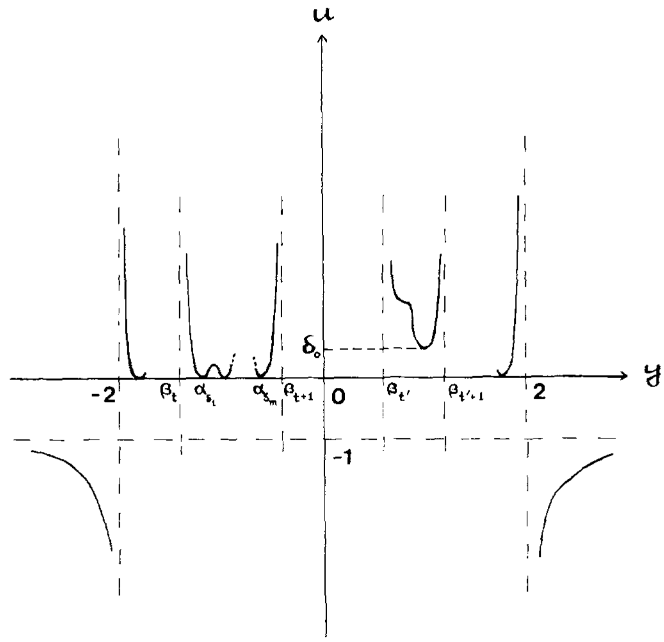


FIG. 1. The curve of the rational polynomial, $u=\Xi(y)$, which has the nonnegative part $[\Xi(y)\ge\delta_0>0]$ for $\beta_{t'}<y<\beta_{t'+1}$.

Theorem 4. Therefore, every zero of the α's and β's must be simple, that is, nondegenerate.

Theorem 2 holds for $\beta\ne 0$. Then the conclusion obtained here holds except for the infinite temperature.

The case of the odd system is almost similarly discussed, and the same conclusion is obtained. (Q.E.D.)

Now let us investigate the distribution of zeros of the spin functions in the limit cases.

First we consider the case of the infinite temperature, i.e., $\beta=0$. In this case the generalized partition function becomes as follows,

$$F_N(z_1,\ldots,z_N,0)=\mathrm{tr}\exp(0)\prod_{i=1}^{N}z_i^{\sigma_i}$$
$$=\prod_{i=1}^{N}(z_i+z_i^{-1}). \tag{5.24}$$

Then it follows from Lemma 1 that the spin function of the nth order is written as

$$f_n(z,0;i_1,\ldots,i_n)$$
$$=(-i)^n\prod_{k'=1}^{N-n}(z+z^{-1})\prod_{k=1}^{n}(iz+(iz)^{-1})$$
$$=(z+z^{-1})^{N-n}(z-z^{-1})^n. \tag{5.25}$$

This shows that the fugacity zeros of the spin function of the nth order degenerate at $z^2=-1$ [$(N-n)$-fold] and at $z^2=1$ (n-fold) at the infinite temperature.

Next we consider the case of the zero temperature, i.e., $\beta=\infty$. In this case the spin function of the nth order is written in the limit

$$A^{-1}f_n(z,\infty;i_1,\ldots,i_n)$$
$$=\lim_{\beta\to\infty}\mathrm{tr}\left(\prod_{k=1}^{n}\sigma_{i_k}\right)\exp\left(-\beta H_0-\beta\sum_{(ij)}J_{ij}\right)\prod_{i=1}^{N}z^{\sigma_i}. \tag{5.26}$$

For any finite system lim and tr can be interchanged.

It is easy to see that

$$\lim_{\beta\to\infty} \exp\left(-\beta H_0 - \beta \sum_{(ij)} J_{ij}\right)$$
$$= \lim_{\beta\to\infty} \exp\left\{-\beta \sum_{(ij)} J_{ij}(1-\sigma_i\sigma_j)\right\}$$
$$= \begin{cases} 1 & \text{for } \sigma_1 = \sigma_2 = \cdots = \sigma_N = +1 \text{ (or } -1), \\ 0 & \text{otherwise.} \end{cases} \quad (5.27)$$

Then we obtain

$$A^{-1} f_n(z, \infty; i_1, \ldots, i_n) = z^N + (-1)^n z^{-N}. \quad (5.28)$$

Therefore, the distribution of zeros can be obtained as

$$A^{-1} f_{2n}(z, \infty; i_1, \ldots, i_{2n}) = \prod_{s=1}^{N} \{z - z^{-1} \exp[i\pi(2s-1)/N]\} \quad (5.29a)$$

and

$$A^{-1} f_{2n+1}(z, \infty; i_1, \ldots, i_{2n+1}) = \prod_{s=1}^{N} \{z - z^{-1} \exp[i\pi(2s)/N]\}. \quad (5.29b)$$

We find that the distribution of the fugacity zeros of the spin function in the zero temperature is determined only by the even–oddness of the order n.

After the above investigations, we obtain the following corollary.

Corollary 1: The fugacity zeros of the spin function of the nth order $(0 \le n \le N)$ distribute as follows: The zeros distribute at the zero temperature uniformly as $z_s^2 = \exp[i\pi(2s-1)/N]$ (for even n) or $z_s^2 = \exp[i\pi(2s)/N]$ (for odd n) $(s = 1, 2, \ldots, N)$, move on the unit circle, isolated from each other, as temperature increases, and finally degenerate at $z^2 = -1$ [$(N-n)$-fold] and $z^2 = 1$ (n-fold) at the infinite temperature.

The feature of the above corollary is shown in Fig. 2. This property is universal for Ising ferromagnets of spin $\frac{1}{2}$, independent of the size, the dimension, the strength, and the range of the spin–spin interactions, and the boundary condition of the system.

6. SEVERAL INEQUALITIES

New inequalities for the correlation functions and the free energy can be proved from the above theorems.

Inequality 1: When $z \ge 1$,

$$f_n(z, \beta; i_1, \ldots, i_n) \ge \frac{z - z^{-1}}{z + z^{-1}} f_{n\pm 1}(z, \beta; i_1, \ldots, i_{n\pm 1}) \quad (n \ge 1). \quad (6.1)$$

Proof: It follows from Lemma 5 and Theorem 5 that the spin function can be expressed for the even system as

$$f_{2n}(z, \beta; i_1, \ldots, i_{2n}) = A \prod_{s=1}^{N'} (y - \alpha_s) \quad (y = z^2 + z^{-2}) \quad (6.2a)$$

and

$$f_{2n\pm 1}(z, \beta; i_1, \ldots, i_{2n\pm 1}) = A(z^2 - z^{-2}) \prod_{s=1}^{N'-1} (y - \beta_s), \quad (6.2b)$$

where

$$2 > \alpha_1 > \beta_1 > \alpha_2 > \beta_2 > \cdots > \beta_{N'-1} > \alpha_{N'} > -2. \quad (6.2c)$$

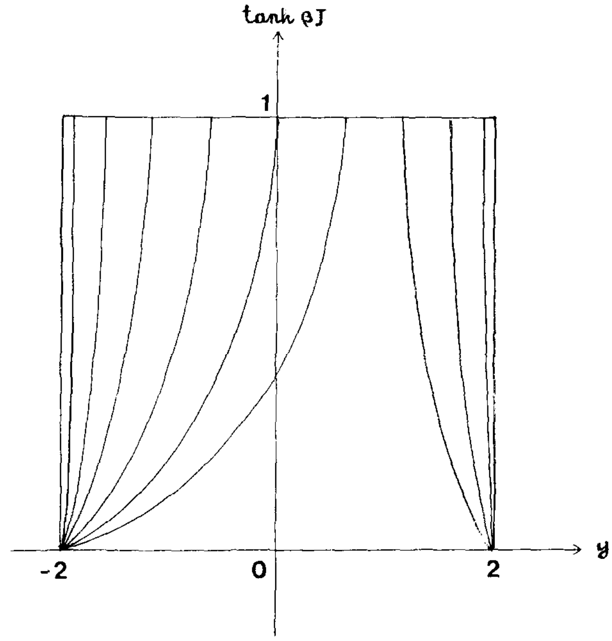


FIG. 2. The trajectories of the fugacity zeros of the spin function of the nth order, $f_n(y, \beta; i_1, \ldots, i_n)$, in the y plane ($y = z^2 + z^{-2}$), for the whole region of the temperature, that is, for $0 \le \tanh\beta J \le 1$, where J is an adequate positive constant of finite value.

From Eq. (6.2b), we have the relations

$$\frac{z - z^{-1}}{z + z^{-1}} f_{2n\pm 1}(z, \beta; i_1, \ldots, i_{2n\pm 1}) = A(y-2) \prod_{s=1}^{N'-1} (y - \beta_s) \quad (6.3a)$$

and

$$\frac{z + z^{-1}}{z - z^{-1}} f_{2n\pm 1}(z, \beta; i_1, \ldots, i_{2n\pm 1}) = A(y+2) \prod_{s=1}^{N'-1} (y - \beta_s). \quad (6.3b)$$

Using Eq. (6.2c), we obtain the inequality

$$(y-2) \prod_{s=1}^{N'-1} (y - \beta_s) < \prod_{s=1}^{N'-1} (y - \beta_s) \quad (6.4)$$
$$< (y+2) \prod_{s=1}^{N'-1} (y - \beta_s) \quad \text{for } y \ge 2.$$

If z is real, we find $y \ge 2$. Since the constant A is positive, we obtain the following relation:

$$\frac{z - z^{-1}}{z + z^{-1}} f_{2n\pm 1} < f_{2n} < \frac{z + z^{-1}}{z - z^{-1}} f_{2n\pm 1} \quad \text{for real } z. \quad (6.5)$$

Thus for $z \ge 1$, the expression (6.1) can be obtained for any n. (It also holds for the odd system.) (Q.E.D.)

In addition, we can easily prove the following inequality,

$$f_n(z, \beta; i_1, \ldots, i_n) \ge 0 \quad \text{for } z \ge 1 \ (n \ge 0). \quad (6.6)$$

Inequality 2: If the external field h is nonnegative ($h \ge 0$), then the correlation function satisfies the following inequality,

$$\langle\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_n}\rangle \geq \tanh^n\left(\frac{mh}{kT}\right) \quad (n\geq 1). \tag{6.7}$$

Proof: The correlation function of the nth order is defined as

$$\langle\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_n}\rangle = \mathrm{tr}\left(\prod_{k=1}^{n}\sigma_{i_k}\right)\exp(-\beta H_N)\Big/\mathrm{tr}\exp(-\beta H_N) = f_n(z,\beta;i_1,\ldots,i_n)/f_0(z,\beta). \tag{6.8}$$

As easily seen, we have the relation

$$\frac{f_n(z,\beta;i_1,\cdots,i_n)}{f_0(z,\beta)} = \frac{f_n}{f_{n-1}}\frac{f_{n-1}}{f_{n-2}}\cdots\frac{f_2}{f_1}\frac{f_1}{f_0}. \tag{6.9}$$

From the relations (6.1) and (6.6), we find

$$\frac{f_{n'+1}}{f_{n'}} \geq \frac{z-z^{-1}}{z+z^{-1}} \quad \text{for } z\geq 1 \quad (n'\geq 0). \tag{6.10}$$

From the relations (6.8)—(6.10), we obtain the inequality

$$\langle\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_n}\rangle \geq \left(\frac{z-z^{-1}}{z+z^{-1}}\right)^n \quad \text{for } z\geq 1 \quad (n\geq 1). \tag{6.11}$$

Thus the expression (6.7) follows from $z=\exp(mh/kT)$.

(Q.E.D.)

Inequality 3: Let the free energy per spin be defined as

$$\mathcal{J}(h,T) = -\frac{kT}{N}\log Z_N(h,T)$$

where $Z_N(h,T)=f_0(z,\beta)$. (6.12)

If the external field h is nonnegative ($h\geq 0$), then the free energy satisfies the following inequality,

$$-mh \leq \mathcal{J}(h,T)-\mathcal{J}(0,T) \leq -kT\log\left\{\cosh\left(\frac{mh}{kT}\right)\right\}. \tag{6.13}$$

Proof: First we write the free energy as a function of z and β as

$$\hat{\mathcal{J}}(z,\beta)=\mathcal{J}(h,T). \tag{6.14}$$

As easily seen,

$$\frac{\partial\hat{\mathcal{J}}(z,\beta)}{\partial\log z} = -\frac{1}{N\beta}\sum_{i=1}^{N}\frac{f_1(z,\beta;i)}{f_0(z,\beta)}. \tag{6.15}$$

From Inequality 2, we obtain

$$-1 \leq \beta\frac{\partial\hat{\mathcal{J}}(z,\beta)}{\partial\log z} \leq -\frac{z-z^{-1}}{z+z^{-1}} \quad \text{for } z\geq 1 \tag{6.16a}$$

or

$$-z^{-1} \leq \beta\frac{\partial\hat{\mathcal{J}}(z,\beta)}{\partial z} \leq -z^{-1}\frac{z-z^{-1}}{z+z^{-1}} \quad \text{for } z\geq 1, \tag{6.16b}$$

where we used the trivial inequality, $\langle\sigma_i\rangle\leq 1$.

The integration of expression (6.16b) with respect to z between $z\geq z'\geq 1$, gives the relation

$$-\log z \leq \beta\{\hat{\mathcal{J}}(z,\beta)-\hat{\mathcal{J}}(1,\beta)\} \leq -\log\frac{z+z^{-1}}{2} \quad \text{for } z\geq 1. \tag{6.17}$$

Using $z=\exp(mh/kT)$ and $\beta=1/kT$, we obtain inequality (6.13). (Q.E.D.)

7. DISCUSSION

In this paper we obtained some rigorous results on the Ising ferromagnets of spin $\frac{1}{2}$. We found that the functions which satisfy the Lee—Yang theorem make some sort of a family. That is, not only the partition function but all other spin functions satisfy the one-circle theorem of Lee and Yang. The fugacity zeros of the spin functions are not known to have a clear physical meaning now, however the whole image obtained here about the correlation functions in the presence of an external field gives some indication for investigating the cooperative phenomena.

The fugacity zeros of the spin functions show some topological invariant feature independent of the strengths of the coupling constants between spins (see Fig. 2). This fact seems to relate to the existence of the scale-invariant properties of the free energy and the correlation functions in the cooperative systems. Moreover this property may assure a certain kind of a dimensional-invariant property for cooperative phenomena, too.

Inequality 2 for the correlation function contains Griffiths' first inequality,

$$\langle\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_n}\rangle \geq 0 \quad \text{for } h\geq 0, \tag{7.1}$$

and our result elaborates it.

Inequality 3 gives some information about the free energy under the uniform external field. The deviation of the free energy in the presence of the external field from that in the absence of the field is limited by the normal analytic function of the external field. This is another expression of Lee—Yang's conclusion that the phase transition may occur only in the absence of the external field, and our result is a little more quantitative.

The results obtained in this article can be extended to the cases of the Ising ferromagnets of arbitrary spin and the anisotropic Heisenberg ferromagnets. Those will be reported in the next papers.

ACKNOWLEDGMENT

The author thanks Professor S. Ono for his kind advice, and also thanks the members of Ono's seminar for valuable discussions.

[1]T.D. Lee and C.N. Yang, Phys. Rev. 87, 410 (1952).
[2]C.N. Yang and T.D. Lee, Phys. Rev. 87, 404 (1952).
[3]C. Kawabata and M. Suzuki, J. Phys. Soc. Jpn. 27, 1105 (1969).
[4]C. Kawabata and M. Suzuki, J. Phys. Soc. Jpn. 28, 861, 1396 (1970).
[5]T. Asano, J. Phys. Soc. Jpn. 25, 1220 (1968).
[6]M. Suzuki, J. Math. Phys. 9, 2064 (1968).
[7]R.B. Griffiths, J. Math. Phys. 10, 1559 (1969).

[8]T. Asano, J. Phys. Soc. Jpn. **29**, 350 (1970).
[9]M. Suzuki, Progr. Theor. Phys. **41**, 1438 (1969).
[10]M. Suzuki and M.E. Fisher, J. Math. Phys. **12**, 235 (1971).
[11]E. Ising, Z. Phys. **31**, 253 (1925).
[12]L. Onsager, Phys. Rev. **65**, 117 (1944).
[13]R.B. Griffiths, J. Math. Phys. **8**, 478 (1967).
[14]For example, H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Clarendon, Oxford, 1971).
[15]Y. Karaki, Phys. Lett. A **53**, 443 (1975).
[16]R.B. Griffiths, J. Math. Phys. **8**, 484 (1967).
[17]S. Ono, Y. Karaki, M. Suzuki, and C. Kawabata, J. Phys. Soc. Jpn. **25**, 54 (1968).
[18]M. Suzuki, C. Kawabata, S. Ono, Y. Karaki, and M. Ikeda, J. Phys. Soc. Jpn. **29**, 837 (1970).
[19]B. Widom, J. Chem. Phys. **43**, 3892, 3898 (1965).
[20]C. Domb and D.L. Hunter, Proc. Phys. Soc. **86**, 1147 (1965).
[21]L.P. Kadanoff, Physics **2**, 263 (1966).
[22]K.G. Wilson, Phys. Rev. B **4**, 3174, 3184 (1971); Rev. Mod. Phys. **47**, 773 (1975).
[23]O.J. Heilmann and E.H. Lieb, Commun. Math. Phys. **25**, 190 (1972).

Nonbijective canonical transformations and their representation in quantum mechanics

P. Kramer

Institut für Theoretische Physik, Tübingen, W. Germany

M. Moshinsky[a] and T. H. Seligman

Instituto de Física, Universidad de México (UNAM), México 20, D.F., Mexico

(Received 13 June 1977)

In the present paper we analyze the representations in quantum mechanics of classical canonical transformations that are nonbijective, i.e., not one-to-one onto. We take as the central example the canonical transformation that changes the Hamiltonian of a one-dimensional oscillator of frequency κ^{-1} into one of frequency k^{-1} where κ, k are relatively prime integers. For the particular case $k = 1$, the mapping of the original phase space (x,p) onto the new one $(\bar{x},\bar{p})$ is κ to 1 and the equivalent points in (x,p) are related by a cyclic group C_κ of linear canonical transformations. When formulating this problem in Bargmann Hilbert space, the canonical transformation can be related to the conformal transformation $w = z^\kappa$, which again is κ to 1 and where a group C_κ also appears. This cyclic group proves fundamental for the determination of representations of the conformal transformation in Bargmann Hilbert space. To begin with, it suggests that while we can take in the original Bargmann Hilbert space a single component function, in the new Bargmann Hilbert space we must take a κ component one. In this way we can map in a one-to-one fashion the states and operators in the old and new Bargmann Hilbert spaces. When translating these results to ordinary Hilbert space, we get in an unambiguous way the quantization of the observables appearing in the equations that determine the representation of the classical canonical transformation relating oscillators of frequencies κ^{-1} and k^{-1}. Furthermore, we also get the solutions of these equations, and the resulting representation is unitary. While our discussion is restricted to the problem mentioned above, in the concluding section we indicate our surmise for deriving systematically the unitary representation in quantum mechanics of arbitrary canonical transformations.

1. INTRODUCTION

The subject of unitary representations of classical canonical transformations has had a long history[1,2] yet even today it is not fully understood.[3] Most physicists are likely to dismiss the whole subject by stating that if we have a set of new canonically conjugate variables

$$\bar{x}=\bar{x}(x,p), \quad \bar{p}=\bar{p}(x,p), \tag{1.1a}$$

$$\{\bar{x},\bar{p}\}_{x,p} \equiv \frac{\partial \bar{x}}{\partial x}\frac{\partial \bar{p}}{\partial p} - \frac{\partial \bar{x}}{\partial p}\frac{\partial \bar{p}}{\partial x} = 1, \tag{1.1b}$$

we can find in quantum mechanics an operator U such that it transforms the original *operators* x, p into the new $\bar{x}, \bar{p}$, i.e.,[2]

$$\bar{x}=U^{-1}xU, \quad \bar{p}=U^{-1}pU. \tag{1.1c}$$

If we choose then a basis in which, for example, the original coordinate is diagonal, i.e.,

$$x|x'\rangle = x'|x'\rangle, \tag{1.2}$$

then

$$\langle x''|U|x'\rangle \tag{1.3}$$

will be the representation of the canonical transformation (1.1a) in this basis.

Before proceeding further a word about notation is in order. We shall make no distinction between classical variables such as x, p and their corresponding operators in quantum mechanics as it is either explicitly mentioned, or it is clear from the context, when we deal with one and when with the other. We shall use Dirac's notation[2] x', x'' for eigenvalues of the operator x and similarly for other observables and denote by the kets $|x'\rangle$, $|x''\rangle$ the eigenstates of the *original* coordinate x. In order to connect our results more readily with the standard language of representation theory,[4] we define U as we did in a previous paper of ours,[5] i.e., we denote by U what Dirac calls U^{-1}.

Turning now our attention to $\langle x''|U|x'\rangle$ we could ask how to determine it explicitly once (1.1a) is given. A case fully discussed from the earliest literature[1,2] concerns the canonical transformation

$$\bar{x}=p, \quad \bar{p}=-x, \tag{1.4a}$$

for which

$$\langle x''|U|x'\rangle = (2\pi)^{-1}\exp(-ix'x''), \tag{1.4b}$$

when one takes units in which $\hbar=1$. For more complex situations the inquirer is usually referred to Dirac's book.[2]

Within certain limits, which Dirac states very carefully, this reference, though cryptic, indicates the procedure to be followed. More recently, Plebanski,[6] Mello and Moshinsky[3] and others have indicated how this procedure could be systematically implemented. We shall briefly review, in the present notation, the analysis of Ref. 3 as it shall provide the ground work for the ideas to be developed in this paper.

First it proves convenient to define the canonical transformations not explicitly but through the implicit equations[3]

[a] Member of the Instituto Nacional de Energía Nuclear and El Colegio Nacional.

 0022-2488/78/1903-0683$1.00

$$H(x,p)=\overline{H}(\overline{x},\overline{p}), \quad (1.5a)$$

$$G(x,p)=\overline{G}(\overline{x},\overline{p}) \quad (1.5b)$$

where

$$\{H,G\}_{x,p}=\{\overline{H},\overline{G}\}_{\overline{x},\overline{p}}. \quad (1.6)$$

The latter equation guarantees that when $\overline{x}$, $\overline{p}$ are obtained explicitly in terms of x, p from (1.5), the Poisson bracket relation (1.1a) is satisfied.[3]

As the relations of the type (1.1c) hold for any function $F(x,p)$ (at least so long as it can be expressed as a power series in x,p) we have

$$U^{-1}F(x,p)U=F(\overline{x},\overline{p})$$

or

$$F(x,p)U=UF(\overline{x},\overline{p}). \quad (1.7)$$

Then making use of (1.5) we obtain

$$\overline{H}(x,p)U=U\overline{H}(\overline{x},\overline{p})=UH(x,p), \quad (1.8a)$$

$$\overline{G}(x,p)U=U\overline{G}(\overline{x},\overline{p})=UG(x,p). \quad (1.8b)$$

Taking the operator relations (1.8) between the bra $\langle x''|$ and the ket $|x'\rangle$ we obtain straightforwardly[3]

$$\overline{H}\left(x'', \frac{1}{i}\frac{\partial}{\partial x''}\right)\langle x''|U|x'\rangle = \left[H^{\dagger}\left(x', \frac{1}{i}\frac{\partial}{\partial x'}\right)\right]^{*}\langle x''|U|x'\rangle, \quad (1.9a)$$

$$\overline{G}\left(x'', \frac{1}{i}\frac{\partial}{\partial x''}\right)\langle x''|U|x'\rangle = \left[G^{\dagger}\left(x', \frac{1}{i}\frac{\partial}{\partial x'}\right)\right]^{*}\langle x''|U|x'\rangle, \quad (1.9b)$$

where $H^{\dagger}$, $G^{\dagger}$ stand for the Hermitian conjugates of the operators H, G; $*$ indicates complex conjugation, and we recall that in the representation where the original coordinate is diagonal x can be replaced by the c-number x', and p by $-i(\partial/\partial x')$. Thus we have a set of two partial differential equations in x', x'' to determine $\langle x''|U|x'\rangle$ which we further restrict by the unitary condition

$$\int\langle x''|U|x'''\rangle dx'''\langle x'''|U^{\dagger}|x'\rangle = \int\langle x''|U|x'''\rangle[\langle x'|U|x'''\rangle]^{*}dx''' = \delta(x''-x'). \quad (1.10)$$

Among the more interesting problems involved in canonical transformations are those in which $H(x,p)$ and $\overline{H}(\overline{x},\overline{p})$ are two Hermitian Hamiltonians. The canonical transformation defined implicitly by (1.5) is then the one that takes us from the Hamiltonian H to $\overline{H}$. To solve Eqs. (1.9) we can then consider the eigenfunctions of H and $\overline{H}$,

$$H\left(x', \frac{1}{i}\frac{\partial}{\partial x'}\right)\phi_{\nu}(x')=E_{\nu}\phi_{\nu}(x'), \quad (1.11a)$$

$$\overline{H}\left(x'', \frac{1}{i}\frac{\partial}{\partial x''}\right)\overline{\phi}_{n}(x'')=\overline{E}_{n}\overline{\phi}_{n}(x''). \quad (1.11b)$$

As these eigenfunctions form a complete set we can then write

$$\langle x''|U|x'\rangle=\sum_{\nu,n}a_{n\nu}\overline{\phi}_{n}(x'')\phi_{\nu}^{*}(x'), \quad (1.12)$$

where the summation is replaced by an integration if the eigenvalues E_{ν}, $\overline{E}_{n}$ in (1.11) have a continuous spectrum. The equations (1.9) and (1.10) can then be used to determine $a_{n\nu}$ up to a constant phase factor.[3]

So far the program has been implemented for specific examples when H and $\overline{H}$ have the same spectrum, i.e., $E_{\nu}=\overline{E}_{n}$ when $\nu=n$. This is also a restriction that Dirac uses in relation with his original variables x, p and new ones $\overline{x}$, $\overline{p}$ though he imposes the further condition that all of them have a continuous spectrum going from $-\infty$ to $+\infty$. Under the above restriction Eq. (1.9a), which leads to the relation

$$a_{n\nu}(E_{\nu}-\overline{E}_{n})=0, \quad (1.13)$$

has the solution

$$a_{n\nu}=b_{n}\delta_{n,\nu}, \quad (1.14)$$

where in turn b_n is fully determined by (1.9b) and (1.10).[3]

What happens though when the spectra of H and $\overline{H}$ are not the same? It seems then that there are many subtle points when we try to find the unitary representation of the classical canonical transformation defined implicitly by the relations (1.5). In this paper we shall not attempt to solve this problem in general, but rather to clarify its structure through the study of a single problem: The canonical transformation that maps a harmonic oscillator whose frequency has an arbitrary rational value onto a harmonic oscillator of unit frequency or, equivalently, a canonical transformation that maps onto each other oscillators of frequencies κ^{-1}, k^{-1} where κ, k are relatively prime integers.

In the next section we shall discuss this problem classically and point out that even there difficulties arise, as the above canonical transformation does not imply a one to one mapping of the two phase spaces, i.e., it is not *bijective*. This difficulty is compounded in quantum mechanics as the translation from Eq. (1.5) to the corresponding operator relations (1.8) or (1.9) is ambiguous.

To be able to overcome these ambiguities we take, in Sec. 3, 4, 5, a long detour thorugh the conformal mapping $w=z^{\kappa}$ and its implications in Hilbert spaces of analytic functions.[7] With the help of this analysis we can return in Sec. 6 to the appropriate formulation and solution of Eq. (1.9) when we have the canonical transformation that maps an oscillator frequency κ^{-1} onto one of unit frequency or, more generally, when we pass from the frequency κ^{-1} to k^{-1}. In the final section we then discuss the implications that this specific problem has for the determination of unitary representations in quantum mechanics of arbitrary classical canonical transformations.

2. THE CLASSICAL CANONICAL TRANSFORMATION RELATING AN OSCILLATOR OF RATIONAL FREQUENCY WITH ONE OF UNIT FREQUENCY. SUMMARY OF THE FOLLOWING ANALYSIS

Let us consider the Hamiltonians of two oscillators, one of them in the original variables x, p and the other in the new ones $\bar{x}$, $\bar{p}$. The first one will have a rational frequency which we denote by

$$(k/\kappa), \tag{2.1}$$

where k, κ are relatively prime integers, and the second one has frequency 1. We take units in which $\hbar$ and the mass of the oscillator are 1 and thus the Hamiltonian of the first oscillator takes the form

$$\tfrac{1}{2}[p^2 + (k/\kappa)^2 x^2]. \tag{2.2}$$

By the simple dilatation

$$x \to (\kappa/k)x, \quad p \to (k/\kappa)p, \tag{2.3}$$

the above Hamiltonian becomes

$$(k/\kappa)\,\tfrac{1}{2}(p^2 + x^2), \tag{2.4}$$

while the one of unit frequency in terms of $\bar{x}$, $\bar{p}$ has the form

$$\tfrac{1}{2}(\bar{p}^2 + \bar{x}^2). \tag{2.5}$$

The functions $H(x,p)$, $\bar{H}(\bar{x},\bar{p})$ discussed in the Introduction can then be defined for this problem as

$$H(x,p) = \tfrac{1}{2}\kappa^{-1}(p^2 + x^2),$$
$$\bar{H}(\bar{x},\bar{p}) = \tfrac{1}{2}k^{-1}(\bar{p}^2 + \bar{x}^2). \tag{2.6}$$

Obviously the canonical transformation that takes H into $\bar{H}$ will be the one, up to the dilatation (2.3), that takes the Hamiltonian of an oscillator of frequency k/κ into one of unit frequency. To find this canonical transformation explicitly we introduce the *classical* annihilation and creation variables as

$$\eta = \frac{1}{\sqrt{2}}(x - ip), \quad \xi = \eta^* = \frac{1}{\sqrt{2}}(x + ip),$$
$$\bar{\eta} = \frac{1}{\sqrt{2}}(\bar{x} - i\bar{p}), \quad \bar{\xi} = \bar{\eta}^* = \frac{1}{\sqrt{2}}(\bar{x} + i\bar{p}). \tag{2.7}$$

We then define the functions $G(x,p)$, $\bar{G}(\bar{x},\bar{p})$ discussed in the Introduction as

$$G(x,p) = \kappa^{-1/2}(\eta\xi)^{(1-\kappa)/2}\eta^\kappa,$$
$$\bar{G}(\bar{x},\bar{p}) = k^{-1/2}(\bar{\eta}\bar{\xi})^{(1-k)/2}\bar{\eta}^k, \tag{2.8a}$$

while from (2.7) and (2.6) we have

$$H(x,p) = \kappa^{-1}\eta\xi, \quad \bar{H}(\bar{x},\bar{p}) = k^{-1}\bar{\eta}\bar{\xi}. \tag{2.8b}$$

We now proceed to show that Eq. (1.5), i.e., in this case

$$\kappa^{-1}\eta\xi = k^{-1}\bar{\eta}\bar{\xi}, \tag{2.9a}$$

$$\frac{\eta^\kappa}{\kappa^{1/2}(\eta\xi)^{(\kappa-1)/2}} = \frac{\bar{\eta}^k}{k^{1/2}(\bar{\eta}\bar{\xi})^{(k-1)/2}}, \tag{2.9b}$$

define implicitly a canonical transformation. For this we note that from (2.7)–(2.9),

$$\{H, G\}_{x,p} = i\left(\frac{\partial H}{\partial \eta}\frac{\partial G}{\partial \xi} - \frac{\partial H}{\partial \xi}\frac{\partial G}{\partial \eta}\right) = -iG. \tag{2.10}$$

As a similar relation holds for $\{\bar{H},\bar{G}\}_{\bar{x},\bar{p}}$ we see from $G(x,p) = \bar{G}(\bar{x},\bar{p})$ that (1.6) is satisfied and this is a necessary and sufficient condition[3] for (2.9) to define a canonical transformation.

We note immediately that the translation of the classical relations (2.9) to the operator form (1.8) or (1.9) is not well defined. To begin with, as in quantum mechanics, ξ, η of (2.7) have the commutation rule $[\xi,\eta]=1$, the expression

$$(1+\sigma)\eta\xi - \sigma\xi\eta, \quad \sigma \text{ real number}, \tag{2.11}$$

which classically is equivalent to $\eta\xi$, in quantum mechanics becomes $\eta\xi - \sigma$. Thus the equation of type (1.9a) resulting from (2.9a) seems to have an arbitrary constant in it. The situation is considerably more complex when we wish to determine the equation of the type (1.9b) resulting from (2.9b), as the $\eta\xi$ in the denominator and η^κ in the numerator do not commute, and thus there is a considerable degree of arbitrariness when we try to write down this equation. Therefore, even in the simple problem we are discussing in this paper the explicit determination of the Eq. (1.9) seems to present complications and thus merits a very careful analysis.

As a first step in this analysis we shall discuss the characteristics of the classical canonical transformation when $k=1$. We then have

$$\bar{\eta} = \kappa^{-1/2}(\eta\xi)^{(1-\kappa)/2}\eta^\kappa, \tag{2.12a}$$
$$\bar{\xi} = \bar{\eta}^* = \kappa^{-1/2}(\eta\xi)^{(1-\kappa)/2}\eta^\kappa, \tag{2.12b}$$

We can think of η, $\bar{\eta}$, defined by (2.7) in terms of the real x,p and $\bar{x},\bar{p}$, as complex variables. We note immediately from (2.12) that when η traces a circular arc, in its complex plane, of angle $(2\pi/\kappa)$, $\bar{\eta}$ traces a full circle and thus the mapping relating $\bar{\eta}$ with η, η^*, is not one to one. Therefore, while (2.12) locally defines a canonical transformation this is no longer true over the whole phase space. In fact we immediately see from (2.12) that $\bar{\eta}$ remains invariant when we replace η by $[\eta\exp(i2\pi q/\kappa)]$, $q=0, 1, \ldots, \kappa-1$ and thus $\xi = \eta^*$ by $[\xi\exp(-i2\pi q/k)]$. Thus there is an ambiguity as to which points x, p in the original phase space are mapped on a given point $\bar{x},\bar{p}$ in the new phase space. The points in x,p are related by the cyclic group $\mathcal{C}_\kappa$ of linear canonical transformations

$$x \to x\cos(2\pi q/\kappa) + p\sin(2\pi q/\kappa),$$
$$p \to -x\sin(2\pi q/\kappa) + p\cos(2\pi q/\kappa). \tag{2.13}$$

The group $\mathcal{C}_k$ will play a fundamental role in the following developments as it will provide a natural and well-defined transition from the relations (2.9) that give the canonical transformation, to the Eq. (1.9) that determine its unitary representation. To display this role in a clearer fashion we shall discuss in the following section the problem of the conformal transformation[8] $w = z^\kappa$ and its unitary representation. This representation will be given in Hilbert spaces of analytic functions with measures differing from the Bargmann measure.[7] In section 4 we then construct the corresponding mapping for two spaces *both* of which have the original Bargmann measure.[7] We next see in section 5 how the operators in these spaces are related, thus paving the way for writing down in section 6, in a well defined manner, the operator form of the equations (2.9). We then immediately find out the unitary representation in quantum mechanics of the classical canonical transformation that maps an oscillator of frequency κ^{-1} onto an oscillator of unit frequency and later extend the analysis to the case when the two oscillators have frequencies κ^{-1}, k^{-1}. Finally in the concluding section we discuss the implications of the present analysis for the unitary representation of arbitrary canonical transformations.

