Covariant n^2-plet mass formulas

A. Davidson

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 3 February 1978)

Using a generalized internal symmetry group analogous to the Lorentz group, we have constructed a covariant n^2-plet mass operator. This operator is built as a scalar matrix in the $(n;n^*)$ representation, and its SU(n) breaking parameters are identified as intrinsic boost ones. Its basic properties are: covariance, Hermiticity, positivity, charge conjugation, quark contents, and a self-consistent n^2-1, 1 mixing. The GMO and the Okubo formulas are obtained by considering two different limits of the same generalized mass formula.

1. INTRODUCTION

In a previous article[1] we presented a possible way to complete the analogy between the external and the internal symmetry groups. We made the hypothesis that the intrinsic SU(n) spin plays, in an internal n^2-dimensional space, a similar role as the ordinary **S** spin does in the Lorentzian space. Following this viewpoint, a generalized SU(n)-based internal symmetry was proposed for elementary particles, such that "broken SU(n)" is automatically contained in it. In this paper, however, we would like to present a modified mass operator which generalizes traditional ones, and has a covariant structure under the so-called hyper-Lorentz group. We have derived it by adopting a former technique,[2] originally used by Weinberg in the construction procedure of covariant Lorentzian propagators.

In order to clarify our motivation, let us summarize the principal properties of the generalized group. In analogy to the Lorentz transformation, an internal hyper-Lorentz transformation R was introduced. This transformation satisfies the fundamental relation

$$g_{\mu_1\cdots\mu_n}R^{\mu_1}_{\nu_1}\cdots R^{\mu_n}_{\nu_n}=g_{\nu_1\cdots\nu_n}, \tag{1.1}$$

and conserves

$$\det\left(\left(\frac{2}{i\hbar}\right)^{1/2}x^0 I+\mathbf{x}\cdot\lambda\right)=g_{\mu_1\cdots\mu_n}x^{\mu_1}\cdots x^{\mu_n}=\text{invariant}, \tag{1.2}$$

for any arbitrary n^2-vector x^μ ($\mu=0, 1, \ldots, n^2-1$). In (1.2) λ are the SU(n) generators. I is the identity, and g is the nth-order "metric tensor" built of the SU(n) symmetric structure constants. The representation corresponding to an infinitesimal transformation $R\simeq I+\omega$, has the form

$$U(1+\omega)\simeq I+i\boldsymbol{\alpha}\cdot\mathbf{F}+i\boldsymbol{\beta}\cdot\mathbf{B}. \tag{1.3}$$

We can identify, apart from the SU(n) rotation generators **F**, another n^2-1 intrinsic boost generators **B**, with commutation relations

$$\begin{aligned}[F_i, F_j]&=+if_{ijk}F_k,\\ [F_i, B_j]&=+if_{ijk}B_k,\\ [B_i, B_j]&=-if_{ijk}F_k,\end{aligned} \tag{1.4}$$

which follow from the group property $U(R_2)U(R_1)=U(R_2R_1)$.

The generators $\frac{1}{2}(\mathbf{F}\pm i\mathbf{B})$ form an SU(n)$\times$SU†(n) algebra, with irreducible representations of the $(a;b)$ type, where a, b are two arbitrary SU(n) representations.[4] In particular, covariant and contravariant vectors transform according to the $(n^*;n)$ and (n,n^*) representations, respectively. These vectorial representations are of a great importance to us, since they can be naturally connected with mesons. Notice that each of these representations consists of the $n^2-1\oplus 1$ SU(n) representations which stay unmixed as long as only pure internal rotations are considered. On the other hand, *any boost induces mixing and destroys the rotational invariants, and therefore can be interpreted as the geometrical*[3] *mechanism of the* SU(n) *breaking*. Furthermore, by completing the analogy to the real space, we define an "intrinsic rest frame," where one expects the conventional SU(n) to be a little group. In any other frame, characterized by the hypermomentum p_μ, there exists a boost penetration into the little-group generators. Thus, p_μ expresses by its magnitude and direction the SU(n) breaking.

In this paper, we have used the above arguments to construct an n^2-plet mass operator built as a scalar matrix under the generalized internal symmetry group. Among its properties we can find: covariance, hermiticity, positivity, charge conjugation, quark contents, and mixing of representations. But the most important one is the possibility to obtain the GMO and the Okubo formulas by considering two different limits of the same generalized mass formula.

2. THE n^2-PLET MASS OPERATOR AND ITS FORMAL PROPERTIES

In the "exact symmetry limit" one expects mass degeneracy within any given SU(n) unitary multiplet, in our language it is to say that the mass operator should be an SU(n) rotational invariant only in the so-called "intrinsic rest frame." This viewpoint naturally suggests the construction of such an operator as a scalar matrix under the internal hyper-Lorentz transformation.

A scalar matrix $\pi(p)$ is defined by the following transformation law,

$$D_{(a;b)}[R]\cdot\pi_{(a;b)}(p)\cdot D^\dagger_{(a;b)}[R]=\pi_{(a;b)}(Rp), \tag{2.1}$$

where $D_{(a;b)}[R]$ is the hyper-Lorentz transformation matrix in the $(a;b)$ representation, and p stands for the hypermomentum p_μ ($\mu=0, 1, \ldots, n^2-1$) describing the magnitude and direction of the unitary symmetry break-

 0022-2488/79/050757-05$01.00

ing. In Appendix A we show that the most general $\pi(p)$ is of the form

$$\pi(p)=\pi(R^0p)=D[R]\{\sum_i c_i\delta_i\}D^\dagger[R], \tag{2.2}$$

where the summation index i runs over all the SU(n) representations contained in the $a\otimes b$ multiplication, and

$$\delta_i=\begin{cases}1: & \text{for the representation } i,\\ 0: & \text{otherwise,}\end{cases}$$

c_i-arbitrary scale coefficients

${}^0p_\mu\sim(1,\ 0)$.

As indicated in the Introduction, one can easily show that

$$D[R]=\exp[i\boldsymbol{\alpha}\cdot(\mathbf{F}_a+\mathbf{F}_b)] \text{ if } R \text{ is a pure rotation,} \tag{2.3a}$$

$$D[R]=\exp[\boldsymbol{\beta}\cdot(\mathbf{F}_a-\mathbf{F}_b)] \text{ if } R \text{ is a pure boost.} \tag{2.3b}$$

The particular R, which transforms ${}^0p$ into p, can always be considered as a pure internal boost transformation, since ${}^0p$ stays unaffected under the SU(n) rotations. Moreover, we are free to choose the orientation of the intrinsic coordinate system in such a way that this R-transformation will be along $n-1$ principal axes (in analogy to external space, where any preferred direction can be chosen as our $\hat{z}$ axis). Thus $D[R]$ takes a simpler form

$$D[R]=\exp\left[\sum_{k=1}^{n-1}\beta_k(H^k_{(a)}-H^k_{(b)})\right], \tag{2.4}$$

where $H^k_{(d)}$ are the $n-1$ diagonal generators of SU(n) in the d representation.

In accordance with the traditional population scheme of the SU(n) representations, it is very popular to connect the mesons with states of the n^2-1, 1 mixed representations. We hereby propose a natural generalization to this naive picture *by letting the mesons transform like an n^2-vector* under the hyper-Lorentz group, i.e., via the $(n;n^*)$ representation.[5] In such a way we shall be able to derive the n^2-plet mass formulas automatically accompanied by the n^2-1, 1 mixing. Apart from the $n-1$ breaking (boost) parameters, the most general n^2-plet mass operator involves two other scale parameters, as indicated by (2.2). Choosing these parameters as m_0^2 and M_0^2, the "bare" masses of the n^2-1 and the singlet representations, respectively, we obtain the desired n^2-plet (mass)^2 operator

$$\hat{m}^2\equiv D[R]\cdot\{m_0^2\delta_{n^2-1}+M_0^2\delta_1\}\cdot D^\dagger[R]. \tag{2.5}$$

The simplest way to reach the explicit form of (2.5) is to use the $q_i\bar{q}_j$ basis. According to the conventional quark classification, given in Table I, one can easily check that

$$\sum_{K=1}^{n-1}\beta_K[H^K_{(n)}-H^K_{(n^*)}]\,|q_i\bar{q}_j\rangle=\tfrac{1}{2}(\gamma_i+\gamma_j)\,|q_i\bar{q}_j\rangle, \tag{2.6}$$

where the new coefficients γ_K are defined by

$$\tfrac{1}{2}\gamma_K\equiv\left(\frac{1}{K}-1\right)\beta_{K-1}+\sum_{l=K}^{n-1}\frac{1}{l+1}\beta l. \tag{2.7}$$

Combining formulas (2.4), (2.5), and (2.6), we can finally present our generalized version to the mesonic mass matrix,

$$\langle q_i\bar{q}_j|\,\hat{m}^2\,|q_k\bar{q}_l\rangle = m_0^2\cdot\exp(\gamma_i+\gamma_j)\cdot\delta_{ik}\delta_{jl}+\frac{1}{n}(M_0^2-m_0^2)\cdot\exp(\gamma_i+\gamma_k)\cdot\delta_{ij}\delta_{kl}. \tag{2.8}$$

Before analyzing the detailed structure of (2.8), let us first list some of its basic properties which are important by themself:

(1) *Covariance*: As indicated by the transformation law (2.1), the m^2 operator has a covariant structure under our generalized internal symmetry group. Furthermore, one can rewrite (2.5) in an alternative form which emphasizes its special form

$$\hat{m}^2=F^\mu F^\nu(k_1p_\mu p_\nu+k_2g_{\mu\nu\lambda_3\cdots\lambda_n}p^{\lambda_3}\cdots p^{\lambda_n}), \tag{2.9}$$

where $F^\mu\equiv(I/\sqrt{2n},\ \mathbf{F}_{(n)})$, $\bar{F}^\nu\equiv(-I/\sqrt{2n},\ \mathbf{F}_{(n^*)})$, k_1 and k_2 are two constants linearly related to m_0^2 and M_0^2, and the tensor g is defined by (1.2).

(2) *Hermiticity*: Notice that the real matrix (2.8) is totally symmetric

$$m^2_{ij,kl}=m^2_{kl,ij}, \tag{2.10}$$

hence the eigenmasses are real quantities.

(3) *Positivity*: One can easily prove that $m_0^2,\ M_0^2>0$ imply definitely positive eigenvalues for any arbitrary breaking parameters.

(4) *Charge conjugation*: It is only in the sense that

$$m^2_{ij,kl}=m^2_{ji,lk}. \tag{2.11}$$

This property assures that polar n^2-plet members will have equal masses.

(5) *Quark Contents*: The mass operator in the $(n;n^*)$ representation is expressed by (2.6) in terms of the quark and antiquark operators $\mathbf{F}_{(n)}$, $\mathbf{F}_{(n^*)}$. To be more precise it follows from the multiplication property $(n;n^*)=(n,1)\otimes(1;n^*)$, and one should notice that the $(n;n^*)$ representation has nothing to do with the $(n^2-1;1)$ or $(1;n^2-1)$ ones.

(6) *The "Exact Symmetry Limit"*: The exact SU(n) limit is obtained by letting the boost parameters vanish. In this limit, the degenerated "bare" mass matrix reads

$$m^2[n^2-1]\to m_0^2,\quad m^2[\text{singlet}]\to M_0^2. \tag{2.12}$$

TABLE I. The conventional quark classification.

	q_1	q_2	q_3	q_4	·	°	q_n
H^1	$\frac{1}{2}$	$-\frac{1}{2}$	0	°	°	·	0
H^2	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{2}{3}$	0	°	°	0
H^3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	0	·	0
°	°	°	·	·	·	·	·
°	°	°	·	·	·	·	·
H^{n-1}	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	·	°	$\frac{1-n}{n}$

(7) *Mixing*: The m^2 matrix (2.8) consists of a diagonal part for the $n(n-1)$ peripheral states $q_i\bar{q}_j$ $(i \neq j)$, as well as of a nondiagonal one for the other n central states $q_i\bar{q}_i$. Thus, the boost transformation induces, *interalia*, a simple mixing among the states which are located at the representation origin.

(8) *Traditional Mass Formulas*: Using the approximation where the leading breaking parameter is small enough in comparison with the $(M_0^2 - m_0^2)/m_0^2$ ratio, we approach the famous GMO mass formula. On the other hand, if M_0^2 and m_0^2 are almost identical, one is able to derive the Okubo type mass formulas for the n^2-plet. These important approximations to the same mass matrix, are considered in Sec. 4.

3. EXACT RELATIONS AMONG THE n^2-PLET MASSES

For the peripheral members of the n^2-plet, it follows directly from (2.8) that

$$m^2(q_i\bar{q}_j) = m_0^2 \exp(\gamma_i + \gamma_j) \quad (i \neq j). \tag{3.1}$$

Hence, we obtain, apart from the basic identity

$$m^2(q_i\bar{q}_j) = m^2(q_j\bar{q}_i), \tag{3.2}$$

an additional multiplicative mass relation

$$m^2(q_i\bar{q}_j) \cdot m^2(q_k\bar{q}_l) = m^2(q_i\bar{q}_l) \cdot m^2(q_k\bar{q}_j). \tag{3.3}$$

This n-independent formula turns out to be an additive one only if γ_k are first-order parameters. (3.3) is apparently useful starting $n = 4$, and for the vectorial mesons one finds

$$\rho F^* = K^* D^* = K^{*\circ} D^{*\circ}, \tag{3.4}$$

where we have used the notation ρ for m_ρ^2, etc. Notice that in a "pure quark model," where $M_0^2 = m_0^2$, (3.3) holds for all the n^2-plet members and not only for the peripheral ones.

For the central states the situation is somewhat more complicated, and the corresponding masses are the eigenvalues of the following nondiagonal matrix,

$$\begin{vmatrix} m_0^2(1+x)\exp(2\gamma_1); & m_0^2 x \exp(\gamma_1+\gamma_2); \ldots; & m_0^2 x \exp(\gamma_1+\gamma_n) \\ m_0^2 x \exp(\gamma_2+\gamma_1); & m_0^2(1+x)\exp(2\gamma_2); \ldots; & m_0^2 x \exp(\gamma_2+\gamma_n) \\ \cdots & \cdots & \cdots \\ m_0^2 x \exp(\gamma_n+\gamma_1); & m_0^2 x \exp(\gamma_n+\gamma_2) \ldots; & m_0^2(1+x)\exp(2\gamma_n) \end{vmatrix}, \tag{3.5}$$

where

$$x \equiv \frac{1}{n}\frac{M_0^2}{m_0^2} - 1 \tag{3.6}$$

measures the deviation from the "pure quark model." As it is shown in Appendix B, the eigenvalues of (3.5) obey the following secular equation,

$$\sum_{K=0}^{n} (1+Kx) \cdot S_n^K(\mu_1^2, \ldots, \mu_n^2) \cdot (-\lambda)^{n-K} = 0, \tag{3.7}$$

where the symmetric coefficients S_n^K are defined by

$$(x_1 + x) \cdot \ldots \cdot (x_1 + x) \equiv \sum_{K=0}^{n} S_n^K(x_1, \ldots, x_n) \cdot x^{n-K},$$

and $\mu_K^2 \equiv m_0^2 \cdot \exp(2\gamma_K)$ are the central masses in the $x = 0$ limit, i.e., in the "pure quark model." Notice that every μ_K^2 can be expressed in terms of the peripheral mass

$$\mu_K^2 \equiv \frac{m^2(q_K\bar{q}_i) \cdot m^2(q_K\bar{q}_j)}{m^2(q_i\bar{q}_j)} \qquad (i \neq j \neq K), \tag{3.8}$$

and therefore Eq. (3.7) leads us to another mass relations of the form

$$S_n^K(m_1^2, \ldots, m_n^2) = (1+Kx) \cdot S_n^K(\mu_1^2, \ldots, \mu_n^2), \quad K = 1, \ldots n \tag{3.9}$$

involving the physical central masses $m_1^2, \ldots, m_n^2$. By eliminating x from (3.9) one gets the following parameter free mass formulas

$$\frac{S_n^K(m_1^2, \ldots, m_n^2) - S_n^K(\mu_1^2, \ldots, \mu_n^2)}{K S_n^K(\mu_1^2, \ldots, \mu_n^2)} = x = \text{const.} \tag{3.10}$$

It should be noted here that *for an $(n-1)^2$-plet, which is contained in the n^2-plet, similar formulas to* (3.9) *and* (3.10) *do not exist.* Solving (3.9) for $S_{n-1}^K(m_1^2, \ldots, m_{n-1}^2)$, using the identity $S_n^K = S_{n-1}^K + m_n^2 S_{n-1}^{K-1}$, we get

$$S_{n-1}^K(m_1^2, \ldots, m_{n-1}^2) = \sum_{l=0}^{K} (1+lx)(-m_n^2)^{K-l} \cdot S_n^l(\mu_1^2, \ldots, \mu_n^2) \neq (1+K\tilde{x}) \cdot S_{n-1}^K(\mu_1^2, \ldots, \mu_{n-1}^2). \tag{3.11}$$

It is to emphasize that different mass formulas are obtained if we try to connect N^2 mesons to various $n \geq N$ cases. However, as we shall later see, identical and familiar mass formulas can be derived for all these cases if we consider only first order breaking of the N^2-plet. Thus, if we are interested in obtaining physically exact mass relations for N^2 mesons, we face the serious problem of determining the correct $n \geq N$. The temporary value for n is 4, but until the real value will be fixed we cannot avoid the penetration of the n^2-plet masses into the N^2-plet mass relations. Meanwhile we must content ourselves with the following approximated mass formulas.

4. THE SINGLE-BREAKING-PARAMETER APPROXIMATIONS

If we believe that the $n-1$ breaking parameters obey

$$\beta_1 \ll \beta_2 \ll \ldots \ll \beta_{n-1},$$

it is justified to consider first the approximation where only the leading parameter does not vanish, i.e.,

$$\beta_{n-1} \equiv -\tfrac{1}{4}\theta, \quad \beta_{n-2} = \ldots = \beta_1 = 0. \tag{4.1}$$

From (2.7) it follows that

$$\gamma_n = \frac{n-1}{2n}\theta, \quad \gamma_{n-1} = \ldots = \gamma_1 = -\frac{1}{2n}\theta. \tag{4.2}$$

Substituting (4.2) into the mass expressions (3.1) and (3.7), one is led to results which describe the SU(n) breaking to the smaller SU($n-1$)$\otimes$U(1) symmetry group. The n^2-plet consists of the following SU($n-1$) subrepresentations: $n-1$, $(n-1)^*$, $(n-1)^2-1$, and two singlets [only one of them is also an SU(n) singlet]. The corresponding masses are given by

$$m_S^2 \equiv m^2\{n-1\} = m^2\{(n-1)^*\} = m_0^2 \cdot \exp\left(\frac{n-2}{2n}\theta\right),$$
$$M_L^2 \equiv m^2[(n-1)^2 - 1] = m_0^2 \cdot \exp\left(-\frac{1}{n}\theta\right), \tag{4.3}$$

and by the relations

$$M^2 + m^2 = m_0^2 \exp\left(-\frac{1}{n}\theta\right)\{(1+e^\theta) + x[(n-1)+e^\theta]\},$$
$$M^2 m^2 = m_0^2 \cdot \exp\left(\frac{n-2}{n}\theta\right) \cdot (1+nx), \tag{4.4}$$

where M^2, m^2 are the eigenmasses of the two mixing singlets, and x was defined by (3.6). We are going to show that (4.3) and (4.4) give under certain conditions a variety of the traditional mass formulas.

Let us now consider some special cases of (4.3) and (4.4),

(1) $\theta \ll 1$, x: In this limit one finds, up to the first order of θ, that

$$m_S^2 \sim m_0^2\left(1 + \frac{n-2}{2n}\theta\right), \quad m_L^2 \sim m_0^2\left(1 - \frac{1}{n}\theta\right),$$
$$m^2 \sim m_0^2\left(1 + \frac{n-2}{n}\theta\right), \quad M^2 \sim M_0^2 = m_0^2(1+nx), \tag{4.5}$$

thus, the GMO mass formula

$$(n-2)m_L^2 + nm^2 = 2(n-1)m_S^2 \tag{4.6}$$

follows immediately.

(2) $x \ll 1$, θ: Here we get, up to the first order of x,

$$m_S^2 = m_0^2 \cdot \exp\left(\frac{n-2}{2n}\theta\right), \quad m_L^2 = m_0^2 \cdot \exp\left(-\frac{1}{n}\theta\right),$$
$$m^2 \sim m_0^2 \cdot \exp\left(-\frac{1}{n}\theta\right) \cdot [1 + (n-1)x],$$
$$M^2 \sim m_0^2 \cdot \exp\left(\frac{n-1}{n}\theta\right) \cdot (1+x), \tag{4.7}$$

which lead to the relation

$$\frac{1}{n-1}\left(\frac{m^2}{m_L^2} + (n-2)\right) \cdot m_S^4 = M^2 m_L^2. \tag{4.8}$$

(3) $x \ll \theta \ll 1$: In this limit

$$m_S^2 \sim m_0^2\left(1 + \frac{n-2}{2n}\theta\right), \quad m_L^2 \sim m_0^2\left(1 - \frac{1}{n}\theta\right),$$
$$m^2 \sim m_0^2\left(1 + (n-1)x - \frac{1}{n}\theta\right), \quad m^2 \sim m_0^2\left(1 + x + \frac{n-1}{n}\theta\right). \tag{4.9}$$

These masses are clearly connected by the Okubo mass formula

$$nm_L^2 + (n-1)M^2 = m^2 + 2(n-1)m_S^2. \tag{4.10}$$

ACKNOWLEDGMENTS

It is my pleasure to express deep gratitude to all the members of the particle group in the Maryland University, especially to Professor O.W. Greenberg, for the wonderful reception and hospitality.

APPENDIX A: THE GENERAL FORM OF THE SCALAR MATRIX

From the transformation law (2.1) it follows immediately that

$$\pi_{(d_1;d_2)}(p) = \pi_{(d_1;d_2)}(R\,{}^0p) = D_{(d_1;d_2)}[R] \cdot \pi_{(d_1;d_2)}({}^0p) \cdot D^\dagger_{(d_1;d_2)}[R], \tag{A1}$$

where ${}^0p \sim (1, 0)$.

If R is a pure rotation, then

$$R\,{}^0p = {}^0p, \quad D[R] = \exp[i\boldsymbol{\alpha} \cdot (\mathbf{F}_{d_1} + \mathbf{F}_{d_2})]. \tag{A2}$$

The substitution of (A2) in (A1) gives the relation

$$\pi({}^0p) = \exp[i\boldsymbol{\alpha} \cdot (\mathbf{F}_{d_1} + \mathbf{F}_{d_2})] \cdot \pi({}^0p) \cdot \exp[-i\boldsymbol{\alpha}(\mathbf{F}_{d_1} + \mathbf{F}_{d_2})], \tag{A3}$$

which is satisfied if and only if

$$[\mathbf{F}_{d_1} + \mathbf{F}_{d_2}, \ \pi({}^0p)] = 0. \tag{A4}$$

Therefore,

$$\pi({}^0p) = \sum_i c_i \delta_i, \tag{A5}$$

where the summation index i runs over all the SU(n) irreducible representations contained in the multiplication of the d_1, d_2 representations, and c_i, δ_i were defined in Sec. 2.

If R is a boost, then

$$R\,{}^0p = p, \quad D[R] = \exp[\boldsymbol{\beta} \cdot (\mathbf{F}_{d_1} - \mathbf{F}_{d_2})]. \tag{A6}$$

(A2) gives $\pi(p)$ its final form

$$\pi(p) = \exp[\boldsymbol{\beta} \cdot (\mathbf{F}_{d_1} - \mathbf{F}_{d_2})] \cdot \left\{\sum_i c_i \delta_i\right\} \exp[\boldsymbol{\beta} \cdot (\mathbf{F}_{d_1} - \mathbf{F}_{d_2})],$$

after another substitution in (A1).

APPENDIX B: THE SECULAR EQUATION FOR THE CENTRAL EIGENMASSES

$$\begin{vmatrix} m_0^2(1+x)\exp(2\gamma_1) - \lambda, & m_0^2 x \exp(\gamma_1+\gamma_2); & \cdot & m_0^2 x \exp(\gamma_1+\gamma_n) \\ m_0^2 x \exp(\gamma_2+\gamma_1); & m_0^2(1+x)\exp(2\gamma_2) - \lambda; & \cdot; & m_0^2 x \exp(\gamma_2+\gamma_n) \\ \circ \quad ; & \cdot \quad ; & \cdot; & \cdot \\ m_0^2 x \exp(\gamma_n+\gamma_1), & m_0^2 x \exp(\gamma_n+\gamma_2); & \circ; & m_0^2(1+x)\exp(2\gamma_n) - \lambda \end{vmatrix} = 0 \tag{B1}$$

can be rewritten as

$$\Delta_n = \begin{vmatrix} 1+\eta_1 & 1 & & 1 \\ 1 & 1+\eta_2 & & \vdots \\ & & \ddots & \\ 1 & & & 1+\eta_n \end{vmatrix} = 0, \tag{B2}$$

where we have defined η_K by

$$\eta_K \equiv \frac{1}{x}\,(m_0^2 - \lambda \exp(-2\gamma_K)) \equiv \frac{\exp(-2\gamma_K)}{x}(\mu_K^2 - \lambda). \tag{B3}$$

From (B2) it follows that:

$$\Delta_n = \eta_n \cdot \Delta_{n-1} + \eta_{n-1} \cdots \eta_1 = \cdots$$
$$= S_n^{n-1}(\eta_1, \ldots, \eta_n) + S_n^n(\eta_1, \ldots, \eta_n), \tag{B4}$$

$$S_n^n(\eta_1, \ldots, \eta_n) \equiv \eta_1 \cdots \eta_n$$
$$= \frac{1}{x^n} \exp[-2(\gamma_1 + \cdots + \gamma_n)] \sum_{K=0}^{n} S_n^K(\mu_1^2, \ldots, \mu_n^2)(-\lambda)^{n-K}, \tag{B5}$$

and

$$S_n^{n-1}(\eta_1, \ldots, \eta_n) = x \frac{d}{dm_0^2} S_n^n(\eta_1, \ldots, \eta_n)$$
$$= \frac{1}{x^{n-1}} \exp[-2(\gamma_1 + \cdots + \gamma_n)] \sum_{K=0}^{n} \frac{K}{m_0^2} S_n^K(\ )(-\lambda)^{n-K}. \tag{B6}$$

Thus (3.7) follows immediately.

[1] A. Davidson and A. Dar, J. Math. Phys. (to be published).
[2] S. Weinberg, Phys. Rev. **133**, 1318 (1964).
[3] For the "external" analogy see, for example S. S. Schweber, *Relativistic Quantum Field Theory*" (Harper and Row, New York, 1964).
[4] Another geometrical treatment of the SU(3) breaking was given by L. Michel, "Symmetry Principles at High-Energy *The Fifth Coral Gables Conference* (Benjamin, New York, 1968).
[5] A nonchiral quark model was considered by L. P. Horwitz, Nucl. Phys. B **34**, 136 (1971).