3. THE CONFORMAL MAPPING $w=z^{\kappa}$

Let us consider two complex variables z, w related by

$$w=z^{\kappa}, \quad \kappa \text{ integer}, \tag{3.1}$$

and inquire about the mapping of *analytic* functions

$$f(z)=\sum_{\nu=0}^{\infty} a_\nu z^\nu, \tag{3.2}$$

expressed by the above series (which we assume absolutely convergent) in the variable z onto *analytic* functions in the variable w.

We first note that if we replace

$$z \to z e^{i\varphi_q}, \quad \varphi_q=(2\pi q/\kappa), \tag{3.3}$$

we still get the same value for w. Thus there exists a cyclic group of transformations $\mathcal{C}_\kappa$ which gives the κ points in the z plane that are mapped onto a single point in the w plane. As $\mathcal{C}_\kappa$ is Abelian, its irreducible representations are one-dimensional and in fact can be given by[5]

$$D^\lambda(\varphi_q)=\exp(i\lambda\varphi_q), \quad \lambda=0,1,\ldots,\kappa-1. \tag{3.4}$$

We can then decompose $f(z)$ in terms of components $f^\lambda(z)$, $\lambda=0,1,\ldots,\kappa-1$ that are irreducible under the group $\mathcal{C}_\kappa$, where the latter are given by[5]

$$\begin{aligned} f^\lambda(z)=\mathcal{P}_\lambda f(z) &\equiv \kappa^{-1}\sum_{q=0}^{\kappa-1} D^\lambda(\varphi_q) f(z e^{-i\varphi_q}) \\ &=\kappa^{-1}\sum_{q=0}^{\kappa-1}\exp(i\lambda\varphi_q) f(z e^{-i\varphi_q}), \end{aligned} \tag{3.5}$$

with the $\mathcal{P}_\lambda$ [defined by (3.5)] being the projection operator.

Returning to the expression (3.2) for $f(z)$ we note that we can write

$$\nu \equiv \lambda \bmod \kappa \quad \text{or} \quad \nu=\mu\kappa+\lambda, \quad \mu=0,1,2,\ldots, \tag{3.6}$$

and from (3.5) we obtain

$$f^\lambda(z)=\sum_{\mu=0}^{\infty} a_{\mu\kappa+\lambda} z^{\mu\kappa+\lambda} \equiv z^\lambda F^\lambda(z^\kappa). \tag{3.7}$$

Thus we finally have

$$f(z)=\sum_{\lambda=0}^{\kappa-1} z^\lambda F^\lambda(w), \tag{3.8}$$

where the $F^\lambda(w)$ are analytic functions of w.

Now we come to the crucial point in our discussion. What is the function corresponding to $f(z)$ in the w plane? At first sight one could say

$$f(w^{\kappa^{-1}})=\sum_{\lambda=0}^{\kappa-1} w^{(\lambda/\kappa)} F^\lambda(w). \tag{3.9}$$

We note though that $w^{(\lambda/\kappa)}$ is only defined if instead of the w complex plane we speak of a Riemann surface with κ sheets joined in the usual fashion.[8] We can enumerate the sheets by an index $\sigma=0,1,\ldots,\kappa-1$ and thus the $f(w^{\kappa^{-1}})$ of (3.9) corresponds to the value of the function on the sheet $\sigma=0$, while on the σ^{th} sheet we have

$$f(w^{\kappa^{-1}})=\sum_{\lambda=0}^{\kappa-1} w^{(\lambda/\kappa)}\exp(i2\pi\lambda\sigma/\kappa)F^\lambda(w). \tag{3.10}$$

We can then speak, as is usually done in the literature,[8] of $f(z)$ as mapped on the κ sheets of a *Riemann surface* in the w plane. From the standpoint of later developments in this paper, it is more appropriate to think of $f(z)$ as being mapped onto the κ component vector

$$\mathbf{F}(w)=\begin{bmatrix} F^0(w) \\ F^1(w) \\ \vdots \\ F^{\kappa-1}(w)\end{bmatrix} \tag{3.11}$$

defined on a single sheet *complex w plane*. We note that the components of $\mathbf{F}(w)$ are all analytic functions of w. Furthermore if $f(z)$ is given we have

$$F^\lambda(w)=[z^{-\lambda}\mathcal{P}_\lambda f(z)]_{z=w^{\kappa^{-1}}}=\sum_\mu a_{\mu\kappa+\lambda} w^\mu, \tag{3.12}$$

where $\mathcal{P}_\lambda$ is the projection operator (3.5). Inversely if $\mathbf{F}(w)$ is given, $f(z)$ is determined by (3.8). Thus we have a one to one mapping between $f(z)$ and $\mathbf{F}(w)$.

Once the mapping between $f(z)$ and $\mathbf{F}(w)$ has been established we can analyze the way that operators acting on $f(z)$ translate into those acting on $\mathbf{F}(w)$ and vice versa.

Let us look first into the operator z that transforms $f(z)$ into $g(z)$ given by

$$g(z)=zf(z). \tag{3.13}$$

From (3.12) the corresponding $G^\lambda(w)$ is given by

$$\begin{aligned} G^\lambda(w) &= [z^{-\lambda}\mathcal{P}_\lambda g(z)]_{z=w^{\kappa^{-1}}} \\ &= [z^{-\lambda}\mathcal{P}_\lambda \sum_{\Lambda=0}^{\kappa-1}\sum_{\mu=0}^{\infty} a_{\mu\kappa+\Lambda} z^{\mu\kappa+\Lambda+1}]_{z=w^{\kappa^{-1}}} \\ &=\sum_{\mu=0}^{\infty} a_{\mu\kappa+\lambda-1} w^\mu \\ &=\begin{cases} wF^{\kappa-1}(w), & \text{if } \lambda=0, \\ F^{\lambda-1}(w), & \text{if } \lambda=1,2,\ldots,\kappa-1. \end{cases} \end{aligned} \tag{3.14}$$

We can express relation (3.14) as a matrix operator acting on the vector $\mathbf{F}(w)$ and thus to the operator z there corresponds the $\kappa\times\kappa$ matrix

$$z \longleftrightarrow \begin{bmatrix} 0 & 0 & \cdots & 0 & w \\ 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{3.15}$$

A similar analysis allows us to obtain the mappings of different operators from z space to w space and vice versa. Among the latter it is trivial to show that wI, where I is the unit $\kappa\times\kappa$ matrix, maps onto z^κ. We shall not carry these mappings of operators in detail here, as in Sec. 5 we obtain the mappings required, when we carry a transformation more directly related to (2.9) than the $w=z^\kappa$ of this section.

In the next section we shall discuss a unitary representation of the transformation $w=z^\kappa$ on Hilbert spaces of analytic functions.

4. THE UNITARY REPRESENTATIONS OF THE MAPPING $w=z^\kappa$

It is well known since the pioneering work of Bargmann[7] that wavefunctions $\psi(x')$ [or $\phi(x')$] in an ordinary Hilbert space can be mapped onto analytic functions $f(z)$ [or $g(z)$] in such a way that scalar products are preserved, i.e.,

$$\int_{-\infty}^{\infty} \phi^*(x')\psi(x')dx' = \int [g(z)]^* f(z)d\tau(z), \tag{4.1a}$$

where $d\tau(z)$ is the Bargmann measure

$$d\tau(z)=\pi^{-1}\exp(-zz^*)du dv, \quad z=u+iv, \tag{4.1b}$$

and the integration takes place over the full complex plane, i.e., $-\infty \le u,\ v\le\infty$. Bargmann[7] then also showed that the following correlation existed between the operators in ordinary and Bargmann Hilbert space,

$$\eta \longleftrightarrow z, \quad \xi \longleftrightarrow \frac{d}{dz}, \tag{4.2}$$

where η, ξ, given classically by (2.7), take the following operator form in ordinary Hilbert space,

$$\eta=\frac{1}{\sqrt2}\left(x'-\frac{\partial}{\partial x'}\right), \quad \xi=\frac{1}{\sqrt2}\left(x'+\frac{\partial}{\partial x'}\right). \tag{4.3}$$

We proceed now to consider another complex plane w related to the previous one by $w=z^\kappa$ and define in it a scalar product[5] involving the vectors $\mathbf{G}(w)$ and $\mathbf{F}(w)$ that corresponds to the one between $g(z)$ and $f(z)$ in the right-hand side of (4.1a). For this purpose we recall that from (3.8) we have

$$f(z)=\sum_{\lambda=0}^{\kappa-1} z^\lambda F^\lambda(z^\kappa), \tag{4.4}$$

and thus

$$\begin{aligned}(g,f) &\equiv \int [g(z)]^* f(z)d\tau(z)\\ &=\sum_{\lambda',\lambda=0}^{\kappa-1}\int [G^{\lambda'}(z^\kappa)]^* F^\lambda(z^\kappa) z^{*\lambda'} z^\lambda d\tau(z).\end{aligned} \tag{4.5}$$

But as $F^\lambda(z^\kappa)$ is an analytic function of z^κ and furthermore[7]

$$\int (z^{\nu'})^* z^\nu d\tau(z)=\nu!\,\delta_{\nu',\nu}, \tag{4.6}$$

we conclude that integral (4.5) vanishes unless $\lambda'=\lambda$ and thus

$$(g,f)=\sum_{\lambda=0}^{\kappa-1}\int [G^\lambda(w)]^* F^\lambda(w)d\sigma^\lambda(w), \tag{4.7}$$

where from the relation $w=z^\kappa$ the measure $d\sigma^\lambda(w)$ in w space becomes[5]

$$\begin{aligned}d\sigma^\lambda(w)&=(\pi\kappa)^{-1}(w^*w)^{[(\lambda+1)/\kappa-1]}\\ &\quad\times\exp[-(w^*w)^{\kappa^{-1}}]d\bar u d\bar v, \quad w=\bar u+i\bar v.\end{aligned} \tag{4.8}$$

Introducing the vector $\mathbf{F}(w)$ of (3.11) whose components are $F^\lambda(w)$, $\lambda=0,1,\ldots,\kappa-1$ and defining a diagonal matrix measure

$$d\boldsymbol{\sigma}(w)=\begin{bmatrix} d\sigma^0(w) & 0 & \cdots\cdots & 0\\ 0 & d\sigma^1(w) & \cdots\cdots & 0\\ \vdots & & & \vdots\\ \cdots & \cdots & \cdots\cdots & \cdots\\ 0 & 0 & \cdots\cdots & d\sigma^{\kappa-1}(w)\end{bmatrix}, \tag{4.9}$$

with $d\sigma^\lambda(w)$ given by (4.8) we can finally write

$$(g,f)=\int [\mathbf{G}(w)]^\dagger d\boldsymbol{\sigma}(w)\mathbf{F}(w)=(\mathbf{G},\mathbf{F}). \tag{4.10}$$

We now wish to find a κ-dimensional column vector $\mathbf{A}(w,z^*)$ of components

$$\mathbf{A}(w,z^*)=\begin{bmatrix} A^0(w,z^*)\\ A^1(w,z^*)\\ \vdots\\ A^{\kappa-1}(w,z^*)\end{bmatrix}, \tag{4.11}$$

such that

$$\mathbf{F}(w)=\int \mathbf{A}(w,z^*)d\tau(z)f(z), \tag{4.12}$$

$$f(z)=\int [\mathbf{A}(w,z^*)]^\dagger d\boldsymbol{\sigma}(w)\mathbf{F}(w). \tag{4.13}$$

The $\mathbf{A}(w,z^*)$ will give us a unitary representation of the conformal mapping $w=z^\kappa$.

To determine $\mathbf{A}(w,z^*)$ we shall only need to take into account the cyclic group $\mathcal{C}_\kappa$ in the z plane which gives the points (3.3) that are mapped on a single point in the w plane.

We consider the unit element in Bargmann Hilbert space,[7] i.e.,

$$\exp(z'z^*), \tag{4.14}$$

which takes a function $f(z)$ into $f(z')$, i.e.,

$$f(z')=\int \exp(z'z^*)f(z)d\tau(z). \tag{4.15}$$

We then proceed to extract from this unit element the desired $A^\lambda(w,z^*)$ in a way similar to the one in which we extracted $F^\lambda(w)$ from $f(z)$ in (3.12). We decompose $\exp(z'z^*)$ into its irreducible parts $\lambda=0,1,\ldots,\kappa-1$ associated with the $\mathcal{C}_\kappa$ group, by applying $\mathcal{P}'_\lambda$ of the type (3.5) to the z' variable and thus define

$$\begin{aligned}A^\lambda(w,z^*) &=[z'^{-\lambda}\mathcal{P}'_\lambda\exp(z'z^*)]_{z'=w^{\kappa^{-1}}}\\ &=\Big\{z'^{-\lambda}\kappa^{-1}\sum_{q=0}^{\kappa-1}\exp(i\lambda\varphi_q)\exp[z'\exp(-i\varphi_q)z^*]\Big\}_{z'=w^{\kappa^{-1}}}\\ &=\sum_{\mu=0}^{\infty}\frac{w^\mu (z^*)^{\mu\kappa+\lambda}}{(\mu\kappa+\lambda)!}.\end{aligned} \tag{4.16}$$

From (4.6), (4.12), and (4.16), we immediately see that if we write

$$f(z)=\sum_{\lambda=0}^{\kappa-1}\sum_{\mu=0}^{\infty}{}' a_{\mu\kappa+\lambda} z^{\mu\kappa+\lambda}, \tag{4.17}$$

it gives for $F^\lambda(w)$ the expression

$$F^\lambda(w)=\sum_{\mu=0}^{\infty} a_{\mu\kappa+\lambda} w^\mu, \tag{4.18}$$

as required in (3.12).

Inversely we would like to have

$$f(z)=\int\sum_{\lambda=0}^{\kappa-1}\{[A^\lambda(w,z^*)]^* d\sigma^\lambda(w)F^\lambda(w)\}. \quad (4.19)$$

Substituting the $A^\lambda(w,z^*)$ of (4.16) and the $F^\lambda(w)$ of (4.18) we get back (4.17) when we make use of the fact that[5]

$$\int w^{*\mu'} d\sigma^\lambda(w) w^\mu = (\mu\kappa+\lambda)!\,\delta_{\mu',\mu}. \quad (4.20)$$

Thus we have determined the vector $A(w,z^*)$ that relates $f(z)$ and $\mathbf{F}(w)$, making use only of the fact that there exists a cyclic group $\mathcal{C}_\kappa$ that gives the κ points in the z space that are mapped on a single one in the w space.

So far we have determined the unitary representation of the conformal transformation $w=z^\kappa$ when in the z space we have the Bargmann measure and in the w space the measure is given by (4.9). In the next section we introduce a new complex variable $\bar{z}$ instead of w, keeping states described by κ-dimensional vectors $\bar{\mathbf{F}}(\bar{z})$, but where now we have a Bargmann measure for each component. The reason for doing this is twofold. On one hand, the classical canonical transformation defined by Eq. (2.12) is real while the conformal mapping $w=z^\kappa$ was shown in Ref. 5 to lead to a complex extension of a canonical transformation. On the other hand, Wolf[9] has shown that for linear canonical transformations, a complex extension leads to a change of measure while for a real canonical transformation the measure remains unchanged. We may hope therefore that by combining the unitary representation we obtained in this section, with the unitary mapping that takes us from a space of measure $d\sigma^\lambda(w)$ to one with the Bargmann measure $d\tau(\bar{z})$, we learn something about the representation of the real canonical transformation (2.12). In Sec. 6 we shall find that this hope is confirmed.

5. A UNITARY MAPPING BETWEEN SPACES WITH BARGMANN MEASURES

In this section we shall proceed to discuss the unitary mapping that relates a Hilbert space associated with the complex variable w, with the unusual measure $d\boldsymbol{\sigma}(w)$ introduced in the previous section, with a Hilbert space associated with a complex variable $\bar{z}$ for which we assume a standard Bargmann measure.[7] One reason for this development is that only in the latter space the oscillator has the simple structure that raising or lowering operators corresponds to multiplication by $\bar{z}$ or to the differentiations $d/d\bar{z}$. We shall construct this mapping for each component $F^\lambda(w)$, $\lambda=0,1,\ldots,\kappa-1$ of the vector $\mathbf{F}(w)$ and thus obtain a diagonal matrix valued kernel that takes $\mathbf{F}(w)$ into $\bar{\mathbf{F}}(\bar{z})$, where the latter is also a κ-dimensional vector for which the measure is of the Bargmann type multiplied by a unit $\kappa\times\kappa$ matrix I. We can then readily combine the mapping thus obtained with the one of the previous section and study the transformation of the operators z, d/dz, and $z(d/dz)$ (the latter corresponding to the number operator) onto the space associated with the complex variable $\bar{z}$. Conversely we shall also start with the operators $\bar{z}I$, $(d/d\bar{z})I$, and $\bar{z}\,d/d\bar{z}\,I$, and see their form in the original Hilbert space associated with the complex variable z. These developments will allow us to return, in Sec. 6, to the ordinary Hilbert spaces associated with the variables x' and x''. We will then see the form that equations (1.9) must take when they are associated with the classical canonical transformation (2.12).

We note that the functions $[(n\kappa+\lambda)!]^{-1/2}w^n$ form an orthonormal basis[5] for square integrable analytic functions under the measure $d\sigma^\lambda(w)$ defined in (4.8), while $(n!)^{-1/2}\bar{z}^n$ forms a similar basis in Bargmann Hilbert space.[7] Thus the kernel

$$B^\lambda(\bar{z},w^*)=\sum_{n=0}^{\infty}\frac{\bar{z}^n(w^*)^n}{[(n\kappa+\lambda)!\,n!]^{1/2}},\quad \lambda=0,1,\ldots,\kappa-1, \quad (5.1)$$

provides us with a unitary mapping between the two spaces. Let us now define the diagonal matrix kernel

$$\mathbb{B}\equiv\begin{bmatrix} B^0(\bar{z},w^*) & 0 & \cdots\cdots & 0\\ 0 & B^1(z,w^*) & \cdots\cdots & 0\\ \cdots & \cdots & \cdots\cdots & \cdots\\ 0 & 0 & \cdots\cdots & B^{\kappa-1}(\bar{z},w^*)\end{bmatrix}, \quad (5.2)$$

which obviously provides us with a unitary mapping between the space of vector valued functions $\mathbf{F}(w)$ defined in the previous section and the space of vector valued functions $\bar{\mathbf{F}}(\bar{z})$ with a measure $d\tau(\bar{z})I$, where $d\tau(\bar{z})$ is given by (4.1b) and I is the $\kappa\times\kappa$ unit matrix. We thus have the relations

$$\bar{\mathbf{F}}(\bar{z})=\int\mathbb{B}(\bar{z},w^*)d\boldsymbol{\sigma}(w)\mathbf{F}(w), \quad (5.3a)$$

$$\mathbf{F}(w)=\int[\mathbb{B}(\bar{z},w^*)]^\dagger d\tau(\bar{z})\bar{\mathbf{F}}(\bar{z}), \quad (5.3b)$$

between the κ component functions in the two spaces.

We now combine this transformation with the $\mathbf{A}(w,z^*)$ of (4.11), (4.16) that takes us from the space associated with the complex variable z to one associated with w, i.e.,

$$\mathbf{C}(\bar{z},z^*)=\int\mathbb{B}(\bar{z},w^*)\,d\boldsymbol{\sigma}(w)\mathbf{A}(w,z^*), \quad (5.4)$$

and thus obtain for the components $\mathbf{C}^\lambda(\bar{z},z^*)$, $\lambda=0,1,\ldots,\kappa-1$, of the κ-dimensional column vector $\mathbf{C}(\bar{z},z^*)$, the expression

$$\mathbf{C}^\lambda(\bar{z},z^*)=\sum_{n=0}^{\infty}\frac{\bar{z}^n(z^*)^{n\kappa+\lambda}}{[n!\,(n\kappa+\lambda)!]^{1/2}}. \quad (5.5)$$

This kernel provides us with a unitary mapping between a Bargmann Hilbert space of analytic functions $f(z)$ and a Hilbert space of vector valued functions $\bar{\mathbf{F}}(\bar{z})$ each of whose components is again defined in a Bargmann Hilbert space.

We now turn our attention to the implications that the unitary mapping (5.4), (5.5) has for the transformation of operators from the z space to $\bar{z}$ space and vice versa. We start with the operator z which in the space associated with the complex variable $\bar{z}$ becomes the $\kappa\times\kappa$ matrix kernel

$$\mathbb{D}(\bar{z},\bar{z}'^*)\equiv\int\mathbf{C}(\bar{z},z^*)z[\mathbf{C}(\bar{z}',z^*)]^\dagger d\tau(z)$$
$$=\left\|\sum_{n,n'=0}^{\infty}\frac{\bar{z}^n(\bar{z}'^*)^{n'}(n'\kappa+\lambda'+1)^{1/2}}{[n!\,n'!]^{1/2}}\times\delta_{n\kappa+\lambda,n'\kappa+\lambda'+1}\right\|, \quad (5.6)$$

where the elements of the matrix on the right-hand side of (5.6), characterized by $\lambda, \lambda' = 0, 1, \ldots, \kappa - 1$, are derived with the help of (4.6) and (5.5). We then immediately see that this matrix can also be written as

$$\mathbb{D}(\bar{z}, \bar{z}'^*) = \begin{bmatrix} 0 & 0 \ldots\ldots & 0 & \kappa^{1/2}\bar{z} \\ \left(\kappa\bar{z}\frac{d}{dz}+1\right)^{1/2} & 0 \ldots\ldots & 0 & 0 \\ \ldots & \ldots\ldots\ldots & \ldots & \ldots \\ 0 & 0 \ldots\ldots & \left(\kappa\bar{z}\frac{d}{d\bar{z}}+\kappa-1\right)^{1/2} & 0 \end{bmatrix} e^{\bar{z}\bar{z}'^*} \tag{5.7}$$

and as $\exp(\bar{z}\bar{z}'^*)$ is a reproducing kernel[7] in Bargmann Hilbert space, we arrive at the conclusion that to the operator z there corresponds, in the space associated with the complex variable $\bar{z}$, the matrix operator

$$z \longleftrightarrow \begin{bmatrix} 0 & 0 \ldots\ldots & 0 & \kappa^{1/2}\bar{z} \\ \left(\kappa\bar{z}\frac{d}{d\bar{z}}+1\right)^{1/2} & 0 \ldots\ldots & 0 & 0 \\ \ldots & \ldots\ldots & \ldots\ldots & \ldots \\ 0 & 0 \ldots\ldots & \left(\kappa\bar{z}\frac{d}{d\bar{z}}+\kappa-1\right)^{1/2} & 0 \end{bmatrix}. \tag{5.8}$$

We have just to replace z in (5.6) by d/dz to get in an entirely similar fashion

$$\frac{d}{dz} \longleftrightarrow \begin{bmatrix} 0 & \left(\kappa\bar{z}\frac{d}{d\bar{z}}+1\right)^{1/2} & \ldots\ldots & 0 \\ \cdot & \ldots & \ldots\ldots & \ldots \\ 0 & 0 & \ldots\ldots & \left(\kappa\bar{z}\frac{d}{d\bar{z}}+\kappa-1\right)^{1/2} \\ \kappa^{1/2}\frac{d}{d\bar{z}} & 0 & \ldots\ldots & 0 \end{bmatrix} \tag{5.9}$$

which could also be obtained by remembering that in Bargmann Hilbert space[7]

$$z^\dagger = \frac{d}{dz}, \quad \bar{z}^\dagger = \frac{d}{d\bar{z}}, \tag{5.10}$$

and thus when we take the Hermitian conjugate of (5.8) we have (5.9). Finally combining (5.8) and (5.9) we obtain that the image of the number operator zd/dz is given by

$$z\frac{d}{dz} \longleftrightarrow \begin{bmatrix} \kappa\bar{z}\frac{d}{d\bar{z}} & 0 & \ldots\ldots & 0 \\ 0 & \kappa\bar{z}\frac{d}{d\bar{z}}+1 & \ldots\ldots & 0 \\ \cdot & \ldots & \ldots\ldots & \ldots \\ 0 & 0 & \ldots\ldots & \kappa\bar{z}\frac{d}{d\bar{z}}+\kappa-1 \end{bmatrix}. \tag{5.11}$$

We now turn our attention to the operators $\bar{z}I$ and $(d/d\bar{z})I$ and look at the operators that correspond to them in the Hilbert space associated with the complex variable z. For this we need the scalar kernel

$$\begin{aligned} D(z, z'^*) &\equiv \int [\mathbf{C}(\bar{z}, z^*)]^\dagger \bar{z}\,\bar{\mathbf{C}}(\bar{z}, z'^*)d\tau(\bar{z}) \\ &= \sum_{\lambda=0}^{\kappa-1}\sum_{n=0}^{\infty} \frac{z^{(n+1)\kappa+\lambda}(z'^*)^{n\kappa+\lambda}(n+1)^{1/2}}{\{[(n+1)\kappa+\lambda]!\,(n\kappa+\lambda)!\}^{1/2}} \\ &= \sum_{\lambda=0}^{\kappa-1}\left[\frac{z(d/dz)-\lambda}{\kappa\prod_{\nu=0}^{\kappa-1}[z(d/dz)-\nu]}\right]^{1/2} z^\kappa \mathcal{P}_\lambda e^{zz'^*}, \end{aligned} \tag{5.12}$$

where again we used (4.6) and (5.5) to obtain the intermediate formula in (5.12) and the final result follows from the fact that the projection operator $\mathcal{P}_\lambda$, defined by (3.5), when applied to $\exp(zz'^*)$, gives

$$\mathcal{P}_\lambda \exp(zz'^*) = \sum_{n=0}^{\infty} \frac{z^{n\kappa+\lambda}(z'^*)^{n\kappa+\lambda}}{(n\kappa+\lambda)!}. \tag{5.13}$$

Again as $\exp(zz'^*)$ is the reproducing kernel in the Bargmann Hilbert associated with the variable z, we conclude that we have the correspondence

$$\bar{z}I \longleftrightarrow \left[\sum_{\lambda=0}^{\kappa-1} \frac{z(d/dz)-\lambda}{\kappa\prod_{\nu=0}^{\kappa-1}[z(d/dz)-\nu]}\right]^{1/2} z^\kappa \mathcal{P}_\lambda, \tag{5.14a}$$

where, to stress the fact that $\bar{z}$ acts on κ component vectors, we multiply it by the unit $\kappa\times\kappa$ matrix. We wish to point out that both z^κ and $z(d/dz) = \kappa z^\kappa(d/dz^\kappa)$ are invariant under the group $\mathcal{C}_\kappa$ of (3.3) and thus they commute with the operator $\mathcal{P}_\lambda$ of (3.5) which projects irreducible representations of this group. Thus we also see that to $\bar{z}I$ there corresponds

$$\bar{z}I \longleftrightarrow \sum_{\lambda=0}^{\kappa-1} \mathcal{P}_\lambda \left(\frac{z(d/dz)-\lambda}{\kappa\prod_{\nu=0}^{\kappa-1}[z(d/dz)-\nu]}\right)^{1/2} z^\kappa, \tag{5.14b}$$

which will be a form particularly useful in the discussion of the next section.

In a similar fashion when we replace $\bar{z}$ by $d/d\bar{z}$ in (5.12), we have that

$$\frac{d}{d\bar{z}}I \longleftrightarrow \sum_{\lambda=0}^{\kappa-1} \mathcal{P}_\lambda \frac{d^\kappa}{dz^\kappa}\left[\frac{z(d/dz)-\lambda}{\kappa\prod_{\kappa=0}^{\kappa-1}[z(d/dz)-\nu]}\right]^{1/2}, \tag{5.15}$$

which could also be obtained by taking the Hermitian conjugate of (5.14a) and using (5.10). Finally combining (5.14) and (5.15) we have

$$\bar{z}\frac{d}{d\bar{z}}I \longleftrightarrow \sum_{\lambda=0}^{\kappa-1}\frac{1}{\kappa}\left(z\frac{d}{dz}-\lambda\right)\mathcal{P}_\lambda. \tag{5.16a}$$

The projector $\mathcal{P}_\lambda$ in this expression appears but once as it commutes with the remaining terms and is idempotent. Again we note that by the same considerations that follow (5.14a) we can also write

$$\bar{z}\frac{d}{d\bar{z}}I \longleftrightarrow \sum_{\lambda=0}^{\kappa-1}\mathcal{P}_\lambda\frac{1}{\kappa}\left(z\frac{d}{dz}-\lambda\right). \tag{5.16b}$$

The commutation relations

$$\left[\frac{d}{dz}, z\right] = 1, \tag{5.17a}$$

$$\left[\frac{d}{d\bar{z}}I, \bar{z}I\right] = I \tag{5.17b}$$

imply corresponding relations for the transformed operators that may readily be verified explicitly. Clearly in (5.17a) we deal with an irreducible and in (5.17b) with a reducible (and explicitly reduced) representation of the Heisenberg algebra. Thus the standard Heisenberg algebras in z and $\bar{z}$ space cannot be mapped into each other.

Having established in this chapter the mapping of operators in the z space on the $\bar{z}$ space and vice versa, we are in a position to actually quantize the classical observables that appear in the Eq. (2.12) through the relations between z, d/dz and the operators η, ξ that are usual in Bargmann Hilbert space. We proceed to carry out this program in the next section.

6. EQUATIONS THAT DETERMINE THE UNITARY REPRESENTATION OF CANONICAL TRANSFORMATIONS AND THEIR SOLUTION

We now return to the task, outlined in Secs. 1 and 2, of finding the unitary representation of the canonical transformation given implicitly by Eqs. (2.9) when $k=1$, i.e., the one that takes us from an oscillator of frequency κ^{-1}, κ integer, to one of unit frequency.

As mentioned in Sec. 2 the classical observables appearing in (2.9) translate into the quantum mechanical operators of Eqs. (1.8) or (1.9) in an ambiguous fashion. The discussion of Secs. 3, 4, and 5 allows us though to resolve this ambiguity in a systematic way.

To begin with the analysis of Secs. 2 and 3 suggests that instead of a scalar unitary representation $\langle x'' | U | x' \rangle$ we deal with a κ component column vector

$$\langle x'' | \mathbf{U} | x' \rangle = \begin{bmatrix} \langle x'' | U^0 | x' \rangle \\ \langle x'' | U^1 | x' \rangle \\ \cdot \\ \cdot \\ \cdot \\ \langle x'' | U^{\kappa-1} | x' \rangle \end{bmatrix}, \tag{6.1}$$

where $\langle x'' | U^\lambda | x' \rangle$, $\lambda = 0, 1, \ldots, \kappa - 1$ are associated with the corresponding irreducible representations of the cyclic group $\mathcal{C}_\kappa$ of linear canonical transformations (2.13). This is in entire analogy with the $C^\lambda(z, z^*)$ of (5.5), in which the components were associated with the irreducible representations of the group $\mathcal{C}_\kappa$ whose elements are the conformal transformations (3.3).

We now turn our attention to Eqs. (2.9) when $k=1$, from which we see that

$$\bar{H}(x,p) = \eta\xi, \quad \bar{G}(x,p) = \eta, \tag{6.2}$$

where η, ξ are given by (2.7) in terms of x, p. We thus have from the relations (1.8) and the operator mappings (5.14b), (5.16b) the expressions

$$\eta\xi\mathbf{U} = \mathbf{U}\sum_{\lambda=0}^{\kappa-1} \mathcal{P}_\lambda \kappa^{-1}(\eta\xi - \lambda), \tag{6.3a}$$

$$\eta\mathbf{U} = \mathbf{U}\sum_{\lambda=0}^{\kappa-1} \mathcal{P}_\lambda \left[\frac{\eta\xi - \lambda}{\kappa \prod_{\nu=0}^{\kappa-1} (\eta\xi - \nu)}\right]^{1/2} \eta^\kappa, \tag{6.3b}$$

where we made use of the correspondences

$$z \longleftrightarrow \eta, \quad \frac{d}{dz} \longleftrightarrow \xi \tag{6.4}$$

from Bargmann to ordinary Hilbert space.

We note that $\mathbf{U}$ is a κ-dimensional column vector operator which, from the remarks following (6.1), has the property that its components under the action of the projection operator $\mathcal{P}_\lambda$ become

$$U^\Lambda \mathcal{P}_\lambda = \delta_\lambda^\Lambda U^\Lambda, \quad \lambda, \Lambda = 0, 1, \ldots, \kappa - 1. \tag{6.5}$$

Furthermore in the classical limit we assume that the eigenvalues of the number operator $\eta\xi$ become very large and thus we can disregard any integer $1, 2, \ldots, \kappa - 1$, as compared with them. In that case, as we also have that the sum of the projection operators is the identity, i.e.,

$$\sum_{\lambda=0}^{\kappa-1} \mathcal{P}_\lambda = 1, \tag{6.6}$$

the right-hand side of Eqs. (6.3) reduces precisely to the classical expressions on the left side of (2.9).

We have in a definite way gone from the classical equations (2.9) when $k=1$, to the quantum mechanical operator relations (6.3). This quantization is by no means trivial as it implies that the unitary operator $\mathbf{U}$ is now a column vector whose components are bases for irreducible representations of the cyclic group $\mathcal{C}_\kappa$ of linear canonical transformations (2.13). Furthermore the operators associated with the classical expressions in (2.9) quantize differently for each irreducible representation $\lambda = 0, 1, \ldots, \kappa - 1$ of the $\mathcal{C}_\kappa$ group, though in the classical limit (i.e., when we assume that the eigenvalues of $\eta\xi \gg \lambda = 0, 1, \ldots, \kappa - 1$) they have the expected form.

It remains now to pass from Eqs. (6.3), which are the appropriate form of (1.8) for our problem, to those corresponding to (1.9). We must then take Eqs. (6.3) between the bra $\langle x'' |$ and the ket $| x' \rangle$ to obtain with the help of (1.9), (6.5), that for the component $\langle x'' | U^\lambda | x' \rangle$, $\lambda = 0, 1, \ldots, \kappa - 1$ we have the equations

$$\eta''\xi'' \langle x'' | U^\lambda | x' \rangle = \kappa^{-1}(\eta'\xi' - \lambda) \langle x'' | U^\lambda | x' \rangle, \tag{6.7a}$$

$$\eta'' \langle x'' | U^\lambda | x' \rangle = \xi'^\kappa \left[\frac{(\eta'\xi' - \lambda)}{\kappa \prod_{\nu=0}^{\kappa-1} (\eta'\xi' - \nu)}\right]^{1/2} \langle x'' | U^\lambda | x' \rangle, \tag{6.7b}$$

where, momentarily disregarding the Dirac notation, we have from (2.7) that the operators appearing in (6.7) have the form

$$\eta' = \frac{1}{\sqrt{2}}\left(x' - \frac{\partial}{\partial x'}\right), \quad \xi' = \frac{1}{\sqrt{2}}\left(x' + \frac{\partial}{\partial x'}\right),$$
$$\eta'' = \frac{1}{\sqrt{2}}\left(x'' - \frac{\partial}{\partial x''}\right), \quad \xi'' = \frac{1}{\sqrt{2}}\left(x'' + \frac{\partial}{\partial x''}\right). \tag{6.8}$$

The solution of Eqs. (6.7) is trivial as, denoting by $\phi_n(x')$ the normalized eigenstate of an harmonic oscillator of unit frequency, we easily see that

$$\langle x'' | U^\lambda | x' \rangle = \sum_{n=0}^{\infty} \phi_n(x'') \phi^*_{n\kappa+\lambda}(x') \tag{6.9}$$

satisfies them. This result could also have been obtained from $C^\lambda(\bar{z}, z^*)$ of (5.5) if we replaced monomials in $z, \bar{z}$ in Bargmann Hilbert space by harmonic oscillator states in the standard manner.

We note furthermore from (6.9) that

$$\int \langle x''' | U^\lambda | x'' \rangle^* \langle x''' | U^\lambda | x' \rangle \, dx'''$$
$$= \sum_{n=0}^{\infty} \phi_{n\kappa+\lambda}(x'') \phi^*_{n\kappa+\lambda}(x')$$
$$\equiv \langle x'' | P_\lambda | x' \rangle \qquad (6.10)$$

is the explicit expression for the projection operator P_λ, $\lambda = 0, 1, \ldots, \kappa - 1$, in the basis in which the original coordinate is diagonal. Thus from (6.6) and (6.10) we have that

$$\int \langle x'' | \mathbf{U}^\dagger | x''' \rangle \, dx''' \langle x''' | \mathbf{U} | x' \rangle$$
$$= \delta(x'' - x'), \qquad (6.11a)$$
$$\int \langle x'' | \mathbf{U} | x''' \rangle dx''' \langle x''' | \mathbf{U}^\dagger | x' \rangle$$
$$= I\delta(x'' - x'), \qquad (6.11b)$$

where I is the $\kappa \times \kappa$ unit matrix.

We can state that in a definite sense we determined the unitary representation in quantum mechanics of the nonlinear canonical transformation (2.12). Before proceeding to discuss these types of representation, let us first generalize them to the case when we have a canonical transformation that takes an oscillator of frequency κ^{-1} into one of frequency k^{-1}, where κ and k are relatively prime integers. We note that this canonical transformation defined implicitly by Eq. (2.9) remains invariant under the cyclic groups C_κ and C_k [whose elements are of the type (2.13)], acting respectively on the left- and right-hand side of these equations. Rather than use this fact to find an adequate quantization for the observables in (2.9), we shall take advantage of the analysis developed in this section for the case $k = 1$, to directly derive the unitary representation required.

To implement the program mentioned we consider three oscillators of frequencies κ^{-1}, $(\kappa k)^{-1}$, k^{-1} given respectively in the phase spaces (x, p), $(\boldsymbol{x}, \boldsymbol{p})$, $(\bar{x}, \bar{p})$. The canonical transformation that takes us from the phase space $(\boldsymbol{x}, \boldsymbol{p})$ to (x, p) is then the one associated with the passage from an oscillator of frequency k^{-1} to one of unit frequency. Thus its representation $\mathbf{V}$ is given by a k-component column vector whose elements are

$$\langle x'' | V^l | x' \rangle = \sum_{m=0}^{\infty} \phi_m(x'') \phi^*_{mk+l}(x'),$$
$$l = 0, 1, \ldots, k - 1. \qquad (6.12)$$

Similarly the canonical transformation that takes us from the phase space $(\boldsymbol{x}, \boldsymbol{p})$ to $(\bar{x}, \bar{p})$ leads to a representation U which is a κ component column vector whose elements are given by (6.9). When we want to go from the phase space (x, p) to $(\bar{x}, \bar{p})$, we first apply the inverse of the canonical transformation that takes us from $(\boldsymbol{x}, \boldsymbol{p})$ to (x, p) and then the one from $(\boldsymbol{x}, \boldsymbol{p})$ to $(\bar{x}, \bar{p})$. Thus the corresponding representation is given by a rectangular $\kappa \times k$ matrix

$$\mathbb{U} = \mathbf{U}\mathbf{V}^\dagger, \qquad (6.13)$$

whose elements are characterized by the indices $\lambda = 0, 1, \ldots, \kappa - 1$ for the row and $l = 0, 1, \ldots, k - 1$ for the column, i.e.,

$$\mathcal{U}^{\lambda l} = U^\lambda V^{l\dagger}. \qquad (6.14)$$

From (6.9) and (6.12) we then obtain

$$\langle x'' | \mathcal{U}^{\lambda l} | x' \rangle$$
$$= \int \langle x'' | U^\lambda | x''' \rangle dx''' \langle x''' | V^{l\dagger} | x' \rangle$$
$$= \sum_{n,m} [\phi_n(x'') \phi^*_m(x') \delta_{n\kappa+\lambda, mk+l}]$$
$$= \sum_{\rho=0}^{\infty} \phi_{\rho k+l'}(x'') \phi^*_{\rho\kappa+\lambda'}(x'), \qquad (6.15)$$

where l', λ' are the unique solutions of the Diophantine equations

$$l'\kappa + \lambda = \lambda' k + l, \qquad (6.16)$$

with the restriction that $0 \leq l' < k$, $0 \leq \lambda' < \kappa$. The reduction to the single summation in the last line of (6.15) is achieved when we notice that from the Kronecker delta we can write

$$n\kappa + \lambda = mk + l = \rho k\kappa + r, \quad r = 0, 1, \ldots, k\kappa - 1. \qquad (6.17)$$

In turn we can express

$$r = \lambda' k + l = l' k + \lambda, \qquad (6.18)$$

giving rise to the above Diophantine equations.

Representation (6.15) could also have been obtained from the equations of the type (1.9) associated with the canonical transformation (2.9). Again, the introduction of an intermediate phase space $(\boldsymbol{x}, \boldsymbol{p})$ in which we have observables associated with an oscillator of frequency $(\kappa k)^{-1}$, i.e.,

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{p}) = (2k\kappa)^{-1}(\boldsymbol{p}^2 + \boldsymbol{x}^2), \qquad (6.19a)$$

$$\mathcal{G}(\boldsymbol{x}, \boldsymbol{p}) = \frac{[(1/2)(\boldsymbol{x} - i\boldsymbol{p})]^{k\kappa}}{(k\kappa)^{1/2}[\frac{1}{2}(\boldsymbol{p}^2 + \boldsymbol{x}^2)]^{(k\kappa-1)/2}}, \qquad (6.19b)$$

allow us to determine the equations of the type (1.9) in a simple fashion. For this we require the canonical transformations defined implicitly by the equations

$$H(x, p) = \mathcal{H}(\boldsymbol{x}, \boldsymbol{p}) = H(x, p),$$
$$G(x, p) = \mathcal{G}(\boldsymbol{x}, \boldsymbol{p}) = \bar{G}(\bar{x}, \bar{p}), \qquad (6.20)$$

in which the transition from x, p to $\bar{x}, \bar{p}$ takes place through $\boldsymbol{x}, \boldsymbol{p}$. We are thus in a position of relating quantum mechanical operators of the original x, p with those of the final $\bar{x}, \bar{p}$ and vice versa, by connecting both of them with operators in the intermediate $\boldsymbol{x}, \boldsymbol{p}$. The latter is already achieved in Sec. 5 as it corresponds again to relating quantum mechanical operators when we deal with a canonical transformation that maps an oscillator of frequency κ^{-1} or k^{-1} onto one of unit frequency. It is thus easy to obtain the equations corresponding to (6.7) for a canonical transformation that takes an oscillator of frequency κ^{-1} onto one of frequency k^{-1}. We do not write them explicitly as we already know that their solution is given by (6.15).

We have found the unitary representation of the nonbijective (i.e., not one to one onto) canonical transformation (2.9). In the concluding section we discuss the implications of this analysis for the unitary representations of arbitrary canonical transformations.

7. CONCLUSION

The discussion of the previous sections suggests the following steps for deriving the equations of the type (1.8) or (1.9) that determine the unitary representations of the classical canonical transformations defined implicitly by Eqs. (1.5).

(1) If the observables $H(x,p)$, $\bar{H}(\bar{x},\bar{p})$, considered as quantum mechanical operators, do not have the same spectra, it is convenient to introduce an auxiliary observable $\mathcal{H}(\boldsymbol{x},\boldsymbol{p})$ in a corresponding phase space, $(\boldsymbol{x},\boldsymbol{p})$, whose spectra contains both that of $H(x,p)$ and $\bar{H}(\bar{x},\bar{p})$. We then can discuss the canonical transformations defined implicitly by the equations

$$H(x,p)=\mathcal{H}(\boldsymbol{x},\boldsymbol{p})=\bar{H}(\bar{x},\bar{p}), \tag{7.1a}$$

$$G(x,p)=\mathcal{G}(\boldsymbol{x},\boldsymbol{p})=\bar{G}(\bar{x},\bar{p}), \tag{7.1b}$$

where

$$\{G,H\}_{x,p}=\{\mathcal{G},\mathcal{H}\}_{\boldsymbol{x},\boldsymbol{p}}=\{\bar{G},\bar{H}\}_{\bar{x},\bar{p}}. \tag{7.1c}$$

We restrict our analysis to the unitary representation of the canonical transformation relating $\boldsymbol{x},\boldsymbol{p}$ with x,p as we face the same type of problem when we relate $\boldsymbol{x},\boldsymbol{p}$ with $\bar{x},\bar{p}$. We can later combine these two representations to get the one associated with direct passage from x,p to $\bar{x},\bar{p}$.

This procedure was implemented explicitly in the last section where $\mathcal{H}(\boldsymbol{x},\boldsymbol{p})$ was the oscillator Hamiltonian associated with the frequency $(\kappa k)^{-1}$ whose spectrum contains the spectra of both $H(x,p)$, $\bar{H}(\bar{x},\bar{p})$ which correspond respectively to oscillators of frequency κ^{-1}, k^{-1}.

(2) We surmise then that the mapping of the phase space $\boldsymbol{x},\boldsymbol{p}$ onto x,p will not be one to one onto, i.e., it will be nonbijective. In fact we expect that there will be a group of canonical transformations (equivalent to the C_κ of the oscillator problem) that provides the number of points in the intermediate phase space $\boldsymbol{x},\boldsymbol{p}$ that map on a single point x,p. As it remains ambiguous, up to the transformations of this group, which is the point in $\boldsymbol{x},\boldsymbol{p}$ space to which a given x,p corresponds, we shall call this group, the *ambiguity group*. As shown in the previous sections the ambiguity group is central to both the classical and quantum mechanical problem. In particular it suggests that the operator **U** that corresponds to a representation of the canonical transformation, is a column vector each of whose components is characterized by an index (the "ambiguity spin" of Plebanski[10]) which is associated with a given irreducible representation of this group.

(3) We then need to derive the quantum mechanical operators associated with the classical observables $\mathcal{H}(\boldsymbol{x},\boldsymbol{p})$, $\mathcal{G}(\boldsymbol{x},\boldsymbol{p})$, $H(x,p)$, $G(x,p)$. We expect that these operators will act differently on states that correspond to different irreducible representations of the ambiguity group. The explicit determination of these operators may be the hardest part of the problem.

In the oscillator example discussed in this paper we took advantage of the relation between its eigenfunctions $\phi_n(x')$ and $z^n/(n!)^{1/2}$ in Bargmann Hilbert space to discuss the problem in the latter. The canonical transformation whose representation we wished to determine, could then be correlated with the conformal transformation $w=z^k$, which was the one that finally provided the explicit form for the operators. It is clear that this procedure cannot be followed when we are not dealing with oscillators. It is still possible though to obtain operators explicitly once we know the ambiguity group, as will be shown in future publications.[12]

Once the operators corresponding to $\mathcal{H}(\boldsymbol{x},\boldsymbol{p})$, are available, the validity of our analysis must be tested by checking whether in the classical limit they reduce to the corresponding observable.

(4) When the operators mentioned above are given we have the equations of the form (1.8) or (1.9), which we need to solve if we wish to obtain $\langle x''|\mathbf{U}|x'\rangle$ explicitly. Again the ambiguity group and its representations are central to this objective as was shown in the present paper for the oscillator problems. Furthermore, after obtaining $\langle x''|\mathbf{U}|x'\rangle$ we must test that the representation is unitary in the sense of Eqs. (6.11) for the oscillator problem.

After outlining our surmise on the general procedure to be followed for determining the unitary representation of a canonical transformation, we should stress some points to avoid misunderstandings. To begin with when we start with an observable $H(x,p)$ in our original phase space, and write the corresponding operator on Hilbert space, the transformation **U** does *not* necessarily take it into the operator obtained by standard quantization of $\bar{H}(\bar{x},\bar{p})$, which incidentally may have a different spectrum. What happens rather is that to the operator associated with $H(x,p)$ there corresponds another one in the final Hilbert space that has the *same* spectrum as illustrated by (5.11) in Bargmann Hilbert space. A similar result applies when we start with the operator in the final Hilbert space that is associated with $\bar{H}(\bar{x},\bar{p})$, i.e., its corresponding operator in the original Hilbert space must have the same spectrum as exemplified in (5.16) and is not the one obtained by standard quantization of $H(x,p)$.

Another point of importance is that equations such as (1.1c) should be interpreted in the sense

$$\bar{x}(x,p)=U^{-1}xU, \tag{7.2}$$

i.e., $\bar{x}(x,p)$ is to be considered as an operator in the original Hilbert space. Thus when the canonical transformation is defined through the implicit equations (1.5) and (1.6), the operators relation we must have are of the form

$$H(x,p)=U^{-1}\bar{H}(x,p)U, \tag{7.3a}$$

$$G(x,p)=U^{-1}\bar{G}(x,p)U, \tag{7.3b}$$

where, as in (7.2), $H(x,p)$ and $G(x,p)$ are defined in the *original* Hilbert space. The whole difficulty of the problem arises when we want to quantize $H(x,p)$ and $G(x,p)$ unambiguously and its there where the ambiguity group shows its full power.

We also wish to stress that the unitary representation of canonical transformations does not necessarily lead to *finite* dimensional rectangular matrices as in (6.13) and (6.14). A simple example occurs when we map an oscillator of frequency $\sqrt{2}$ onto one of unit frequency. As $\sqrt{2}$ can be approximated by a ratio k/κ, where k and κ are ever increasing relatively prime integers, this will lead to ambiguity groups $\mathcal{C}_\kappa$, $\mathcal{C}_k$ that become $\mathcal{C}_\infty$.

Finally we wish to remark on an analogy between canonical transformations and conformal transformations. The former are defined on a two-dimensional phase space while the latter are given on a two-dimensional complex plane. The structure of the complex plane was at first not fully understood even by mathematicians of the early nineteenth century who derived so many theorems of analysis connected with it. This structure became much deeper through the work of Riemann who introduced the fundamental concept of Riemann surface,[8] later refined by Klein and Weyl.[11]

In the phase space plane we seem to be still at the pre-Riemann stage waiting for a deeper insight, such as the one initiated in the work of Souriau,[13] that will allow us to understand it more fully. Once this is achieved we may be able to implement the phrase of Souriau at the 1975 Bonn meeting on geometrical quantization[14] which, freely reconstructed from memory, stated the following: "Physicists assume that the representation of canonical transformations in quantum mechanics is fully discussed in Dirac's book, while mathematicians think that with luck and great effort they may understand the subject in ten years."

ACKNOWLEDGMENT

The authors would like to thank Professor J. Plebanski for many helpful discussions on the subject of this article. In particular we are indebted to him for a better understanding of the role of projection operators in the representation of canonical transformations.

[1]M. Born, W. Heinsenberg, and P. Jordan, Z. Physik **33**, 879 (1925).
[2]P.A.M. Dirac, *Quantum Mechanics* (Clarendon, Oxford, 1958), 4th ed., pp. 103–7.
[3]P.A. Mello and M. Moshinsky, J. Math. Phys. **16**, 2017 (1976).
[4]E.P. Wigner, *Group Theory* (Academic, New York, 1953).
[5]P. Kramer, M. Moshinsky, and T.H. Seligman, "Complex extensions of canonical transformations and quantum mechanics" in *Group Theory and Applications*, edited by E.M. Loebl (Academic, New York, 1975).
[6]J. Plebanski, Bull. Acad. Pol. Sci. **III 2**, 219 (1954); Acta Phys. Pol. **14**, 275 (1955); L. Infeld and J. Plebanski, Bull. Acad. Pol. Sci. **III 3**, 95 (1955); Acta Phys. Pol. **14**, 41 (1955); J. Plebanski and T.H. Seligman (to be published).
[7]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[8]Z. Nehari, *Conformal Mapping* (McGraw-Hill, New York, 1952), pp. 74–80.
[9]B. Wolf, J. Math. Phys. **15**, 1295 (1974).
[10]J. Plebanski (private communication).
[11]Felix Klein, *On Riemann's Theory of Algebraic Functions and Their Integrals*, edited by Felix Klein (Dover, New York, 1963); Herman Weyl, *Die Idee der Riemannschen Fläche* (Teubner, Stuttgart, 1955).
[12]M. Moshinsky and T.H. Seligman, Ann. Phys. (N.Y.) (1978) (to be published).
[13]J. Souriau, *Les systemes dynamiques* (Dunod, Paris, 1970).
[14]*Proceedings of the Symposium on Differential Geometrical Methods in Mathematical Physics*, *Bonn* 1975, Lecture Notes in Mathematics **570** (Springer-Verlag, Berlin, 1977).

New variational bounds on generalized polarizabilities[a)]

Peter D. Robinson[b)]

Mathematics Research Center and Theoretical Chemistry Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706
(Received 5 July 1977)

New variational bounds are derived on the generalized polarizabilities of a quantum-mechanical system, for arbitrary complex frequencies $\zeta = \nu + i\omega$ and two different perturbations u and v. No power of the Hamiltonian h higher than h^2 is involved in the bounding functionals. For a certain range of ν-values, upper and lower bounding functionals are obtained which contain merely a single trial vector but also introduce an inverse operator like h^{-1}. This impractical feature can be avoided with a subsidiary variational principle, leading to bivariational upper and lower bounds. Explicit bivariational bounds are also derived which are valid for all values of ζ. Both theoretical and practical aspects of the bounds are discussed.

1. INTRODUCTION

Let a quantum-mechanical system be described by a self-adjoint Hamiltonian operator h in a complex Hilbert space $\mathcal{H}$, and suppose that h possesses a complete set of orthonormal eigenvectors $\{\theta_k\}$ with corresponding energy eigenvalues $\{E_k\}$. If the system is in a state θ_n, its dynamic polarizability $\alpha(\zeta)$ at complex frequency $\zeta = \nu + i\omega$ associated with a perturbation u can be defined as

$$\alpha(\zeta) = \beta(\zeta) + \beta(-\zeta), \quad (1.1)$$

where

$$\beta(\zeta) = \mathrm{Re}\sum_{k\neq n}(E_k - E_n + \zeta)^{-1}\langle u\theta_n, \theta_k\rangle\langle\theta_k, u\theta_n\rangle, \quad (1.2)$$

the summation being over all states different from θ_n. The notation $\langle\ ,\ \rangle$ denotes the complex inner product, so that for all Φ and Ψ in $\mathcal{H}$ and complex numbers s we have

$$\langle\Phi,\Psi\rangle = \overline{\langle\Psi,\Phi\rangle}, \quad \langle s\Phi,\Psi\rangle = \bar{s}\langle\Phi,\Psi\rangle,$$
$$\langle\Phi, s\Psi\rangle = s\langle\Phi,\Psi\rangle, \quad (1.3)$$

a bar denoting complex conjugate. Previous authors [see, for example, Refs. 1–8 and 25–27] have presented bounding variational functionals on $\alpha(\nu)$ or $\alpha(i\omega)$, the dynamic polarizabilities when ζ is wholly real or wholly imaginary. Often there has been a restriction to real u and real θ_n. With $\omega\neq 0$, shortcomings of many of these bounding functionals have been the high powers of h involved, and a multiplicative factor of ω^{-1} which is unfortunate for small ω.

Expressions similar to that in (1.2), but with inner products $\langle v\theta_n, \theta_k\rangle\langle\theta_k, u\theta_n\rangle$ involving different perturbations u and v, define quantities arising, for example, in theories of optical rotatary power[9] and nuclear-magnetic shielding and chemical shifts.[10,11] Such more general expressions can also arise in double perturbation theory.[12] With generalizations such as this in mind, as well as the desirability of admitting arbitrary complex frequencies, perturbations, and unperturbed states, we show how to derive upper and lower bounding variational functionals on the quantity

$$Z(f,g;\zeta) = 2\,\mathrm{Re}\sum_{k\neq n}(E_k - E_n + \zeta)^{-1}\langle g,\theta_k\rangle\langle\theta_k, f\rangle. \quad (1.4)$$

No separate consideration is necessary for quantities defined as imaginary parts of summations like that in (1.4), for g can merely be replaced by $-ig$ if necessary. Without significant loss of generality, the arbitrary complex vectors f and g are taken as members of the reduced Hilbert space $\mathcal{H}_n \subset \mathcal{H}$, containing all vectors in $\mathcal{H}$ which are orthogonal to θ_n, i.e.,

$$f, g \in \mathcal{H}_n = \{\Phi\,|\,\Phi\in\mathcal{H},\ \langle\Phi,\theta_n\rangle = 0\}. \quad (1.5)$$

Apart from their intrinsic theoretical interest, bounding variational functionals can in principle lead (with suitably artificial choice of trial vector) to bounds on unknown quantities in terms of certain known quantities, like sum rules or moments.[4,6] However, if they are to be a viable practical tool, bounding functionals must not present exceptionally severe problems of evaluation when reasonable trial vectors are employed. It is for this reason that we do not much concern ourselves with functionals which involve powers of the operator h higher than the second.

2. A BIVARIATIONAL APPROXIMATION TO Z

Variational approaches to the task of bounding Z stem from its alternative but equivalent specification in terms of the solution-vector ϕ of the equation in $\mathcal{H}_n$

$$(h - E_n + \zeta)\phi = f, \quad \phi, f\in\mathcal{H}_n. \quad (2.1)$$

This is simply

$$Z(f,g;\zeta) = 2\,\mathrm{Re}\langle g,\phi\rangle = \langle g,\phi\rangle + \langle\phi, g\rangle. \quad (2.2)$$

Setting $\zeta = \nu + i\omega$ (with ν and ω real) and defining for convenience

$$H = h - E_n + \nu, \quad (2.3)$$

Eq. (2.1) is

$$A\phi = f, \quad \phi, f\in\mathcal{H}_n, \quad (2.4)$$

with

$$A = H + i\omega. \quad (2.5)$$

This decomposition of the linear operator A as the sum of a self-adjoint part H and a skew-self-adjoint part $i\omega$ is important for the establishment of the bounds in Sec. 3.

[a)]Sponsored by the United States Army under Contract No. DAAG29-75-C-0024, and by a Grant from the University of Wisconsin-Madison Graduate School.

[b)]On leave from Bradford University, England.

Along with equation (2.4) we consider the auxiliary equation

$$A^*\psi = g, \quad \psi, g \in \mathcal{H}_n, \tag{2.6}$$

where

$$A^* = H - i\omega \tag{2.7}$$

is the Hilbert-space adjoint of A. We note that

$$\langle \phi, g\rangle = \langle \phi, A^*\psi\rangle = \langle A\phi, \psi\rangle = \langle f, \psi\rangle, \tag{2.8}$$

so that we can express Z in the form

$$Z(f,g;\zeta) = \langle g, \phi\rangle + \langle f, \psi\rangle \tag{2.9}$$

in terms of the solution vectors ϕ and ψ of Eqs. (2.4) and (2.6).

Associated with this pair of equations is the bivariational approximation to $\langle g, \phi\rangle$ given by

$$R(\Psi,\Phi) = -\langle \Psi, A\Phi\rangle + \langle \Psi, f\rangle + \langle g, \Phi\rangle,$$
$$\Psi, \Phi \in \mathcal{H}_n, \tag{2.10}$$

with the trial vector Ψ playing the role of a kind of Lagrange multiplier. In terms of the difference vectors

$$\delta\psi = \Psi - \psi \in \mathcal{H}_n, \quad \delta\phi = \Phi - \phi \in \mathcal{H}_n, \tag{2.11}$$

we have the relation

$$R(\Psi,\Phi) = \langle g, \phi\rangle - \langle \delta\psi, A\delta\phi\rangle. \tag{2.12}$$

The complex conjugate of $R(\Psi,\Phi)$ is

$$\bar{R}(\Psi,\Phi) = -\langle \Phi, A^*\Psi\rangle + \langle \Phi, g\rangle + \langle f, \Psi\rangle \tag{2.13}$$

$$= \langle f, \psi\rangle - \langle \delta\phi, A^*\delta\psi\rangle, \tag{2.14}$$

which is a bivariational approximation to $\langle f, \psi\rangle$ (or $\langle \phi, g\rangle$). Thus, by addition, the real functional

$$J(\Psi,\Phi) = R(\Psi,\Phi) + \bar{R}(\Psi,\Phi)$$
$$= -\langle \Psi, A\Phi\rangle - \langle \Phi, A^*\Psi\rangle + \langle \Psi, f\rangle + \langle f, \Psi\rangle$$
$$+ \langle g, \Phi\rangle + \langle \Phi, g\rangle \tag{2.15}$$

is a bivariational approximation to

$$J(\psi,\phi) = \langle g, \phi\rangle + \langle f, \psi\rangle = Z(f,g;\zeta) \tag{2.16}$$

with the property

$$J(\Psi,\Phi) = Z(f,g;\zeta) - \langle \delta\psi, A\delta\phi\rangle - \langle \delta\phi, A^*\delta\psi\rangle. \tag{2.17}$$

3. MIXED VARIATIONAL BOUNDS ON Z

In the event that H (the self-adjoint part of A) is a positive operator, with an inverse H^{-1}, it is possible to construct two special cases of the bivariational functional $J(\Psi,\Phi)$ which provide complementary (upper and lower) variational bounds on $Z(f,g;\zeta)$. To do this, we think in terms of the "mixed" vectors

$$x = \lambda\phi + \mu\psi, \tag{3.1}$$

$$y = \lambda\phi - \mu\psi, \tag{3.2}$$

where the scalar multipliers λ and μ are real and satisfy

$$2\lambda\mu = 1. \tag{3.3}$$

Combining Eqs. (2.4) and (2.6), we see that x and y are the solution vectors of the simultaneous equations

$$Hx + i\omega y = \lambda f + \mu g = p \quad \text{(say)} \tag{3.4}$$

and

$$Hy + i\omega x = \lambda f - \mu g = q \quad \text{(say)}. \tag{3.5}$$

If we now choose trial vectors

$$X = \lambda\Phi + \mu\Psi, \tag{3.6}$$

$$Y = \lambda\Phi - \mu\Psi \tag{3.7}$$

and let

$$\delta x = X - x = \lambda\delta\phi + \mu\delta\psi, \tag{3.8}$$

$$\delta y = Y - y = \lambda\delta\phi - \mu\delta\psi, \tag{3.9}$$

we find that, using (3.3),

$$J(\Psi,\Phi) = I(X,Y) = -\langle X, HX\rangle + \langle Y, HY\rangle - \langle X, i\omega Y\rangle$$
$$+ \langle Y, i\omega X\rangle + \langle X, p\rangle + \langle p, X\rangle - \langle Y, q\rangle$$
$$- \langle q, Y\rangle \tag{3.10}$$

and

$$I(X,Y) = Z(f,g;\zeta) - \langle \delta x, (H\delta x + i\omega\delta y)\rangle$$
$$+ \langle \delta y, (H\delta y + i\omega\delta x)\rangle. \tag{3.11}$$

Accordingly, if the trial vectors X and Y are constrained to satisfy the equation

$$HX + i\omega Y = p \tag{3.12}$$

in line with (3.4), so that

$$H\delta x + i\omega\delta y = 0, \tag{3.13}$$

it follows that the resulting functional $I_+(X,Y)$ has the property

$$I_+(X,Y) = Z(f,g;\zeta) + \langle \delta y, (H + \omega^2 H^{-1})\delta y\rangle, \tag{3.14}$$

and is thus a variational upper bound on Z. Similarly, if

$$HY + i\omega X = q, \tag{3.15}$$

in line with (3.5), so that

$$H\delta y + i\omega\delta x = 0, \tag{3.16}$$

it follows that the resulting function $I_-(X,Y)$ has the property

$$I_-(X,Y) = Z(f,g;\zeta) - \langle \delta x, (H + \omega^2 H^{-1})\delta x\rangle, \tag{3.17}$$

and is thus a variational lower bound on Z.

In terms of an arbitrary trial vector Y, we have

$$I_+(X,Y) = I(H^{-1}(p - i\omega Y), Y) = K_+(Y) \quad \text{say}, \tag{3.18}$$

with

$$K_+(Y) = \langle Y, HY\rangle + \langle (p - i\omega Y), H^{-1}(p - i\omega Y)\rangle$$
$$- \langle Y, q\rangle - \langle q, Y\rangle. \tag{3.19}$$

Similarly,

$$I_-(X,Y) = I(X, H^{-1}(q - i\omega X)) = K_-(X) \quad \text{say}, \tag{3.20}$$

with

$$K_-(X) = -\langle X, HX\rangle - \langle (q - i\omega X), H^{-1}(q - i\omega X)\rangle$$
$$+ \langle X, p\rangle + \langle p, X\rangle. \tag{3.21}$$

Thus we obtain the complementary upper and lower variational bounds

$$K_-(X) \leq K_-(x) = Z(f,g;\zeta) = K_+(y) \leq K_+(Y), \qquad (3.22)$$

with the bounding functionals each depending on a single "mixed" complex vector. Hence we call them "mixed" variational bounds. It is a straightforward matter to optimize the functionals with respect to parameters multiplying the trial vectors, and if H is real there is separation of contributions from the real and imaginary parts of the trial vectors. The ratio $\lambda;\mu$ is also a disposable parameter.

These mixed variational bounds hold good whenever

$$E_0 - E_n + \nu > 0, \quad n \neq 0, \qquad (3.23)$$

E_0 being the lowest energy eigenvalue of h. In the case $n=0$, they hold whenever

$$E_1 - E_0 + \nu > 0, \qquad (3.24)$$

E_1 being the closest eigenvalue to E_0, since $\mathcal{H}_0$ does not contain θ_0. If E_0 is degenerate, then $E_1 = E_0$. Given that (3.23) or (3.24) holds, we have

$$\langle \Phi, (h - E_n + \nu)\Phi \rangle = \langle \Phi, H\Phi \rangle \geq b \langle \Phi, \Phi \rangle, \quad b > 0, \qquad (3.25)$$

for all $\Phi \in \mathcal{H}_n$, with

$$b = E_0 - E_n + \nu, \quad n \neq 0, \quad b = E_1 - E_0 + \nu, \quad n = 0. \qquad (3.26)$$

It follows from (3.25) that H is positive, and since it is bounded below away from zero the inverse operator H^{-1} exists with domain the whole of $\mathcal{H}_n$.

The idea of introducing a mixture of equations like (2.4) and (2.6) in order to obtain bounds has been exploited for dissipative systems in real spaces by Collins[13]; see also Herrera.[14,15] The bivariational functional $I(X,Y)$ shows the saddle-type dependence on X and Y which is appropriate for complementary bounds.[16,17]

4. AVOIDANCE OF H^{-1}; IMPLICIT BIVARIATIONAL BOUNDS

A practical disadvantage of the bounding functionals $K_+(Y)$ and $K_-(X)$ is that they involve the inverse H^{-1}, which only in elementary cases is likely to have a representation simple enough to permit the evaluation of the relevant inner products. One way of avoiding H^{-1} in $K_+(Y)$ is to write K_+ as a functional of X via (3.12) giving

$$\begin{aligned} K_+(Y(X)) &= I(X, (i/\omega)(HX - p)) = (1/\omega^2)\langle (HX-p), H(HX-p)\rangle \\ &\quad + \langle X, HX\rangle - (i/\omega)\langle (HX-p), q\rangle \\ &\quad + (i/\omega)\langle q, (HX-p)\rangle. \end{aligned} \qquad (4.1)$$

Similarly,

$$\begin{aligned} K_-(X(Y)) &= I((i/\omega)(HY-q), Y) \\ &= -(1/\omega^2)\langle (HY-q), H(HY-q)\rangle - \langle Y, HY\rangle \\ &\quad + (i/\omega)\langle (HY-q), p\rangle - (i/\omega)\langle p, (HY-q)\rangle. \end{aligned} \qquad (4.2)$$

However, these forms involve H^3 (as well as factors of ω^{-1}), and we rule them out as impractical.

A better way of avoiding the H^{-1} terms in (3.19) and (3.21) is to bound them separately, using individual variational bounds of the type

$$\langle l, H^{-1} l\rangle \leq -\langle \chi, H\chi\rangle + \langle \chi, l\rangle + \langle l, \chi\rangle + (1/b)\| H\chi - l\|^2,$$
$$l, \chi \in \mathcal{H}_n, \quad b > 0, \qquad (4.3)$$

which follows from the positivity hypothesis (3.25). Taking $l = p - i\omega Y$ and $\chi = \tilde{X}$ in (4.3), we find that (3.19) gives, after simplification,

$$K_+(Y) \leq I(\tilde{X}, Y) + (1/b)\|H\tilde{X} + i\omega Y - p\|^2. \qquad (4.4)$$

Similarly, putting $l = q - i\omega X$ and $\chi = \tilde{Y}$ in (4.3), we obtain from (3.21) the result

$$K_-(X) \geq I(X, \tilde{Y}) - (1/b)\|H\tilde{Y} + i\omega X - q\|^2. \qquad (4.5)$$

Dropping the tildes in (4.4) and (4.5), we see from (3.22) that for arbitrary vectors X and Y in $\mathcal{H}_n$,

$$\begin{aligned} I(X,Y) - (1/b)\|HY + i\omega X - q\|^2 &\leq Z(f,g;\zeta) \\ &\leq I(X,Y) + (1/b)\|HX + i\omega Y - p\|^2. \end{aligned} \qquad (4.6)$$

These are *bivariational* bounds on $Z(f,g;\zeta)$, which with prescience we might have derived directly from (3.11) and (3.25). They hold whenever H is positive and bounded below away from zero, so that a suitable positive b can be found according to (3.26).