A critique of the major approaches to damping in quantum theory

Daniel M. Greenberger

City College of the City University of New York, New York
(Received 5 January 1977; revised manuscript received 5 April 1978)

We examine the two major approaches that have been suggested for the quantum mechanical treatment of the damped motion of a particle as a one-body problem. These are the linear, but time dependent, Kanai Hamiltonian, and the more recent nonlinear potentials which have been introduced to simulate the damping force. The most important criticism that has been leveled at the Kanai Hamiltonian is that its solutions seem to violate the uncertainty relations. We show that this Hamiltonian actually represents a particle of variable mass, whose classical behavior is identical to that of a damped particle of constant mass. But quantum mechanically, its changing mass does lead to unphysical behavior when misinterpreted as a constant mass particle. So this Hamiltonian cannot directly describe a constant mass damped quantum particle. The nonlinear model has been interpreted in terms of the hydrodynamical analogy of quantum theory, and a well behaved decaying wavepacket solution has been produced. However we generalize this result to produce solutions that "decay" to arbitrarily high energy. Thus it is not clear that this model specifically treats dissipation. Rather it seems to seek out any stationary state. At any rate, its physical interpretation is obscure at present. However we show, by analyzing the physical problem of damping at low energies, that one can modify the Kanai Hamiltonian to eliminate its unphysical features, so that this modified Kanai Hamiltonian can in fact be interpreted as representing a constant mass damped particle with physically reasonable solutions.

I. INTRODUCTION

Recently, there has been an upsurge of interest in the quantum mechanical problem of a particle subject to a damping force proportional to its velocity. The impetus has come from nuclear physics, but the problem is very interesting in its own right. An excellent, short, review article by Hasse[1] covers much of what has been done on the subject to date, and the reader is referred there for a very detailed list of references.

The reason the subject is intrinsically interesting is that one believes that at a microscopic level there is no damping, and that effectively, damping is a collective phenomenon produced by the interaction of a particle with the sea of background particles in the medium. It is the dissipation of energy to this background that causes the damping, and at the same time the particle receives energy from fluctuations in the background.

Yet classically, the end result of this many-body interaction is the existence of an effective one-body problem, namely that of the particle subject to a damping force, and a great deal of literature has been generated by the problem in statistical mechanics of producing this one-body force.[2]

Now, since this one-body force does come about, it is only natural to expect both classical and quantum theory to be able to cope with it, and that the solution should blend convicingly with the results of the many-body approach to the problem, to whatever extent the physical situations themselves overlap. This becomes especially important when one realizes that in fact a truly undamped oscillator never occurs in nature, and one must always include some damping in order to obtain realistic results near resonance. and yet it is the undamped oscillator that is easily treated as a one-body problem.

The classic approach to the one-body problem for damped motion is through Kanai's Hamiltonian[3] (in one dimension)

$$H = (p^2/2m_0)e^{-\gamma t} + V(x)e^{\gamma t}. \tag{1}$$

This Hamiltonian has received a considerable amount of attention,[4] but it produces some results that are considered unphysical, and this has led to a second approach, due originally to Kostin,[5] which replaces the linear but time-dependent Hamiltonian above, by a rather ingeniously constructed nonlinear one, which we shall also discuss. We will see that the nonlinear approach also has a severe problem with interpretation.

We will first examine the solutions obtained from the Kanai Hamiltonian, and will discuss explicitly those features which have been criticized as unphysical. These criticisms fall into two classes, the first of which we will show to be unjustified, but the second of which is very valid, and will necessitate a modification of the Kanai Hamiltonian, if it is to have a physical interpretation.

First, in the quantum mechanical case of a damped free particle, the spread of the Kanai wavefunction, $\Delta x(t)$, remains finite for all times, even in the limit $t \to \infty$. This is in sharp contrast to the case of the undamped free particle, where the wavefunction ultimately disperses through all space. It also conflicts with the solutions to all other proposed damping Hamiltonians in this respect. Nontheless, by examining the behavior of a swarm of classical particles in phase space, we will show that their behavior in the damped case is very different from that in the undamped case, and in fact $\Delta x(t)$ always remains finite. Thus the behavior of the Kanai solution becomes very plausible in this regard, as the properties of the undamped case turn out to be a very poor guide as to what to expect in the damped case.

The second criticism is much more meaningful, as the Kanai solution appears to violate the uncertainty principle. Of course formally, the uncertainty principle is satisfied, $\Delta p \Delta x \sim \hbar$. However, the canonical momentum, $p = m_0 v e^{\gamma t}$,

 0022-2488/79/050762-09$01.00

is not equal to the physical "kinetic" momentum, $p_k = m_0 v$, for this problem, and for the kinetic momentum one has $m_0 \Delta v \Delta x \sim \hbar e^{\gamma t} \to 0$, for large times. This behavior has been a long-standing puzzle for the interpretation of the solution. We show that the problem is caused by the fact that the Kanai Hamiltonian actually refers to a particle of variable mass, $m(t) = m_0 e^{\gamma t}$. This increasing mass causes an effect equivalent classically to a damping force proportional to the velocity.

So the particle with increasing mass classically has an equation of motion identical to that of a particle of constant mass subject to a damping force. Thus classically, these two problems are equivalent. But quantum mechanically they are not. Because even though the expectation values of the quantum problem are the same as those of its classical counterpart, the spread of the wavefuntion will be governed by the specific form of the Hamiltonian. And in this case, the increasing mass absorbs most of the momentum fluctuations, so that the spread in velocity goes to zero.

We would like to point out the rather fascinating mathematical implications of this situation. Normally, one has a given, unique, physical situation, and one can describe it by various different mathematical formulations. These formulations are generally connected by unitary (or classically, canonical) transformations, and the totality of different formulations is usually directly related to the amount of symmetry inherent in the physical problem. The situation here is almost diametrically opposite. Here we have two completely different physical situations—a constant mass particle subject to a damping force, and a particle of varying mass—both of which give rise to the same classical mathematical description. Yet quantum mechanically, the Kanai Hamiltonian is only consistent with the variable mass interpretation, and the constant mass interpretation is inconsistent, and violates the uncertainty principle.

Thus in this respect the Kanai Hamiltonian is truly unphysical. But does this mean that it is therefore not possible to give a constant mass reinterpretation of the Kanai Hamiltonian? In the classical case, since the equation of motion for the two interpretations are identical, one can certainly interpret the Hamiltonian as representing a damped constant mass particle, as has always been done. Quantum mechanically, one cannot do so with the Hamiltonian in its present form.

However, we shall show that there is a further physical ambiguity in both the quantum and classical problems, and one can exploit this fact to modify the Kanai Hamiltonian in such a way as to restore a reasonable constant mass interpretation even quantum mechanically. The nature of this ambiguity has to do with the fact that when a damped particle approaches closely enough to equilibrium, the damping becomes sufficiently small so that one can no longer detect that the motion is being damped, and thus beyond this point one can effectively switch off the damping force. Quantum mechanically, this has the effect of limiting the shrinkage of the wavefunction, thus making it possible to preserve the constant mass interpretation.

For the nonlinear potentials, there are at present some physically reasonable solutions known, which decay to the ground state of the unperturbed oscillator. However it is also known that any stationary state of the unperturbed oscillator is also a solution, so that not all solutions show damping. We will produce a set of solutions which generalize both these results, and which "decay" to any arbitrary stationary state of the unperturbed oscillator. Thus it is not clear that the nonlinear damping potential specifically damps. Rather it seeks out any stationary state at all. Thus the entire physical basis of the method remains difficult to interpret, a situation which is not helped by the failure of superposition amongst known solutions, or the lack of information as to how many unknown solutions exist, both problems being consequences of the nonlinearity of the problem.

In Sec. II we shall produce some solutions to the classical and quantum mechanical damping problem, as formulated with the Kanai Hamiltonian. We will show that this Hamiltonian actually represents a particle of variable mass, which is responsible for its strange behavior when misinterpreted as describing a particle of constant mass. Thus the Kanai Hamiltonian has definite unphysical features. In Sec. III, we will show that a classical distribution of damped particles in phase space does not spread indefinitely, but remains finite, as does the Kanai wavefunction. This behavior is completely different from the undamped case, and argues for the plausibility of the Kanai solution in this particular respect.

In Sec. IV, we discuss the nonlinear approach and point out the difficulties of interpretation it runs into. In Sec. V we point out that it is unrealistic to expect a particle to keep damping beyond a certain minimum energy, both classically and quantum mechanically. We then use this fact to modify the Kanai Hamiltonian by eliminating the damping beyond this point. This eliminates the unphysical qualities of the solutions and then they can be reasonably reinterpreted as describing the damped motion of a constant mass particle. We explain exactly what such a reinterpretation entails, in terms of the formalism. While it may prove somewhat disturbing to have to cope with a variable mass particle, the fact is that this property is already built into the Kanai Hamiltonian. By modifying the Hamiltonian we are actually allowing it to describe the physical situation of a constant mass particle, and it then becomes the first linear model to successfully handle the problem of damping. (In fact, in the companion paper of Ref. 6 it is shown how one can introduce the concept of a "dissipation variable" and thereby eliminate all mention of a variable mass, if one prefers.) Finally, Sec. VI contains a summary of the paper.

Before we begin, we should point out that when one discusses damping forces, there are at least three types of dissipation involved. First, the Schrödinger equation has an intrinsic dissipation built into it, in the sense that a free particle wave packet will spread spatially. This is due to the uncertainty principle, and for a Gaussian packet, $(\Delta x)^2 = (\Delta x)_0^2 + (\Delta v)^2 t^2$. This diffusion effect does not destroy the coherence of the packet (i.e., its capacity to produce diffraction effects). Furthermore, given any encouragement,

a wave packet will not diffuse, as when bound in a stationary state of an attractive force.

Second, if one puts in a damping force, via some Hamiltonian like that above, the particle center of mass will reproduce its classical motion, and so energy will be dissipated. But nonetheless, the packet will not lose coherence, though there will be diffusion effects due to the uncertainty principle, as before. This is important, and it is unfair to expect such a wavefunction to reproduce effects of say, dissipation in a heat bath.

Third, there are the incoherent effects due to the chaotic interactions with the surrounding medium. This can be handled via a density matrix of statistical mechanics. A particle in a medium at finite temperature, whether in equilibrium or not, has no memory beyond its relaxation time. But a pure quantum mechanical wavefunction has perfect recall until a measurement is made. These considerations affect what one can rationally expect from the behavior of a solution to Eq. (1), and they indicate a certain nonequivalence between the one-body and many-body approaches.

II. INTERPRETATION OF SOLUTIONS TO THE KANAI HAMILTONIAN

The Hamiltonian of Eq. (1) leads classically to the equations

$$v = \frac{\partial H}{\partial p} = (p/m_0)e^{-\gamma t}, \tag{2}$$

$$\dot{p} = (m_0 v e^{\gamma t})^{\cdot} = -\frac{\partial H}{\partial x} = -\left(\frac{\partial V}{\partial x}\right)e^{\gamma t}, \tag{3}$$

and the equation of motion

$$\dot{v} + \gamma v + m_0^{-1}\frac{\partial V}{\partial x} = 0, \tag{4}$$

which adds a damping force, $F_d = -\gamma m_0 v$, to the problem of a particle moving in a potential V. It should be noted that the canonical momentum,

$$p = m_0 v e^{\gamma t}, \tag{5}$$

is not equal to the "kinetic" momentum,

$$p_k = m_0 v. \tag{6}$$

For the free particle case, $V = 0$, the classical solutions are

$$x = x_0 + (v_0/\gamma)(1 - e^{-\gamma t}), \tag{7}$$

$$v = v_0 e^{-\gamma t} \tag{8}$$

$$(x - x_0) = (v_0 - v)/\gamma. \tag{9}$$

The form of the momentum, Eq. (5), shows that the Kanai Hamiltonian actually refers to a particle of increasing mass, $m(t) = m_0 e^{\gamma t}$. To see this in more detail, let us examine a particle whose mass is increasing, but for which the extra mass carries no extra momentum—such as the case of a raindrop moving through a mist which is at rest. For such a particle, momentum conservation gives

$$(m + \delta m)(v + \delta v) - mv = F\delta t, \tag{10}$$

where F represents any force present on the particle, so that the equation of motion becomes

$$\left(\frac{d}{dt}\right)(m(t)v) = F. \tag{11}$$

The Lagrangian for this equation is

$$L = \tfrac{1}{2}m(t)v^2 - V \tag{12}$$

and the canonical momentum is

$$p = \frac{\partial L}{\partial v} = m(t)v. \tag{13}$$

Furthermore, if the potential is of the form

$$V = m(t)\varphi(x) \tag{14}$$

(which has the formal interpretation of an external gravitational potential, being proportional to the mass), then the Hamiltonian will be of the form

$$H = pv - L = p^2/2m(t) + m(t)\varphi. \tag{15}$$

Therefore, the Hamiltonian equation of motion becomes

$$\begin{aligned} \dot{p} &= \left(\frac{d}{dt}\right)(m(t)v) = -m(t)\frac{\partial \varphi}{\partial x}, \\ \dot{v} &+ (\dot{m}(t)/m(t))v = -\frac{\partial \varphi}{\partial x}. \end{aligned} \tag{16}$$

Now, the Kanai Hamiltonian, Eq. (1), is exactly of the form of Eq. (15), where

$$m(t) = m_0 e^{\gamma t}, \quad \dot{m}(t)/m(t) = \gamma, \tag{17}$$

so that the canonical momentum of Eq. (5), is exactly that of Eq. (13), and the equation of motion, Eq. (4), is exactly that of Eq. (16). Thus it follows that the Kanai Hamiltonian actually represents a particle of variable mass, Eq. (17), which produces a classical equation of motion that is identical to that of a damped particle with constant mass.

So the problem arises that if the variable mass particle of the Kanai Hamiltonian produces a classical motion which cannot be distinguished from a constant mass damped particle, then in what way does it differ from such a particle—how does its variable mass express itself? Of course the momentum of the particle in the two interpretations is different, one being p and the other being p_k, but classically, the problems are formally identical, and either interpretation is valid. However in the quantum mechanical case, while the expectation values obey the classical equations, the fluctuations are nonetheless controlled by the detailed nature of the Hamiltonian, and this is precisely where the variable mass shows up,[6] as we shall show.