The mixed vectors (x,y) and (X,Y) were introduced with the object of deriving variational bounds depending on a single trial vector, and so their usefulness has evaporated in (4.6). Referring back to the original vectors (ψ,ϕ) and (Ψ,Φ), the bivariational bounds (4.6) become

$$\begin{aligned} J(\Psi,\Phi) &- (1/b)\{\lambda^2\|A\Phi - f\|^2 + \mu^2\|A^*\Psi - g\|^2 \\ &- \mathrm{Re}\langle A\Phi - f, A^*\Psi - g\rangle\} \leq Z(f,g;\zeta) \\ &\leq J(\Psi,\Phi) + (1/b)\{\lambda^2\|A\Phi - f\|^2 + \mu^2\|A^*\Psi - g\|^2 \\ &+ \mathrm{Re}\langle A\Phi - f, A^*\Psi - g\rangle\}. \end{aligned} \qquad (4.7)$$

With the optimal choice for the ratio $\lambda:\mu$ of

$$\lambda\|A\Phi - f\| = \mu\|A^*\Psi - g\|, \quad (2\lambda\mu = 1), \qquad (4.8)$$

we obtain the result

$$\boxed{\begin{aligned} J(\Psi,\Phi) + (1/b)S(\Psi,\Phi) - (1/b)C(\Psi,\Phi) &\leq Z(f,g;\zeta) \\ &\leq J(\Psi,\Phi) + (1/b)S(\Psi,\Phi) + (1/b)C(\Psi,\Phi) \end{aligned}}, \qquad (4.9)$$

where, in terms of H,

$$\begin{aligned} J(\Psi,\Phi) &= -\langle \Psi, H\Phi\rangle - \langle \Phi, H\Psi\rangle + i\omega\{\langle \Phi,\Psi\rangle - \langle \Psi,\Phi\rangle\} \\ &\quad + \langle \Psi, f\rangle + \langle f, \Psi\rangle + \langle g, \Phi\rangle + \langle \Phi, g\rangle, \end{aligned} \qquad (4.10)$$

$$\begin{aligned} S(\Psi,\Phi) &= \mathrm{Re}\langle (H + i\omega)\Phi - f, (H - i\omega)\Psi - g\rangle \\ &= \mathrm{Re}\langle A\delta\phi, A^*\delta\psi\rangle, \end{aligned} \qquad (4.11)$$

and

$$C(\Psi,\Phi) = \|(H + i\omega)\Phi - f\|\,\|(H - i\omega)\Psi - g\| = \|A\delta\phi\|\,\|A^*\delta\psi\|. \qquad (4.12)$$

We call the bivariational bounds (4.9) *implicit*, because they are contained in the mixed bounds (3.22).

In the special case of zero ω, when A becomes self-adjoint, the implicit bounds (4.9) are actually tighter than others previously derived for self-adjoint operators in real spaces (Ref. 18; see also Ref. 19).

5. EXPLICIT BIVARIATIONAL BOUNDS

The bounds in (4.9) only hold when the operator H is positive and bounded below away from zero by a positive b, given by (3.26). However, irrespectively of whether this condition is met, it is possible to derive explicitly alternative bivariational bounds which merely require the condition

$$\|A\Phi\| \geq a\|\Phi\|, \quad a>0, \quad \text{for all } \Phi \in \mathcal{H}_n. \tag{5.1}$$

Since

$$\|A\Phi\|^2 = \langle A\Phi, A\Phi\rangle = \langle \Phi, A^*A\Phi\rangle = \langle \Phi, H^2\Phi\rangle + \omega^2\langle \Phi,\Phi\rangle \geq (E_{n'} - E_n + \nu)^2\|\Phi\|^2 + \omega^2\|\Phi\|^2, \tag{5.2}$$

where $E_{n'}$ is the energy eigenvalue of h which minimizes $(E_{n'} - E_n + \nu)^2$, we see that condition (5.1) is satisfied by taking

$$a^2 = (E_{n'} - E_n + \nu)^2 + \omega^2, \quad a>0. \tag{5.3}$$

We bear in mind that $E_{n'} \neq E_n$ (unless E_n is degenerate), since $\mathcal{H}_n$ does not contain θ_n. Thus, disregarding the exceptional case of zero ζ and degenerate E_n, we can always find a constant a to satisfy (5.1).

Applying (5.1) to (4.12) and then using Schwarz's inequality, we have

$$C(\Psi,\Phi) \geq a\|\delta\phi\|\,\|A\delta\psi\| \geq a\left|\langle \delta\phi, A^*\delta\psi\rangle\right|. \tag{5.4}$$

The magnitude of the complex number on the right of (5.4) is greater than or equal to the magnitude of its real part, and so from (2.17) it follows that

$$C(\Psi,\Phi) \geq \tfrac{1}{2}a\left|\langle \delta\phi, A^*\delta\psi\rangle + \langle \delta\psi, A\delta\phi\rangle\right| = \tfrac{1}{2}a\left|Z - J(\Psi,\Phi)\right|. \tag{5.5}$$

Rearranging (5.5) we obtain at once the explicit bivariational bounds

$$\boxed{J(\Psi,\Phi) - (2/a)C(\Psi,\Phi) \leq Z(f,g;\zeta) \leq J(\Psi,\Phi) + (2/a)C(\Psi,\Phi)}. \tag{5.6}$$

Existence theorems for bounds of this type have recently been presented, together with some applications.[20,21]

6. THE CASE $f = g$

When $f=g$, as is the case for the dynamic polarizability $\alpha(\zeta)$ in (1.1), Eqs. (2.4) and (2.6) become

$$(H+i\omega)\phi = f = (H-i\omega)\psi. \tag{6.1}$$

If a similar relationship is imposed on the trial vectors Ψ and Φ by taking

$$\Phi = (H-i\omega)\Theta, \quad \Psi = (H+i\omega)\Theta, \tag{6.2}$$

for some trial vector Θ which is supposed to approximate $(H^2+\omega^2)^{-1}f$, the functionals in (4.9) and (5.6) become

$$J = -2\langle \Theta, H(H^2+\omega^2)\Theta\rangle + 2\langle f, H\Theta\rangle + 2\langle H\Theta, f\rangle \tag{6.3}$$

and

$$C = S = \|(H^2+\omega^2)\Theta - f\|^2. \tag{6.4}$$

The bivariational bounds (4.9) and (5.6) become variational bounds, involving high powers of H, of a type previously obtained.[3,5] The H^3 and H^4 rule them out for practical purposes. Thus there seems little point in trying to impose a constraint like (6.2). However, when choosing trial vectors, it would be sensible to have in mind the relationship between the respective ω dependence of Φ and Ψ which is implied by (6.2).

There are some simplifications to be made if f is a real vector, and the Hamiltonian h (and hence H) is a real operator. Then it follows from (6.1) that $\psi = \overline{\phi}$. Accordingly, let us set

$$\Phi = \Phi_1 + i\Phi_2, \quad \Psi = \Phi_1 - i\Phi_2 = \overline{\Phi}, \tag{6.5}$$

in the functionals J, C, and S, where Φ_1 and Φ_2 are real vectors. We obtain the functionals

$$J(\overline{\Phi},\Phi) = -2\langle \Phi_1, H\Phi_1\rangle + 2\langle \Phi_2, H\Phi_2\rangle + 4\omega\langle \Phi_1, \Phi_2\rangle + 4\langle \Phi_1, f\rangle, \tag{6.6}$$

$$C(\overline{\Phi},\Phi) = \|H\Phi_1 - \omega\Phi_2 - f\|^2 + \|H\Phi_2 + \omega\Phi_1\|^2, \tag{6.7}$$

and

$$S(\overline{\Phi},\Phi) = \|H\Phi_1 - \omega\Phi_2 - f\|^2 - \|H\Phi_2 + \omega\Phi_1\|^2. \tag{6.8}$$

The implicit bivariational bounds (4.9) become

$$J(\overline{\Phi},\Phi) - (2/b)\|H\Phi_2 + \omega\Phi_1\|^2 \leq Z(f,f;\zeta) \leq J(\overline{\Phi},\Phi) + (2/b)\|H\Phi_1 - \omega\Phi_2 - f\|^2, \tag{6.9}$$

and the explicit bounds (5.6) take the form

$$J(\overline{\Phi},\Phi) - (2/a)C(\overline{\Phi},\Phi) \leq Z(f,f;\zeta) \leq J(\overline{\Phi},\Phi) + (2/a)C(\overline{\Phi},\Phi), \tag{6.10}$$

which at $\nu=0$ is essentially that given by Burrows.[8] The mixed bounds $K_+(Y)$ and $K_-(X)$ simplify in this situation, too, particularly if we take $\lambda = \mu = 2^{-1/2}$, so that $X = \Phi_1\sqrt{2}$, $Y = i\Phi_2\sqrt{2}$, $p = f\sqrt{2}$ and $q=0$, yielding

$$K_+(Y) = 2\langle \Phi_2, H\Phi_2\rangle + 2\langle (f+\omega\Phi_2), H^{-1}(f+\omega\Phi_2)\rangle \tag{6.11}$$

and

$$K_-(X) = -2\langle \Phi_1, H\Phi_1\rangle - 2\omega^2\langle \Phi_1, H^{-1}\Phi_1\rangle + 4\langle \Phi_1, f\rangle. \tag{6.12}$$

Goscinski[4] has given the amplitude-optimized version of the bound in (6.12) for $\nu=0$.

7. DISCUSSION

Interest in the mixed variational bounds $K_+(Y)$ and $K_-(X)$ is primarily theoretical. Not only do they contain the implicit bivariational bounds, but also they can lead to bounds in terms of other known quantities. To take a simple example, when ω is small the solution to Eq. (2.4) is approximately $H^{-1}f - i\omega H^{-2}f$. Thus if we take

$$\Phi_1 = c_1 H^{-1}f, \quad \Phi_2 = -i\omega c_2 H^{-2}f \tag{7.1}$$

in the simplified functionals (6.11) and (6.12), and optimize with respect to c_1 and c_2, we should obtain bounds on $Z(f,f;\zeta)$ which are accurate for small ω. The bounds are

$$s_{-2}-\omega^2\left(\frac{1}{s_{-4}}+\frac{\omega^2}{s_{-2}}\right)^{-1}\le \tfrac{1}{2}Z(f,f;\zeta)$$

$$\le s_{-2}-\omega^2\left(\frac{1}{s_{-4}}+\frac{\omega^2 s_{-6}}{(s_{-4})^2}\right)^{-1} \quad (7.2)$$

in terms of the "sum rules"

$$s_n=\langle f,H^{n+1}f\rangle. \quad (7.3)$$

For ground-state hydrogen, the bounds (7.2) give the result (in atomic units)

$$4.2490\le\alpha(i\omega)\le 4.2503 \quad (7.4)$$

for the dipole polarizability at $\nu=0,\ \omega=0.1$.

When H is positive the quantity $Z(f,f;\zeta)$ is a series-of-Stieltjes-representable function of ω^2, and with suitable choices of trial vector different families of Padé approximant bounds can be derived from K_+ and K_-.[6,22] More generally, whenever $\langle f,(H^2+\omega^2)^{-1}g\rangle$ is real, it can be shown that

$$Z(f,g;\zeta)=\langle (H^{1/2}p),(H^2+\omega^2)^{-1}(H^{1/2}p)\rangle$$

$$-\langle (H^{1/2}q),(H^2+\omega^2)^{-1}(H^{1/2}q)\rangle, \quad (7.5)$$

whence $Z(f,g;\zeta)$ is the difference of two series-of-Stieltjes-representable functions. The K_+ and K_- functionals are again the appropriate ones to yield the Padé approximant bounds for this kind of situation,[23] rather than bivariational functionals which lead to Padé approximants plus correction terms.[21,24]

From a practical standpoint, it is unlikely that a convenient representation of H^{-1} will be available, in which case the mixed bounds $K_+(Y)$ and $K_-(X)$ are not of direct interest even though they only involve a single trial vector. Likewise bounding functionals containing H^3 and higher powers can be ignored, because of the consequent difficulties in evaluating the inner products when sensible trial vectors are employed. Thus for a practical tool we are left with the bivariational bounds, in either the implicit form (4.9) (only valid when H is bounded below away from zero by a positive number b) or the explicit form (5.6) (always valid). Although an extra trial vector is involved, only H and H^2 appear in the inner products, and this advantage is crucial. The bivariational bounds should still be used even when $f=g$; the simpler versions (6.9) or (6.10) are relevant when $f=g$ is a real vector and H is real.

Even when the implicit bounds (4.9) are avilable, they will not necessarily yield better results than the explicit bounds (5.6). For example, when ω is large, the number a [given by (5.3)] is of order ω, whereas the number b [given by (3.26)] is not. With the correct asymptotic choices

$$\Phi=-if/\omega,\quad \Psi=ig/\omega, \quad (7.6)$$

the explicit bounds give

$$J\pm\frac{2}{a}C=(i/\omega)(\langle f,g\rangle-\langle g,f\rangle)+(1/\omega^2)(\langle f,Hg\rangle$$

$$+\langle g,Hf\rangle)\pm O\left(\frac{1}{\omega^3}\right), \quad (7.7)$$

whereas the bounds (4.9) leave some uncertainty in the ω^{-2} term. However, in cases like (6.4) when $(S-C)$ is zero, or very small, it is clear that the lower bound in (4.9) is better than the lower bound in (5.6). We notice also that the explicit bounds can never have quite the correct ω dependence because of the square-root defining the number a. Thus, when both are available, the implicit and the explicit bivariational bounds should each be investigated in any given situation to see which gives the better results.

Extensive calculations of bounds on dynamic polarizabilities ($f=g$) for two-electron atoms at zero ω or zero ν have recently been carried out by Glover and Weinhold.[25-27] The functionals of Braun and Rebane[2] which they employed led to inaccuracies for small values of ω. Applications of the Glover–Weinhold techniques to the bivariational bounds developed here are being explored.[28] The numerical results exhibit the significant advantages of the bivariational techniques over previously available methods. A pleasing feature of the present approach is that the same bivariational functionals provide bounds whether of not $\omega=0$, $\nu=0$, or $f=g$.

ACKNOWLEDGMENTS

I should like to thank Professor Frank Weinhold for suggesting this work, and both him and Professor Saul Epstein for helpful discussions about it. I am grateful to Professor J. O. Hirschfelder and Professor Ben Noble for the hospitality and facilities of their institutes at the University of Wisconsin—Madison.

[1]T.K. Rebane, Opt. Spectry. 21, 66–67 (1966).
[2]P.A. Braun and T.K. Rebane, Int. J. Quant. Chem. 4, 639–46 (1972).
[3]S.T. Epstein, J. Chem. Phys. 48, 4716–17 (1968).
[4]O. Goscinski, Int. J. Quant. Chem. 2, 761–84 (1968).
[5]P.D. Robinson, J. Phys. A (Gen. Phys.) 2, 295–303 (1969).
[6]P.W. Langhoff, R.G. Gordon, and M. Karplus, J. Chem. Phys. 55, 2126–45 (1971).
[7]R. Ahlberg, Int. J. Quant. Chem. 8, 363–77 (1974).
[8]B.L. Burrows, J. Phys. B (Atom. Molec. Phys.) 9, 1–7 (1976).
[9]E.V. Condon, Rev. Mod. Phys. 9, 432–57 (1937).
[10]N.F. Ramsey, *Molecular Beams* (Clarendon, Oxford, 1956).
[11]D.K. Hindermann and C.D. Cornwell, J. Chem. Phys. 48, 4148–54 (1968).
[12]J.O. Hirschfelder, W. Byers Brown, and S.T. Epstein, Adv. Quant. Chem. 1, 255–374 (1964).
[13]W.D. Collins, "Dual extremum principles for dissipative systems," Technical Report No. 1624 (Mathematics Research Center, University of Wisconsin—Madison, 1976).
[14]I. Herrera, "A general formulation of variational principles," Report E.10 (Instituto de Ingenieria, Universidad Nacional Autónoma de México, 1974).
[15]I. Herrera and J. Bielak, Quart. Appl. Math. 34, 85–102 (1976).

[16]B. Noble and M.J. Sewell, J. Inst. Math. Appl. 9, 123–93 (1972).
[17]M.J. Sewell, J. Struct. Mech. 2, 1–32 (1973).
[18]M.F. Barnsley and P.D. Robinson, Proc. Roy. Soc. A 338, 527–33 (1974).
[19]R.J. Cole and D.C. Pack, Proc. Roy. Soc. A 347, 239–52 (1975).
[20]M.F. Barnsley and P.D. Robinson, Proc. Roy. Soc. Edinburgh A 75, 109–18 (1976).
[21]M.F. Barnsley and G.A. Baker, Jr., J. Math. Phys. 17, 1019–27 (1976).
[22]M.F. Barnsley and P.D. Robinson, J. Inst. Math. Appl. 14, 229–49 (1974).
[23]M.F. Barnsley, J. Math. Phys. 17, 559–65 (1976).
[24]M.F. Barnsley, J. Math. Phys. 16, 918–28 (1975).
[25]R.M. Glover and F. Weinhold, J. Chem. Phys. 65, 4913–26 (1976).
[26]R.M. Glover and F. Weinhold, J. Chem. Phys. 66, 185–90 (1977).
[27]R.M. Glover and F. Weinhold, J. Chem. Phys. 66, 191–98 (1977).
[28]F. Weinhold and P.D. Robinson, "Bivariational calculations of bounds on complex-frequency polarizabilities," submitted for publication.

Charge conservation in metric–torsion gravitational theories

S. J. Aldersley

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3Gl
(Received 14 February 1977)

The concepts of charge conservation and dimensional consistency are employed to derive conditions which uniquely characterize the field equations of electromagnetism and gravitation in a vector–metric–torsion field theory which contains no cosmological constant. Using these results, it is shown that the Einstein–Maxwell field equations are unique among all possible field equations of gravitation and electromagnetism involving a vector field in a metric gravitational theory.

1. INTRODUCTION

The purpose of this paper is to apply the techniques of dimensional analysis to the consideration of electromagnetic and gravitational theories which involve a vector field in addition to the usual metric tensor and, for added generality, a torsion tensor.[1]

We shall be interested in field equations which take the form

$$A^{ij} = 8\pi\kappa_1\sqrt{g}T^{ij}, \tag{1.1}$$

$$B^i = 16\pi\kappa_2\sqrt{g}J^i, \tag{1.2}$$

and

$$C^{lm}{}_k = 8\pi\kappa_1\sqrt{g}\mu_k{}^{ml}, \tag{1.3}$$

where A^{ij}, B^i, and $C^{lm}{}_k$ are assumed to be tensor density concomitants of the metric g_{ij}, the torsion $S_{jk}{}^i$, and the vector field ψ_i along with their partial derivatives to some order. T^{ij} is the (symmetric) energy—momentum tensor of all nonelectromagnetic matter fields, J^i is the charge-current vector of these fields and $\mu_k{}^{ml}$ is their pseudospin tensor. It is assumed that the tensor densities A^{ij}, B^i, and $2C^{lm}{}_k$ are variational derivatives of a suitable scalar density Lagrangian with respect to g_{ij}, ψ_i, and $S_{lm}{}^k$.

An important feature of electromagnetic theories is the physical assumption that charge is conserved. Mathematically this takes the form[2]

$$(\sqrt{g}J^i)_{,i} = 0. \tag{1.4}$$

Hence it appears reasonable to demand that in (1.2), the concomitant B^i satisfy

$$B^i{}_{,i} = 0,$$

identically. The principle of conservation of charge has been investigated by Horndeski[3,4] and the technique of dealing with (1.4) which will be employed here is a generalization of his approach.

Probably the most significant physical assumption made here is that of dimensional consistency. This assumption has been developed into a technique of dimensional analysis by the author and been employed in the consideration of a number of concomitant problems related to relativistic gravitational theories. The axiom for dimensional analysis and a discussion of its consequences can be found in Refs. 5 and 6.

2. THE MAIN THEOREM

In this section we state the main theorem of the paper and discuss some of its consequences. The proof will be given in Sec. 3.

The dimensions of the functions involved are determined as follows. From a previous paper[6] we have

$$g_{ij,k_1\cdots k_\alpha} \sim L^{-\alpha}, \quad S_{ij}{}^k{}_{,l_1\cdots l_\beta} \sim L^{-(1+\beta)},$$

$T^{ij} \sim L^{-2}$, and $\mu_k{}^{ml} \sim L^{-1}$. In conventional electromagnetic theory $F_{ij} \sim L^{-1}$ where $F_{ij} := \psi_{j,i} - \psi_{i,j}$, thus we have $\psi_i \sim L^0$ and $\psi_{i,j_1\cdots j_\gamma} \sim L^{-\gamma}$.

Since J^i is a charge-current 3-density we must have $J^i \sim L/L^3 \sim L^{-2}$. By demanding that the field equations (1.1), (1.2), and (1.3) be dimensionally consistent we find that $A^{ij} \sim L^{-2}$, $B^i \sim L^{-2}$, and $C^{lm}{}_k \sim L^{-1}$,

Theorem: Suppose that A^{ij} (of class C^3), B^i, (C^3), and $C^{lm}{}_k$, (C^2), are tensor density concomitants of g_{ab}, ψ_r, $S_{rs}{}^t$ and their various partial derivatives to some finite order and satisfy the following conditions:

(a) $A^{ij} \sim L^{-2}$, $B^i \sim L^{-2}$, $C^{lm}{}_k \sim L^{-1}$ and they satisfy the axiom of dimensional analysis;

(b) there exists a scalar density $\mathcal{L}$ which is a concomitant (both tensorially and dimensionally) of the form

$$\mathcal{L} = \mathcal{L}(g_{ab}; \cdots; g_{ab,c_1\cdots c_\alpha}; \psi_a; \cdots; \psi_{a,b_1\cdots b_\beta}; S_{ab}{}^c; \cdots; S_{ab}{}^c{}_{,d_1\cdots d_\gamma})$$

such that $A^{ij} = \delta\mathcal{L}/\delta g_{ij}$, $B^i = \delta\mathcal{L}/\delta\psi_i$, and $C^{lm}{}_k = \frac{1}{2}\delta\mathcal{L}/\delta S_{lm}{}^k$;

(c) $B^i{}_{,i} = 0$.

Then in a 4-space,

$$A^{ij} = a_1\sqrt{g}G^{ij} + \tfrac{1}{2}g^{ij}C^{rs}{}_tS_{rs}{}^t + C^{rsi}S_{rs}{}^j - 2C^{rj}{}_sS_r{}^{is} - \tfrac{1}{2}a_2\sqrt{g}(F^i{}_lF^{jl} - \tfrac{1}{4}g^{ij}F^{rs}F_{rs}), \tag{2.1}$$

$$B^i = a_2\sqrt{g}F^{ij}{}_{|j}, \tag{2.2}$$

and

$$C^{lm}{}_k = \sqrt{g}\{a_3\delta^{[l}{}_kS_b{}^{m]b} + a_4S^{lm}{}_k + a_5S_k{}^{[lm]}\} + a_6\epsilon^{lmbc}S_{bck} + a_7\epsilon^{lmab}S_{kab} - a_7\epsilon^{abc[m}S_{ab}{}^{l]}g_{ck}, \tag{2.3}$$

where a_σ, $\sigma = 1, \ldots, 7$ are arbitrary unitless constants. Moreover, a Lagrangian which satisfies condition (b) for these expressions is given by

$$\mathcal{L} = -a_1\sqrt{g}R + C^{rs}{}_tS_{rs}{}^t + \tfrac{1}{4}a_2\sqrt{g}F^{rs}F_{rs}. \tag{2.4}$$

Condition (b) excludes the consideration of dependence upon constants with nonzero dimension ($\sim L^\alpha$, $\alpha \neq 0$). Hence the theorem applies to those theories of electro-

 0022-2488/78/1903-0700$1.00

magnetism and gravitation which do not involve universal constants (such as the cosmological constant) other than c and K. Although this prevents the elementary charge e from occurring explicitly in the field equations, it may arise implicitly in a source term (e.g., in J^i).

The above theorem leads to a number of interesting and quite general conclusions about electromagnetic fields in a metric—torsion gravitational theory. Provided one prescribes the electromagnetic and gravitational fields via equations of the form (1.1), (1.2), and (1.3) and if one accepts the restrictions of the theorem as physically reasonable, then A^{ij}, B^i, and $C^{lm}{}_k$ must be given by (2.1), (2.2), and (2.3) respectively. As is apparent from (2.2) and (2.3) no direct interaction terms occur between the electromagnetic field ψ_i and the torsion field $S_{ij}{}^k$. In the absence of sources the field equations reduce to the vacuum Einstein—Maxwell equations since for most values of the constants in (2.3), one can deduce[6] $S_{ij}{}^k=0$ (proving that[1] "photons do not produce torsion").

Within the context of conventional metric gravitational theories we have the following:

Corollary: Suppose A^{ij} and B^i satisfy the conditions of the theorem but are also independent of the torsion and its derivatives. Then in a 4-space

$$A^{ij}=a_1\sqrt{g}G^{ij}-\tfrac{1}{2}a_2\sqrt{g}(F^i{}_l F^{jl}-\tfrac{1}{4}g^{ij}F^{rs}F_{rs})$$

and B^i is given by (2.2).

A Lagrangian satisfying condition (b) of the theorem is given by

$$L=-a_1\sqrt{g}R+\tfrac{1}{4}a_2\sqrt{g}F^{rs}F_{rs}.$$

It appears that any generalization of the above results will most likely involve dimensional constants in a direct manner. Along these lines, Horndeski[3] has investigated second-order electromagnetic theories (within a vector—metric context) and found that the Einstein—Maxwell equations (with sources and cosmological term) could be modified by the addition of terms involving a dimensioned constant and still satisfy the law of conservation of charge.

3. PROOF OF THE MAIN THEOREM

The main theorem is proven with the aid of several propositions, beginning with one which is based upon the concept of dimensional analysis.

Proposition 1: Suppose A^{ij}, B^i, and $C^{lm}{}_k$ are tensor density concomitants of g_{ab}, ψ_r, $S_{rs}{}^t$ and their partial derivatives to some order, where A^{ij} and B^i are of class C^3 and $C^{lm}{}_k$ is of class C^2. In addition we assume that $A^{ij}\sim L^{-2}$, $B^i\sim L^{-2}$, and $C^{lm}{}_k\sim L^{-1}$ and each concomitant satisfies the axiom of dimensional analysis. Then A^{ij} and B^i are linearly homogeneous in $g_{ab,cd}$, $S_{ab}{}^c{}_{,d}$ and $\psi_{a,bc}$ and quadratically homogeneous in $g_{ab,c}$, $\psi_{a,b}$, and $S_{ab}{}^c$ while $C^{lm}{}_k$ is linearly homogeneous in the latter set, all with coefficients which are (zeroth order) concomitants of g_{ab} and ψ_r. ∎

The proof is similar to that used to deduce (2.5) in Ref. 6.

In the next proposition the symbol ρ_A is used to denote a collection of (not necessarily tensorial) field functions [e.g., $\rho_A=(g_{ab};S_{ab}{}^c)$].

Proposition 2: Let L be a concomitant of the form

$$L=L(\rho_A;\dots;\rho_{A,i_1\cdots i_\alpha};\psi_h;\dots;\psi_{h,k_1\cdots k_\beta}).$$

If

$$\frac{\delta L}{\delta\psi_h}{}_{,h}=0, \tag{3.1}$$

identically; then

$$\frac{\partial}{\partial\psi_h}\left(\frac{\delta L}{\delta\psi_k}\right)=0, \tag{3.2}$$

and

$$\frac{\partial}{\partial\psi_h}\left(\frac{\delta L}{\delta\rho_A}\right)=0. \tag{3.3}$$

Moreover, if L is a scalar density and the fields ρ_A are tensorial then (3.2) and (3.3) imply (3.1). ∎

The proof of (3.2) and (3.3) can be deduced from Theorem (3.1) of Ref. 4 by substituting ρ_A for g_{ij} and $\delta/\delta\rho_A$ for $\delta/\delta g_{ij}$. The converse follows similarly from Theorem (2.2) of the same paper. [The required generalization of Eq. (2.15) of Ref. 4 is obtained using Eq. (4.2) of Ref. 7.] Proposition 2 implies that the theorem requires the coefficients described in Proposition 1 to be independent of ψ_i and thus concomitants of g_{ab} only. At this point we can construct $C^{lm}{}_k$.

Proposition 3: If $C^{lm}{}_k$ is a class C^1 tensor density concomitant of the form

$$C^{lm}{}_k=C^{lm}{}_k(g_{rs};\cdots;g_{rs,t_1\cdots t_\alpha};\psi_{r,s};\cdots;\psi_{r,s_1\cdots s_\beta};S_{rs}{}^t;\cdots;S_{rs}{}^t{}_{,u_1\cdots u_\gamma})$$

such that $C^{lm}{}_k\sim L^{-1}$ and satisfies the axiom of dimensional analysis, then in a 4-space $C^{lm}{}_k$ is given by (2.3).

Proof: Evidently the coefficients of $g_{ab,c}$, $S_{rs}{}^t$, and $\psi_{r,s}$ in $C^{lm}{}_k$ depend upon g_{ab} only. Since $\partial C^{lm}{}_k/\partial\psi_{r,s}$ is a tensor density concomitant of g_{ab}, it vanishes in a 4-space by a well-known result.[8] Using the replacement theorems of classical tensor analysis,[9] we deduce that $C^{lm}{}_k=\eta^{lm}{}_k{}^{rs}{}_t(g_{ab})S_{rs}{}^t$. Equation (2.3) follows from the construction given in Ref. 6. ∎

Since $C^{lm}{}_k=\tfrac{1}{2}\delta(C^{rs}{}_tS_{rs}{}^t)/\delta S_{lm}{}^k$, where $C^{lm}{}_k$ is given by (2.3), that part of the theorem is proven.

Bearing in mind the results of Propositions 1—3 along with the identities[10]

$$\frac{\partial}{\partial S_{rs}{}^t{}_{,u}}\left(\frac{\delta L}{\partial\psi_h}\right)=-\frac{\partial}{\partial\psi_{h,u}}\left(\frac{\delta L}{\delta S_{rs}{}^t}\right)$$

and

$$\frac{\delta}{\delta\psi_h}\left(\frac{\delta L}{\delta S_{lm}{}^k}\right)=\frac{\partial}{\partial S_{lm}{}^k}\left(\frac{\delta L}{\delta\psi_h}\right),$$

we see that B^i is independent of $S_{lm}{}^t$ and $S_{lm}{}^k{}_{,i}$. Hence B^i is linearly homogeneous in $g_{ab,cd}$ and $\psi_{r,st}$ and quadratically homogeneous in $g_{ab,c}$ and $\psi_{r,s}$ with coefficients depending upon g_{ab} only. B^i is then given by the following result.

Proposition 4: If B^i is a class C^2 vector density concomitant of the form

$$B^i = B^i(g_{rs};\cdots;g_{rs,t_1\cdots t_\alpha};\psi_{r,s};\cdots;\psi_{r,s_1\cdots s_\beta})$$

and suppose $B^i \sim L^{-2}$ and satisfies the axiom of dimensional analysis, then in a 4-space B^i is given by (2.2).

Proof: The usual dimensional technique[5] implies that B^i must have the form described in the previous paragraph. Since the tensor density $\partial B^i/\partial g_{ab,cd}$ is a concomitant of g_{ab} (only) it vanishes in a 4-space. The invariance identity[7] $\partial B^i/\partial\psi_{(r,st)}=0$ enables one to deduce that[11]

$$\frac{\partial B^i}{\partial \psi_{r,st}} = a\sqrt{g}\,[g^{ir}g^{st} - \tfrac{1}{2}(g^{is}g^{rt} + g^{it}g^{sr})].$$

Hence $B^i = a\sqrt{g}F^{ij}{}_{|j} + \tilde{B}^i$ where $\tilde{B}^i$ is a vector density (quadratically homogeneous in $\psi_{r,s}$ and $g_{ab,c}$ with coefficients depending upon g_{ab} only). As above, the tensor density $\partial^2\tilde{B}^i/\partial\psi_{a,b}\partial\psi_{r,s}=0$. The replacement theorem then implies that $\tilde{B}^i=0$, from which the proposition follows. ■

Since $a\sqrt{g}F^{ij}{}_{|j} = \delta(\frac{1}{4}a\sqrt{g}F^{rs}F_{rs})/\delta\psi_i$ and $(\sqrt{g}F^{ij}{}_{|j})_{|i} \equiv 0$, B^i, as given by (2.2), satisfies the conditions of the theorem.

The proof of the theorem can now be completed. Taking into account the previous propositions along with the identities[10]

$$\frac{\partial}{\partial g_{ab,c}}\left(\frac{\delta \mathcal{L}}{\partial S_{lm}{}^k}\right) = -\frac{\partial}{\partial S_{lm}{}^k{}_{,c}}\left(\frac{\delta \mathcal{L}}{\delta g_{ab}}\right)$$

and

$$\frac{\partial}{\partial g_{ab,cd}}\left(\frac{\delta \mathcal{L}}{\delta \psi_h}\right) = \frac{\partial}{\partial \psi_{h,cd}}\left(\frac{\delta \mathcal{L}}{\delta g_{ab}}\right),$$

we find that $\partial A^{ij}/\partial S_{lm}{}^k{}_{,a}=0$ and $\partial A^{ij}/\partial\psi_{h,cd}=0$. We also have $\partial^2A^{ij}/\partial\psi_{r,s}\partial S_{ab}{}^c=0$ and $\partial^2A^{ij}/\partial g_{ab,c}\partial\psi_{r,s}=0$ since they are both tensor density concomitants of g_{ab} (only) in a 4-space. Differentiating the second invariance identity[7] for A^{ij} with respect to ψ_h yields $\partial A^{ij}/\partial\psi_{(r,s)}=0$, implying that

$$\frac{\partial^2 A^{ij}}{\partial\psi_{r,s}\partial\psi_{a,b}} = -\frac{\partial^2 A^{ij}}{\partial\psi_{s,r}\partial\psi_{a,b}}.$$

Hence we can write A^{ij} as

$$A^{ij} = \tilde{A}^{ij} + \tfrac{1}{4}\eta^{ijabcd}(g_{rs})F_{ab}F_{cd}, \tag{3.4}$$

where $\tilde{A}^{ij}$ is a tensor density concomitant linear in $g_{ab,cd}$ and quadratic in $g_{ab,c}$ and $S_{lm}{}^k$. The identity[10]

$$\frac{\partial}{\partial\psi_{r,s}}\left(\frac{\delta \mathcal{L}}{\delta g_{ij}}\right) = -\frac{\partial}{\partial g_{ij,s}}\left(\frac{\delta \mathcal{L}}{\delta\psi_r}\right)$$

implies that

$$\eta^{ijabcd} = -\frac{1}{2}\frac{\partial^2 B^c}{\partial\psi_{a,b}\partial g_{ij,d}},$$

and thus from (2.2) we deduce that

$$\tfrac{1}{4}\eta^{ijabcd}F_{ab}F_{cd} = -\tfrac{1}{2}a\sqrt{g}\,(F^i{}_lF^{jl} - \tfrac{1}{4}g^{ij}F^{rs}F_{rs}). \tag{3.5}$$

It can be shown that[6]

$$\tilde{A}^{ij} = b\sqrt{g}\,G^{ij} + \tfrac{1}{2}g^{ij}C^{rs}{}_tS_{rs}{}^t + C^{rsi}S_{rs}{}^j - 2C^{rj}{}_sS_r{}^{is},$$

which together with (3.4) and (3.5) implies (2.1). That the Lagrangian (2.4) satisfies condition (b) of the theorem is straightforward. This proves the theorem.

ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. G.W. Horndeski for his comments and suggestions and to the referee for his remarks. This work was supported in part by the National Research Council of Canada.

[1]F.W. Hehl, Gen. Rel. Grav. 4, 333 (1973); 5, 491 (1974).
[2]Notation and mathematical conventions are the same as those employed in S.J. Aldersley, Phys. Rev. D 15, 3507 (1977).
[3]G.W. Horndeski, J. Math. Phys. 17, 1980 (1976).
[4]G.W. Horndeski, "Gauge Invariance and Charge Conservation," (submitted for publication).
[5]S.J. Aldersley, Phys. Rev. D 15, 370 (1977).
[6]S.J. Aldersley, Gen. Rel. Grav. 8, 397 (1977).
[7]G.W. Horndeski, Utilitas Mathematica 9, 3 (1976).
[8]S. Golab and M. Kucharzewski, Acta Math. Acad. Sci. Hungar. 11, 173 (1960).
[9]See T.Y. Thomas, *The Differential Invariants of Generalized Spaces* (Cambridge U.P., Cambridge, 1934), p. 109.
[10]These identities are due to I.M. Anderson, "Tensorial Euler-Lagrange Expressions and Conservation Laws," to appear in Aeq. Math. (see in particular the remark made following Theorem 2 of this paper.)
[11]See D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles* (Wiley-Interscience, New York, 1975), p. 320.