To acquire some feeling for the quantum mechanical behavior of the Kanai wavefunctions, we can examine the Schrödinger equation for the free particle case, $V = 0$,

$$-\frac{\hbar^2}{2m_0}e^{-\gamma t}\frac{\partial^2 \psi}{\partial x^2} = i\hbar\frac{\partial \psi}{\partial t}. \tag{18}$$

There is a natural unit of length $\lambda = (\hbar/m_0\gamma)^{1/2}$, in terms of which the equation becomes

$$\frac{1}{2}\frac{\partial^2 \psi}{\partial y^2} = (i\gamma e^{-\gamma t})^{-1}\frac{\partial \psi}{\partial t}, \quad y = x/\lambda. \tag{19}$$

Also, in terms of the variable u, given by

$$u = 1 - e^{-\gamma t}, \tag{20}$$

the equation reduces to the free particle equation

$$\frac{1}{2}\frac{\partial^2\psi}{\partial y^2} = -i\frac{\partial\psi}{\partial u}. \tag{21}$$

Introducing the Green's function $G(y,u)$, for this problem,

$$G(y,u) = (2\pi i u)^{-1/2} e^{iy^2/2u}, \tag{22}$$

which obeys Eq. (21) as well as

$$G(y,0) = \delta(y), \tag{23}$$

the general wavepacket solution obeying the boundary condition

$$\psi(y,0) = \psi_0(y), \tag{24}$$

becomes

$$\psi(y,u) = \int dy' G(y-y',u)\psi_0(y'). \tag{25}$$

Specifically, for the Gaussian packet,

$$\psi_0 = e^{-y^2/a^2}, \tag{26}$$

we find that ψ evolves in such a way that

$$|\psi(y,u)|^2 \sim e^{-2y^2/a^2(u)}, \tag{27}$$

where the width $a(u)$ is

$$a^2(u) = a^2 + 4u^2/a^2 = a^2 + (\Delta v)^2 u^2, \tag{28}$$

in terms of our dimensionless variables (v here is also in dimensionless units). Remembering that $u = 1 - e^{-\gamma t}$, we see that the wavepacket is finite both at $t = 0$, and at $t = \infty$, in contrast to the case of an undamped particle wavepacket, $(\Delta x)^2 = (\Delta x)_0^2 + (\Delta v)^2 t^2$.

In fact, had we made instead the substitution $u_1 = e^{-\gamma t}$, representing the solution for a wavepacket that becomes Gaussian at $t = \infty$, its width would also be given by Eq. (28), with u_1 replacing u. In this case the wavepacket actually *shrinks* in time. So the wavepacket can shrink or expand, but the important feature is that it maintains a finite width at all times. We shall see in the next section that this behavior is perfectly plausible from a classical point of view, although it differs drastically from both the free particle case, and from all the nonlinear potentials that have been proposed.

The unphysical feature of the Kanai Hamiltonian shows up in the following way. In the case of a bound particle, say in a harmonic oscillator, if one calculates the spread of the wavefunction in time, one finds[5]

$$(\Delta x)^2 \simeq (\Delta x)_0^2 e^{-\gamma t} \tag{29}$$

and

$$(\Delta p)^2 \simeq (\Delta p)_0^2 e^{\gamma t}, \tag{30}$$

so that the uncertainty principle is formally satisfied,

$$(\Delta p)^2(\Delta x)^2 \sim (\Delta p)_0^2(\Delta x)_0^2 \sim \hbar^2. \tag{31}$$

If we remember that the Hamiltonian represents a particle of increasing mass, there is nothing strange about this. But if one wants to force a constant mass interpretation on the problem, then in terms of the kinetic momentum, $p_k = m_0 v$,

$$\begin{aligned} p_k &= pe^{-\gamma t}, \\ (\Delta p_k)^2 &= (\Delta p)^2 e^{-2\gamma t}, \\ (\Delta p_k)^2(\Delta x)^2 &\sim (\Delta p)^2(\Delta x)^2 e^{-2\gamma t} \to 0, \end{aligned} \tag{32}$$

we have the rather perplexing situation that the uncertainty principle seems to be violated. It would be incorrect to attribute this vanishing of the product $\Delta p_k \Delta x$, as given by Eq. (32), to the lack of an extra fluctuation term in the Hamiltonian, to match the dissipation present (as has been done in the literature). That would be true in a macroscopic treatment of the problem via statistical methods, but the one-body Schrödinger equation contains all the fluctuations it needs to guarantee the uncertainty principle.

This strange behavior, which has proved very puzzling, is a direct consequence of our attempt to impose a constant mass interpretation on the quantum mechanical solution. The quantity $p_k/m_0 = p/m(t)$ represents the velocity, whose fluctuations are decreasing, $(\Delta v)^2 \sim (\Delta v)_0^2 e^{-\gamma t}$, while p represents the momentum, whose fluctuations are increasing. The difference between the two is provided by the mass, whose increase furnishes the added momentum fluctuations. So the uncertainty principle is not violated, but only appears to be violated if one insists on the incorrect constant mass interpretation of the Hamiltonian. But of course an actual, physical damped oscillator does have constant mass, and so the Kanai Hamiltonian as written does not apply to such a system. Nonetheless, we shall show in Sec. IV that it is physically valid to modify the Hamiltonian at low energy to achieve a reasonable constant mass interpretation of the problem, eliminating the problems associated with the uncertainty principle. There will then be no further obstacles to a constant mass interpretation.

III. THE DISPERSION OF THE KANAI WAVEFUNCTION

We saw in the last section that for a damped free particle, the Kanai wavefunction does not spread indefinitely, like that for an undamped free particle, but maintains a finite width in the limit $t \to \infty$, and it has been criticized as being unphysical on these grounds. We shall examine the behavior of a distribution of classical damped particles in phase space, and show that it behaves in this regard like the Kanai wavefunction, while a distribution of undamped classical particles spreads indefinitely like the undamped free particle wavefunction. So in this respect the Kanai wavefunction actually behaves very plausibly.

The first thing to notice is that there is no Galilean invariance in the problem of damped motion—Newton's first law does not hold. All particles slow down and asymptotically come to rest. So there is something very special about speed zero. It is the end point of all motion, but it is never reached in a finite time by moving particle. On the other hand, a particle at rest remains at rest always.

Furthermore, there is no such thing as a "little bit of damping." Either the motion is undamped, and all speeds are equivalent, or it is damped, and speed zero is special. If it is damped, the time scale is set by $1/\gamma = t_0$, the damping constant. In units of this time, $\gamma t = t/t_0$, all damped motion is scaled identically. So there is no smooth passage to the limit $\gamma \to 0$. All finite values of γ are formally identical, and the value $\gamma = 0$ represents a different problem. [This is not true if other forces are present. For example in the harmonic oscillator there are two time scales, $1/\gamma$ and $1/\omega_0$, and it makes sense to speak of the limit $\gamma \ll \omega_0$.)

Note too that the form of the damping force, $F_d = -\gamma m_0 v$, acts as a converging lens in velocity space. This means that if one releases a swarm of particles simultaneously, with a spread in velocities, $\rho(v)$, the faster particles slow up more rapidly than the slower ones, so that the distribution tends to shrink in velocity space. Ultimately, of course, all velocities tend to zero.

And so, as can be seen from Eq. (7), all particles travel only a finite distance before coming to rest. A particle originally at x_0, with velocity v_0, will end up at $t = \infty$, with

$$x_f = x_0 + v_0/\gamma. \tag{33}$$

Therefore, of course, the distribution will have a finite spread in Δx as $t \to \infty$. Thus, far from being unphysical, classically this behavior is mandatory. And while the classical behavior only controls the motion of the expectation values in quantum theory, and the quantum fluctuations are not necessarily related to classical considerations, nonetheless, this argument shows the plausibility of having Δx stay finite for the quantum wavefunction, and the Kanai wavefunction has this property.

This is totally different from the behavior of a free undamped wave packet, which will spread indefinitely in Δx. There the spread is due to an uncertain knowledge of the velocity, $(\Delta v)_0$. But with damping, we know that every velocity component of Δv will decrease separately, as will Δv itself.

In line with our comments above, this finite spread of the wave packet in x space will take place for any nonzero value of γ. Depending on the original shape of the packet, Δx_f can be smaller or larger than Δx_0, so that the behavior in the limit $\gamma \to 0$ is no guide to the behavior at $\gamma = 0$.

We can quantitatively illustrate these remarks. A distribution of particles which starts off at $t = 0$ with a shape given by $\rho_0(x,v)$, will at time t be governed by the equation

$$\frac{D\rho}{Dt} = v\frac{\partial \rho}{\partial x} + a\frac{\partial \rho}{\partial v} + \frac{\partial \rho}{\partial t} = 0. \tag{34}$$

For a free undamped distribution, $a = 0$, and so the distribution at time t will be

$$\rho_t^{(un)}(x,v) = \rho_0(x - vt, v). \tag{35}$$

For the case under consideration, damping with $a = -\gamma v$, the solution can be constructed from the particle motion as given by Eqs. (7), (8), and (9),

$$\rho_t^{(d)}(x,v) = \rho_0(x - (v/\gamma)(e^{\gamma t} - 1), ve^{\gamma t}). \tag{36}$$

So for example, if the velocity part of the distribution was initially $\rho(v) = A\exp[-(v - v_0)^2/\Delta^2]$, then at time t the distribution would become

$$\rho_t(v) = A\exp[-(ve^{\gamma t} - v_0)^2/\Delta^2]$$

$$= A\exp[-(v - v_0 e^{-\gamma t})^2/(\Delta e^{-\gamma t})^2], \tag{37}$$

and would have its center at $\bar{v} = v_0\exp(-\gamma t)$, and width $\Delta(t) = \Delta\exp(-\gamma t)$.

To study a simple example in detail, consider the following special distribution. At $t = 0$, the distribution is uniform within a region of (x,v) space,

$$\rho_0(x,v) = \begin{cases} \text{const}, & x_1 \leqslant x \leqslant x_2, \quad v_1 \leqslant v \leqslant v_2, \\ 0, & \text{elsewhere.} \end{cases} \tag{38}$$

If we normalize by

$$\int dx\, dv\, \rho_0 = 1, \tag{39}$$

then the constant will be

$$\rho_0 = 1/\Delta x \Delta v, \quad \Delta x = (x_2 - x_1), \quad \Delta v = (v_2 - v_1). \tag{40}$$

For undamped motion, each particle will move at constant speed, or horizontally in Fig. 1(a), and the distribution $\rho(x) = \int dv\, \rho(x,v)$ will spread as shown in Fig. 1(b).

If now we consider the case of damped motion for the same initial distribution, ρ_0, of Eq. (38), then each particle will move along the straight line given by Eq. (9), ultimately

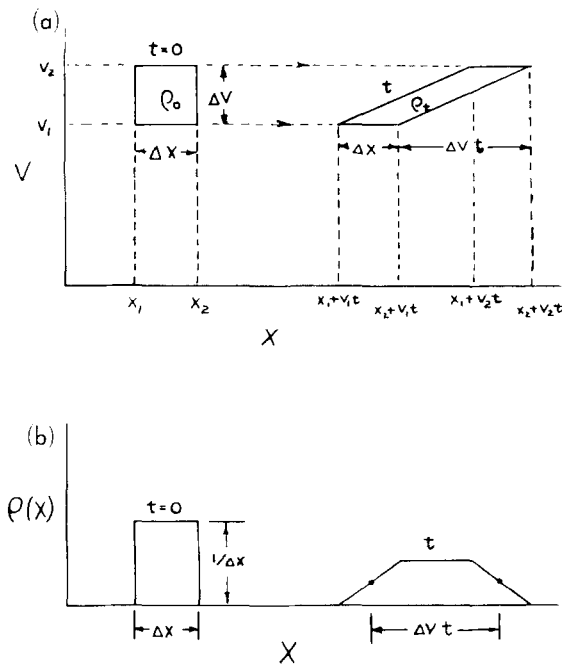

FIG. 1. Spread of a uniform distribution of undamped particles. (a) Evolution of $\rho(x,v)$ in time. An originally uniform rectangular distribution of particles spreads out as the faster particles outspeed the slower. (b) The distribution $\rho(x) = \int dv\, \rho(x,v)$ as a function of time. The half-width remains Δx until $\Delta x = \Delta v t$, and thereafter becomes $\Delta v t$, expanding indefinitely.

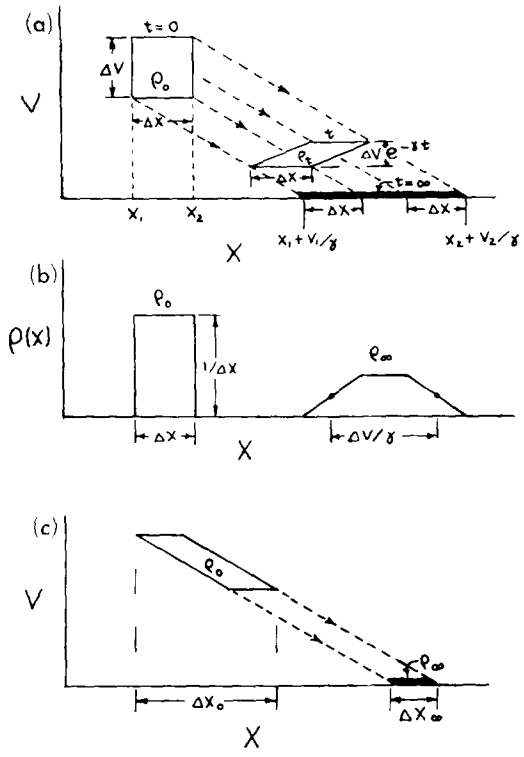

FIG. 2. Spread of a uniform distribution of damped particles. (a) Evolution of $\rho(x,v)$ in time. An originally uniform rectangular distribution of particles squeezes together in v, as all particles move along parallel slanting straight lines, exponentially approaching the x axis at $t = \infty$. (b) The distribution $\rho(x) = \int dv\, \rho(x,v)$ as a function of time. The width as $t \to \infty$ is finite, $\Delta x_\infty = \Delta v/\gamma$. (c) A distribution where Δx_∞ is actually less than Δx_0.

settling on the horizontal axis as $t \to \infty$. The spread of the distribution will be as shown in Fig. 2(a), and the spread in $\rho(x)$ will be as shown in Fig. 2(b). In Fig. 2(c) we show a distribution that will actually shrink in size as time goes on. However the important point is that the spread at $t = \infty$ will always be finite if the initial distribution was. In this respect the solutions to the Kanai Hamiltonian show the same behavior as a classical distribution of particles. Other quantum mechanical wavefunctions, representing other approaches to the damping problem which behave quite differently in this respect, will have to establish their own plausibility.