On the Weyl coefficients of SO(n)

M. K. F. Wong

Fairfield University, Fairfield, Connecticut 06430
(Received 5 April 1977)

It is shown that the representation functions of SO(n) can be obtained from those of SO($n-1$) by calculating the Weyl coefficient of SO(n), which is equal to the representation function of SO(n) calculated at $\theta=\pi/2$. Thus the problem of calculating the representation functions of SO(n) is reduced to its calculation at a particular value of θ ($=\pi/2$), instead of a whole range of values of θ. The Weyl coefficients of SO(3), SO(4), and SO(5) are explicitly obtained and discussed in detail. It is shown that the Weyl coefficients of SO(n) are all expressible as sums and products of 3-j symbols of SO(3) and normalization factorials. We also conclude that the representation functions of all orthogonal groups (as well as unitary groups) are ultimately reducible to the representation function of SO(2), i.e., $e^{im\theta}$, multiplied by corresponding Weyl coefficients. Therefore, all representation functions of SO(n) and U(n) are ultimately expressible as Fourier series, whose coefficients can be explicitly calculated if one so wishes.

1. INTRODUCTION

It is interesting to note that our knowledge of the representation functions of U(n) surpasses our knowledge of the representation functions of SO(n); in other words, the representation functions of the two classical groups have not developed at an equal pace. For the unitary group, it was first pointed out by Gel'fand and Graev[1] that the representation functions can be expressed as generalized beta functions. A different approach was made by Chacón and Moshinsky[2] while calculating the finite transformation matrix of U(3). They calculated the Weyl coefficients of U(3), and found that this was connected with the 6-j symbols of U(2). Holman[3] and Wong[4] then showed that the Weyl coefficients of U(n) are basically 6-j symbols of U($n-1$). We intend to show in a future publication that all the Weyl coefficients of U(n), or equivalently all multiplicity-free 6-j symbols of U(n), can be explicitly evaluated, and not merely as sums and products of four 3-j symbols. Louck and Biedenharn[5] showed that the generalized beta functions of Gel'fand and Graev are expressible as a product of an isoscalar factor in U(n) and an isoscalar factor in U($n-1$). Wong[4] showed that it can also be expressed as a product of a stretched 6-j symbol in U($n-1$) and an isoscalar factor in U($n-1$). Thus we can at least say that our present knowledge of the representation functions of U(n) has reached a satisfactory stage.

We cannot say the same with regard to the representation functions of SO(n). So far there are basically two approaches, one by Maekawa[6] and the other by Vilenkin[7] and Wolf.[8] Maekawa's method is to write the d function for the highest weight, and use lowering operators of Pang and Hecht[9] (or Wong[10]) operating on the highest weight to get the general state. Vilenkin and Wolf's method is to write the d function as an integral over its maximum compact subgroup, using the multiplier representation for SO(n,1), and analytically continue to SO($n+1$).

There remain a few questions unanswered by these two methods. The first is: What is the structure of the d functions for SO(n) in general? Is it expressible in terms of elementary functions, as we know the d functions of U(n) are? And if so, do the coefficients have some meaning? The second problem is: Why are there different expressions for the representation functions of the same group, and how are they related to each other? For example, Wolf has shown that the d functions of SO(n) are all expressible as a sum over $\sin^p\theta \times\exp(iq\theta)$. But in the case of SO(4) for example, we also have the result of Freedman and Wang,[11] where only $\exp(iq\theta)$ appears. A similar situation is found in the case of the noncompact group SO(3.1), where the representation functions can either be expressed as a sum over a hypergeometric function,[12–16] or as a Fourier series, as Smorodinskii and Shepelev[17] and recently Wong and Yeh[18] have shown. A third problem concerns the practical calculation of the d function using the above methods. Since both methods treat θ as a variable, the calculation becomes very complicated as n increases.

It is at this point that we look at the corresponding d functions of U(n) and see if we can obtain some help from there. We find that there is indeed an approach common to both U(n) and SO(n): the Weyl coefficient approach. In this approach it is not necessary to calculate the d function for the whole range of θ. For SO(n) it is sufficient to calculate the value at $\theta=\pi/2$ (or $\theta=-\pi/2$). The rest can be obtained by recurrence with SO($n-1$). Since the lowest order groups are well known: for SO(2) the d function is equal to $\exp(im\theta)$, for SO(3), it is the well-known Wigner $d^J_{mm'}(\theta)$ function, we can say that the Weyl coefficient approach reduces the calculation of the d function of SO(n) to a single value of θ, i.e., when $\theta=\pi/2$.

Thus the first problem is answered. There is indeed a structure in the d functions of SO(n). It can be related to any of its subgroups, multiplied by the corresponding Weyl coefficients. Thus ultimately we can say that the d functions of all SO(n) can be written as Fourier series, whose coefficients can be explicitly calculated. The second problem is also answered; since the d function of SO(4) can be written either in terms of SO(2) or SO(3), therefore we can express it either as a sum over $\exp(iq\theta)$ (SO(2)), or $\sin^p\theta\cos^q\theta$ (SO(3)).

The third problem is answered in the following way: Whereas the d function of SO(n), n large, still remains

 0022-2488/78/1903-0703$1.00

complicated, we have subdivided its complexity, and have posed the problem in a simpler way: I.e., What is the d function of SO(n) at $\theta=\pi/2$? If we can answer that question, then we can answer the whole question. However, we have to admit that the final expression for the representation function of SO(n) is still quite lengthy and complicated, because all the Weyl coefficients have to be multiplied together.

In Sec. 2 we present the formalism for the Weyl coefficients of SO(n). In Secs. 3, 4, and 5, we discuss the Weyl coefficients of SO(3), SO(4), and SO(5), respectively. In Sec. 6 we discuss the Weyl coefficients of SO(n), for $n>5$. In Sec. 7 we draw some conclusions from the above discussions.

2. WEYL COEFFICIENTS OF SO(n)

The special orthogonal group SO(n) is equivalent to the rotation group in an n-dimensional Euclidean space. If one uses the familiar J_{ij} as generators, one can regard its finite transformation $\exp(iJ_{ij}\theta)$ as a rotation in the $i-j$ plane, represented by the matrix

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

in the $i-j$ plane.

A group element in SO(n) is made up of $\frac{1}{2}n(n-1)$ such rotations, and in order to obtain the representation function of SO(n) it is only necessary to know the matrix element of the rotation $D_{n-1,n}(\theta)$. The rest can be obtained by recurrence, or induction on n. $D_{n-1,n}(\theta)$ is the matrix d with $d_{ii}=1$ for $i=1,2,\ldots,n-2$. $d_{n-1,n-1}=d_{n,n}=\cos\theta$, $d_{n-1,n}=\sin\theta$, $d_{n,n-1}=-\sin\theta$, and all other matrix elements equal to zero. We shall introduce the general rotation matrix $D_{ij}(\theta)$, whose matrix elements are $d_{kk}=1$ for $k\neq i\neq j$, $d_{ii}=d_{jj}=\cos\theta$, $d_{ij}=\sin\theta$, $d_{ji}=-\sin\theta$, all other matrix elements equal to zero. Furthermore we define the Weyl coefficient $W_{ij}=D_{ij}(\pi/2)$. It is then easy to see that $W_{ij}^{-1}=D_{ij}(-\pi/2)$. We then have the following relation:

$$\begin{aligned} D_{n-1,n}(\theta) &= W^{-1}_{n-2,n}D_{n-2,n-1}(-\theta)W_{n-2,n} \\ &= W^{-1}_{n-2,n-1}W_{n-1,n}W_{n-2,n-1}D_{n-2,n-1}(-\theta) \\ &\quad\times W^{-1}_{n-2,n-1}W^{-1}_{n-1,n}W_{n-2,n-1} \\ &= W^{-1}_{n-2,n-1}W^{-1}_{n-1,n}D_{n-2,n-1}(\theta)W_{n-1,n}W_{n-2,n-1}. \end{aligned} \tag{2.1}$$

Therefore, if we know the representation function $D_{n-2,n-1}(\theta)$, then we can obtain the representation function $D_{n-1,n}(\theta)$ provided we can calculate

$$W_{n-1,n}=D_{n-1,n}(\pi/2). \tag{2.2}$$

This means that the representation function of SO(n) for arbitrary θ can be found if we know its value at a particular value of θ, i.e., when $\theta=\pi/2$. Historically this has not been the approach used to find the representation functions of SO(n). In the methods used by Maekawa, and Vilenkin and Wolf, θ is treated as a variable. The result is that the actual expression for the d function of SO(n) for large n becomes very complicated, and very few structural properties can be discovered from the expression.

The advantages of using the Weyl coefficient approach are the following:

(1) It reduces the calculation of the d function of SO(n) to a particular value of θ, i.e., $\theta=\pi/2$. Though the final expression for the d functions of SO(n) is still quite lengthy, at least the θ-dependent part is quite clear, i.e., it is entirely restricted to the subgroup of SO(n), while the other factors are recognized as the Weyl coefficients in accordance with (2.1).

(2) It shows its connection with all the different subgroups of SO(n), thereby clarifying the structural properties of the d function of SO(n). As a result, one can make a very sweeping statement such as the following: All representation functions of the orthogonal group (and in fact the unitary group as well) are ultimately expressible as Fourier series, whose coefficients can be explicitly calculated.

(3) One can also show that the Weyl coefficients are all expressible as sums and products of 3-j symbols of SO(3) and normalization factorials.

3. WEYL COEFFICIENTS OF SO(3)

It is interesting to note that the Weyl coefficient of SO(3) was treated by Wigner[19] and can be found in Edmonds.[20] However, it has not received much attention. Using Eq. (2.1), we find

$$D_{23}(\theta)=W_{12}^{-1}W_{23}W_{12}D_{12}(-\theta)W_{12}^{-1}W_{23}^{-1}W_{12}, \tag{3.1}$$

where

$$D_{12}(\theta)=\exp(im\theta) \tag{3.2}$$

$$W_{23}=D_{23}(\pi/2)=d^{J}_{m',m}(\pi/2), \tag{3.3}$$

where $d^{J}_{m',m}(\theta)$ is the Wigner function of SO(3).

Substituting the values in (3.1) we obtain

$$d^{J}_{m',m}(\theta)=\sum_{m''}(i)^{m'-m}d^{J}_{m'',m}(\pi/2)d^{J}_{m'',m'}(\pi/2)\exp(-im''\theta). \tag{3.4}$$

This agrees with Edmonds[20] (p. 62).

Let us now make a brief survey of the different ways $d^{J}_{m',m}(\theta)$ can be written. Basically, there are two different ways it can be written. The first way is to express it directly in terms of $\sin\theta/2$ and $\cos\theta/2$, e.g., according to Edmonds,[20]

$$d^{J}_{m',m}(\theta)=\left[\frac{(j+m')!\,(j-m')!}{(j+m)!\,(j-m)!}\right]^{1/2}\sum_{\sigma}\binom{j+m}{j-m'-\sigma}\binom{j-m}{\sigma}$$
$$\times(-1)^{j-m'-\sigma}(\cos\theta/2)^{2\sigma+m'+m}(\sin\theta/2)^{2j-2\sigma-m'-m}. \tag{3.5}$$

The second way is to express it as Jacobi polynominals, or equivalently as hypergeometric functions. Thus, e.g.,

$$\begin{aligned} d^{J}_{m',m}(\theta) &= \left[\frac{(J+m')!\,(J-m')!}{(J+m)!\,(J-m)!}\right]^{1/2}(\cos\theta/2)^{m'+m} \\ &\quad\times\sin\theta/2)^{m'-m}\,P^{(m'-m,\,m'+m)}_{J-m'}(\cos\theta) \\ &= \left[\frac{(J-m)!\,(J+m')!}{(J+m)!\,(J-m')!}\right]^{1/2}(\cos\theta/2)^{2J+m-m'} \end{aligned}$$

$$\times \frac{(\sin\theta/2)^{m'-m}}{(m'-m)!}\ {}_2F_1(m'-J,\,-m-J;$$

$$\times\, m'-m+1;\,-\tan^2\theta/2). \tag{3.6}$$

We now show a third way that $d^J_{m'm}(\theta)$ can be written, i.e., in terms of the Clebsch–Gordan coefficients of SO(3). As far as we know, this expression has not been found in the literature.

We start from the normalized double boson polynomials in $U(2) * U(2)$:

$$\mathcal{N}^{-1/2} B\begin{pmatrix} \mu_{11} \\ m_{12} m_{22} \\ m_{11}\end{pmatrix} = \sum_s \frac{[(m_{11}-m_{22})!(m_{12}-m_{11})!(\mu_{11}-m_{22})!(m_{12}-\mu_{11})!]^{1/2}}{s!(\mu_{11}-m_{22}-s)!(m_{11}-m_{22}-s)!(m_{12}-m_{11}-\mu_{11}+m_{22}+s)!}\left[\frac{(m_{12}+1)!\,m_{22}!}{(m_{12}-m_{22}+1)!}\right]^{-1/2}$$

$$\times (a_1^1a_2^2-a_1^2a_2^1)^{m_{22}}\, a_1^{1\,s}\, a_1^{2\,\mu_{11}-m_{22}-s}\, a_2^{1\,m_{11}-m_{22}-s}\, a_2^{2\,m_{12}-m_{11}-\mu_{11}+m_{22}+s}$$

$$=\sum_\alpha C\begin{pmatrix} \frac{1}{2}(m_{11}+m_{22}-\mu_{11}) & \frac{1}{2}\mu_{11} & \frac{1}{2}(m_{12}-m_{22}) \\ m_{11}-\alpha-\frac{1}{2}(m_{12}+m_{22}-\mu_{11}) & \alpha-\frac{1}{2}\mu_{11} & m_{11}-\frac{1}{2}(m_{12}+m_{22})\end{pmatrix}$$

$$\times \frac{a_1^{1\,\alpha}\, a_1^{2\,m_{11}-\alpha}\, a_2^{1\,\mu_{11}-\alpha}\, a_2^{2\,m_{12}+m_{22}-m_{11}-\mu_{11}+\alpha}}{[\alpha!(m_{11}-\alpha)!(\mu_{11}-\alpha)!(m_{12}+m_{22}-m_{11}-\mu_{11}+\alpha)!]^{1/2}} \tag{3.7}$$

For the derivation of (3.7), see Wong.[4] We now make the following substitution:

$m_{11}=2J,\quad \mu_{11}=m'+J,\quad m_{11}=m+J,\quad m_{22}=0,\quad a_1^1=a_2^2=\cos\theta/2,\quad a_1^2=\sin\theta/2,\quad a_2^1=-\sin\theta/2.$

Then (3.7) becomes

$$d^J_{m'm}(\theta)=\sum_\alpha C\begin{pmatrix} \frac{1}{2}(m-m') & \frac{1}{2}(J+m') & J \\ m+\frac{1}{2}(m'+J)-\alpha & \alpha-\frac{1}{2}(m'+J) & m\end{pmatrix}(-1)^{m'+J-\alpha}\frac{(\cos\theta/2)^{2\alpha-m-m'}(\sin\theta/2)^{m+m'+2J-2\alpha}}{[\alpha!(m+J-\alpha)!(m'+J-\alpha)!(\alpha-m-m')!]^{1/2}}. \tag{3.8}$$

The Weyl coefficient of SO(3), W_{23}, can be easily obtained from these expressions by putting $\theta=\pi/2$ in $d^J_{mm'}(\theta)$.

4. WEYL COEFFICIENTS OF SO(4)

From the Freedman–Wang result, we have

$$W_{34}=d^{m_{41}m_{42}}_{jj'm}(\pi/2)=\sum_\mu C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j \\ \mu & m-\mu & m\end{pmatrix} C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j' \\ \mu & m-\mu & m\end{pmatrix}(-i)^{2\mu-m}. \tag{4.1}$$

Thus using (2.1) again, we can write the "boost" matrix as

$$D_{34}(\theta)=W_{23}^{-1}W_{34}W_{23}D_{23}(-\theta)W_{23}^{-1}W_{34}^{-1}W_{23}, \tag{4.2}$$

where $D_{23}(-\theta)$, as we have seen in the previous sections, can be expressed as a hypergeometric function. Continuing to the Lorentz group SO(3,1), one can now understand why the earlier work on the boost matrix of the Lorentz group by Ström,[12] Duc and Hieu,[13] Verdiev and Dadashev,[14] Sciarrino and Toller,[15] Makarov and Shepelev,[16] etc., to quote but a few, all showed that it is a sum over a hypergeometric function. This is because they have expanded the boost according to (4.2). Then it was shown by Smorodinskii and Shepelev[17] and recently by Wong and Yeh[18] that it can also be written as a Fourier series, i.e., in terms of the Freedman and Wang expression. From the Weyl coefficient point of view, it is clear that we can also write $D_{34}(\theta)$ as

$$D_{34}(\theta)=(W_{34}^{-1}W_{23}W_{34}W_{12}^{-1}W_{23}W_{12})D_{12}(\theta)(W_{12}^{-1}W_{23}^{-1}W_{12}W_{34}^{-1}W_{23}^{-1}W_{34}). \tag{4.3}$$

Then (4.3) will give the Freedman and Wang result.

Let us now try to express the Freedman–Wang result in the form of (4.3). We have

$$d^{m_{41}m_{42}}_{jj'm}(\theta)=\sum_\mu C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(m-\mu) & m\end{pmatrix}\exp(i\theta\mu)C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(m-\mu) & m\end{pmatrix}$$

$$\times\sum_{j''} C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j'' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(\mu-m) & \mu\end{pmatrix} C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j'' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(\mu-m) & \mu\end{pmatrix}$$

$$=\sum_{\mu,j''} C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(m-\mu) & m\end{pmatrix} C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j'' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(\mu-m) & \mu\end{pmatrix}$$

$$\times\exp(i\theta\mu)C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j'' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(\mu-m) & \mu\end{pmatrix}C\begin{pmatrix}\frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(m-\mu) & m\end{pmatrix}$$

$$= \sum_{j''\mu} \left\langle \begin{matrix} m_{41}\ m_{42} \\ j \\ m \end{matrix} \middle| W^{-1} \middle| \begin{matrix} m_{41}\ m_{42} \\ j'' \\ \mu \end{matrix} \right\rangle \exp(i\mu\theta) \left\langle \begin{matrix} m_{41}\ m_{42} \\ j'' \\ \mu \end{matrix} \middle| W \middle| \begin{matrix} m_{41}\ m_{42} \\ j' \\ m \end{matrix} \right\rangle, \tag{4.4}$$

where

$$\left\langle \begin{matrix} m_{41}\ m_{42} \\ j'' \\ \mu \end{matrix} \middle| W \middle| \begin{matrix} m_{41}\ m_{42} \\ j' \\ m \end{matrix} \right\rangle = C\begin{pmatrix} \frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j'' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(\mu-m) & \mu \end{pmatrix} C\begin{pmatrix} \frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(m-\mu) & m \end{pmatrix}. \tag{4.5}$$

Comparing (4.4) and (4.5) with (4.3), we conclude

$$\left\langle \begin{matrix} m_{41}\ m_{42} \\ j'' \\ \mu \end{matrix} \middle| W_{12}^{-1} W_{23}^{-1} W_{12} W_{34}^{-1} W_{23}^{-1} W_{34} \middle| \begin{matrix} m_{41}\ m_{42} \\ j' \\ m \end{matrix} \right\rangle = C\begin{pmatrix} \frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j'' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(\mu-m) & \mu \end{pmatrix} C\begin{pmatrix} \frac{1}{2}(m_{41}+m_{42}) & \frac{1}{2}(m_{41}-m_{42}) & j' \\ \frac{1}{2}(m+\mu) & \frac{1}{2}(m-\mu) & m \end{pmatrix}. \tag{4.6}$$

(4.6) is a remarkably simple result, though it involves the product of six Weyl coefficients.

5. WEYL COEFFICIENTS OF SO(5)

The d function of SO(5) has been obtained by Holman.[21] We wish to make three corrections on Holman's expression for the normalization factor $\mathcal{J}$. These can be checked either from direct calculation or from consistency arguments, e.g., by calculating the matrix elements of $\exp(iL_{23}\theta)$ and showing that it must reduce to the SO(3) d function. The final result is

$$\left\langle \begin{matrix} J_m\ \Lambda_m \\ J'\ \Lambda' \\ L' \\ M' \end{matrix} \middle| D_{45}(\theta) \middle| \begin{matrix} J_m\ \Lambda_m \\ J\ \Lambda \\ L \\ M \end{matrix} \right\rangle = \delta_{LL'}\delta_{MM'}(-1)^{\Lambda-\Lambda'} \sum_{\substack{j_1 j_2 j_1' j_2' \\ K_1 K_2}} (2K_1+1)(2K_2+1)\mathcal{J}(J_m\Lambda_m;J'\Lambda';j_1'j_2')[(2J'+1)(2\Lambda'+1)]^{1/2}$$

$$\times \begin{Bmatrix} j_1' & \lambda_1' & K_1 \\ j_2' & \lambda_2' & K_2 \\ J' & \Lambda' & L' \end{Bmatrix} d^{K_1}_{j_1'-\lambda_1',\, j_1-\lambda_1}(\theta)\, d^{K_2}_{j_2'-\lambda_2',\, j_2-\lambda_2}(\theta)\, \mathcal{J}(J_m\Lambda_m;J\Lambda;j_1j_2)[(2J+1)(2\Lambda+1)]^{1/2} \begin{Bmatrix} j_1 & \lambda_1 & K_1 \\ j_2 & \lambda_2 & K_2 \\ J & \Lambda & L \end{Bmatrix}, \tag{5.1}$$

where

$$j_1+\lambda_1 = j_1'+\lambda_1' = J_m, \quad j_2+\lambda_2 = j_2'+\lambda_2' = \Lambda_m, \tag{5.2}$$

$$\mathcal{J}(J_m\Lambda_m J\Lambda j_1 j_2) = (-1)^{2j_2}\Bigg[\frac{(J+j_1-j_2)!\,(\Lambda+J_m-\Lambda_m-j_1+j_2)!}{(J+j_2-j_1)!\,(j_1+j_2-J)!\,(J+j_1+j_2+1)!}$$

$$\times \frac{(J_m+\Lambda_m-J-\Lambda)!\,(J_m+\Lambda_m-J+\Lambda+1)!\,(J_m+\Lambda_m+J-\Lambda+1)!\,(J_m+\Lambda_m+J+\Lambda+2)!}{(J_m-\Lambda_m+J+\Lambda+1)!\,(J_m-\Lambda_m+J-\Lambda)!\,(J_m-\Lambda_m-J+\Lambda)!\,(2J_m+2\Lambda_m+2)!}$$

$$\times \frac{(2J_m-2\Lambda_m+1)!\,(\Lambda_m-J_m+\Lambda+J)!}{(\Lambda-J_m+j_1+\Lambda_m-j_2)!\,(J_m-j_1+\Lambda_m-j_2-\Lambda)!\,(\Lambda+J_m+\Lambda_m-j_1-j_2+1)!}\Bigg]^{1/2}. \tag{5.3}$$

The Weyl coefficient W_{45} can be obtained from (5.1), (5.2), and (5.3) by putting $\theta=\pi/2$.

Again we wish to point out that (5.1) is not the only way $D_{45}(\theta)$ can be written. It can be written in terms of the subgroup SO(4), in which case $D_{45}(\theta)$ can be expressed as a Fourier series.

From the Weyl coefficient point of view, what is interesting about Holman's expression (5.1) is that the Weyl coefficient of SO(5) can be expressed as 9-j symbols multiplied by normalization factorials. This could be the key for higher SO(n) D functions. In the next section we shall show that all Weyl coefficients of SO(n) are at least expressible as sums and products of 3-j symbols of SO(3) and normalization factorials.

6. WEYL COEFFICIENTS OF SO(n), $n>5$

From the results of the previous three sections, we see that the Weyl coefficients of SO(3), SO(4), and SO(5) can all be expressed as sums and products of 3-j symbols of SO(3) and normalization factorials. We shall show in this section that this is true for all n.

We shall prove this by using the integration method of Vilenkin and Wolf. This method has the advantage that both the pseudo-orthogonal group SO(n,1) and the compact group SO($n+1$) can be treated in the same way.

It has been shown by Vilenkin and Wolf that the d matrix for SO(n,1) or SO($n+1$) can be written as [Wolf, Eqs. (5.11), (5.12)]

$$d^{L}_{JL',J'}(\zeta)=\frac{(\dim_{n}J\dim_{n}J')^{1/2}\Gamma(n/2)}{\dim_{n-1}L\dim_{n-1}L'\pi^{1/2}\Gamma(\frac{1}{2}(n-1))}\sum_{M}\dim_{n-2}M$$

$$\times\int_{0}^{\pi}\sin^{n-2}\theta d\theta\, \overline{d^{J}_{LML'}(\theta)}\left(\frac{\sin\theta}{\sin\theta'}\right)^{\lambda} d^{J'}_{LML'}(\theta'),\quad(6.1)$$

where for the noncompact group SO(n,1),

$$\frac{\sin\theta}{\sin\theta'}=\cosh\zeta-\cos\theta\sinh\zeta,\quad(6.2)$$

$$\cos\theta'=\frac{\cos\theta\cosh\zeta-\sinh\zeta}{\cosh\zeta-\cos\theta\sinh\zeta}.\quad(6.3)$$

For the compact group SO($n+1$): $\zeta=-i\delta$

$$\frac{\sin\theta}{\sin\theta'},=\cos\delta+i\cos\theta\sin\delta,\quad(6.4)$$

$$\cos\theta'=\frac{\cos\theta\cos\delta+i\sin\delta}{\cos\delta+i\cos\theta\sin\delta}.\quad(6.5)$$

Also to continue from SO(n,1) to SO($n+1$), one should multiply by the phase factor W_1^{-1} of Maekawa.[22]

Now for the group SO($n+1$), $n\geq 5$, we can always express the $d^{J}_{LML'}(\theta)$ in (6.1) in terms of the d function of SO(5), which, as Holman[21] has shown, can be expressed in terms of the d function of SO(3). We are thus required to evaluate an integral of the form:

$$\int_{0}^{\pi}\sin^{n-2}\theta d\theta\, \overline{d^{J}_{mn}(\theta)}\left(\frac{\sin\theta}{\sin\theta'}\right)^{\lambda} d^{J'}_{m'n'}(\theta').\quad(6.6)$$

Our purpose is to show that (6.6) can be expressed as sums and products of 3-j symbols of SO(3) and normalization factorials. To show this we first use Vilenkin's[23] formula for the d function of SO(3):

$$d^{J}_{mn}(z)=i^{-m-n}\left[\frac{(J-m)!\,(J-n)!}{(J+m)!\,(J+n)!}\right]^{1/2}\left(\frac{1+z}{1-z}\right)^{1/2(m+n)}$$

$$\times\sum_{j}\frac{(i)^{2j}(J+j)!}{(J-j)!\,(j-m)!\,(j-n)!}\left(\frac{1-z}{2}\right)^{j},\quad(6.7)$$

where $z=\cos\theta$. Then we use the following transformation:

$$1+z'=\exp(-\zeta)(1+z)(\cosh\zeta-\sinh\zeta z)^{-1},$$

$$1-z'=\exp(\zeta)(1-z)(\cosh\zeta-\sinh\zeta z)^{-1},$$

$$(\cosh\zeta-\sinh\zeta z)^{-1}=\sinh^{-1}\zeta\left(\frac{1+\exp(-2\zeta)}{1-\exp(-2\zeta)}-x\right)^{-1}.\quad(6.8)$$

Finally putting $z=-y+1$, we can evaluate the integral (6.6) by (3.259.2) of GR[24]:

$$\int_{0}^{u}y^{\nu-1}(u-y)^{\mu-1}(y^{m}+\beta^{m})dy$$
$$=\beta^{m\lambda}u^{\mu+\nu-1}B(\mu\nu)_{m+1}F_{m}(-\lambda,\nu/m,\ldots(\nu+m-1)/m;$$
$$(\mu+\nu)/m\cdots(\mu+\nu+m-1)/m;(-u/\beta)^{m}).\quad(6.9)$$

Since $m=1$ in (6.9), we obtain a ${}_2F_1$ function, with arguments $1-\exp(-2\zeta)$. The next step is to express the hypergeometric function by the Barnes integral representation and the Barnes lemma.[25]

$${}_2F_1(abc;z)=\frac{\Gamma(c)}{\Gamma(c-a)\Gamma(c-b)\Gamma(a)\Gamma(b)}\frac{1}{2\pi i}\int_{-k-i\infty}^{-k+i\infty}$$

$$dt\,\Gamma(a+t)\Gamma(b+t)\Gamma(c-a-b-t)\Gamma(-t)(1-z)^{t}.\quad(6.10)$$

It then turns out that the summation over j and j' resulting from the two d functions of SO(3) can be performed giving two ${}_3F_2$ functions with unit argument, which in turn, can be expressed as 3-j symbols of SO(3). Finally the contour integral can be performed by closing the contour by an infinite semicircle in the right half-plane. It can be checked that by Stirling's formula for the asymptotic behavior of the gamma function the integral on the semicircle vanishes. The integral can then be expressed as a Fourier series, whose coefficients are sums over the residues of two series of poles. Finally by putting $\delta=\pi/2$, we obtain the desired Weyl coefficient. The details of the above calculation for SO(3,1) and therefore for SO(4) have been given elsewhere by Wong and Yeh.[18]

Another way of showing the above result is by projection. Since for SO(n), $n>3$, the d function can always be expressed in terms of the d function of SO(3), we can calculate the Weyl coefficients by projection, i.e., the Weyl coefficients can be obtained from (6.6) by evaluating the integral

$$\int_{-1}^{1}D_{n,n+1}(x)d^{J}_{mm'}(x)dx,\quad(6.11)$$

where $x=\cos\theta$. But all the terms in (6.6) can eventually be expressed as functions of $(1+x)$ and $(1-x)$. Now using Eq. (3), p. 284, Vol. 2 of Erdelyi *et al.*,[26] we can evaluate the integral in (6.11) through

$$\int_{-1}^{1}(1-x)^{\rho}(1+x)^{\sigma}P_{n}^{\alpha\beta}(x)dx$$

$$=2^{\rho+\sigma+1}\frac{\Gamma(\rho+1)\Gamma(\sigma+1)}{\Gamma(\sigma+\rho+2)}$$

$$\times{}_3F_2(-n,\alpha+\beta+n+1,\rho+1,\alpha+1,\rho+\sigma+2,1).\quad(6.12)$$

Equation (6.12) is in the desired form, i.e., in terms of 3-j symbols of SO(3) and normalization factorials.

7. CONCLUSION

We have treated the representation functions of SO(n) from the Weyl coefficient point of view. The advantages are at least twofold. First, one finds that the representation function of SO(n) for all values of θ is known once it is known for a particular value of θ, i.e., $\theta=\pi/2$. Second, one obtains a connection between the d functions of SO(n) and all its subgroups. When applied to noncompact groups such as SO(3,1), this explains why there have been so many different forms for the boost matrix of the Lorentz group. But basically there are two different forms: one in terms of the hypergeometric function (i.e., with SO(3) as the subgroup), and the other according to the Freedman and Wang form, i.e., in terms of the subgroups SO(2).

We have obtained the Weyl coefficients of SO(3), SO(4), and SO(5) (up to a substitution). It is tempting to ask whether, for SO(5), it is possible to obtain a simple form, such as Eq. (4.6), when SO(5) is decomposed in terms of SO(3). So far we have not succeeded in doing so. The problem is worth further study.

We have shown that all Weyl coefficients are expressible as sums and products of 3-*j* symbols of SO(3) and normalization factorials. In the case of SO(5), we know from Holman's work that the Weyl coefficients are expressible in terms of 9-*j* symbols of SO(3). Whether this is true for higher order groups remains to be seen.

Finally, we conclude that all representation functions of SO(n) [and U(n) as well] are ultimately expressible as Fourier series, whose coefficients can be explicitly calculated if one so wishes.

[1]I.M. Gel'fand and M.I. Graev, Izv. Akad. Nauk SSSR Mekh. Mashinostr. **29**, 1329 (1965) [Am. Math. Soc. Transl. Ser. 2 **65**, 116 (1967)].
[2]E. Chacón and M. Moshinsky, Phys. Lett. **23**, 567 (1966).
[3]W.J. Holman III, Nuovo Cimento A **4**, 904 (1971).
[4]M.K.F. Wong, J. Math. Phys. **17**, 1558 (1976).
[5]J.D. Louck and L.C. Biedenharn, J. Math. Phys. **14**, 1336 (1973).
[6]T. Maekawa, J. Math. Phys. **16**, 334 (1975).
[7]N. Ja Vilenkin, Dokl. Akad. Nauk SSSR **113**, 16 (1957).
[8]K.B. Wolf, J. Math. Phys. **12**, 197 (1971).
[9]S.C. Pang and K.T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[10]M.K.F. Wong, J. Math. Phys. **8**, 1899 (1967).
[11]D.Z. Freedman and J.M. Wang, Phys. Rev. **160**, 1560 (1967).
[12](a) S. Ström, Ark. Fys. **29**, 467 (1965); (b) S. Ström, *ibid.* **33**, 465 (1966); (c) S. Ström, *ibid* **38**, 373 (1968).
[13]D.W. Duc and N.V. Hieu, Dokl. Akad. Nauk SSSR **172**, 1281 (1967) [Sov. Phys.-Dokl. **12**, 312 (1967)].
[14]I.A. Verdiev and L.A. Dadashev, Yad. Fiz. **6**, 1094 (1967) [Sov. J. Nucl. Phys. **6**, 795 (1968)].
[15]A. Sciarrino and M. Toller, J. Math. Phys. **9**, 1252 (1967).
[16]A.A. Makarov and G.I. Shepelev, Yad. Fiz. **12**, 1092 (1970) [Sov. J. Nucl. Phys. **12**, 596 (1971)].
[17]Ya.A. Smorodinskii and G.I. Shepelev, Yad. Fiz. **13**, 441 (1971) [Sov. J. Nucl. Phys. **13**, 248 (1971)].
[18]M.K.F. Wong and H.Y. Yeh, J. Math. Phys. **18**, 1768 (1977).
[19]E.P. Wigner, "On Matrices which Reduce Kronecker Products of Representations of S.R. group," 1951 (unpublished), quoted by Edmonds (Ref. 20).
[20]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University, Princeton, N.J., 1960).
[21]W.J. Holman, III, J. Math. Phys. **10**, 1710 (1969).
[22]T. Maekawa, J. Math. Phys. **16**, 2372 (1975).
[23]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).
[24]I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[25]E.T. Whittaker and J.N. Watson, *A Course of Modern Analysis* (Cambridge University, Cambridge, England, 1935).
[26]A. Erdelyi, W. Magnus, F. Oberhetting, and F.G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954).