IV. THE NONLINEAR APPROACH TO DAMPING

The alternate one-particle approach that has been tried for the problem of damping uses nonlinear potentials. Kostin[5] first introduced such a class of potentials with the property that the expectation values of the dynamical variables reproduce their classical behavior. We agree with the conclusion of Immele, Kan, and Griffin,[7,8] that the natural interpretation of such potentials is in terms of the fluid mechanical interpretation of quantum mechanics, where the resistance is proportional to the "velocity" of the fluid, which is related to the phase of the wavefunction, or $\ln(\psi/\psi^*)$. They showed that there exists a solution to this potential that behaves in just the manner one would expect intuitively from a damped wavepacket. An extension of the type of potential allowed was given by Albrecht,[9] and numerous general wavepacket solutions were produced by Hasse.[1]

At present, it is known that these potentials possess some intriguing, physically reasonable solutions. However, they also possess extra, undamped solutions, which are difficult to interpret. Furthermore, because of their nonlinearity, they may possess other, nonintuitive solutions (in fact, we will produce one such class), as there is no completeness theorem for them. Other consequences of their nonlinearity are that solutions cannot be superimposed, and until an investigation of their regions of stability is carried out, it would be difficult to apply perturbation theory methods to them with any confidence. These difficult problems have barely begun to be attacked,although Ref. 8 has a discussion of superposition. In another vein, some general properties of non-Hermitian Hamiltonians have been discussed by Eck and Thompson.[10]

However, beyond all these difficulties, the entire approach suffers from the fact that its physical interpretation is very obscure. To underline this point, we would like to show that one can generalize the known solutions to the problem by producing a class of solutions whose interpretation is very puzzling. The nonlinear Hamiltonian, in the form of Kan and Griffin[7], is

$$H\psi = [H_0 + (\gamma\hbar/2i)\ln(\psi/\psi^*) - W(t)]\psi = i\hbar\dot{\psi}. \quad (41)$$

Here H_0 is the Hamiltonian without damping, and the ln term produces the velocity-dependent damping, with γ the damping constant ($F_d = \gamma m_0 v$). The function $W(t)$ is chosen as

$$W(t) = (\gamma\hbar/2i)\int d^3r\,|\psi|^2 \ln(\psi/\psi^*), \quad (42)$$

the expectation value of the damping term, in order to make the expectation value of the total energy, $H(t)$, well behaved.

If H_0 is chosen as the Hamiltonian for the undamped Harmonic oscillator,

$$H_0 = p^2/2m_0 + \tfrac{1}{2}m_0\omega_0^2 x^2, \quad (43)$$

then Kan and Griffin noted that one solution of Eq. (41) is

$$\begin{aligned}\psi &= \psi_0 e^{-g_0(t)} \\ &= N_0 \exp[-(m_0\omega_0/2\hbar)(x - X(t))^2] \\ &\quad \times \exp[ixP(t)/\hbar]e^{-ig_0(t)}, \end{aligned} \quad (44)$$

where

$$\begin{aligned} X(t) &= X_0 e^{-\gamma t}\cos(\omega t - \delta), \\ P(t) &= m_0\dot{X}(t), \quad \omega = (\omega_0^2 - \gamma^2)^{1/2}, \end{aligned} \quad (45)$$

the classical solutions for the motion, while

$$g_0(t) = \omega_0 t/2 + \int_0^t dt'(P^2/2m_0 - m_0\omega_0^2 X^2/2 - \gamma PX)/\hbar, \quad (46)$$

and N_0 is a normalization constant. This solution has the

classical solutions for its expectation values, and as $t\to\infty$, approaches the Gaussian ground state of the undamped oscillator, a behavior which would appear to be quite reasonable. However they also noted that

$$u_n = N_n \mathscr{H}_n(x) \exp[-(m_0\omega_0/2\hbar)x^2] e^{-i(n+\frac{1}{2})\omega_0 t}, \quad (47)$$

where $\mathscr{H}_n$ are the Hermite polynomials, and N_n are normalization constants, are also solutions of Eq. (41). These are merely the stationary solutions of H_0, the undamped oscillator, and their appearance as solutions to what purports to be a damping problem is rather strange.

We can in fact generalize both of these solutions. The functions

$$\psi_n = N_n \mathscr{H}_n(x - X(t))\psi_0 e^{-ig_n(t)}, \quad (48)$$

where

$$g_n(t) = g_0(t) + n\omega_0 t, \quad (49)$$

are also solutions to the Hamiltonian, Eq. (41). The special case, $n = 0$, is the solution ψ_0 of Kan and Griffin. Also, these solutions "decay" to the unperturbed stationary states,

$$\psi_n(x,t) \underset{t\to+\infty}{\longrightarrow} u_n, \quad (50)$$

which have arbitrarily high energy $E_n = (n + \frac{1}{2})\hbar\omega_0$.

One can see that the damping force, the gradient of the logarithmic damping potential, vanishes in any stationary state of the form

$$\varphi(x,t) = f(x)e^{i\beta t}, \quad (51)$$

where $f(x)$ is real, and β is a constant, regardless of whether or not φ is a solution of the Hamiltonian. So it is clear that this potential seeks out stationary states, rather than low energy states. [In fact it seeks out any separable function $f(x)g(t)$, with $f(x)$ real.] The individual highly excited stationary states are very nonclassical objects, and one normally makes classical wave packets by superimposing them. But of course one cannot superimpose nonlinear solutions. Thus we feel that it would be fair to characterize the present status of the nonlinear approach by saying that it offers intriguing possibilities, but that its entire physical basis and interpretation remain quite obscure. And of course, one has no guarantee that other, even stranger, solutions do not exist.

V. MODIFICATION OF THE KANAI HAMILTONIAN

We pointed out in Sec. II that because the Kanai Hamiltonian represents a variable mass particle, its interpretation as representing a constant mass particle breaks down as $t\to\infty$, because the fluctuations become unrealistic, and both $\Delta x\to 0$ and $\Delta p_k = m_0 v\to 0$. Now there is nothing implausible in both Δx and Δp_k decreasing as the particle damps. The problem arises because the damping never stops.

On the other hand, there is always a physical limit to such damping. For example, if the motion is that of a damped harmonic oscillator initially in a highly excited state, where $\Delta p_k \Delta x \gg \hbar$, there will be no further damping once it reaches its ground state. If the motion is that of a damped free particle, once its energy decreases to $E = (3/2)kT$, the damping will be swamped by thermal vibrations. Also, if the particle is composite, with internal energy ϵ_0, the damping will be overwhelmed by internal motions when its energy has decreased to $E\sim\epsilon_0$. In all these examples the damping mechanism will still be working, but there will be no further decrease in amplitude of the wavefunction. So the solution to the problem of unlimited damping depends on the realization that beyond a certain lower energy limit the damping force is balanced out by other, not explicitly stated, forces and for the purposes of the equivalent one-body problem may just as well be discontinued.

These statements have their counterparts even for a classical particle. Such a particle at rest in the medium will remain indefinitely at rest, even with damping present, and so behaves like a free particle. On the other hand, a moving particle will continue to move forever, though with ever decreasing speed. Thus the classical problem could be restated by saying that the damping force is $F_d = -\gamma m_0 v$, for $v\neq 0$, and $F_d = 0$, for $v = 0$. We write it in this manner merely to emphasize the special role played by $v = 0$. However from a physical point of view one can go further and say that there is a point beyond which, for $v\neq 0$, the motion effectively ceases. In the general statement of the problem, this point is arbitrary. But in any specific problem, it will be determined by the nature of the system, as in the above examples. So for practical purposes, one might replace the damping force by

$$F_d = \begin{cases} -\gamma m v, & \epsilon > \epsilon_0, \\ 0, & \epsilon \leqslant \epsilon_0. \end{cases} \quad (52)$$

So we see that even classically, the particle keeps damping long beyond the time when there is any physical meaning to the notion, and this is exacerbated by the mathematical problem that this motion does not smoothly blend into the $v = 0$ case. The replacement of the damping force by Eq. (52) represents a primitive attempt to remedy the situation.

A different way to say this is that actually, in the classical case one solves the Langevin equation, which not only has a damping force present, but also a random force due to the collisions with other particles in the medium. This random force averages out to zero, but keeps the fluctuations of the particle motion from vanishing. When the velocity decreases to the point where the damping is no greater than the random force, then we have reached the energy ϵ_0, given above.

Quantum mechanically, these problems are brought into a much sharper focus. If we consider, for example, the harmonic oscillator, there is a natural energy, ϵ_0, determined by the uncertainty principle, beyond which internal fluctuations will prevent the energy from decreasing. But if one tries to interpret the Kanai Hamiltonian as describing a constant mass particle then Eq. (1), with $V = \frac{1}{2}m_0\omega_0^2 x^2$, makes the "physical energy," $E_k = He^{-\gamma t}$, decrease continuously beyond this minimum (since the time average of the "canonical energy" $E = H$, is constant in this case).

So the unphysical quality of the solutions, as given by Eq. (32), is really partly a reflection of the classically unreal-

istic solutions to the problem. The most primitive method for dealing with the quantum mechanical problem would be to replace the damping force by one similar to that of Eq. (52), so that for the oscillator

$$H = \begin{cases} (p^2/2m_0)e^{-\gamma t} + \frac{1}{2}m_0\omega_0^2 x^2 e^{\gamma t}, & \overline{H}e^{-\gamma t} > \frac{1}{2}\hbar\omega, \\ (p^2/2m_0) + \frac{1}{2}m_0\omega^2 x^2, & \overline{H}e^{-\gamma t} = \frac{1}{2}\hbar\omega, \end{cases} \tag{53}$$

where

$$\omega^2 = \omega_0^2 - \gamma^2/4, \quad \overline{H} = \frac{1}{T}\int_{t-T/2}^{t+T/2} dt\,\langle H\rangle, \tag{54}$$

ω being the displaced frequency, and $\overline{H}$ being the Hamiltonian averaged over a period.

This Hamiltonian would keep the oscillator frequency at ω for all times. It is primitive in that it introduces excessive transient effects in matching the solutions during the discontinuous switching of the Hamiltonians. (Because H oscillates in time about a fixed average, one must average over a period to determine the appropriate switching energy.)

A more refined procedure would be to consider the particular problem on hand, whether it is statistical or otherwise, and attempt to modify the Hamiltonian to preserve the correct correlations $\langle\psi(t)\psi(t+\tau)\rangle$. However, one must be careful not to confuse the statistical correlations with the quantum mechanical ones, as mentioned in the introduction. For example, in the case of the damped free particle, if the particle comes to rest and no further observations are made upon it, its initial velocity uncertainty will cause an ambiguity that only a clear insight into the problem at hand can resolve, if one wants a truly realistic model. Because of the damping, the initial velocity uncertainty will certainly damp, unlike a true free particle, in the sense that in whatever direction it takes off in, it will surely slow down. On the other hand, constant interaction with the damping medium will serve to continually relocalize the particle, as in a cloud chamber, and thus constantly renew its velocity spread, converting the problem into a many-body statistical one.

The important point is that one is not doing any violence to the physics by modifying the Hamiltonian in such a fashion. One is merely removing a classical problem, which becomes much more bothersome in the quantum case. It might be pointed out that this procedure makes γ a function of energy, rather than time, since for a highly excited initial state, it takes longer to decay to the ground state energy—and it is the energy, not the time, that sets the criterion for the change in Hamiltonian.

Whatever model one chooses, one can now interpret m_0 as representing the constant mass of the particle. Once the particle has decayed to its physically lowest energy, the fluctuations will cease to shrink, because of the modified Hamiltonian, and the problems associated with the original Kanai Hamiltonian no longer occur.

The specific procedure for reinterpreting the solution as representing a constant mass damped particle depends on noting that while the variable x still denotes the position of the particle, the canonical momentum, $p = m(t)v \to (\hbar/i)\partial/\partial x$, no longer represents the physical momentum. Rather, in the constant mass reinterpretation the physical momentum is represented by

$$p_k = e^{-\gamma t}p \to (\hbar/i)e^{-\gamma t}\partial/\partial x \equiv m_0 v_{op}. \tag{55}$$

Thus if the system is, for example, a damped harmonic oscillator originally in a highly excited state, it will originally have $\Delta p_k \Delta x \gg \hbar$. As it decays, the uncertainty product will decrease exponentially until such a time that it would take the particle to decay to the ground state, and at this time $\Delta p_k \Delta x \sim h$. Beyond this time, if one used the original Kanai Hamiltonian, the system would keep decaying and offer no possibilities for reinterpretation. But with the modified Hamiltonian, there will be no further decay, and for later times, one uses $p_k = p = m_0 v \to (\hbar/i)\partial/\partial x$. The Hamiltonian changes discontinuously at the critical moment, but the wavefunction is kept continuous. As an example, if one wanted the kinetic energy in the constant mass interpretation, at some time before the critical time: classically, one would take $p_k^2/2m_0 = \frac{1}{2}m_0 v^2 = e^{-2\gamma t}p^2/2m_0$; quantum mechanically, one would calculate

$$\text{K.E.} = e^{-2\gamma t}\langle p^2/2m_0\rangle$$

$$= -(\hbar^2/2m_0)e^{-2\gamma t}\int dx\,\psi^*\,\partial^2\psi/\partial x^2. \tag{56}$$

So the modified Kanai Hamiltonian offers one viable method for treating the problem of a constant mass particle subject to damping. Its solutions represent the effects of the damping up until a certain physically reasonable cutoff point, after which the effects due to the damping cease. It is also true that once one accepts the modified Hamiltonian, with a constant mass interpretation, the entire concept of a variable mass becomes superfluous, and one can restate the problem so that one need not think in terms of a changing mass at all. This is best done by introducing a new dynamical variable, the "dissipation variable." This line of thought will be pursued further in Ref. 6.