Elementary construction of graded Lie groups

V. Rittenberg

Physikalisches Institut der Universität Bonn, Bonn, F. R. Germany

M. Scheunert[a)]

Dublin Institute for Advanced Studies, Dublin, Ireland

(Received 30 June 1977)

We show how the definitions of the classical Lie groups have to be modified in the case where Grassmann variables are present. In particular we construct the general linear, the special linear and the orthosymplectic graded Lie groups. Special attention is paid to the question of how to formulate an adequate "unitarity condition."

1. INTRODUCTION

During the last two years considerable progress has been made in the theory of graded Lie algebras. For example all complex simple graded Lie algebras are now known.[1-3] Furthermore, a general theory of graded manifolds and of graded Lie groups is going to be developed.[4,5] Nevertheless we think that is is worthwhile to contruct some sequences of graded Lie groups (and the Lie algebras associated with them) in an elementary matrix notation. This is the more true since these matrix groups and algebras have made their appearance in various branches of theoretical physics (supersymmetric field theory,[6] supergravity,[7] theory of classical spinning particles[8]).

To begin with we describe in Sec. 2 some topics of the corresponding matrix algebra. All these results are more or less well-known, however, it might be useful to have them collected at one place. Section 2 also contains the definition of the general linear graded Lie groups as well as a discussion of some of their elementary properties.

The matrix algebra being established, we construct in Sec. 3 the special linear and the orthosymplectic graded Lie groups.

To obtain "compact forms" of our groups we have to introduce the appropriate adjoint operations in our matrix algebra (Sec. 4). It turns out that there exist (at least) two essentially different possibilities. Using these operations we can construct "compact forms" by a unitarity condition.

The last section contains a short discussion of our results as well as some final remarks.

2. THE MATRIX ALGEBRAS $M(n,m)$ AND THE GENERAL LINEAR GRADED LIE GROUPS PL(n,m)

Let W be any (complex) vector space and let $A=\wedge W$ be the exterior algebra constructed over W,

$$A=\wedge W=\underset{r\geq 0}{\oplus}\overset{r}{\wedge} W. \tag{2.1}$$

It is well-known that $\wedge W$ is an associative Z-graded algebra. We define

$$A_0=\underset{r\geq 0}{\oplus}\overset{2r}{\wedge} W,\quad A_1=\underset{r\geq 0}{\oplus}\overset{2r+1}{\wedge} W. \tag{2.2}$$

The elements of A_0 (resp. of A_1) are called the even (resp. odd) elements of A. Recall that A is graded commutative in the sense that

$$ab=(-1)^{\alpha\beta}ba, \tag{2.3}$$

if $a\in A_\alpha$, $b\in A_\beta$.

Most of our results remain valid for more general associative graded commutative algebras A.

A. The algebra $M(n,m)$

Now let $n,m\geq 1$ be some natural numbers which are assumed to be fixed in the following. Let $M(n,m)$ be the set of all block matrices of the form $X=\begin{pmatrix} a & \xi \\ \eta & b\end{pmatrix}$ with a an $n\times n$ matrix and b an $m\times m$ matrix whose elements are taken from A_0, furthermore, ξ an $n\times m$ matrix and η an $m\times n$ matrix whose elements are taken from A_1. More precisely, for any natural number r let $M_{2r}(n,m)$ be the subset of $M(n,m)$ consisting of the diagonal block matrices $\begin{pmatrix} a & 0 \\ 0 & b\end{pmatrix}$ with elements taken from $\wedge^{2r} W$ and let $M_{2r+1}(n,m)$ be the subset of $M(n,m)$ consisting of the off-diagonal block matrices $\begin{pmatrix} 0 & \xi \\ \eta & 0\end{pmatrix}$ with elements taken from $\wedge^{2r+1} W$. Equipped with the usual addition and multiplication $M(n,m)$ is an associative algebra, and the subspaces $M_r(n,m)$, $r\geq 0$, define a grading of $M(n,m)$ in the sense that

$$M(n,m)=\underset{r\geq 0}{\oplus}M_r(n,m),$$

$$M_r(n,m)M_s(n,m)\subset M_{r+s}(n,m), \tag{2.4}$$

for all $r,s\geq 0$. Every element $X\in M(n,m)$ has a unique decomposition

$$X=\sum_{r\geq 0}X_r, \tag{2.5}$$

with $X_r\in M_r(n,m)$ (only finitely many X_r being different zero). The element X_r is called the (homogeneous) component of X of degree r. Note that X_0 is a complex block-diagonal matrix.

Furthermore, let us introduce a subspace $M_+(n,m)$ of $M(n,m)$ as follows:

$$M_+(n,m)=\underset{r\geq 1}{\oplus}M_r(n,m). \tag{2.6}$$

a)Work supported by the Deutsche Forschungsgemeinschaft.

Obviously $M_+(n,m)$ is an ideal of $M(n,m)$. Every element X of $M_+(n,m)$ is nilpotent, i.e., there exists a natural number q such that $X^q = 0$.

Next we want to transcribe the notions of transpose, trace, and determinant of a matrix to the graded case. It turns out that the ordinary definitions are not appropriate for our purposes. Let us begin with the transposition. For any matrix B we denote by ${}^tB$ the transposed matrix of B. Of course, for all $X \in M(n,m)$ the transposed matrix ${}^tX$ is also a well-defined element of $M(n,m)$. However, if $X, Y \in M(n,m)$, then in general

$$ {}^t(XY) \neq {}^tY\,{}^tX. \tag{2.7}$$

The theory of graded vector spaces and algebras suggests the "correct" definition. If $X = \begin{pmatrix} a & \xi \\ \eta & b \end{pmatrix}$ is any element of $M(n,m)$ we define the *graded transpose* ${}^TX$ of X by

$$ {}^TX = \begin{pmatrix} {}^ta & -{}^t\eta \\ {}^t\xi & {}^tb \end{pmatrix}. \tag{2.8}$$

It is then easy to check that

$$ {}^T(XY) = {}^TY\,{}^TX \tag{2.9}$$

for all $X, Y \in M(n,m)$. Note, however, that in general ${}^{TT}X \neq X$. The situation for the trace is similar. The (usual) trace $\mathrm{Tr}(X)$ of an element $X \in M(n,m)$ is well-defind, but in general, for $X, Y \in M(n,m)$

$$ \mathrm{Tr}(XY) \neq \mathrm{Tr}(YX). \tag{2.10}$$

Again the theory of graded vector spaces and algebras helps to cure this disease. If $X = \begin{pmatrix} a & \xi \\ \eta & b \end{pmatrix}$ is any element of $M(n,m)$ we define the *graded trace* $\mathrm{Trg}(X)$ of X by

$$ \mathrm{Trg}(X) = \mathrm{Tr}(a) - \mathrm{Tr}(b). \tag{2.11}$$

It follows that

$$ \mathrm{Trg}(XY) = \mathrm{Trg}(YX) \tag{2.12}$$

for all $X, Y \in M(n,m)$. The definition of the graded determinant will be given below.

B. The general linear graded Lie group PL(n,m)

Once the matrix algebra $M(n,m)$ has been introduced it is evident how to define the *general linear graded Lie group* $\mathrm{PL}(n,m)$: This is the multiplicative subgroup of $M(n,m)$ consisting of all those elements which have an inverse.

To discuss the structure of $\mathrm{PL}(n,m)$ we remark that an element $X \in M(n,m)$ lies in $\mathrm{PL}(n,m)$ if and only if its component X_0 of degree zero has an inverse (recall that X_0 is a complex block-diagonal matrix). If this is the case, we have

$$ X_0^{-1}X \in 1 + M_+(n,m). \tag{2.13}$$

Now all elements of $M_+(n,m)$ are nilpotent. Hence

$$ \exp: M_+(n,m) \to 1 + M_+(n,m) \tag{2.14}$$

is a bijective mapping and

$$ \log: 1 + M_+(n,m) \to M_+(n,m) \tag{2.15}$$

is its inverse. Both exp and log are defined by their power-series expansions; problems of convergence do not arise since in the present situation only finitely many terms of these series are nonzero. [Of course, $\exp X$ is defined for all $X \in M(n,m)$ and is an element of $\mathrm{PL}(n,m)$.]

Let $\mathrm{PL}_0(n,m)$ be the group of all elements of $M_0(n,m)$ which have an inverse; by definition $\mathrm{PL}_0(n,m)$ is the group of all block matrices $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ with $a \in GL(n)$ and $b \in GL(m)$. Then we conclude from our remarks made above that every element $U \in \mathrm{PL}(n,m)$ has a decomposition of the form

$$ U = U_0 \exp U_+, \tag{2.16}$$

where the elements $U_0 \in \mathrm{PL}_0(n,m)$ and $U_+ \in M_+(n,m)$ are *uniquely* determined. In particular $\mathrm{PL}(n,m)$ is the semidirect product of the subgroup $\mathrm{PL}_0(n,m)$ with the normal subgroup $1 + M_+(n,m) = \exp[M_+(n,m)]$.

This decomposition turns out to be a useful tool for the discussion of $\mathrm{PL}(n,m)$ and of its subgroups. For, on the one hand, $\mathrm{PL}_0(n,m)$ is an ordinary Lie group. On the other hand, the normal subgroup $1 + M_+(n,m)$ is easily treated by purely algebraic means, using the bijectivity of the exponential map (2.14) as well as the Baker–Campbell–Hausdorff series H. In fact, if $X, Y \in M_+(n,m)$ then the series representing $H(X,Y)$ breaks off after finitely many terms and we have $H(X,Y) \in M_+(n,m)$ and

$$ \exp[H(X,Y)] = (\exp X)(\exp Y). \tag{2.17}$$

It follows, for example, that every element $\exp(X)$, $X \in M_+(n,m)$, has a unique decomposition of the form

$$ \exp X = (\exp Y_0)(\exp Y_1), \tag{2.18}$$

with $Y_j \in M_+(n,m)$, Y_0 even and Y_1 odd, i.e., $Y_j \in \oplus_{r\geq 1} \times M_{2r-j}(n,m)$.

Next we remark that the associative algebra $M(n,m)$ can be converted into a Lie algebra by defining the commutator bracket as usual. The Lie algebra which emerges will be denoted by $\mathrm{pl}(n,m;A)$. The argument A (indicating the algebra A) is added in order to distinguish this Lie algebra from the complex general linear *graded* Lie algebra $\mathrm{pl}(n,m)$ which has been defined in Ref. 9. Recall that the elements of $\mathrm{pl}(n,m)$ are *complex* $(n+m)\times(n+m)$ matrices written in the same block form as the elements of $M(n,m)$. The algebra $\mathrm{pl}(n,m;A)$ is obtained from $\mathrm{pl}(n,m)$ according to the well-known rule: Multiply even elements of $\mathrm{pl}(n,m)$ by even elements of A, odd elements of $\mathrm{pl}(n,m)$ by odd elements of A, and consider all finite sums of the matrices thus obtained. Formally this means that

$$ \mathrm{pl}(n,m;A) = (A_0 \otimes \mathrm{pl}(n,m)_0) \oplus (A_1 \otimes \mathrm{pl}(n,m)_1), \tag{2.19}$$

where the commutator of two elements from the right-hand side is defined by

$$ [b \otimes B, d \otimes D] = bd \otimes \langle B, D \rangle, \tag{2.20}$$

with $b \in A_\beta$, $B \in \mathrm{pl}(n,m)_\beta$, $d \in A_\delta$, $D \in \mathrm{pl}(n,m)_\delta$; $\beta, \delta \in \{0,1\}$. The bracket $\langle\ ,\ \rangle$ denotes the graded commutator in $\mathrm{pl}(n,m)$. Note that according to this definition odd elements from A "commute" with odd elements from $\mathrm{pl}(n,m)$. Our discussion of the group $\mathrm{PL}(n,m)$ suggests that $\mathrm{pl}(n,m;A)$ is the Lie algebra to be associated with the graded Lie group $PL(n,m)$. This is the more true since in the case $\dim W < \infty$ the group $PL(n,m)$ is an ordinary Lie group and $pl(n,m;A)$ *is* its Lie algebra.

We close this section with the definition of the *graded determinant*.[10] Starting from the decomposition (2.16) it is not difficult to see that there exists a unique mapping detg of $PL(n,m)$ into A_0 which has the following properties:

$$\operatorname{detg}(UV) = \operatorname{detg}(U)\operatorname{detg}(V) \tag{2.21}$$

for all $U, V \in PL(n,m)$ and

$$\operatorname{detg}(\exp X) = \exp(\operatorname{Trg} X) \tag{2.22}$$

for all $X \in M(n,m)$. It should be observed that the graded determinant $\operatorname{detg}(U)$ is defined only for elements of $PL(n,m)$ and not on the whole of $M(n,m)$.

If $U = \begin{pmatrix} a & \xi \\ \eta & b \end{pmatrix}$ is an element of $PL(n,m)$ and if $U^{-1} = \begin{pmatrix} a' & \xi' \\ \eta' & b' \end{pmatrix}$ is its inverse then

$$\operatorname{detg}(U) = (\det a)(\det b'). \tag{2.23}$$

In particular we conclude that

$$\operatorname{detg}\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = (\det a)(\det b)^{-1} \tag{2.24}$$

(if a and b have an inverse), which is in general *different* from $\det \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$. Finally we observe that for all $U \in PL(n,m)$

$$\operatorname{detg}({}^T U) = \operatorname{detg}(U), \tag{2.25}$$

where ${}^T U$ denotes the graded transpose of U which has been defined in (2.8).

3. GRADED LIE GROUPS ASSOCIATED WITH THE SPECIAL LINEAR AND THE ORTHOSYMLECTIC GRADED LIE ALGEBRAS (COMPLEX CASE)

Among the classical simple graded Lie algebras[1,3] there exist two double sequences of algebras which are particularly important for applications[6-8]; these are the special linear graded Lie algebras $spl(n,m)$ and the orthosymplectic graded Lie algebras $osp(n,m)$, m even. Using the results of the preceding section it is easy to define the subgroups of $PL(n,m)$ which are to be associated with these algebras.

To begin with we define the *special linear graded Lie group* $SPL(n,m)$ by

$$SPL(n,m) = \{U \in PL(n,m) \mid \operatorname{detg}(U) = 1\} \tag{3.1}$$

and a subalgebra $spl(n,m;A)$ of $pl(n,m;A)$ by

$$spl(n,m;A) = \{X \in pl(n,m;A) \mid \operatorname{Trg}(X) = 0\}. \tag{3.2}$$

As an abbreviation let us also introduce the (ordinary) Lie group

$$SPL_0(n,m) = \{U \in PL_0(n,m) \mid \operatorname{detg}(U) = 1\} \tag{3.3}$$

and the ideal $spl_+(n,m;A)$ of $spl(n,m;A)$ by

$$spl_+(n,m;A) = M_+(n,m) \cap spl(n,m;A). \tag{3.4}$$

Note that $SPL_0(n,m)$ consists of the block matrices $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ with $a \in GL(n)$, $b \in GL(m)$, $\det(a) = \det(b)$. The algebra $spl(n,m;A)$ is obtained from $spl(n,m)$ in the same way as $pl(n,m;A)$ has been constructed out of $pl(n,m)$, i.e.,

$$spl(n,m;A) = (A_0 \otimes spl(n,m)_0) \oplus (A_1 \otimes spl(n,m)_1). \tag{3.5}$$

From (2.22) it follows that

$$\exp(X) \in SPL(n,m) \quad \text{for all } X \in spl(n,m;A). \tag{3.6}$$

Conversely an element $U \in PL(n,m)$ lies in $SPL(n,m)$ if and only if in the decomposition (2.16) we have

$$\operatorname{detg}(U_0) = 1, \quad \operatorname{Trg}(U_+) = 0, \tag{3.7}$$

i.e., if and only if $U_0 \in SPL(n,m)$ and $U_+ \in spl(n,m;A)$. We conclude that $spl(n,m;A)$ is the Lie algebra to be associated with the graded Lie group $SPL(n,m)$ and that $SPL(n,m)$ is the semi-direct product of the ordinary Lie group $SPL_0(n,m)$ with normal subgroup $\exp[spl_+(n,m;A)]$.

The orthosymplectic case is somewhat more interesting. In this case we suppose that m is even, $m = 2r$. Let g be a symmetric nonsingular complex $n \times n$ matrix, let h be a skew-symmetric nonsingular complex $2r \times 2r$ matrix, and let G be block matrix

$$G = \begin{pmatrix} g & 0 \\ 0 & h \end{pmatrix}. \tag{3.8}$$

A natural choice is

$$g = I_n, \quad h = \begin{pmatrix} 0 & I_r \\ -I_r & 0 \end{pmatrix},$$

where I_s denotes the $s \times s$ unit matrix.

Then the complex orthosymplectic graded Lie algebra $osp(n, 2r)$[1-3,9] is isomorphic to

$$osp(G) = \{B \in pl(n,2r) \mid {}^T B G + G B = 0\}, \tag{3.9}$$

where ${}^T B$ denotes the graded transpose of B [see (2.8)]. Similarly we define the *orthosymplectic graded Lie group* $OSP(G)$ by

$$OSP(G) = \{U \in PL(n,2r) \mid {}^T U G U = G\} \tag{3.10}$$

and a subalgebra $osp(G;A)$ of $pl(n,2r;A)$ by

$$osp(G;A) = \{X \in pl(n,2r;A) \mid {}^T X G + G X = 0\}. \tag{3.11}$$

Furthermore we introduce the ordinary Lie group

$$OSP_0(G) = \{U \in PL_0(n,2r) \mid {}^T U G U = G\} \tag{3.12}$$

[this group is isomorphic to $O(n,C) \times SP(2r,C)$] and the ideal $osp_+(G;A)$ of $osp(G;A)$

$$osp_+(G;A) = M_+(n,2r) \cap sp(G;A). \tag{3.13}$$

The equation ${}^T U G U = G$ yields

$$\operatorname{detg}(U) = \pm 1 \text{ for all } U \in OSP(G). \tag{3.14}$$

For any $X = \begin{pmatrix} a & \xi \\ \eta & b \end{pmatrix} \in pl(n,2r;A)$ the condition ${}^T X G + G X = 0$ is equivalent to

$$\begin{aligned} {}^t a g + g a &= 0, \\ {}^t b h + h b &= 0, \\ {}^t \xi g + h \eta &= 0. \end{aligned} \tag{3.15}$$

The following discussion is now completely analogous to the one for SPL. First we have

$$\operatorname{osp}(G;A) = (A_0 \otimes \operatorname{osp}(G)_0) \oplus (A_1 \otimes \operatorname{osp}(G)_1). \tag{3.16}$$

Next it is easy to see that

$$\exp(X) \in \mathrm{OSP}(G) \quad \text{for all } X \in \operatorname{osp}(G;A). \tag{3.17}$$

Finally an element $U \in \mathrm{PL}(n, 2r)$ lies in $\mathrm{OSP}(G)$ if and only if in the decomposition (2.16) we have

$$U_0 \in \mathrm{OSP}_0(G), \quad U_+ \in \operatorname{osp}_+(G;A). \tag{3.18}$$

Hence $\operatorname{osp}(G;A)$ is the Lie algebra to be associated with the graded Lie group $\mathrm{OSP}(G)$. [Let us stress once again that for $\dim W < \infty$ the group $\mathrm{OSP}(G)$ is an ordinary Lie group and that $\operatorname{osp}(G;A)$ is its Lie algebra.] Furthermore, we see that $\mathrm{OSP}(G)$ is the semidirect product of the ordinary Lie group $\mathrm{OSP}_0(G)$ with the normal subgroup $\exp(\operatorname{osp}_+(G;A))$.

The elements of $\mathrm{OSP}(G)$ may also be described by the properties of their Cayley transform. In fact, let $U \in \mathrm{PL}(n, 2r)$ and suppose that the Cayley transform $S = (1-U)(1+U)^{-1}$ does exist. Then U is an element of $\mathrm{OSP}(G)$ if and only if S is an element of the Lie algebra $\operatorname{osp}(G;A)$.

4. ADJOINT OPERATIONS IN *M(n,m)* AND "COMPACT FORMS" OF THE GRADED LIE GROUPS

Our next aim is to construct certain "compact forms" of the graded Lie groups which we have obtained above. These "compact forms" will be characterized by a "unitarity condition." Hence we shall consider first some adjoint operations in the algebra $M(n,m)$.

To begin with we choose an adjoint operation (of the first kind) in the algebra A. This is a semilinear mapping $a \to a^*$ of A into itself which satisfies

$$A_\alpha^* \subset A_\alpha \quad \text{for } \alpha \in \{0,1\} \tag{4.1}$$

and

$$(ab)^* = b^* a^* \tag{4.2}$$

$$a^{**} = a, \tag{4.3}$$

for all $a, b \in A$. It follows that $1^* = 1$.

To construct such an adjoint operation, we may choose any semilinear mapping $y \to y^*$ of W onto itself which satisfies $y^{**} = y$ (i.e., an involution of the first kind). Then there exists a unique adjoint operation of A which extends this involution.

Given the adjoint operation $a \to a^*$ of A we define as usual an adjoint operation $X \to X^+$ of $M(n,m)$ by

$$X^+ = {}^t X^* \quad \text{for all } X \in M(n,m). \tag{4.4}$$

Indeed, $X \to X^+$ is a semilinear mapping of $M(n,m)$ into itself which satisfies

$$(XY)^+ = Y^+ X^+, \tag{4.5}$$

$$X^{++} = X \tag{4.6}$$

for all $X, Y \in M(n,m)$.

Of course, all this is well known and is included only in order to contrast it with a second type of adjoint operations in $M(n,m)$ that will be discussed next. To define these operations we start with an adjoint operation "of the second kind" in the algebra A. This is a semilinear mapping $a \to a^x$ of A into itself which satisfies

$$A_\alpha^x \subset A_\alpha \quad \text{for } \alpha \in \{0,1\} \tag{4.7}$$

and

$$(ab)^x = a^x b^x, \tag{4.8}$$

$$c^{xx} = (-1)^\gamma c \tag{4.9}$$

for all $a, b \in A$ and $c \in A_\gamma$; $\gamma \in \{0,1\}$. Note that $1^x = 1$.

Such an operation does not necessarily exist. However, if the dimension of W is even (which includes the case where $\dim W$ is infinite), then the following construction is possible. Choose any semilinear mapping $y \to y^x$ of W onto itself which satisfies $y^{xx} = -y$ (i.e., an involution of the second kind; it is well-known that this yields the structure of a quaternionic vector space on W). Then there exists a unique adjoint operation of the second kind in A which extends this involution.

Suppose now that the operation $a \to a^x$ is given. Then we define an adjoint operation $X \to X^\neq$ in $M(n,m)$ by

$$X^\neq = {}^T X^x \quad \text{for all } X \in M(n,m) \tag{4.10}$$

(where ${}^T X$ denotes the graded transpose of X). In fact, it is easy to see that $X \to X^\neq$ is indeed a semilinear mapping which satisfies

$$(XY)^\neq = Y^\neq X^\neq \tag{4.11}$$

$$X^{\neq\neq} = X \tag{4.12}$$

for all $X, Y \in M(n,m)$. Note the close connection of this operation with the grade adjoint operations as defined in Ref. 11. The definitions (4.4) and (4.10) are easily generalized to give an adjoint operation with respect to some indefinite Hermitian scalar product.

It is now obvious to define the *unitary graded Lie groups* $\mathrm{UPL}(n,m)$ [and, similarly, the *special unitary graded Lie groups* $\mathrm{USPL}(n,m)$]. In fact, we introduce the subgroup $\mathrm{UPL}(n,m)$ of $\mathrm{PL}(n,m)$ by

$$\mathrm{UPL}(n,m) = \{U \in \mathrm{PL}(n,m) \mid U^+ = U^{-1}\}, \tag{4.13}$$

the real subalgebra $\operatorname{upl}(n,m;A)$ of $\operatorname{pl}(n,m;A)$ by

$$\operatorname{upl}(n,m;A) = \{X \in \operatorname{pl}(n,m;A) \mid X^+ = -X\}, \tag{4.14}$$

the ordinary real Lie group $\mathrm{UPL}_0(n,m)$ by

$$\mathrm{UPL}_0(n,m) = \{U \in \mathrm{PL}_0(n,m) \mid U^+ = U^{-1}\}, \tag{4.15}$$

and an ideal $\operatorname{upl}_+(n,m;A)$ of $\operatorname{upl}(n,m;A)$ by

$$\operatorname{upl}_+(n,m;A) = M_+(n,m) \cap \operatorname{upl}(n,m;A). \tag{4.16}$$

As in the preceding section we conclude then that $\operatorname{upl}(n,m;A)$ is the Lie algebra associated with the group $\mathrm{UPL}(n,m)$. Furthermore, this group is the semidirect product of the ordinary Lie group $\mathrm{UPL}_0(n,m)$ [which is isomorphic to $U(n) \times U(m)$] with the normal subgroup $\exp(\operatorname{upl}_+(n,m;A))$. Finally, if U is any element of $\mathrm{PL}(n,m)$ whose Cayley transform S does exist then U belongs to $\mathrm{UPL}(n,m)$ if and only if S lies in $\operatorname{upl}(n,m;A)$.

Literally the same discussion can be carried out if we work with the adjoint operation $X \to X^\neq$ instead of $X \to X^+$. Hence we obtain two different types of unitary graded Lie groups. At present we do not know of any connection between these two cases. Dealing with the unitary graded Lie groups one might prefer the adjoint operation $X \to X^+$ however.

In the construction of a "compact form" of the orthosymplectic groups only the adjoint operation $X \to X^{\neq}$ will work. The reason is that on the graded Lie algebra $\mathrm{osp}(n, 2r)$ there does not exist an adjoint operation defining a compact form of the Lie algebra $\mathrm{osp}(n, 2r)_0 \approx \mathrm{O}(n) \times \mathrm{sp}(2r)$ (see Ref. 11).

Let us now construct a "compact form" of the orthosymplectic group OSP(G). To avoid the possible complication of having to modify our definition of the adjoint operation we shall assume that the (complex) metric matrix $G = \begin{pmatrix} g & 0 \\ 0 & h \end{pmatrix}$ is *unitary*. Then it is easy to see that the adjoint operation $X \to X^{\neq}$ (which is assumed to exist) *maps* OSP(G) *and* osp(G;A) *into themselves*.

Hence we define the group

$$\mathrm{UOSP}(G) = \{U \in \mathrm{PL}(n, 2r) \mid {}^T UGU = G, U^{\neq} = U^{-1}\}, \quad (4.17)$$

the Lie algebra.

$$\mathrm{uosp}(G;A) = \{X \in \mathrm{pl}(n, 2r;A) \mid {}^T XG + GX = 0, X^{\neq} = -X\} \quad (4.18)$$

and, similarly to the earlier cases, the ordinary Lie group $\mathrm{UOSP}_0(G)$ and the ideal $\mathrm{uosp}_+(G;A)$.

Then uosp(G;A) [which is a real form of osp(G;A)] is the Lie algebra associated with the group UOSP(G) and this group is, once again, the semidirect product of the ordinary Lie group $\mathrm{UOSP}_0(G)$ [which is isomorphic to the compact Lie group $\mathrm{O}(n) \times \mathrm{SP}(2r)$] with the normal subgroup $\exp(\mathrm{uosp}_+(G;A))$. Of course, for fixed dimensions $n, 2r$ all the groups UOSP(G) (with G unitary) are isomorphic.

5. CONCLUSION

In this work we have constructed the general linear, the special linear, and the orthosymplectic graded Lie group, their "compact forms," and the corresponding Lie algebras. The "parameters" were taken from an exterior algebra $A = \wedge W$. To prepare our constructions we have first extended the usual matrix operations to the algebra $M(n, m)$. Once this had been done we could proceed as in the ordinary cases. All the graded Lie groups that we have obtained are semidirect products of an ordinary Lie group with a normal subgroup that "depends only on the Grassmann variables". It should be stressed that in the case $\dim W < \infty$ our groups are ordinary Lie groups. In the construction of "compact forms" two special features make their appearance. First, we find two types of adjoint operations and, consequently, two types of unitary groups. Second, in this construction it is not the real forms of the graded Lie algebras which are important but rather the adjoint resp. grade adjoint opeations as defined in Ref. 11. In fact, in the orthosymplectic case a real form of $\mathrm{osp}(n, 2r)$ containing a compact form of the complex Lie algebra $\mathrm{osp}(n, 2r)_0 \approx \mathrm{o}(n) \times \mathrm{sp}(2r)$ does not even exist.

We would like to suggest to the reader to consider the group OSP(1,2) as an example. In fact, in this case everything can be worked out explicitly in a trivial way.

It is obvious from our considerations that the definitions of the general linear, the special linear and the orthosymplectic graded Lie groups are completely analogous to those of the general linear, the special linear, and the orthogonal[12] and symplectic Lie groups, respectively. The same holds true for the "compact forms" apart from the fact that two different types of unitarity conditions are possible.

From the multitude of questions which are suggested by this work let us only mention the following. Is there any connection between the two types of unitary graded Lie groups? What is the bearing of the unitary graded Lie groups on the representation theory of graded Lie groups?[13]

[1]V.G. Kac, Commun. Math. Phys. 53, 31 (1977) and references therein.
[2]P.G.O. Freund and I. Kaplansky, J. Math. Phys. 17, 228 (1976).
[3]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. 17, 1626, 1640 (1976). Note a slight change in notation as compared to this work. In the present paper the Lie algebra of $\mathrm{osp}(n, 2r)$ is equal to $\mathrm{o}(n) \times \mathrm{sp}(2r)$.
[4]F.A. Berezin and D.A. Leites, Dokl. Akad. Nauk SSSR 224, 505 (1975) [Sov. Phys. Dokl. 16, 1218 (1975)].
[5]B. Kostant, in *Differential Geometrical Methods in Mathematical Physics*, Bonn, 1975, Lecture Notes in Mathematics 570 (Springer-Verlag, Berlin, 1977).
[6]For a recent survey of the subject see P. Fayet and S. Ferrara, "Supersymmetry," Phys. Rep. C 32, 250 (1977).
[7]R. Arnowitt and P. Nath, Phys. Lett. B 56, 117 (1975); P.K. Tounsend and P. Nieuwenhuizen, Stony Brook Report ITP-SB-77-18 (1977).
[8]R. Casalbuoni, Nuovo Cimento A 33, 389 (1976); F.A. Berezin and M.S. Marinov, Ann. Phys. (NY) 104, 336 (1977).
[9]W. Nahm and M. Scheunert, J. Math. Phys. 17, 868 (1976). See also Ref. 3.
[10]R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B 56, 81 (1975).
[11]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. 18, 146 (1977).
[12]Note that there are no "orthogonal" graded Lie groups. The usual transpose operation is incompatible with the group multiplication rule [see (2.7)]. For the graded transpose operation (2.8) one gets ordinary Lie groups.
[13]For the case of the "usual" adjoint operation see F.A. Berezin, Funct. Anal. Appl. 10, 70 (1976).

Asymptotic simplicity is stable[a)]

Robert Geroch and Basilis C. Xanthopoulos

Laboratory for Astrophysics and Space Research, 933 E. 56th Street, Chicago, Illinois 60637

Consider an asymptotically simple solution of Einstein's equation. It is shown that any internally generated, first-order perturbation of the metric, as a consequence of the linearized Einstein equation, preserves asymptotic simplicity to first order.

1. INTRODUCTION

There is available[1] in general relativity the notion of an asymptotically simple space—time. The definition requires, roughly speaking, that it be possible to attach to the space—time manifold $\tilde{M}$ a boundary "at null infinity" such that a certain conformal rescaling of the physical metric $\tilde{g}_{ab}$ results in a metric having a smooth extension to that boundary, and such that the corresponding conformal factor have specified asymptotic behavior, i.e., specified behavior on that boundary.

It is intended that asymptotic simplicity capture the physical idea that one has an isolated system—that in particular any deviations of the space—time metric from flatness can in some sense be attributed to the presence of that system. Thus, for example, it is a consequence of the definition that the space—time metric approach a flat metric in the asymptotic limit far from the system. There are essentially two pieces of evidence that the detailed definition is in fact an appropriate one for the physics it is intended to represent. First, asymptotic simplicity has turned out to provide a particularly natural framework for a number of important and useful notions in general relativity: Bondi energy—momentum,[2] radiation fields and in particular their peeling behavior, the Newman—Penrose conserved quantities,[3] cosmic censorship and the global structure of black holes,[4] the BMS asymptotic symmetry group, etc. Second, it has been found that various known exact solutions which seem intuitively to represent "isolated systems"—such as Minkowski space, the Schwarzschild, Kerr, and certain Weyl solutions (all with appropriate sources)—in fact satsify the conditions for asymptotic simplicity. Unfortunately, this evidence taken as a whole is perhaps not as strong as one would like. In particular, as a result of the relative scarcity of known exact solutions of Einstein's equation, the class of examples on which one can test the definition is not large. Thus, for example, even though the formalism is to provide the basic framework for the description of radiation in general relativity, there is no known exact, radiating, asymptotically simple solution of Einstein's equation.

We here consider a third test of the definition. Consider an asymptotically simple solution of Einstein's equation, with bounded source. Let there be introduced a first-order perturbation of the metric, generated, say, from a bounded source. Then, as a consequence of the linearized Einstein equation, this perturbation will radiate to infinity. One can now ask whether, under this arrangement, asymptotic simplicity will continue to be satisfied to first order in the perturbation.[2] One is asking, then, whether asymptotic simplicity is stable in a certain sense.[5] One is vastly enlarging the class of examples on which to test the definition, at the price of carrying out that test only to first order.[6] Were asymptotic simplicity unstable to such internally generated perturbations, then the definition would presumably have to be modified. We shall show stability.

Our result has a couple of other implications. Each of the various asymptotic constructions can be carried out only in the presence of a certain degree of smoothness of the conformally scaled metric. The Newman—Penrose quantities, for example, require C^5. The issue of which of these constructions are likely to have physical significance thus becomes: What degree of smoothness is physically realistic? It appears to be difficult to answer this question directly, for the translation of "the unphysical metric is n times continuously differentiable" in terms of the physical metric results in an awkward statement having little direct physical meaning. The present result will rather suggest that a reasonable differentiability condition on the conformally scaled metric is "C^∞". Second, we provide a gauge for the metric perturbation along with a guarantee that, in this gauge, the perturbation will be well-behaved asymptotically. It is easy to derive from this other gauges within which one can work, and to show that in certain other gauges one cannot work, in "linearizing" various asymptotic constructions.

2. THE STABILITY THEOREM

Let $\tilde{M}$, $\tilde{g}_{ab}$ be a space—time, i.e., $\tilde{M}$ is a smooth (C^∞) 4-manifold and $\tilde{g}_{ab}$ is a smooth metric of Lorentz signature on $\tilde{M}$. This space-time is said to be *asymptotically simple*[7] if there exists a smooth manifold with boundary, $M = \tilde{M} \cup I$, consisting of $\tilde{M}$ with boundary I attached, together with a smooth Lorentz metric g_{ab} on M and a smooth scalar field Ω on M, such that:

1. On $\tilde{M}$, $g_{ab} = \Omega^2 \tilde{g}_{ab}$.

2. At points of I, Ω vanishes and its gradient is nonzero and null.

3. Every maximally extended null geodesic in $\tilde{M}$ has, in M, two end points on I.

The first condition requires that the unphysical metric, g_{ab}, be a conformal scaling of the physical, $\tilde{g}_{ab}$; the second specifies the asymptotic behavior of the con-

a)Supported in part by the National Science Foundation under contract number PHY 76-81102 with the University of Chicago.

formal factor (essentially, that Ω vanish asymptotically "as $1/r$"); the third ensures that the entire boundary at null infinity has been attached to $\tilde{M}$ in obtaining M.