V. SUMMARY

We have shown that the most straightforward approach to the problem of a damped particle, namely via the Kanai Hamiltonian, actually yields unphysical results. This is because the Hamiltonian does not really represent the system it is expected to. While in the classical limit the theory can be interpreted as representing either a constant mass damped particle, or a particle of increasing mass, quantum mechanically, only the variable mass interpretation is correct, and this leads to the often noted violation of the uncertainty principle when it is misinterpreted as representing a constant mass particle.

Nonetheless, we have shown that as the particle continues to damp and its energy decreases sufficiently, there is an ambiguity in the physical situation itself, when treated as a one-body problem, both classically and quantum mechanically. We have also shown that one may modify the Hamil-

tonian so as to eliminate any further damping. Once this is done, one may reinterpret the Hamiltonian so that it really does represent a particle of constant mass. From this point of view, the mass of the particle is constant, and it is rather the dissipation process that is varying.

The alternative point of view, that of the nonlinear potentials, produces a set of solutions that behave quite reasonably as representing damped particle motion. Unfortunately this method produces other solutions as well, whose physical interpretation seems very strange. There are totally undamped solutions, as well as solutions which damp to states of arbitrarily high energy, and width. In fact the entire "damping" procedure seems to single out stationary states, rather than low energy states. Thus the method is plagued not only by the intrinsic difficulties of nonlinearity, but by very basic questions of interpretation as to what the Hamiltonian represents.

Therefore, the modified Kanai Hamiltonian, despite its disturbing pedigree—having evolved from the original variable mass Kanai Hamiltonian—actually yields (along with the treatment in the companion paper[6]) the only one-particle treatment of the damping problem to date whose solutions can be thoroughly understood and represent physically well-behaved damped particles of constant mass.

ACKNOWLEDGMENT

Research supported (in part) by a grant from the PSC–BHE Research Award Program of the City University of New York.

ADDENDA

We have recently come across a paper by Feshbach and Tikochinsky,[11] based on a Lagrangian suggested by Morse and Feshbach,[12] which attacks the problem of damping from an apparently different point of view. However, their approach is actually equivalent to that of Kanai.

Their Lagrangian,

$$L_0 = m_0\dot{x}\dot{y} + (m_0\gamma/2)(x\dot{y} - y\dot{x}) - m_0\omega_0^2 xy, \tag{57}$$

leads to the equations of motion for the independent variables, x and y,

$$\begin{aligned} &\ddot{y} + \gamma\dot{y} + \omega_0^2 y = 0, \\ &\ddot{x} - \gamma\dot{x} + \omega_0^2 x = 0, \end{aligned} \tag{58}$$

where y is a damped oscillator, and x is an "antidamped" one (i.e., of exponentially increasing amplitude). The apparent time independence is due to the canceling of the damping and antidamping exponentials. However, if one adds to L_0 a total time derivative,

$$\begin{aligned} &L = L_0 - \frac{df}{dt}, \\ &f = (m_0\gamma/4)[(xe^{\gamma t/2})^2 + (ye^{-\gamma t/2})^2], \end{aligned} \tag{59}$$

and then makes the substitution

$$\begin{aligned} &x = (\xi + i\eta e^{-\gamma t})/\sqrt{2}, \\ &y = (\xi e^{\gamma t} - i\eta)/\sqrt{2}, \end{aligned} \tag{60}$$

both of which operations formally represent canonical transformations which do not change the physical problem, then one obtains

$$L = (m_0/2)(\dot{\xi}^2 - \omega_0^2\xi^2)e^{\gamma t} + (m_0/2)(\dot{\eta}^2 - \omega_0^2\eta^2)e^{-\gamma t}. \tag{61}$$

Comparison with Eqs. (12), (14), and (17), for the Kanai Lagrangian, and Eq. (15) for the Kanai Hamiltonian, shows that Eq. (61) is just the Kanai Lagrangian for a damped particle (ξ), plus an antidamped one (η). The Kanai form represents the uncoupled normal coordinates for the problem. And in fact, while (elegantly) solving the problem quantum mechanically, the authors[11] introduce raising and lowering operators which decouple the modes. Their wavefunctions are therefore equivalent to the Kanai solutions (and also exhibit the attendant physical difficulties with the uncertainty principle).

[1]R.W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[2]Two interesting examples of such calculations are I.R. Senitzky, Phys. Rev. **119**, 670 (1960); and G.W. Ford, M. Kac, and P. Mazur, J. Math. Phys. **6**, 504 (1965).
[3]E. Kanai, Prog. Theor. Phys. **3**, 440 (1948).
[4]Some of the various solutions are discussed in Ref. 1, along with an extensive list of references. However we shall solve it explicitly in Sec. II for a damped free particle, and in Ref. 6 for the damped harmonic oscillator.
[5]M.D. Kostin, J. Chem. Phys. **57**, 3589 (1972); J. Stat. Phys. **12**, 145 (1975).
[6]These observations form the basis for a new approach to the entire problem, in the companion paper, D.M. Greenberger, J. Math. Phys. **20**, 771 (1979).
[7]K. Kan and J.J. Griffin, Phys. Lett. B **50**, 241 (1974).
[8]J.D. Immele, K. Kan, and J.J. Griffin, Nucl. Phys. A **241**, 47 (1975).
[9]K. Albrecht, Phys. Lett. B **56**, 127 (1975).
[10]J. Eck and W.J. Thompson, Am. J. Phys. **45**, 161 (1977).
[11]H. Feshbach and Y. Tikochinsky, contribution to *A Festschrift for I.I. Rabi*, edited by L. Motz (New York Academy of Sciences, New York, 1977).
[12]P. Morse and H. Feshbach, *Methods of Mathematical Physics* (McGraw-Hill, New York, 1953), Vol. I, p. 298.

A new approach to the problem of dissipation in quantum mechanics

Daniel M. Greenberger

City College of the City University of New York, New York
(Received 5 January 1977, revised manuscript received 5 April 1978)

The usual treatment of damping forces in quantum mechanics starts from the introduction of the explicitly time dependent Kanai Hamiltonian, which actually represents a variable mass particle, and the misinterpretation of this Hamiltonian as representing a particle of constant mass leads to certain physical difficulties. However the Hamiltonian can be modified so that it can be reinterpreted as describing a constant mass particle. Here we explicitly introduce the mass as a new dynamical variable, which allows us to write a linear, time independent Hamiltonian for the system, which can be solved by conventional methods. The damped harmonic oscillator and damped free particle are treated in detail, both for the Kanai Hamiltonian and for our case, and the solutions are compared. Our solution can be reduced to the Kanai one in appropriate circumstances, but in general it has a much greater versatility, as a result of which it can be more easily reinterpreted as describing a constant mass particle subject to a damping force, which reinterpretation is of course necessary if the method is to have practical applicability. We also show how such a reinterpretation can be carried out in detail by introducing a "dissipation variable", in terms of which one may avoid the concept of a variable mass altogether.

I. INTRODUCTION

In the companion paper[1] we discussed some of the physical considerations relevant to the treatment of a quantum mechanical system with a damping proportional to its velocity. The most direct treatment is through Kanai's time-dependent Hamiltonian[2,3] [see Eq. (3.6) for the harmonic oscillator case], which reproduces the classical equation of motion. We noted during that paper that the Kanai Hamiltonian actually represents a particle of varying mass, $m = m_0 e^{2\gamma t}$, which fact explains the peculiar quantum mechanical features of the solution, when misinterpreted as representing a particle of fixed mass subject to a damping force. The problems that arise concern the fact that as the energy of the system decreases, so do the fluctuations in position and in the physical momentum, until ultimately they become so small that they violate the uncertainty principle. We also showed that one could alter the Kanai Hamiltonian at low energies so as to provide a reasonable constant mass reinterpretation of the problem.

Here, we are going to exploit the varying mass to introduce a new mathematical technique for solving the problem. It turns out that the mass and proper time of a particle can easily be treated as conjugate dynamical variables, obeying their own equations of motion.[4] We happen to believe that since the formalism is so simple and suggestive, it most likely provides a clue to a more fundamental application of the ideas involved. However, we will not pursue such ideas here. Rather, we will show that from a purely computational point of view, it provides a linear, time-independent treatment of a problem where the physical energy is clearly time-dependent, and where previous treatments are either explicitly time-dependent, or nonlinear. Furthermore, the general idea of taking a parameter of the problem, and making it into a dynamical variable in its own right, should prove a versatile technique applicable to other problems.

There is another advantage to specifically treating the mass as a dynamical variable. Normally, if one wants to treat an interacting system, one introduces a potential which produces the desired interaction. Then if one is interested in only part of the system, say an almost isolated atom, the potential can be viewed as an external force which causes transitions between states of the atom in question, and which gives rise to an energy spread in the states of the atom.

But what if one's basic isolated entity is itself a composite system, with internal degrees of freedom? Any energy spread of the system's states can normally only be treated by introducing a model which includes these internal degrees of freedom, even if one is not specifically interested in them. For example, we might have a nucleus subject to a force we want to treat, but this nucleus may happen to be unstable to, say γ radiation to a lower energy state. If we do not care about the details of the γ decay, our formalism allows one to treat the excited nucleus as a particle of indefinite mass (a "mass" wave packet) without having to inquire about its internal structure. This represents a definite advantage over the conventional formalism.

A further advantage of our method is the flexibility of the solutions. We will show that they reduce to the Kanai solution in special cases, but in general they can encompass a much wider range of behavior of the physical system.

A feature that our Hamiltonian shares with the Kanai one is that it can also be interpreted in the classical limit as representing a damped particle of constant mass. However the specific Hamiltonian we introduce to approximate the Kanai Hamiltonian, has the same quantum problems associated with the constant mass interpretation, so that the wave-packets shrink beyond the limit prescribed by the uncertainty relation. But in our method, one may also introduce a class of Hamiltonians which eliminate the damping beyond a certain point, in order to provide a quantum mechanical constant mass reinterpretation. As an example of the flexibility of the method, we show that one can do this in a continuous manner just as easily as in our discontinuous treatment of the Kanai Hamiltonian.

 0022-2488/79/050771-10$01.00

Once the problem is reinterpreted in terms of constant mass, the concept of a "variable mass" actually has no further relevance. It becomes the vestige that remains of our search for an interpretation of the Kanai Hamiltonian. So one can at this stage again reinterpret the formalism in such a way as to eliminate all mention of a variable mass. This is done by introducing in its stead a "dissipation variable" which keeps track for us of the damping mechanism.

In this paper we will primarily treat the damped harmonic oscillator, as a representative problem in damped motion. In Sec. II, we shall present a brief account of the general idea of our approach. In Sec. III, we discuss the classical damped oscillator from both the conventional Kanai approach and from ours. We treat the Kanai solution of the damped quantum mechanical oscillator in Sec. IV, in some detail, so as to be able to compare and contrast the solution with that of our treatment, which is presented in Sec. V. In these two sections we also show how to reinterpret these solutions so as to describe damped particles of constant mass. We also introduce here the dissipation variable as a method to avoid the concept of a variable mass altogether.

With both the Kanai Hamiltonian and our Hamiltonian, there are problems with the uncertainty principle in such a reinterpretation. However, in Sec. VI, we show how to modify the Hamiltonian at low energies in such a way as to circumvent these physical problems, and thus make possible a consistent constant mass reinterpretation of the solution. Finally, in Sec. VII, we briefly solve the damped free particle by our approach, and we close with a short summary in Sec. VIII.

II. MASS AS A DYNAMICAL VARIABLE

It is very easy to extend the formalism of classical physics to treat systems whose mass is variable.[4] The general procedure merely consists of considering the mass, m, of a system and its proper time, τ, as conjugate dynamical variables, so that the Hamiltonian, which was formerly of the form $H = H(x,p;t)$ now takes the form

$$H = H(x,p;\tau,m;t). \tag{2.1}$$

(We will always work in one dimension—the results are easily generalizable. Also, we take the mass in units of energy, mc^2, where we will always take $c = 1$.) The "velocities" are given by

$$v = \dot{x} = \frac{\partial H}{\partial p}, \quad \dot{\tau} = \frac{\partial H}{\partial m}. \tag{2.2}$$

The equations of motion, one for p, the other for m, become,

$$\dot{p} = -\frac{\partial H}{\partial x}, \quad \dot{m} = -\frac{\partial H}{\partial \tau}. \tag{2.3}$$

Thus a potential depending on x provides a force which alters the momentum, and similarly, a potential depending on τ provides a "force" which changes the mass of the particle. The relation between p and v is dynamically determined by the equation of motion for x, Eq. (2.2). Similarly, $\dot{\tau}$ is no longer given kinematically, as in special relativity $[d\tau = (1 - v^2)^{1/2}dt]$, but through its own equation of motion, Eq. (2.2).