We shall be concerned principally with regions in which the physical metric, $\tilde{g}_{ab}$, satisfies Einstein's equation with zero source. This equation, expressed in terms of the unphysical metric g_{ab} using condition 1, is

$$\Omega R_{ab} + 2\nabla_a n_b + (\nabla_m n^m - 3\Omega^{-1} n_m n^m) g_{ab} = 0, \quad (1)$$

where indices are raised and lowered with g_{ab} and its inverse, ∇_a is the derivative operator compatible with g_{ab}, R_{ab} is its Ricci tensor,[8] and where we have set $n_a = \nabla_a \Omega$.

Let $\tilde{\gamma}_{ab}$ be a first-order perturbation of the physical metric, $\tilde{g}_{ab}$. The linearized Einstein equation on $\tilde{\gamma}_{ab}$, expressed in terms of the corresponding perturbation, $\gamma_{ab} = \Omega^2 \tilde{\gamma}_{ab}$, of the unphysical metric, is

$$\begin{aligned}\nabla^2 \gamma_{ab} = {} & 2\nabla_{(a}\nabla^m \gamma_{b)m} - \nabla_a \nabla_b \gamma^m{}_m - 2R_{ambn}\gamma^{mn} + 2R^m{}_{(a}\gamma_{b)m} \\ & - 2\Omega^{-1} n^m \cdot (2\nabla_{(a}\gamma_{b)m} - \nabla_m \gamma^n{}_n) - (2\Omega^{-2} n^m n_m \\ & + \tfrac{1}{3}R) \cdot (\gamma_{ab} + \tfrac{1}{2}\gamma^m{}_m \cdot g_{ab}) + (R^{mn}\gamma_{mn} \\ & + \Omega^{-1} n^m \nabla_m \gamma^n{}_n - 2\Omega^{-1} n^m \nabla^n \gamma_{mn}) \cdot g_{ab} + 6\Omega^{-2} n^m n^n \gamma_{mn} \cdot g_{ab},\end{aligned} \quad (2)$$

where $\nabla^2 = g^{ab}\nabla_a \nabla_b$. We have the freedom to perform gauge transformations: The fields $\tilde{\gamma}_{ab}$ and $\tilde{\gamma}_{ab} + \tilde{\nabla}_{(a}\tilde{\xi}_{b)}$, where $\tilde{\xi}_b$ is any vector field on $\tilde{M}$, represent the same physical perturbation. In terms of unphysical fields, a gauge transformation replaces γ_{ab} by

$$\gamma_{ab} + \Omega^2 \nabla_{(a}\xi_{b)} + 2\Omega n_{(a}\xi_{b)} - \Omega n^m \xi_m g_{ab}. \quad (3)$$

A symmetric $\tilde{\gamma}_{ab}$ on an asymptotically simple space–time will be said to be *asymptotically regular* if $\gamma_{ab} = \Omega^2 \tilde{\gamma}_{ab}$ has smooth extension from $\tilde{M}$ to M and, under this extension, $\gamma_{ab} n^a n^b$ vanishes at I. These conditions will be recognized as the linearizations of the conditions in the definition of asymptotic simplicity, i.e., they guarantee that "asymptotic simplicity is preserved to the first order in the perturbation." Indeed, condition 1 is reflected in the definition of γ_{ab}; the vanishing of Ω and the nonvanishing of its gradient at I in condition 2 are metric-independent; the nullness of the gradient of Ω in condition 2 is reflected in the vanishing of $\gamma_{ab} n^a n^b$; that condition 3 be preserved to first order follows from the others, since null geodesics are conformally invariant.[9]

Our result is:

Theorem: Let $\tilde{M}$, $\tilde{g}_{ab}$ be an asymptotically simple space–time (with M, I, g_{ab}, Ω), and let $\tilde{\gamma}_{ab}$ be a smooth symmetric tensor field on $\tilde{M}$. Let, in some neighborhood of I, $\tilde{g}_{ab}$ satisfy Einstein's equation with zero source and $\tilde{\gamma}_{ab}$ the linearized Einstein equation. Let $\tilde{\gamma}_{ab}$ vanish outside of some compact subset of some slice of $\tilde{M}$. Then, possibly after a gauge transformation, $\tilde{\gamma}_{ab}$ is asymptotically regular.[2]

That $\tilde{\gamma}_{ab}$ satisfy the linearized Einstein equation, i.e., that γ_{ab} satisfy (2), in a neighborhood of I ensures that the perturbation field reaching I will be "a true radiation field, which evolved to I from within the space–time." The theorem is, of course, false without this condition: Choose, in Minkowski space–time, any $\tilde{\gamma}_{ab}$ very badly behaved asymptotically. The condition that $\tilde{\gamma}_{ab}$ vanish outside a compact subset of a slice prevents one from sending in $\tilde{\gamma}$-radiation from past infinity whose intensity increases into the future. Such radiation would "pile up" on future infinity, and could (e.g., in Minkowski space–time) destroy asymptotic regularity there. The theorem would also be false without "after a gauge transformation": E.g., $\tilde{\gamma}_{ab} = \tilde{\nabla}_{(a}\tilde{\xi}_{b)}$ in Minkowski space–time, with $\tilde{\xi}_b$ badly behaved asymptotically, will not be asymptotically regular. We shall actually prove slightly more than is stated in the theorem, namely that, possibly after a gauge transformation, one can have not only asymptotic regularity, but even the vanishing of γ_{ab} at I, with, furthermore, $\gamma_{am}n^m$ vanishing one order faster and $\gamma_{mn}n^m n^n$ vanishing one order faster still. That is to say, the metric perturbation can be made to satisfy somewhat stronger peeling conditions than one might *a priori* have expected.

Statements analogous to that of the theorem can be made for other fields. For the conformally invariant fields, the analogous statements are both true and easy to prove. We sketch, for the scalar case $[(\tilde{\nabla}^2 - \frac{1}{6}\tilde{R})\tilde{\varphi} = 0]$, the proof.[2] Let asymptotically simple $\tilde{M}$, $\tilde{g}_{ab}$ be given. First fix an extension of the manifold with boundary M through its boundary I to a manifold without boundary $\hat{M}$, and smooth extensions of g_{ab} and Ω to $\hat{M}$. Then $n^a = \nabla^a \Omega$ is, at I, a nonzero null normal to I, and in particular is tangent to I. Let p be any point of, say, future null infinity, I^+ (Fig. 1). The integral curve of $-n^a$ from p is directed into the past, and remains in I^+. Since $\tilde{\varphi}$ vanishes outside a compact subset of a slice, this integral curve will eventually reach a point q of I^+ in some neighborhood of which $\tilde{\varphi}$ vanishes. Fix a small neighborhood S in I^+ of this segment of the integral curve between p and q, so S is a null 3-submanifold of $\hat{M}$. We next "tip" S slightly to obtain a spacelike 3-submanifold S_0, as shown in the figure, which meets q and which is well away from any source for $\tilde{\varphi}$. Since our equation on $\tilde{\varphi}$ is conformally invariant, we have, on $\varphi = \Omega^{-1}\tilde{\varphi}$, the equation $(\nabla^2 - \frac{1}{6}R)\varphi = 0$. By construction, $\tilde{\varphi}$ vanishes in a neighborhood of $S_0 \cap I$, i.e., Ω is bounded away from zero in the intersection of the support of $\tilde{\varphi}$ and S_0. Hence, φ and its first normal derivative on S_0, the initial data for our wave equation, will

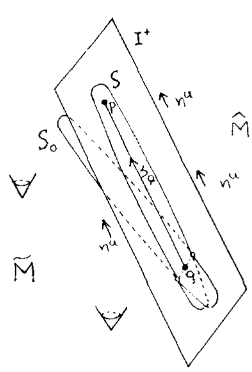


FIG. 1. The geometrical arrangement for establishing smoothness of certain fields at I. The null, three-dimensional surface S lies in future null infinity I^+, while S_0 is a nearby spacelike submanifold passing through point q.

be smooth on S_0. We thus have smooth initial data on S_0 for a field satisfying an equation with a well-posed initial-value formulation, and so, since p is in the future domain of dependence of S_0 in $\hat{M}$, φ must be smooth at p. Since p on I is arbitrary, asymptotic regularity (meaning, for this case, smoothness of φ at I) follows. The situation for other fields, and for certain coupled fields, is discussed in Sec. 4.

The idea is to prove the theorem for the present case, gravitation, by an argument similar to that above. What makes the gravitational case more difficult than the scalar is two features: (i) The equation for the perturbation is not conformally invariant, and (ii) one must show regularity of the "potential," γ_{ab}, and not just the "field," or, what is the same thing, one must contend with the freedom of gauge transformations. In more detail, what one must do is find a set of fields constructed from (say, linear in) γ_{ab}, together with a set of gauge conditions on γ_{ab}, such that the following conditions are satisfied: (i) Smoothness of the fields at I implies asymptotic regularity, (ii) the gauge conditions can in fact be realized by means of a gauge transformation (3), and (iii) when Eq. (2) is rewritten, possibly using the gauge conditions, as a system of differential equations on these fields, the resulting system admits a well-posed initial-value formulation in some sense which makes the proof work. The exhibition of a set of fields and gauge conditions with these three properties, we claim, will complete the proof.

It turns out that the choice of fields and of gauge conditions is a rather delicate business. Suppose, for example, that one made the obvious choice: Let the field be γ_{ab} itself, and impose on γ_{ab} its Lorentz gauge condition, $\nabla^m(\gamma_{am}-\frac{1}{2}\gamma^p{}_p g_{am})=0$. Then Eq. (2), since the first terms on the right now vanish, is a hyperbolic differential equation for γ_{ab}. Unfortunately, this equation does not have a well-posed initial-value formulation in the necessary sense, for some coefficients on the right involve inverse powers of Ω, while we must apply the equation in a region of $\hat{M}$ in which Ω goes through zero. One might therefore proceed by introducing additional fields in which these inverse powers are incorporated, e.g., for the last term on the right, the field $u=\Omega^{-2}\gamma_{mn}n^m n^n$, in terms of which this last term becomes simply $6\,u g_{ab}$. Now, however, we require an equation for u. Contracting (2) with $n^a n^b$, we indeed obtain an expression for $\nabla^2 u$ in terms of our fields and their first derivatives, but, again, more terms in inverse powers of Ω appear. One might therefore introduce additional fields and/or further gauge conditions. As far as we are aware, this process does not terminate. Our naive initial choice, in short, does not work.

3. THE PERTURBATION FIELDS

A choice which, as we shall show, does work is the following. Let the fields be

$$\tau_{ab}=\Omega^{-1}\gamma_{ab},\quad \tau_a=\Omega^{-1}n^m\tau_{am},\quad \tau=g^{mn}\tau_{mn},$$
$$\sigma=\Omega^{-1}(n^m\tau_m+\tfrac{1}{2}n^m\nabla_m\tau+\tfrac{1}{4}f\tau), \tag{4}$$

and let the gauge conditions be

$$y_a=\nabla^m\tau_{am}-\nabla_a\tau-3\tau_a=0, \tag{5}$$
$$(n^m\nabla_m+\tfrac{1}{6}\Omega R+\tfrac{3}{2}f)\nabla^2\tau$$
$$=\tfrac{1}{12}Rf\tau-\tfrac{1}{2}\tau\nabla^2 f-\tfrac{1}{3}Rn^m\tau_m+4\Omega^{-1}C_{ambn}\tau^{mn}n^a n^b, \tag{6}$$

where we have set $f=\Omega^{-1}n^a n_a$, C_{ambn} is the Weyl tensor.

We now claim that these fields and these gauge conditions do satisfy the three properties required in the previous section. Indeed, the first condition (that smoothness of the fields implies asymptotic regularity) is immediate, since smoothness of τ_{ab} alone implies smoothness of (and also vanishing of) γ_{ab} at I, so, vanishing of $n^a n^b\gamma_{ab}$ at I too. For the second condition (realizability of the gauge conditions), we first compute, from (3), the changes in the fields y_a and τ under a gauge transformation:

$$\delta y_a=\Omega\nabla^m\nabla_{[m}\xi_{a]}, \tag{7}$$
$$\delta\tau=\Omega\nabla_m\xi^m-2n_m\xi^m. \tag{8}$$

Taking the trace of (2), substituting (5), we obtain $\nabla_a(\Omega^{-1}y^a)=0$. Hence, from (7), one can always find a gauge vector ξ_a such that $y_a=0$, i.e., such that our gauge condition (5) is satisfied.[10] We must still realize (6). The gauge condition (5) is preserved by any further gauge transformation (3) with $\xi_a=\nabla_a\xi$. We now demand of this scalar field ξ that (6) be satisfied. There results, by (8) and (3), a differential equation of the form:

$$(n^m\nabla_m+\tfrac{1}{6}\,\Omega R+\tfrac{3}{2}f)[\nabla^2[\Omega^2(\nabla^2-\tfrac{1}{6}R)(\Omega^{-1}\xi)]]$$

$=$ terms involving ξ only linearly in its value and the values of its first two derivatives. (9)

But this equation admits a solution ξ, by the Appendix. [Let, for application of that theorem, the fields be $F_1=\Omega^{-1}\xi$, $F_2=\nabla F_1$, $F_3=\nabla F_2$, $F_4=\Omega^2(\nabla^2-\frac{1}{6}R)\,F_1$, $F_5=\nabla F_4$, and $H_1=\nabla^2 F_4$. The equations for the F's are those which result from these definitions; for H_1, Eq. (9).] We conclude, then, that our gauge conditions, (5) and (6), can in fact be realized by means of a gauge transformation.

There remains only the verification of the third condition [that our fields (4) satisfy, under (5) and (6), equations with a well-posed initial-value formulation]. The situation here is, of course, somewhat more delicate than that of the previous paragraph: Whereas the gauge conditions need only be imposed in the physical space–time, the present initial-value formulation must be applicable in the extended unphysical space–time, i.e., even as Ω goes through zero. We claim that, from (2) using (5) and (6), the fields (4) satisfy the following system of equations:

$$\nabla^2\tau_{ab}=\nabla_a\nabla_b\tau+4\nabla_{(a}\tau_{b)}-2C_{ambn}\tau^{mn}-\tfrac{1}{6}R\tau_{ab}$$
$$+\tfrac{1}{12}R\tau g_{ab}-\tfrac{1}{2}\tau R_{ab}+2R_{m(a}\tau_{b)}{}^m-2\sigma g_{ab}, \tag{10}$$

$$\nabla^2\tau_a=2\nabla_a\sigma+\tfrac{1}{2}R_{am}\nabla^m\tau+\tfrac{1}{12}R\nabla_a\tau-R^{mn}\nabla_m\tau_{an}-\tfrac{1}{3}\tau_{ab}\nabla^b R$$
$$+2\tau^{mn}\nabla_{[m}R_{a]n}+2\tau^m R_{am}+\tfrac{1}{2}R\tau_a+\tfrac{1}{6}\,\tau\nabla_a R, \tag{11}$$

$$\nabla^2\sigma=-\tfrac{1}{2}R^{mn}\nabla_m\nabla_n\tau-2R^{mn}\nabla_m\tau_n-\tfrac{1}{12}(\nabla^m R)(\nabla_m\tau)+R\sigma$$
$$+\tfrac{1}{72}R^2\tau-\tfrac{1}{2}\tau_{ab}R^a{}_m R^{bm}-\tfrac{1}{3}\tau^m\nabla_m R. \tag{12}$$

Indeed, (10) follows from (2), eliminating γ_{ab}, using (4) and then using the gauge condition (5). Equation (11) follows from (10), contracting with n^b; and (12) in turn from (11), contracting with n^a and using the gauge condition (6). We next claim that the system of equations (10), (11), (6), and (12) for the respective fields (4) has a well-posed initial-value formulation. First note that all of the coefficient fields on the right in these equations are smooth in the extended unphysical space—time: Smoothness of f follows from the definition of asymptotic simplicity; of $\Omega^{-1}C_{abcd}n^d$ by taking the curl of (1). But we now have a system to which the Appendix is applicable. Let, for that theorem, the fields be $F_1 = \tau_{ab}$, $F_2=\tau_a$, $F_3=\sigma$, $F_4=\tau$, $F_5=\nabla F_4$, and $H_1=\nabla^2 F_4$. The equations for F_1, F_2, F_3, and H_1 are (10), (11), (12), and (6), respectively, while the equations for F_4 and F_5 come from the definition of H_1. By the Appendix, then, our system has a well-posed initial-value formulation.

This completes the proof of the theorem.

The present choices of perturbation fields and gauge conditions are rather complicated. We describe briefly how these choices arise. First, one knows that, in an asymptotically simple space—time, the value of the unphysical metric g_{ab} at I is purely kinematical, i.e., that it tells one nothing about the particular space—time under consideration. This observation suggests that one demand that the perturbation, γ_{ab}, of the unphysical metric vanish at I, i.e., that one choose for one's perturbation field the τ_{ab} given in (4). Replacing γ_{ab} in (2) in favor of τ_{ab}, and eliminating divergences of τ_{ab} in favor of y_a, one obtains (10)—but with some additional terms on the right involving negative powers of Ω. These terms can, however, be eliminated by the gauge choice (5)—a choice further suggested by the observation that if τ_{ab} and $\Omega^{-1}y_a$ are smooth at I under some gauge choice, then (5) will not destroy smoothness of τ_{ab}. We now have our equation, (10), for τ_{ab}, an equation which, however, has two unsatisfactory features: (i) Its right side involves τ_a and σ, fields the smoothness of which does not follow from that of τ_{ab}, and (ii) its right side involves second derivatives of τ, whence the equation is not even hyperbolic. The resolution of (i) is to introduce τ_a and σ as fields in their own right, subject to their own equations. On τ_a, one has Eq. (11), and so there remains only the introduction of an equation on σ and the resolution of (ii). Contracting (11) with n^a, we obtain the equation given by the sum of (12) and -2Ω times (6). One now wishes to resolve this equation into two, by separating its terms into two groups and equating each group to zero. One of these equations is to be a gauge condition and the equation for τ [ultimately, (6)]; the other, the equation for σ [ultimately, (12)]. This grouping of terms, however, must be carried out in such a way that the Appendix be applicable, i.e., such that (i) only terms having an external factor of Ω be grouped with "$\Omega\nabla^2\sigma$," in order that the $\nabla^2\sigma$ equation have smooth coefficients on the right, and (ii) no terms involving derivatives of our fields be grouped with "$(n^m\nabla_m + \frac{1}{6}\Omega R + \frac{3}{2}f)\nabla^2\tau$," in order that the presence of second derivatives of τ in (10) not destroy hyperbolicity. But such a grouping turns out to be possible. [There are many.]

4. CONCLUSION

Statements analogous to that of the theorem of Sec. 2 can be formulated for fields other than gravitation. Their general form is that, given the field satisfying its equation, then, subject to suitable global safeguards, an appropriate unphysical (i.e., rescaled) version of the physical field admits a smooth extension to I. As we have seen in Sec. 2, for the conformally invariant fields (scalar, neutrino, electromagnetic—there are, of course, no higher spin fields in curved space—times), these statements are true. A similar argument works for the conformally noninvariant, mass-zero scalar field ($\tilde{\nabla}^2\tilde{\phi}=0$).[11] For the case of the various massive fields (spins 0, $\frac{1}{2}$, and 1), the situation is less clear. One would have expected, both on physical grounds (massive particles cannot reach null infinity) and from an examination of behavior in Minkowski space, that, not only would such fields be asymptotically regular, but also that the unphysical field itself would vanish on I. However, a naive attempt to adapt the present arguments to these cases fails. The mass-term acquires, under the conformal transformation, a coefficient involving a negative power of Ω, while there seems to be no obvious mechanism for eliminating it. Thus, while surely asymptotic regularity must hold for these massive fields, a proof is apparently lacking.

One can also formulate statements analogous to that of the theorem for various systems of coupled fields. The general rule seems to be that the equations used to show asymptotic regularity of the individual noninteracting fields, when modified to include the coupling terms, suffice to show asymptotic regularity of the system of interacting fields. Consider, as an example, the coupled Einstein—Maxwell system. Since the Maxwell stress—energy has vanishing trace, one can still impose the gauge condition (5) on the gravitational perturbation. Impose also (6). Then the system of equations for the gravitational perturbation, (6), (10), (11), and (12), is modified only by the inclusion of additional terms on the right involving the background Maxwell field and its perturbation. Similarly, as a consequence of Maxwell's equations, the perturbation of the Maxwell field satisfies an equation in which only the perturbation of the gravitational field and its derivative appear on the right. No coefficients on the right of either equation, however, involve inverse powers of Ω, i.e., these coefficients are smooth at I. The result, then, is a system of equations on the unphysical perturbations to which the Appendix is again applicable. One thus shows asymptotic regularity for this system. One can apparently treat in a similar way more complicated coupled systems. There is, however, one anomalous case: that of the coupled Einstein-conformally invariant scalar field. The problem here is that the stress—energy of the scalar field involves second derivatives of that field, whence, since that stress—energy appears on the right in the gravitational equations, one does not even obtain a hyperbolic system. This difficulty, however, seems to be inherent in the structure of conformally invariant scalar fields: Indeed, it is apparently not known whether the coupled Einstein-conformally invariant scalar system admits a well-posed initial-value formulation in the physical space—time.

We have here treated only the case of first-order perturbations in the gravitational case, i.e., it is only to this order that we have shown preservation of asymptotic simplicity. One might imagine looking for a generalization to the full non-linear case. It seems, however, to be difficult to even formulate the question. One might, for example, try to introduce a suitable topology on a space of initial-data sets for Einstein's equation, and then attempt to show that, in this topology, the collection of initial-data sets which evolve to an asymptotically simple space—time is open. The problem seems to be that, given the nonlinear character of Einstein's equation, it is extremely difficult to characterize structural properties of the evolved space—time in terms of just the initial data. A much more tractable question would be to ask whether or not asymptotic simplicity is preserved in the gravitational case to orders higher than the first. It seems likely that one could obtain a positive result to any given order n, although—at least in the absence of some understanding of what makes the first-order case work—the computations, even at the second order, will become extremely messy.

The present choice of fields and gauge-conditions is rather complicated—and in particular this choice seems to shed little light on why any choice at all should work or how asymptotic simplicity came to enjoy this stability property. The discussion at the end of Sec. 3 suggests that at least to some extent these complications are inherent in the problem. But Einstein's equation is basically so simple: There must be some way to see what mechanism is operating here.

ACKNOWLEDGMENTS

We would like to thank Abhay Ashtekar, Roy Kerr, and Hans Seifert for helpful discussions.

APPENDIX

Theorem: Let M, g_{ab} be a smooth (C^∞) space—time, t a smooth scalar field on M which has everywhere timelike gradient and which assumes all values in the interval $(-1, +1)$, and n^a a smooth vector field on M with $n^a\nabla_a t < 0$. Let it be the case that along every maximally extended timelike curve in M and along every maximally extended n^a-integral curve in M, t assumes all values in $(-1, +1)$.

Consider, on tensor fields $F_1, \ldots, F_p$ and $H_1, \ldots, H_q$ on M, the following system of partial differential equations:

$$\nabla^2 F_i = \{1, F_m, H_n, \nabla F_m, \nabla H_n\} \; (i = 1, \ldots, p), \tag{A1}$$

$$n^a \nabla_a H_j = \{1, F_m, H_n\} \quad (j = 1, \ldots, q), \tag{A2}$$

where ∇ denotes the g_{ab}-derivative operator, ∇^2 the g_{ab}-wave operator, and where the curly brackets on the right denote linear combinations (possibly different combinations for each i and j), with coefficients given smooth fields on M, of the enclosed fields.

Then this system of equations has a well-posed initial-value formulation, in the following sense: Given smooth values for the F's and their first normal derivatives, and for the H's, on the spacelike 3-submanifold S_0 given by $t = 0$, there is a unique smooth solution of (A1) and (A2) on M which induces on S_0 this initial data.

The conditions of the first paragrah require essentially that the submanifolds of constant t be spacelike, transverse to n^a, and finally suitable initial-value surfaces for our system of equations. The proof is a standard iteration argument using Sobelev spaces. We merely sketch it.

For any $-1 < a < 0 < b < +1$, denote $t^{-1}[a, b]$ by $U_{a,b}$. Fix any smooth positive-definite metric on M. Denote by $W^k_{a,b}$, where k is any nonnegative integer, the Sobelev spaces of index k in $U_{a,b}$, i.e., the Banach spaces of tensor fields of various ranks defined in $U_{a,b}$, where the norm takes the integral over $U_{a,b}$ of the sum of squares, using the positive-definite metric, of the fields and their first k derivatives.[12] The solution at any point of M will depend only on the initial data in a compact subset of S_0, and the coefficients on the right in (A1) and (A2) in a compact subset of M. No generality is lost, therefore, by assuming that M is an open submanifold with compact closure of a larger manifold, that the metrics, n^a, t, the coefficients on the right in (A1) and (A2), and the initial data are induced from corresponding fields in this larger manifold, and that the conditions of the first paragraph of the theorem are there satisfied. Fix sufficiently large k.

Consider first the system which results from (A1) and (A2) by restricting to $U_{a,b}$ and replacing the right sides with given fields f_i in $W^{k-1}_{a,b}$ and h_j in $W^k_{a,b}$, respectively. Then, with our given initial data, this system has a unique solution $(F_i, H_j) \in W^k_{a,b}$.[7] Furthermore, there is a constant c, depending only on the metrics, n^a, and t, such that, for (F_i', H_j') the solution for right sides (f_i', h_j'),

$$\|F_i' - F_i\| \le c(b-a)\|f_i' - f_i\|,$$
$$\|H_j' - H_j\| \le c(b-a)\|h_j' - h_j\|, \tag{A3}$$

where the norms are in the appropriate Sobelev spaces.

We next note that, for (F_i, H_j) in $W^k_{a,b}$, the right sides of (A1) (resp., of (A2)) are in $W^{k-1}_{a,b}$ (resp., in $W^k_{a,b}$) and that the norms in these Sobelev spaces of the respective differences of these right sides for (F_i', H_j') and (F_i, H_j) do not exceed $d\,\|(F_i', H_j') - (F_i, H_j)\|$, for some constant d. Choose $(b-a)$ sufficiently small that $dc(b-a) < 1$. Denote by $\overline{W}$ the Sobelev space of (F_i, H_j) in $W^k_{a,b}$. Let ψ be the mapping, whose existence is now guaranteed, from $\overline{W}$ to $\overline{W}$ which sends (F_i, H_j) to that (F, H)-pair which solves (A1) and (A2) with the right sides evaluated on the given (F_i, H_j). Our choice of $(b-a)$ ensures that this ψ is a contraction mapping on the Banach space $\overline{W}$. Hence, there is a fixed point.

We conclude, then, that for sufficiently small $(b-a)$ [this size determined by the metrics, t, n^a, and the coefficients in (A1) and (A2)], there is a unique solution of (A1), (A2) in $W^k_{a,b}$. Repeating the argument for successively larger and smaller initial t values, there is a unique solution in M in W^k. Since k is arbitrary, there is a unique smooth solution in M.

Note added in proof: In light of R. Geroch, G.T. Horowitz, Phys. Rev. Lett. (to appear), it would perhaps have been more appropriate to include in the present definition of asymptotic simplicity the additional condition that, in the Ω-gauge in which $\nabla_a n_b = 0$, n^a be complete on I. In fact, the present perturbation, as a consequence of the vanishing of γ_{ab} and of $\Omega^{-1} n^a n^b \gamma_{ab}$ on I, preserves to first order both the completeness of n^a on I and this gauge condition.

[1]R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).
[2]R. Geroch, in *Asymptotic Structure of Spacetime*, edited by F.P. Esposito and L. Witten (Plenum, New York, 1977), p. 60, also, pp. 12, 39.
[3]E.T. Newman and R. Penrose, Proc. Roy. Soc. A **305**, 175 (1968).
[4]S.W. Hawking, Comm. Math. Phys. **25**, 152 (1972).
[5]This "dynamical stability" contrasts with what might be called the "kinematical stability" of D. Lerner and J.R. Porter, J. Math. Phys. **15**, 1416 (1974), in which one seeks a topology on a certain space of metrics with respect to which the set of asymptotically simple metrics is open.
[6]This situation may be contrasted with that at spatial infinity [Ref. 2, p. 72 or A. Ashtekar and R.O. Hansen (preprint 1977)]. No similar test is available there, since perturbations do not propagate to, and hence do not affect, spatial infinity. The result is that, on the whole, the evidence in favor of various definitions of asymptotic flatness at spatial infinity is less firm than at null, and, indeed, there still remains some controversy regarding the correct conditions for the former.
[7]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., London, 1973), pp. 222, 243.
[8]Our conventions are, for the Riemann tensor $\nabla_{[a}\nabla_{b]}K_c = \frac{1}{2}R_{abc}{}^m K_m$ and $R_{ab} = R_{amb}{}^m$ and $R = R_m{}^m$.
[9]One might have thought it necessary to allow also a perturbation of the conformal factor. This, however, is unnecessary, for a first-order change in Ω can be absorbed into γ_{ab} by gauge.
[10]That is, one can solve for a vector potential in the presence of a divergence-free current. A proof is immediate, for example, from the appendix, demanding $\nabla_a \xi^a = 0$ and using the single field $F_1{}^a = \xi^a$.
[11]In the electromagnetic case, there is available an alternative argument which is perhaps more analogous to that of the gravitational case. One introduces a vector potential as the "field," and imposes on it a suitable (the Lorentz, in the unphysical space—time) gauge condition.
[12]J. Marsden, *Application of Global Analysis in Mathematical Physics* (Publish or Perish, Boston, 1974), p. 50.

Branching rules for the subgroups of the unitary group[a)]

D. Braunschweig and K. T. Hecht

Department of Physics, University of Michigan, Ann Arbor, Michigan 48109
(Received 31 August 1977)

Expressions are given in terms of simple matrices d for the reduction of the Kronecker (outer) product of two or more irreducible representations which can be characterized by Young patterns. These are then used to obtain practical formulas for branching rules. The needed matrices d can be constructed by a very efficient recursive process.

1. INTRODUCTION

The effective use of group theoretical methods in atomic and nuclear shell model calculations requires that the many-particle states be classified according to the irreducible representations (IR) of the unitary group and its subgroups.[1] These IR may conveniently be labeled by their permutation symmetry in terms of Young patterns.[2] The knowledge of the relevant branching rules appropriate to the subgroup chain under consideration is usually needed at some step in the calculations.

Therefore, it is not surprising that considerable effort has been expended in finding practical ways of determining these branching rules, and numerous tabulations are available.[3] The plethysm of S functions has been used to obtain branching rules,[4] but except for simple cases the procedure becomes cubersome.[5]

The problem of calculating branching rules and the resolution of the Kronecker (outer) product can be simplified greatly if advantage is taken of the systematic decomposition of Young patterns into their completely symmetric or antisymmetric components. The simplicity gained thereby allows us to obtain the resolution of the Kronecker (outer) product of more than two patterns directly, which becomes an essential part in the evaluation of branching rules.

Our method is applicable to any set of subgroups whose IR can be labeled by Young patterns, and furthermore it is well suited for desk top calculations.[6]

2. NOTATION

We denote by $[\lambda]^n$ the partition

$$\lambda_1+\lambda_2+\cdots+\lambda_n=n, \quad \lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n\geq 0 \tag{1}$$

of the integer n. With this partition $[\lambda]^n$ we associate in the usual manner[7] a Young pattern with λ_1 boxes in the first row, λ_2 boxes in the second row, ..., etc. We say that a partition $[\mu]^n$ is of lower symmetry than $[\lambda]^n$ if the first nonzero difference $(\mu_t-\lambda_t)$ is negative. This allows us to order the $p(n)$ partitions of n in order of decreasing symmetry, such that if $i>j$, than $[\lambda]_i^n$ is of lower symmetry than $[\lambda]_j^n$. In what follows we denote by λ_{si}^n the number of boxes in row s of partition number i.

We denote by $[\tilde{\lambda}]_i^n$ the partition conjugate of $[\lambda]_i^n$ obtained from $[\lambda]_i^n$ by interchanging rows and columns.

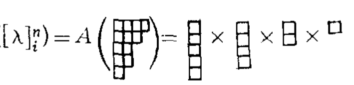
3. DECOMPOSITION OF PATTERNS

In this section we introduce the decomposition of patterns into their completely symmetric or antisymmetric components.

With each pattern $[\lambda]_i^n$ we associate an antisymmetric factor pattern $A([\lambda]_i^n)$ such that

$$A([\lambda]_i^n)=[1^{\tilde{\lambda}_{1i}}]\times[1^{\tilde{\lambda}_{2i}}]\times\cdots\times[1^{\tilde{\lambda}_{\lambda_1 i}}], \tag{2}$$

For example, if $[\lambda]_i^n=[43221]$, then

$A([\lambda]_i^n)=$  (3)

The products on the right are the Kronecker (outer) products of pattern multiplication which can be evaluated using Littlewood's rules.[7]

Given a pattern $[\lambda]_i^n$ we may set up the system of linear equations

$$A([\lambda]_j^n)=\sum_{k=i}^{p(n)} d_{jk}^{nA}[\lambda]_k^n, \quad j=i, i+1, \ldots, p(n), \tag{4}$$

or inversely,

$$[\lambda]_k^n=\sum_{j=i}^{p(n)} c_{kj}^{nA} A([\lambda]_j^n), \quad k=i, i+1, \ldots, p(n), \tag{5}$$

where

$$c_{kj}^{nA}=(-)^{k+j}\,|d_{jk}^{nA}| \tag{6}$$

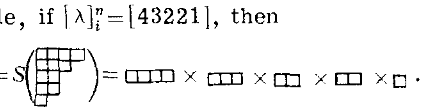
is the cofactor of the matrix element d_{jk}^{nA}.

As an example we give in Fig. 1 the matrices d_{jk}^{nS} and c_{kj}^{nS} for $n=5$.

In complete analogy with the above we can associate with each pattern $[\lambda]_i^n$ a symmetric factor pattern $S([\lambda]_i^n)$ such that

$$S([\lambda]_i^n)=[\lambda_{1i}]\times[\lambda_{2i}]\times\cdots\times[\lambda_{ni}]. \tag{7}$$

For example, if $[\lambda]_i^n=[43221]$, then

$S([\lambda]_i^n)=$  (8)

Again we may set up the system of equations

$$S([\lambda]_j^n)=\sum_{k=1}^{i} d_{jk}^{nS}[\lambda]_k^n, \quad j=1, 2, \ldots, i, \tag{9}$$

or inversely,

[a)]Work supported by U.S. National Science Foundation.

 0022-2488/78/1903-0720$1.00

$$
\begin{array}{l}
[11111] \\ [2111] \\ [221] \\ [311] \\ [32] \\ [41] \\ [5]
\end{array}
\quad
d^{5S}_{ij} = \begin{pmatrix}
1 &&&&&& \\
1 & 1 &&&&& \\
1 & 1 & 1 &&&& \\
1 & 2 & 1 & 1 &&& \\
1 & 2 & 2 & 1 & 1 && \\
1 & 3 & 3 & 3 & 2 & 1 & \\
1 & 4 & 5 & 6 & 5 & 4 & 1
\end{pmatrix}
\quad
c^{5S}_{ij} = \begin{pmatrix}
1 &&&&&& \\
-1 & 1 &&&&& \\
0 & -1 & 1 &&&& \\
1 & -1 & -1 & 1 &&& \\
0 & 1 & -1 & -1 & 1 && \\
-1 & 1 & 2 & -1 & -2 & 1 & \\
1 & -2 & -2 & 3 & 3 & -4 & 1
\end{pmatrix}
$$

FIG. 1. Matrices d^{nS}_{ij} and c^{nS}_{ij} for $n=5$.

$$[\lambda]^n_k = \sum_{j=1}^{i} c^{nS}_{kj} S([\lambda]^n_j), \quad k=1,2,\ldots,i. \tag{10}$$

If we adhere to the convention that whenever we use antisymmetric decomposition we order the patterns in order of decreasing symmetry, while when using symmetric decomposition we order them in order of increasing symmetry, we obtain the result

$$c^{nA}_{kj} = c^{nS}_{kj}, \quad \text{and} \quad d^{nA}_{jk} = d^{nS}_{jk}, \tag{11}$$

which allows us to drop the superscripts S and A from here on.

Since all our results are expressed in terms of the matrices c_{ij} and d_{ij} we summarize in the Appendix some of their properties and give as well a recursion relation for the matrix d_{ij}.

4. RESOLUTION OF THE KRONECKER (OUTER) PRODUCT

For arbitrary $[\lambda]^n_i$ and $[\lambda]^m_j$ we have

$$[\lambda]^n_i \times [\lambda]^m_j = \Big(\sum_{k=i}^{p(n)} c^n_{ik} A([\lambda]^n_k)\Big) \times \Big(\sum_{l=j}^{p(m)} c^m_{jl} A([\lambda]^m_l)\Big). \tag{12}$$

The $\lambda_{1k} + \lambda_{1l}$ factors in the product

$$[1^{\tilde{\lambda}_{1k}}] \times \cdots \times [1^{\tilde{\lambda}_{\lambda_1 k}}] \times [1^{\tilde{\lambda}_{1l}}] \times \cdots \times [1^{\tilde{\lambda}_{\lambda_1 l}}] \tag{13}$$

result from the decomposition of the pattern $[\lambda]^{n+m}_p$, where

$$\lambda^{n+m}_{sp} = \lambda^n_{sk} + \lambda^m_{sl}, \quad s=1,2,\ldots,\max(\tilde{\lambda}^n_{1k}, \tilde{\lambda}^m_{1l}). \tag{14}$$

Therefore,

$$\begin{aligned}
[\lambda]^n_i \times [\lambda]^m_j &= \sum_{k,l} c^n_{ik} c^m_{jl} A([\lambda]^{n+m}_p) \\
&= \sum_{k,l,q} c^n_{ik} c^m_{jl} d^{n+m}_{p,q} [\lambda]^{n+m}_q \\
&= \sum_q R^q_{i,j} [\lambda]^{n+m}_q,
\end{aligned} \tag{15}$$

where

$$R^q_{i,j} = \sum_{\substack{k=1,i \\ l=1,j}} c^n_{ik} c^m_{jl} d^{n+m}_{p,q} \tag{16}$$

gives the number of times the pattern $[\lambda]^{n+m}_q$ is contained in $[\lambda]^n_i \times [\lambda]^m_j$.

This result can be easily generalized to the case where there are more than two factors in the product,

$$[\lambda]^{n_1}_{i_1} \times [\lambda]^{n_2}_{i_2} \times \cdots \times [\lambda]^{n_f}_{i_f} = \sum_q R^q_{i_1 i_2 \cdots i_f} [\lambda]^{n_1+n_2+\cdots+n_f}_q, \tag{17}$$

where

$$R^q_{i_1 i_2 \cdots i_f} = \sum_{k_1 k_2 \cdots k_f} c^{n_1}_{i_1 k_1} c^{n_2}_{i_2 k_2} \cdots c^{n_f}_{i_f k_f} d^{n_1+n_2+\cdots+n_f}_{pq}. \tag{18}$$

It should be noted that the use of Eq. (18) for $R^q_{i_1 i_2 \cdots i_f}$ involves considerably less computation than the repeated use of Eq. (16) if the number of factors in Eq. (17) is greater than two.

5. BRANCHING RULES

The use of canonical subgroup chains greatly simplifies the calculation of branching rules. For example, the branching rules for the chain

$$\mathrm{U}(N) \supset \mathrm{U}(N-1) \supset \cdots \supset \mathrm{U}(1) \tag{19}$$

are given by the "betweenness" conditions[8] of Weyl's branching theorem. However the physically relevant subgroup chains are determined by the symmetries of the physical problem and seldom coincide with a canonical chain. Therefore, we consider below the branching rule problem for an arbitrary subgroup chain.

Let G and H be two arbitrary subgroups of the full linear group such that $G \supset H$. To obtain the branching rules for $G \supset H$ we shall assume that the IR's of H contained in the totally antisymmetric IR of G are known; that is, we assume that the coefficient α_{gh} are known in

$$[1^g] = \sum_h \alpha_{gh} [\lambda]'_h, \tag{20}$$

where $[1^g]$ is a totally antisymmetric IR of G and $[\lambda]'_h$ are IR of H.

Then

$$\begin{aligned}
[\lambda]^n_i &= \sum_{j=i}^{p(n)} c^n_{ij} A([\lambda]^n_j) \\
&= \sum_j c^n_{ij} [1^{\tilde{\lambda}_{1j}}] \times [1^{\tilde{\lambda}_{2j}}] \times \cdots \times [1^{\tilde{\lambda}_{\lambda_1 j}}] \\
&= \sum_j c^n_{ij} \sum_{h_1 h_2 \cdots h_{\lambda_1}} \alpha_{\tilde{\lambda}_{1j} h_1} \cdots \alpha_{\tilde{\lambda}_{\lambda_1 j} h_{\lambda_1}} [\lambda]'_{h_1} \\
&\quad \times \cdots \times [\lambda]'_{h_{\lambda_1}}.
\end{aligned} \tag{21}$$

Introducing Eq. (17) into Eq. (21) we get

$$[\lambda]_i^n = \sum_q S_i^q [\mu]_q' , \tag{22}$$

where

$$S_i^q = \sum_{\substack{j \\ h_1 h_2 \cdots h_{\lambda_1}}} c_{ij}^n \tilde{\alpha}_{\lambda_{1j} h_1} \cdots \alpha_{\lambda_{\lambda_1 j} h_{\lambda_1}} R_{h_1 h_2 \cdots h_{\lambda_1}}^q \tag{23}$$

gives the number of times the IR $[\mu]_q'$ of H is contained in the IR $[\lambda]_i^n$ of G.

In general, Eq. (22) can be simplifed further by using modification rules appropriate to the subgroup H.

6. SPECIAL CASES

A. Two column patterns

Assume that $[\lambda]_i^n$ is such athat $\tilde{\lambda}_{si}^n = 0$ for $s > 2$ while $\tilde{\lambda}_{1i}^n \neq 0 \neq \tilde{\lambda}_{2i}^n$. It is clear that in this case

$$d_{ij}^n = 0 \quad \text{for } j > i \quad \text{and} \quad d_{ij}^n = 1 \quad \text{for } j \leq i, \tag{24}$$

which in turn means that

$$c_{ij}^n = (-)^{i+j} |d_{ji}^n| = \delta_{ij} - \delta_{i+1j}. \tag{25}$$

Introducing this into Eq. (5) we get

$$[\lambda]_i^n = [1^{\tilde{\lambda}_{1i}}] \times [1^{\tilde{\lambda}_{2i}}] - [1^{\tilde{\lambda}_{1i}+1}] \times [1^{\tilde{\lambda}_{2i}-1}]. \tag{26}$$

This formula is useful in atomic spectroscopy where the electron being a fermion of spin $\frac{1}{2}$ requires that the spatial symmetry of a many-electron wavefunction be described by a two-columned pattern.

B. Two-rowed patterns

Assume that $[\lambda]_i^n$ is such that $\lambda_{si}^n = 0$ for $s > 2$ while $\lambda_{1i}^n \neq 0 \neq \lambda_{2i}^n$. It is clear that for this case Eqs. (24) and (25) hold. Introducing Eq. (25) into Eq. (10) we get

$$[\lambda]_i^n = [\lambda_{1i}] \times [\lambda_{2i}] - [\lambda_{1i} + 1] \times [\lambda_{2i} - 1]. \tag{27}$$

This formula is useful when dealing with SU(3) since all the patterns in this case are at most two rowed. In Elliot's language[9]

$$(\lambda \mu) = (\lambda + \mu, 0) \times (\mu, 0) - (\lambda + \mu + 1, 0) \times (\mu - 1, 0). \tag{28}$$

The product of two SU(3) representations is easily expressed in terms of the appropriate d matrices,

$$\begin{aligned} R_{i,j}^q &= \sum_{\substack{k=1,i \\ l=1,j}} c_{ik}^n c_{jl}^m d_{p,q}^{n+m} \\ &= \sum_{k,l} (\delta_{ik} - \delta_{i+1,k})(\delta_{jl} - \delta_{j+1,l})\, d_{p,q}^{n+m} \\ &= d_{p(i,j)q}^{n+m} + d_{p(i+1,j+1)q}^{n+m} - d_{p(i+1,j)q}^{n+m} - d_{p(i,j+1)q}^{n+m}. \end{aligned} \tag{29}$$

APPENDIX

We summarize below the main properties of the matrices c_{ij}^n and d_{ij}^n:

(1) $c_{ij}^n = (-)^{i+j} |d_{ji}^n|$, (A1)

(2) $d_{ij}^n = c_{ij}^n = 0$ for $j > i$, (A2)

(3) $d_{ii}^n = c_{ii}^n = 1$ for $i = 1, \ldots, p(n)$, (A3)

(4) $d_{ij}^n \geq 0$ for $i, j = 1, \ldots, p(n)$, (A4)

(5) $\mathrm{Tr}(D) = \mathrm{Tr}(C) = p(n)$, (A5)

(6) $\mathrm{Det}(D) = \mathrm{Det}(C) = 1$, (A6)

(7) $\sum_{j=1}^{p(n)} c_{ij}^n = 0$ for $i = 1, 2, \ldots, [p(n) - 1]$, (A7)

(8) both the matrix elements c_{ij}^n and d_{ij}^n are integers.

The matrix d_{ij}^n may be obtained directly from its definition in Eqs. (4) and (9); however, in practice it is more convenient to build the necessary matrix elements from the recursion relation we give below. This is particularly important for the matrix d_{ij}^n since normally only a few matrix elements are needed.

Let $[\lambda]_i^n$ and $[\lambda]_j^m$ be partitions of the integers n and m respectively, with say $n > m$. If

$$(\lambda_{si}^n - \lambda_{sj}^m) \geq 0, \quad s = 1, \ldots, n. \tag{A8}$$

then the set of positive integers $(\lambda_{si}^n - \lambda_{sj}^m)$ form a partition of the integer $n - m$ which we denote by $[\lambda_i - \lambda_j]^{n-m}$. For $n - m = 3$, e.g., the numbers $\lambda_{si}^n - \lambda_{sj}^m = 210000$ or 020010 both define the same partition $[21]_{k=2}^3$. We define

$$H([\lambda]_i^n, [\lambda]_j^m, [\lambda]_k^{n-m}) = \begin{cases} 1 & \text{if } [\lambda]_k^{n-m} = [\lambda_i - \lambda_j]^{n-m}, \\ 0 & \text{otherwise}, \end{cases} \tag{A9}$$

then the following recursion relation holds,

$$d_{jk}^n = \sum_{j'=1}^{x} H([\lambda]_k^n, [\lambda]_{j'}^{n-\tilde{\lambda}_{1j}}; [1^{\tilde{\lambda}_{1j}}])\, d_{xj'}^{n-\tilde{\lambda}_{1j}}, \tag{A10}$$

where x is the index of the partition of $n - \tilde{\lambda}_{1j}$ obtained by removing the first column of partition number j. If x labels a one- or two-columned pattern, the recursive process has come to an end since then clearly $d_{xj'}^{n-\tilde{\lambda}_{1j}} = 1$.

For most applications in nuclear physics where the space symmetry is restricted to patterns of four columns or less the calculation of d_{jk}^n using (A10) will require at most two steps.

[1]K.T. Hecht, Ann. Rev. Nucl. Sci. **23**, 123 (1973).
[2]M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1964), Chap. 10.
[3]B.G. Wybourne, *Symmetry Principles in Atomic Spectroscopy* (Addison-Wesley, Reading, Mass., 1964); E.M. Ibrahim, Roy. Soc. Dep. of Math. Tables, No. 1 (1950); D.E. Littlewood, J. Lond. Math. Soc. **11**, 49 (1936); R. Perez and J. Flores, Nucl. Data Tables A **4**, 265 (1968).
[4]K.V. Krishna Brahmam and R.D. Ratna Raju, Atomic and Nucl. Data Tables **16**, 165 (1975).
[5]V.K.B. Kota, J. Phys. A **10**, L39 (1977).
[6]Univ. of Michigan internal report No. 257.
[7]D.E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford U.P., Oxford, 1940).
[8]W.J. Holman and L.C. Biedenharn, in *Group Theory and its Applications*, edited by E.M. Loebl (Academic, New York, 1971), Chap. 1.
[9]J.P. Elliot, Proc. Roy. Soc. A **245**, 128 (1958).

Self-gravitating irrotational barotropic fluid in a conformally flat space

Dipankar Ray

Department of Physics, New York University, New York, New York 10003
(Received 16 January 1977)

Solutions to Einstein's equation for a conformally flat metric $e^{\psi}(dx^{1\,2}+dx^{2\,2}+dx^{3\,2}-dx^{4\,2})$ are sought for a self-gravitating irrotational barotropic fluid. It is found that velocity potential ϕ can be expressed in one of the two forms $x^{1\,2}+x^{2\,2}+x^{3\,2}-x^{4\,2}$ or x^4, and ψ, p, and ρ are functions of ϕ.

1. INTRODUCTION

Exact solutions to Einstein's equations in general relativity for self-gravitating irrotational fluids have been obtained by several authors; Taub and Tabensky[1] for a plane-symmetric metric, Letelier and Tabensky[2] for the Einstein–Rosen metric and Letelier[3] and the present author[4] for Marder's metric. However in all these cases, the equation of state used has been a somewhat artificial one, namely, $p=\rho c^2$ and the extension to other physically interesting equations of state have not yet been possible.

In the present note, we seek exact solutions to Einstein's equations for an irrotational fluid with a conformally flat metric; partly because in this case considerable progress can be made under a very general assumption that the fluid is barotropic (i.e., pressure is a function of density only), but also because conformally flat metrics have cosmological interest.[5]

2. FIELD EQUATIONS

We write the conformally flat metric as

$$ds^2=e^{\psi}(dx^{1^2}+dx^{2^2}+dx^{3^2}-dx^{4^2}),$$

i.e.,

$$g_{\mu\nu}=e^{\psi}\begin{bmatrix}1&&&\\&1&&\\&&1&\\&&&-1\end{bmatrix}\equiv e^{\psi}\eta_{\mu\nu}\quad\text{(say)}. \tag{1}$$

For fluid with pressure p, density ρ, and velocity v_μ, the field equations are

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=-(\rho+p)v_\mu v_\nu-pg_{\mu\nu}. \tag{2}$$

For an irrotational fluid there exists σ and ϕ such that

$$v_\mu=\sigma\phi_{,\mu}.$$

Also, as always, (3)

$$v_\mu v^\mu=-1.$$

For Eq. (3), Eq. (2) can be reduced to

$$R_{\mu\nu}=-(\rho+p)\sigma^2\phi_{,\mu}\phi_{,\nu}-\tfrac{1}{2}(\rho-p)g_{\mu\nu}. \tag{4}$$

For the metric (1), $R_{\mu\nu}$ is given by

$$R_{\mu\nu}=\psi_{,\mu,\nu}-\tfrac{1}{2}\psi_{,\mu}\psi_{,\nu}+\tfrac{1}{2}\eta_{\mu\nu}\chi,$$

where (5)

$$\chi=\psi_{,1,1}+\psi_{,2,2}+\psi_{,3,3}-\psi_{,4,4}+\psi_{,1}^2+\psi_{,2}^2+\psi_{,3}^2-\psi_{,4}^2.$$

From (1), (4), and (5)

$$-2e^{\psi/2}(e^{-\psi/2})_{,\mu,\nu}=-(\rho+p)\sigma^2\phi_{,\mu}\phi_{,\nu}\quad\text{for}\quad\mu\neq\nu. \tag{6}$$

From (6) we note that the derivatives of $(e^{-\psi/2})_{,4}$ with respect to x^1, x^2, and x^3 are proportional to the derivatives of ϕ with respect to x^1, x^2, and x^3. Thus if x^4 is treated as a constant, $(e^{-\psi/2})_{,4}$ and ϕ are functionally dependent[6] and since ϕ is not a constant, $(e^{-\psi/2})_{,4}$ is a function of ϕ if x^4 is treated as a constant; in other words,

$$(e^{-\psi/2})_{,4}=\delta(x^4,\phi)\text{ where }\delta\text{ is some function.}$$

Similarly

$$\left.\begin{aligned}&(e^{-\psi/2})_{,1}=\alpha(x^1,\phi),\\&(e^{-\psi/2})_{,2}=\beta(x^2,\phi),\\&(e^{-\psi/2})_{,3}=\gamma(x^3,\phi),\end{aligned}\right\} \tag{7}$$

where α,β,γ are some functions.

From (6) and (7)

$$\begin{aligned}\alpha_\phi&=\tfrac{1}{2}(\rho+p)\sigma^2e^{-\psi/2}\phi_{,1},\\\beta_\phi&=\tfrac{1}{2}(\rho+p)\sigma^2e^{-\psi/2}\phi_{,2},\\\gamma_\phi&=\tfrac{1}{2}(\rho+p)\sigma^2e^{-\psi/2}\phi_{,3},\\\delta_\phi&=\tfrac{1}{2}(\rho+p)\sigma^2e^{-\psi/2}\phi_{,4},\end{aligned} \tag{8}$$

where

$$\alpha_\phi\equiv\left.\frac{\partial\alpha}{\partial\phi}\right|_{x^1\text{ as constant}},\ \text{etc.}$$

From (1), (4), and (5)

$$\begin{aligned}&-2e^{\psi/2}(e^{-\psi/2})_{,i,i}+\tfrac{1}{2}\chi=-(\rho+p)\sigma^2\phi_{,i}^2-\tfrac{1}{2}(\rho-p)e^{\psi},\\&-2e^{\psi/2}(e^{-\psi/2})_{,4,4}-\tfrac{1}{2}\chi=-(\rho+p)\sigma^2\phi_{,4}^2+\tfrac{1}{2}(\rho-p)e^{\psi}.\end{aligned} \tag{9}$$

Using (7), (8), and (9)

$$\alpha_{x^1}=\beta_{x^2}=\gamma_{x^3}=-\delta_{x^4}=\frac{e^{-\psi/2}}{4}[(\rho-p)e^{\psi}+\chi], \tag{10}$$

where

$$\alpha_{x^1}\equiv\left.\frac{\partial\alpha}{\partial x^1}\right|_{\phi\text{ as constant}}\text{ and so on.}$$

However, α_{x^1} is a function of ϕ and x^1; β_{x^2} is a function of ϕ and x^2, etc. Thus (10) is possible only if α_{x^1}, β_{x^2}, γ_{x^3}, and $-\delta_{x^4}$ are all equal to a function of ϕ alone. Thus

$$\alpha=x^1v+p,\quad\beta=x^2v+q,\quad\gamma=x^3v+r,\quad\delta=-x^4v+s, \tag{11}$$

where v, p, q, r, and s are functions of ϕ and $v=(e^{-\psi/2}/4)[(\rho-p)e^{\psi}+\chi]$.

From (8) and (11)

$$\frac{\phi_{,1}}{x^1 v_\phi + p_\phi} = \frac{\phi_{,2}}{x^2 v_\phi + q_\phi}. \tag{12}$$

From (12) we see that along a $x^3 = \text{constant}$, $x^4 = \text{con-}$ stant surface, $\phi = \text{constant}$ curves are $(x^1 v_\phi + p_\phi)\,dx^1 + (x^2 v_\phi + q_\phi)\,dx^2 = 0$ curves, i.e.,

$$\frac{v_\phi}{2}(x^{1^2} + x^{2^2}) + p_\phi x^1 + q_\phi x^2 = \text{constant curves.}$$

Thus along $x^3 = \text{constant}$, $x^4 = \text{constant surface}$;

$$\frac{v_\phi}{2}(x^{1^2} + x^{2^2}) + p_\phi x^1 + q_\phi x^2 = \text{function of } \phi.$$

Similarly along $x^1 = \text{constant}$, $x^4 = \text{constant surface}$,

$$\frac{v_\phi}{2}(x^{2^2} + x^{3^2}) + q_\phi x^2 + r_\phi x^3 = \text{function of } \phi$$

and so on.

Comparing and combining

$$\frac{v_\phi}{2}(x^{1^2} + x^{2^2} + x^{3^2} - x^{4^2}) + p_\phi x^1 + q_\phi x^2 + r_\phi x^3 + s_\phi x^4 + t_\phi = 0, \tag{13}$$

where t_ϕ is some function of ϕ.

From (7), (11), and (13)

$$e^{-\psi/2} = \frac{v}{2}(x^{1^2} + x^{2^2} + x^{3^2} - x^{4^2}) + px^1 + qx^2 + rx^3 + sx^4 + t, \tag{14}$$

where $t = \int t_\phi\, d\phi$. Also from (8), (11), and (13)

$$-\frac{(\rho+p)}{2}\sigma^2 e^{-\psi/2} = \frac{v_{\phi\phi}}{2}(x^{1^2} + x^{2^2} + x^{3^2} - x^{4^2}) + (p_{\phi\phi}x^1 + q_{\phi\phi}x^2 + r_{\phi\phi}x^3 + s_{\phi\phi}x^4 + t_{\phi\phi}). \tag{15}$$

From (13), (14), and (15)

$$\begin{aligned} &e^{-\psi/2} = Px^1 + Qx^2 + Rx^3 + Sx^4 + T \equiv L_0 \quad \text{(say)} \\ &-\frac{(\rho+p)}{2}\sigma^2 e^{-\psi/2} = Ax^1 + Bx^2 + Cx^3 + Dx^4 + E \equiv L \quad \text{(say)}, \end{aligned} \tag{16}$$

where $P, Q, R, S, T, A, B, C, D, E$ are functions of ϕ, given by

$$\begin{aligned} &P = p - \frac{vp_\phi}{v_\phi}, \quad Q = q - \frac{vq_\phi}{v_\phi}, \quad R = r - \frac{vr_\phi}{v_\phi}, \\ &S = s - \frac{vs_\phi}{v_\phi}, \quad T = t - \frac{vt_\phi}{v_\phi}, \quad A = p_{\phi\phi} - \frac{v_{\phi\phi}p_\phi}{v_\phi}, \\ &B = q_{\phi\phi} - \frac{v_{\phi\phi}q_\phi}{v_\phi}, \quad C = r_{\phi\phi} - \frac{v_{\phi\phi}r_\phi}{v_\phi}, \\ &D = s_{\phi\phi} - \frac{v_{\phi\phi}s_\phi}{v_\phi}, \quad E = t_{\phi\phi} - \frac{v_{\phi\phi}t_\phi}{v_\phi}. \end{aligned} \tag{17}$$

From (8),

$$(\rho+p)^2\sigma^4 e^{-\psi}(\phi_{,1}^2 + \phi_{,2}^2 + \phi_{,3}^2 - \phi_{,4}^2) = 4(\alpha_\phi^2 + \beta_\phi^2 + \gamma_\phi^2 - \delta_\phi^2). \tag{18}$$

Using (11) and (13), it is easy to see that the right-hand side of (18) is a function of ϕ alone. Also from (1) and (3)

$$\sigma^2 e^{\psi}(\phi_{,1}^2 + \phi_{,2}^2 + \phi_{,3}^2 - \phi_{,4}^2) = -1.$$

Thus, from (18) we get

$$(\rho + p)\,\sigma = \eta(\phi), \tag{19}$$

where η is some function.

From (16) and (19)

$$\sigma = -\frac{2L}{\eta L_0}, \quad (\rho + p) = -\frac{\eta^2 L_0}{2L}. \tag{20}$$

Also from (2), (3), and Bianchi identities

$$\frac{\{(\rho+p)\sigma^2\phi_{,\nu}\}^{;\nu}}{\rho+p}\,\phi_{,\mu} = \left|\int \frac{dp}{\rho+p} - \log\sigma\right|_{,\mu}. \tag{21}$$

We have used $\phi_{,\nu;\mu}\phi^{,\nu} = \frac{1}{2}(\phi_{,\nu}\phi^{,\nu})_{,\mu}$, $\int dp/(\rho+p)$ is meaningful since p and ρ are functionally related.

From (21)

$$\int \frac{dp}{\rho+p} - \log\sigma = \xi(\phi), \tag{22}$$

where ξ is some function.

From (20), (22), and $p = p(\rho)$ we get that $-\eta^2/2LL_0^3$ and $-(2LL_0/\eta)\,e^\xi$ are functionally related. Let

$$-\frac{\eta^2 L_0}{2L} = \Phi\left(-\frac{2L}{\eta L_0}e^\xi\right). \tag{23}$$

Equations (13) and (23) are two relations between x^1, x^2, x^3, x^4, and ϕ. In fact one can get four relations among x^1, x^2, x^3, x^4 and by differentiating (23) with respect to x^1, x^2, x^3, x^4 and replacing in those equations, $\phi_{,1}$, $\phi_{,2}$, $\psi_{,1}$, $\psi_{,2}$, etc., by using (8), (7), (9), and (16). However since x^1, x^2, x^3, x^4 are independent of each other there cannot be more than one independent relation between x^1, x^2, x^3, x^4, and ϕ. Thus comparing the relations obtained by differentiating (23) with relation (13) it is easy to see that (23) can be true only if identically true and we get

$$\frac{L}{L_0} = \zeta(\phi),$$

i.e.,

$$\zeta = \frac{p_{\phi\phi} - v_{\phi\phi}p_\phi/v_\phi}{p - vp/v} = \frac{q_{\phi\phi} - v_{\phi\phi}q_\phi/v_\phi}{q - vq_\phi/v_\phi} = \frac{r_{\phi\phi} - v_{\phi\phi}r_\phi/v_\phi}{r - vr_\phi/v_\phi} = \frac{s_{\phi\phi} - v_{\phi\phi}s_\phi/v_\phi}{s - vs_\phi/v_\phi} = \frac{t_{\phi\phi} - v_{\phi\phi}t_\phi/v_\phi}{t - vt_\phi/v_\phi},$$

where ζ is some function of ϕ; (24)

which can be rewritten as

$$\frac{v_{\phi\phi} - \zeta v}{v_\phi} = \frac{p_{\phi\phi} - \zeta p}{p_\phi} = \frac{q_{\phi\phi} - \zeta q}{q_\phi} = \frac{r_{\phi\phi} - \zeta r}{r_\phi} = \frac{s_{\phi\phi} - \zeta s}{s_\phi} = \frac{t_{\phi\phi} - \zeta s}{t_\phi} = \lambda \quad \text{(say)}.$$

Thus v, p, q, r, s are solutions of an ordinary linear homogeneous second order differential equation and hence at most two of them are independent. If only one of them is independent

$$v = Kg, \quad p = K_1 g, \quad q = K_2 g, \quad r = K_3 g, \quad s = K_4 g, \tag{25}$$

where g is a function of ϕ satisfying the same linear homogeneous differential equation and K, K_1, K_2, K_3, and K_4 are constants.

By (11), (16), (17), (24), and (25), Eqs. (7) and (8) reduce to

$$-\tfrac{1}{2}L_0\psi_{,\mu}=gG_{,\mu}, \quad L_0\zeta(\phi)\,\phi_{,\mu}=g_\phi G_{,\mu},$$

where

$$G=\frac{K}{2}(x^{1^2}+x^{2^2}+x^{3^2}-x^{4^2}) + (K_1x^1+K_2x^2+K_3x^3+K_4x^4). \tag{26}$$

Thus

$$\psi=\psi(G), \quad \phi=\phi(G). \tag{27}$$

Now if possible let two of v, p, q, r, and s be independent, so let

$$v=Kg+K'h, \quad p=K_1g+K_1'h, \quad q=K_2g+K_2'h, \quad r=K_3g+K_3'h, \quad s=K_4g+K_4'h, \tag{28}$$

where g and h are two independent solutions of the same differential equation and K, K', K_1, K_1', K_2, K_2', K_3, K_3', K_4, and K_4' are constants. From (11), (16), (17), (24), and (28), Eqs. (7) and (8) reduce to

$$-\tfrac{1}{2}L_0\psi_{,\mu}=gG_{,\mu}+hH_{,\mu}, \quad L_0\zeta(\phi)\,\phi_{,\mu}=g_\phi G_{,\mu}+h_\phi H_{,\mu}, \tag{29}$$

where G is given by (26) and

$$H=\frac{K'}{2}(x^{1^2}+x^{2^2}+x^{3^2}-x^{4^2}) + (K_1'x^1+K_2'x^2+K_3'x^3+K_4'x^4).$$

From (29)

$$-L_0\left(\tfrac{1}{2}\psi+\int\frac{h\zeta}{h_\phi}\,d\phi\right)_{,\mu}=\left(g-\frac{hg_\phi}{h_\phi}\right)G_{,\mu}.$$

Thus

$$\tfrac{1}{2}\psi+\int\frac{\zeta h}{h_\phi}\,d\phi=f(G), \quad -\frac{g-hg_\phi/h_\phi}{L_0}=f_G(G). \tag{30}$$

From (16), (17), and (30)

$$-\left(g-\frac{hg_\phi}{h_\phi}\right)\exp\left[-\int(\zeta h/h_\phi)\,d\phi\right]=f_G(G)\exp[-f(G)]. \tag{31}$$

However, the left-hand side of (30) is a function of ϕ and the right-hand side of (31) is a function of G. Therefore, from (26) and (31), we get (27). Thus (27) is true either way. [The fact that (27) is true also means that only one of v, p, q, r, s is independent.]

From (3) we note that the velocity potential ϕ is undetermined to the extent that any arbitrary function of ϕ can also serve as velocity potential provided σ is adjusted accordingly. Thus without loss of generality we set

$$\phi=K(x^{1^2}+x^{2^2}+x^{3^2}-x^{4^2})+K_1x^1 + K_2x^2+K_3x^3+K_4x^4 \tag{32}$$

and

$$\psi=\psi(\phi), \tag{33}$$

where as before K, K_1, K_2, K_3, and K_4 are constants.

Case 1: $K\neq 0$

Here, without loss of generality, we can set

$$\phi=x^{1^2}+x^{2^2}+x^{3^2}-x^{4^2} \tag{34}$$

and using (1), (3), and (33), we get

$$4\phi\sigma^2\exp(-\psi)=-1.$$

Therefore, we must have

$$x^{1^2}+x^{2^2}+x^{3^2}-x^{4^2}<0 \tag{I}$$

and the field equations reduce to

$$\phi(\psi_{\phi\phi}-\tfrac{1}{2}\psi_\phi^2)=\frac{e^\psi}{4}(p+\rho), \quad \phi(\psi_{\phi\phi}+\tfrac{1}{2}\psi_\phi^2)+3\psi_\phi=\frac{e^\psi}{4}(p-\rho). \tag{35}$$

The condition that $\rho>0$ gives is

$$\phi\psi_\phi^2+2\psi_\phi<0. \tag{II}$$

The condition that $p>0$, i.e., there is pressure, rather than tension, gives

$$2\phi\psi_{\phi\phi}+3\psi_\phi>0. \tag{III}$$

Case 2: $K=0$

Using (1), (3), and (33),

$$\sigma^2\exp(-\psi)(K_1^2+K_2^2+K_3^2-K_4^2)=-1.$$

Thus

$$K_1^2+K_2^2+K_3^2-K_4^2<0.$$

Therefore, without loss of generality, we can set

$$K_1=0, \quad K_2=0, \quad K_3=0, \quad K_4=1,$$

i.e.,

$$\phi=x^4 \tag{36}$$

and the field equations reduce to

$$\psi_{\phi\phi}-\tfrac{1}{2}\psi_\phi^2=-e^\psi(p+\rho), \quad \psi_{\phi\phi}+\psi_\phi^2=-e^\psi(p-\rho). \tag{37}$$

$\rho>0$ is automatically satisfied and $p>0$ gives

$$\psi_{\phi\phi}+\tfrac{1}{4}\psi_\phi^2<0. \tag{IV}$$

3. CONCLUSION

Thus any solution of (2) and (3) for metric (1) is of the form (33), where ϕ is given by (34) or (36).

In Case 1, when ϕ is given by (34), the field equations reduce to (35) and the inequalities (I), (II), and (III) are to be satisfied for a fluid with $\rho>0$, $p>0$. Also, in this case p and ρ are functions of $x^{1^2}+x^{2^2}+x^{3^2}-x^{4^2}$.

This leads to an interesting model of the universe or a star, where at any value of x^4, the entire system is confined in a "sphere" $x^{1^2}+x^{2^2}+x^{3^2}=x^{4^2}$. For $x^4<0$, this "sphere" keeps on shrinking, until at $x^4=0$, the system collapses to a point $x^1=0=x^2=x^3$. For $x^4>0$, the system keeps on expanding. The metric here admits six Killing vectors $(0,x^3,-x^2,0)$, $(x^2,-x^1,0,0)$,

$(x^3, 0, -x^1, 0)$, $(x^4, 0, 0, -x^1)$, $(0, x^4, 0, x^2)$, and $(0, 0, x^3, x^4)$ corresponding to six rotations. The metric is however not invariant under any of the four translations.

In Case 2, when ϕ is given by (36), the metric is of the well-known Robertson—Walker form and hence need not be discussed in detail. Here also, the metric admits six Killing vectors, which are $(0, x^3, x^2, 0)$, $(x^2, -x^1, 0, 0)$, $(x^3, 0, -x^1, 0)$, $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, and $(0, 0, 1, 0)$, and correspond to three translations and three rotations in the (x^1, x^2, x^3) space.

It may also be pointed out that both the metrics discussed here belong to a category of metrics that Petrov[7] has called Kagan subprojective space.

The author thanks the referee for suggesting the need for studying symmetry and physical significance.

[1] R. Tabensky and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).
[2] P. S. Letelier and R. Tabensky, Nuovo Cimento B **28**, 407 (1975).
[3] P. S. Letelier, J. Math. Phys. **16**, 1488 (1975).
[4] D. Ray, J. Math. Phys. **17**, 1171 (1976).
[5] J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), p. 322.
[6] It is a well known result of the theory of functions that if two functions of several variables have their derivatives proportional to each other, then the two functions are functionally dependent.
[7] A. A. Petrov, *Einstein Space* (Pergamon, New York, 1969), p. 252.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*