A simple example, to illustrate the procedure, is given by a relativisitc free particle. We have chosen this example because the difference between t and τ is familiar here, and has physical significance, and also comes immediately from the formalism. However, there is nothing intrinsically relativistic about the formalism itself, as we use it here, and we will be applying it to nonrelativistic problems only. The Hamiltonian takes the form

$$H = (p^2 + m^2)^{1/2}. \tag{2.4}$$

The velocity is given by

$$v = \frac{\partial H}{\partial p} = p(p^2 + m^2)^{-1/2}. \tag{2.5}$$

This equation can be inverted to give the momentum, p,

$$p = mv\gamma, \quad \gamma = (1 - v^2)^{-1/2} = (p^2 + m^2)^{1/2}/m. \tag{2.6}$$

This is the conventional approach to the problem. However we can also, in our formulation, write

$$\dot{\tau} = \frac{\partial H}{\partial m} = m(p^2 + m^2)^{-1/2} = (1 - v^2)^{1/2}, \tag{2.7}$$

which dynamically defines the behavior of the proper time. Of course, for a free particle, $\dot{p} = \dot{m} = 0$. So, from the point of view of our formalism, we can say that conventional classical mechanics leaves out half the subject, since the kinematical background for the problem must be independently postulated. For nonrelativistic problems this is usually done implicitly, by assuming that the mass is constant, and that $\tau = t$.

The question arises as to what new physics is introduced by the formalism. If the mass remains constant, as in our example above, there is no new physics in the classical case. However even here there will be a difference in the quantum case. This is because the very introduction of the mass and proper time as dynamical variables guarantees the uncertainty principle $\Delta m \Delta\tau \sim \hbar$. So one can compose a wave packet of different mass states, whose expectation value obeys the classical equations. Only in the limit $\Delta\tau \to \infty$, $\Delta m \to 0$, does the system approach the conventional classical limit. (We might point out that the Bargmann theorem, preventing the superposition of different mass states in conventional nonrelativistic quantum theory, does not apply to our formalism[4].)

If the Hamiltonian has a τ dependence, so that the mass changes, then the problem will be formally equivalent in the classical case to a classical problem with the mass postulated to change in the specified manner. In the quantum case, the expectation value of the mass will change in this manner, and there will also be fluctuations in the mass governed by the uncertainty principle.

We shall apply the formalism in the next section to the classical damped harmonic oscillator.

III. THE CLASSICAL DAMPED HARMONIC OSCILLATOR

The Hamiltonian for an undamped harmonic oscillator is

$$H_0 = p^2/2m_0 + \tfrac{1}{2}m_0\omega_0^2x^2. \tag{3.1}$$

Here m_0 is the constant mass and ω_0 the (angular) frequency. The velocity, from Eq. (2.2), is

$$\dot{x} = v = p/m_0, \tag{3.2}$$

and the equation of motion, from Eq. (2.3), is

$$\begin{aligned} \dot{p} &= m_0\dot{v} = -m_0\omega_0^2x, \\ \dot{v} &= \ddot{x} = -\omega_0^2x. \end{aligned} \tag{3.3}$$

H_0 also plays the role of the physical energy,

$$H_0 = E = \text{const.} \tag{3.4}$$

The solutions take the form

$$x = x_0\cos\omega_0 t + (v_0/\omega_0)\sin\omega_0 t, \tag{3.5}$$

where x_0 and v_0 are the initial position and velocity.

The damped harmonic oscillator can be described by the Hamiltonian due to Kanai,[2]

$$H_1 = (p^2/2m_0)e^{-2\gamma t} + \tfrac{1}{2}m_0\omega_0^2x^2e^{2\gamma t}. \tag{3.6}$$

The velocity in this case is

$$\dot{x} = v = (p/m_0)e^{-2\gamma t}, \quad p = m_0ve^{2\gamma t}. \tag{3.7}$$

The equation of motion becomes

$$\begin{aligned} \dot{p} &= m_0(\dot{v} + 2\gamma v)e^{2\gamma t} = -m_0\omega_0^2xe^{2\gamma t}, \\ \ddot{x} &+ 2\gamma\dot{x} + \omega_0^2x = 0. \end{aligned} \tag{3.8}$$

The solutions are given by

$$x = x_0e^{-\gamma t}\cos\omega t + \omega^{-1}(v_0 - \gamma x_0)e^{-\gamma t}\sin\omega t. \tag{3.9}$$

The frequency of oscillation is ω,

$$\omega^2 = \omega_0^2 - \gamma^2 > 0. \tag{3.10}$$

We will assume throughout this paper that the system is underdamped.

The Hamiltonian H_1 is not constant, however it is approximately constant and varies only to order γ/ω_0. The physical energy, E is time-dependent and is given by

$$E = H_1e^{-2\gamma t}. \tag{3.11}$$

Note that the "kinetic" momentum, $p_k = m_0v$, is given by

$$p_k = m_0v = pe^{-2\gamma t}, \tag{3.12}$$

so that the canonical momentum, p, is not the same as the physical momentum. It was pointed out in Ref. 1 that this is because the Kanai Hamiltonian actually refers to a particle of increasing mass,

$$m(t) = m_0e^{2\gamma t}. \tag{3.13}$$

The canonical momentum refers to the momentum of this varying mass particle. If one interprets the system instead, as applying to a constant mass particle of mass m_0, subject to a damping force, then its momentum is given by p_k. Classically one can use either interpretation, because both systems have the same equation of motion. But quantum mechanically, the constant mass interpretation runs into trouble with the uncertainty principle, and cannot be maintained unless the Hamiltonian is modified.

One can also treat the problem by writing a specifically chosen variable mass Hamiltonian,

$$\mathscr{H} = p^2/2m + \tfrac{1}{2}m\omega_0^2x^2 - 2\gamma m\tau. \tag{3.14}$$

We have left out the term relating to the rest mass of the particle, which will in turn rob τ of any interpretation as a physical time. The Hamiltonian above should be interpreted as a formal treatment of the problem, which is closely related to the above time-dependent treatment. For the "velocities," we have

$$\begin{aligned} v &= p/m, \\ \dot{\tau} &= \frac{\partial H}{\partial m} = \frac{-p^2}{2m^2} + \tfrac{1}{2}\omega_0^2x^2 - 2\gamma\tau \\ &= -\tfrac{1}{2}v^2 + \tfrac{1}{2}\omega_0^2x^2 - 2\gamma\tau. \end{aligned} \tag{3.15}$$

The equation of motion for m becomes

$$\dot{m} = \frac{-\partial H}{\partial\tau} = +2\gamma m, \tag{3.16}$$

with the solution

$$m = m_0e^{2\gamma t}, \tag{3.17}$$

which is the same as that of the "proper" interpretation of the Kanai Hamiltonian, Eq. (3.13). The equation of motion for x is

$$\begin{aligned} \dot{p} &= (mv)^{\cdot} = -m\omega_0^2x, \\ \ddot{x} &+ (\dot{m}/m)\dot{x} + \omega_0^2x = 0. \end{aligned} \tag{3.18}$$

Together with Eq. (3.16) this reduces to Eq. (3.8).

The Hamiltonian, since it is explicitly independent of the time, will be a constant of the motion. The physical energy of a particle of constant mass m_0 would be given by

$$E = m_0\mathscr{H}_0/m, \quad \mathscr{H}_0 = p_2/2m + \tfrac{1}{2}m\omega_0^2x^2. \tag{3.19}$$

It should be pointed out that just as the Kanai Hamiltonian can be interpreted in the classical case as applying to either a constant mass or variable mass particle, so too can the same be said of our Hamiltonian. For our variable mass Hamiltonian, since the velocity is given by $v = p/m$, then if one wanted to interpret the theory as applying to a particle of constant mass m_0, subject to a damping force, one would use for the physical momentum in this case

$$p_0 = m_0v = m_0p/m. \tag{3.20}$$

However, since we have chosen this specific Hamiltonian to do exactly what the Kanai Hamiltonian does, we shall see that quantum mechanically, it runs into the same problems in the attempt to give it a constant mass interpretation, and for exactly the same reasons. But again, we shall be able to modify it to make such an interpretation possible.

IV. QUANTUM TREATMENT OF THE DAMPED OSCILLATOR

We present here a brief quantum treatment of the Kanai Hamiltonian, since we will want the solution for detailed comparison purposes, both for the similarities and differences it presents to our treatment. We will solve it by using operator methods, which we present in some detail since it is not covered in standard texts.

First we write the solutions of the undamped oscillator, to establish our notation. The Schrödinger equation becomes, using the Hamiltonian of Eq. (3.1),

$$H_0(x,p)\psi = -\frac{\hbar}{i}\frac{\partial\psi}{\partial t}, \quad p = \frac{\hbar}{i}\frac{\partial}{\partial x}. \tag{4.1}$$

Then, introducing the length scale defined by

$$\lambda_0 = (\hbar/m_0\omega_0)^{1/2}, \tag{4.2}$$

we introduce the dimensionless variable

$$y = \frac{x}{\lambda_0}, \quad p_y = \frac{1}{i}\frac{\partial}{\partial y}. \tag{4.3}$$

The Hamiltonian then becomes

$$H_0 = \frac{\hbar\omega_0}{2}\left(-\frac{\partial^2}{\partial y^2} + y^2\right) = \frac{\hbar\omega_0}{2}(p_y^2 + y^2). \tag{4.4}$$

The step-up operator a^+, and the step-down operator a, are defined as

$$a^+ = (1/\sqrt{2})(p_y + iy), \quad a = (1/\sqrt{2})(p_y - iy), \tag{4.5}$$

and obey the following commutation rules:

$$[p_y,y] = 1/i, \quad [a,a^+] = 1,$$
$$[H_0,a^+] = \hbar\omega_0 a^+, \quad [H_0,a] = -\hbar\omega_0 a. \tag{4.6}$$

The Hamiltonian then takes the form

$$H_0 = \hbar\omega(a^+a + \tfrac{1}{2}).$$

The stationary state solutions are

$$\psi = u_n e^{-iE_n t/\hbar}, \quad E_n = (n + \tfrac{1}{2})\hbar\omega_0, \tag{4.7}$$

obeying

$$\left(-\frac{\partial^2}{\partial y^2} + y^2\right)u_n = 2\epsilon_n u_n, \quad \epsilon_n = n + \tfrac{1}{2} \tag{4.8}$$

with solutions

$$u_n = N_n H_n(y)e^{-y^2/2}, \quad N_n = (\sqrt{\pi}\,2^n n!\lambda_0)^{-1/2}, \tag{4.9}$$

where the H_n are the Hermite polynomials, and N_n is chosen so that $\int u_n^2\,dx = 1$.

The Kanai Hamiltonian is given by Eq. (3.6). In this equation m_0 is a constant, but it should be borne in mind that physically, this equation represents a particle with mass $m(t) = m_0 e^{2\gamma t}$. The canonical momentum, p, represents the classical quantity $m(t)v$. This has no effect on the technique of solution, which proceeds formally. Making the substitutions (4.2) and (4.3), we have

$$H_1\psi = \frac{\hbar\omega_0}{2}(e^{-2\gamma t}p_y^2 + e^{2\gamma t}y^2)\psi = -\frac{\hbar}{i}\frac{\partial\psi}{\partial t}. \tag{4.10}$$

We will also need to introduce the first quadrant angle θ, defined by

$$\tan\theta = \gamma/\omega. \tag{4.11}$$

Defining the operators

$$b^+ = (1/\sqrt{2})(e^{-\gamma t}p_y + e^{\gamma t}iy),$$
$$b = (1/\sqrt{2})(e^{-\gamma t}p_y - e^{\gamma t}iy), \tag{4.12}$$

which obey the commutation rules

$$[b,b^+] = 1, \quad [H_1,b^+] = \hbar\omega_0 b^+, \quad [H_1,b] = -\hbar\omega_0 b, \tag{4.13}$$

we have, for the Hamiltonian,

$$H_1 = \hbar\omega_0(b^+b + \tfrac{1}{2}). \tag{4.14}$$

Unfortunately, here the similarity with the undamped case ends, as the b's are explicitly time-dependent, and there are no stationary states. However, we have

$$\left[-\frac{\hbar}{i}\frac{\partial}{\partial t}, b\right] = \frac{\hbar}{i}\gamma b^+,$$
$$\left[-\frac{\hbar}{i}\frac{\partial}{\partial t}, b^+\right] = \frac{\hbar}{i}\gamma b, \tag{4.15}$$

and introducing the complete Schrödinger operator U,

$$U = H_1 + \frac{\hbar}{i}\frac{\partial}{\partial t}, \tag{4.16}$$

in terms of which the Schrödinger equation becomes

$$U\psi = 0, \tag{4.17}$$

we find

$$[U,b^+] = \hbar\omega_0 b^+ - (\hbar/i)\gamma b,$$
$$[U,b] = -\hbar\omega_0 b - (\hbar/i)\gamma b^+. \tag{4.18}$$

We can still find raising and lowering operators for U as linear combinations of b and b^+,

$$[U,d^+] = \hbar\omega d^+, \quad [U,d] = -\hbar\omega d, \tag{4.19}$$

where

$$d = b - i\tan(\theta/2)b^+, \quad d^+ = b^+ + i\tan(\theta/2)b, \tag{4.20}$$

with θ defined by Eq. (4.11). Using these definitions, d and d^+ are not normalized, and

$$[d,d^+] = \cos\theta/\cos^2(\theta/2), \tag{4.21}$$

so that we could introduce new operators

$$D = \cos(\theta/2)(\cos\theta)^{-1/2}d, \quad [D,D^+] = 1, \tag{4.22}$$

but for our purposes we will not need them.[5]

The operators d^+ and d are step-up and step-down operators for U, but they have no simple relation to H_1. Nonetheless, we can use them to construct eigenfunctions for U satisfying

$$U\varphi_n = \lambda_n\varphi_n, \tag{4.23}$$

and from these we easily construct solutions of the Schrödinger equation, since U contains an explicit time derivative. In fact, if

$$\psi_n = \varphi_n e^{-i\lambda_n t/\hbar}, \tag{4.24}$$

then ψ_n satisfies the Schrödinger equation, $U\psi_n = 0$.

To find the lowest state, we set

$$d\varphi_0 = 0. \tag{4.25}$$

Then

$$U(d\varphi_0) - d(U\varphi) = -d(U\varphi_0) = -\hbar\omega(d\varphi_0) = 0, \tag{4.26}$$

which implies

$$U\varphi_0 = \lambda_0\varphi_0, \quad \psi_0 = \varphi_0 e^{-i\lambda_0 t/\hbar}. \tag{4.27}$$

One cannot find a lower state, since reapplying d will give zero. Using definitions (4.20) and (4.12), the solution to Eq. (4.25) is

$$\varphi_0 = N\exp(-\tfrac{1}{2}y^2 e^{i\theta + 2\gamma t}), \tag{4.28}$$

and

$$U\varphi_0 = (\hbar\omega_0/2)e^{i\theta}\varphi_0. \tag{4.29}$$

Finally

$$\psi_0 = N\exp[-\tfrac{1}{2}e^{i\theta}(y^2e^{2\gamma t} + i\omega_0 t)]$$

$$= N\exp[-\tfrac{1}{2}e^{i\theta}y^2e^{2\gamma t} + \tfrac{1}{2}(\gamma - i\omega)t]. \tag{4.30}$$

The general solution (unnormalized) is

$$\psi_n = (d^+)^n\psi_0 e^{-in\omega t}, \tag{4.31}$$

where from Eq. (4.20),

$$d^+ = (1/\sqrt{2})\sec(\theta/2)[p_z + iz], \tag{4.32}$$

$$z = ye^{-i\theta/2 + \gamma t}, \quad p_z = (1/i)\frac{\partial}{\partial z}.$$

To express the answer in Hermite polynomials, we note that for the undamped oscillator,

$$a^+H_n(y)e^{-(1/2)y^2} = (\sqrt{2}i)^{-1}\left(\frac{\partial}{\partial y} - y\right)H_n(y)e^{-y^2/2}$$

$$= (\sqrt{2}i)^{-1}\left(\frac{\partial H_n}{\partial y} - 2yH_n\right)e^{-y^2/2}$$

$$= (\sqrt{2}i)^{-1}H_{n+1}e^{-y^2/2}, \tag{4.33}$$

from the recurrence relation for the H_n's. Similarly, because of the $(-\frac{1}{2}z^2e^{2i\theta})$ of Eq. (4.30), we can write

$$d^+[H_n(cz)\exp(-\tfrac{1}{2}z^2e^{2i\theta})]$$
$$= A\left(\frac{\partial}{\partial z} - z\right)H_n(cz)\exp(-\tfrac{1}{2}z^2e^{2i\theta})$$
$$= H_{n+1}(cz)\exp(-\tfrac{1}{2}z^2e^{2i\theta}), \tag{4.34}$$

for the specific choice of c,

$$c = e^{i\theta/2}(\cos\theta)^{1/2} \quad cz = (\cos\theta)^{1/2}ye^{\gamma t}. \tag{4.35}$$

Here A is a constant whose value we need not determine. The factor $(\cos\theta)^{1/2} = (\omega/\omega_0)^{1/2}$ has the effect of changing the ω_0 in λ_0 and y, into ω [Eqs. (4.2) and (4.3)]. So finally, we can express the normalized solutions, Eq. (4.31), as

$$\psi_n = N'_nH_n((\cos\theta)^{1/2}ye^{\gamma t})$$

$$\times \exp[-\tfrac{1}{2}e^{i\theta}y^2e^{2\gamma t} + \tfrac{1}{2}\gamma t - i(n + \tfrac{1}{2})\omega t],$$

$$N'_n = N_n(\cos\theta)^{-1/4}, \tag{4.36}$$

where N_n is given in Eq. (4.9).

The unphysical feature of this solution comes from the fact that the decreasing exponential determines the spread of the function, so that

$$|\psi|^2 \sim \exp(-y^2\cos\theta e^{2\gamma t}) \sim e^{-y^2/\Delta^2}, \tag{4.37}$$

which implies

$$\Delta^2 \sim e^{-2\gamma t}. \tag{4.38}$$

For the momentum

$$(\Delta p)^2 \sim e^{4\gamma t}\Delta^2 \sim e^{2\gamma t}. \tag{4.39}$$

Thus the solution formally satisfies the uncertainty relation $(\Delta p)^2(\Delta x)^2 \sim \hbar^2$.

But here, in the quantum case, if one wanted as an alternative to be able to maintain a constant mass interpretation for the solution, as one can do classically, and consider m_0 as the mass of the particle, one would run into trouble. Because in this case the "kinetic" momentum, $p_k = m_0 v = pe^{-2\gamma t}$, is not equal to the canonical momentum, so that

$$(\Delta p_k)^2 \sim e^{-2\gamma t}, \tag{4.40}$$

and so the product

$$(\Delta p_k)(\Delta y) \sim \hbar e^{-2\gamma t}. \tag{4.41}$$

Thus, as $t \to \infty$, the uncertainty product goes to zero. Therefore, as was discussed in Ref. 1, the interpretation as a constant mass particle must break down, because ultimately the uncertainty relation will be violated.

One can however save the constant mass interpretation by modifying the Kanai Hamiltonian, as indicated in Ref. 1. The way one would apply a constant mass reinterpretation of the problem would be to note that x still represents the position, but for a constant mass particle, the momentum would be p_k, $p_k = e^{-2\gamma t}p \to e^{-2\gamma t}(\hbar/i)\partial/\partial x$. The expectation values of this operator, or functions of it, will give values that one would interpret as representing $m_0 v$. We will see at the end of the next section how this fits into a general scheme for such a constant mass reinterpretation.

V. QUANTUM TREATMENT WITH VARIABLE MASS

Now we are going to solve the variable mass Hamiltonian of Eq. (3.14). Remember that here m is a dynamical variable, which quantum mechanically becomes the "mass operator." Its classical value, $m_0e^{2\gamma t}$, will have meaning only as an expectation value. We will actually find it expedient to work in the "m representation," where τ becomes a differential operator.

The variable mass Hamiltonian of Eq. (3.14) is independent of time and one can find stationary state solutions. If

one represents the τ dependence in the form[4]

$$\varphi(x,\tau) = (2\pi)^{-1/2}\int dm\, \psi(x,m)e^{im\tau/\hbar}, \tag{5.1}$$

then the operator m becomes

$$m = (\hbar/i)\frac{\partial}{\partial\tau}, \tag{5.2}$$

while in the m representation, we have

$$\tau = -(\hbar/i)\frac{\partial}{\partial m}. \tag{5.3}$$

The solutions to the problem will only make sense for $m > 0$, so we will place this restriction on m here, for purely mathematical reasons. This also has consequences for the analyticity in τ, which we will not need.

The operator product $(m\tau)$ must be symmetrized,

$$m\tau \to \tfrac{1}{2}(m\tau + \tau m) = -(\hbar/i)(m\partial/\partial m + \tfrac{1}{2}) \tag{5.4}$$

so that the Schroedinger equation becomes

$$\mathscr{H}\psi = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \frac{1}{2}m\omega_0^2x^2 + \frac{2\gamma\hbar}{i}\left(m\frac{\partial}{m} + \frac{1}{2}\right)\right]\psi$$

$$= -\frac{\hbar}{i}\frac{\partial\psi}{\partial t}. \tag{5.5}$$

Again we introduce the length λ, but now λ is a function of m,

$$\lambda = (\hbar/m\omega_0)^{1/2} = f(m), \tag{5.6}$$

so that

$$\mathscr{H} = \frac{\hbar\omega_0}{2}\left(-\lambda^2\frac{\partial^2}{\partial x^2} + \frac{1}{\lambda^2}x^2\right)$$

$$+ \frac{2\gamma\hbar}{i}\left(m\frac{\partial}{\partial m} + \frac{1}{2}\right). \tag{5.7}$$

Once again we take $y = x/\lambda$, but now the change in variables from (x,m) to $(y = x/\lambda\,(m),\, m)$ implies

$$y = x/\lambda = x(\omega_0/\hbar)^{1/2}m^{1/2},$$

$$\frac{\partial}{\partial x} \to \frac{1}{\lambda}\frac{\partial}{\partial y},$$

$$\frac{\partial}{\partial m} \to \frac{y}{2m}\frac{\partial}{\partial y} + \frac{\partial}{\partial m}, \tag{5.8}$$

$$\int dx\, dm \to \int dy\, \lambda\,(m)dm.$$

The relevant operators become

$$\lambda\frac{\partial}{\partial x} \to \frac{\partial}{\partial y}, \quad m\frac{\partial}{\partial m} \to \frac{1}{2}y\frac{\partial}{\partial y} + m\frac{\partial}{\partial m}, \tag{5.9}$$

and finally, the Hamiltonian becomes

$$\mathscr{H} = \frac{\hbar\omega_0}{2}\left(-\frac{\partial^2}{\partial y^2} + y^2 + \frac{2\gamma}{i\omega_0}y\frac{\partial}{\partial y}\right)$$

$$+ \frac{2\gamma\hbar}{i}\left(m\frac{\partial}{\partial m} + \frac{1}{2}\right), \tag{5.10}$$

and is separable. It should be noted that in the variable mass case the switch to dimensionless coordinates is no longer merely convenient, but is necessary to make the equation separable.

We write

$$\psi = Y(y)M(m)e^{-i\epsilon t}, \tag{5.11}$$

where

$$\left(-\frac{\partial^2}{\partial y^2} + y^2 + \frac{2\gamma}{i\omega_0}y\frac{\partial}{\partial y}\right)Y = 2\lambda Y, \tag{5.12}$$

$$\left[m\left(\frac{\partial}{\partial m}\right) + \frac{1}{2}\right]M = (i/2)\beta M \tag{5.13}$$

and

$$\hbar\epsilon = \hbar\omega_0\lambda + \hbar\gamma\beta. \tag{5.14}$$

The solution to Eq. (5.13) is

$$M = (4\pi m)^{-1/2}(m/m_0)^{i\beta/2}. \tag{5.15}$$

The normalization will be explained at the end of this section.

To solve Eq. (5.12) we write

$$Y = e^{\kappa y^2}U(y). \tag{5.16}$$

The choice

$$\kappa = \gamma/i\omega \tag{5.17}$$

eliminates the first-order term, giving

$$\left(-\frac{\partial^2}{\partial y^2} + \frac{\omega^2}{\omega_0^2}y^2\right)U = \left(2\lambda + \frac{\gamma}{i\omega_0}\right)U, \tag{5.18}$$

which is nothing but the undamped oscillator equation, Eq. (4.4), in the variable $u = y(\omega/\omega_0)^{1/2} = x(m\omega/\hbar)^{1/2}$, which converts the frequency appearing in λ from ω_0 to ω. The solution is

$$U = N'_nH_n(u)e^{-u^2/2}, \tag{5.19}$$

where N' is the same numerical normalization factor appearing in Eq. (4.36), and where

$$2\lambda + \gamma/i\omega_0 = 2(\omega/\omega_0)(n + \tfrac{1}{2}), \tag{5.20}$$

which gives for the energy factor, $\hbar\epsilon$,

$$\hbar\epsilon(n,\beta) = (n + \tfrac{1}{2})\hbar\omega - \hbar\gamma/2i + \beta\hbar\gamma. \tag{5.21}$$

Thus the complete solution becomes

$$\psi(x,m,t) = \sum_n \int d\beta\, a_n(\beta)(4\pi m)^{-1/2}(m/m_0)^{i\beta/2}$$

$$\times e^{-x^2(m/\hbar)(\omega + i\gamma)/2}$$

$$\times N'_nH_n(x(m\omega/\hbar)^{1/2})e^{-i\epsilon(n,\beta)t}. \tag{5.22}$$

The exponential factor $\omega + i\gamma$ can be written $\omega_0e^{i\theta}$. For each value of n, the mass dependence becomes a wavepacket,

$$m^{-1/2}\int d\beta\, a_n(\beta)(m/m_0)^{i\beta/2}e^{-i\beta\gamma t}$$
$$= m^{-1/2}\int d\beta\, a_n(\beta)\exp[i\beta(\tfrac{1}{2}\ln(m/m_0)-\gamma t)]$$
$$= (2\pi m)^{-1/2}A_n(\tfrac{1}{2}\ln(m/m_0)-\gamma t)$$
$$= m^{-1/2}f_n(m-m_0e^{2\gamma t}), \tag{5.23}$$

where A_n is the Fourier transform of a_n,

$$A_n(\xi) = (2\pi)^{-1/2}\int d\beta\, a_n(\beta)e^{i\beta\xi}. \tag{5.24}$$

The functions A_n must be normalized in accordance with Eq. (5.32), which in turn determines the normalization of the $a_n(\beta)$. Each of these packets is centered about the "classical" value $m = m_0e^{2\gamma t}$. However the superposition of different masses, plus the independence of the shape of these packets for each value of n, gives the solution far greater flexibility than the solution of the Kanai Hamiltonian.

We can recover the Kanai solution by requiring the A_n to be Dirac δ functions. This is done by choosing the $a_n(\beta)$ to be constant, for all β. If we take $a_n = c/m_0^{1/2}$, then

$$\int \lambda(m)dy\, dmc(m_0m)^{-1/2}\int d\beta\exp[i\beta(\tfrac{1}{2}\ln(m/m_0)-\gamma t)]$$
$$= (\hbar/\omega_0m_0)^{1/2}2\pi c\int dy(dm/m)\delta(\tfrac{1}{2}\ln(m/m_0)-\gamma t)$$
$$= \lambda_0 4\pi c\int dy\, dm\delta(m-m_0e^{2\gamma t})$$
$$= 4\pi c\int dx\Big|_{m=m_0e^{2\gamma t}} \tag{5.25}$$

and if one replaces m by $m_0e^{2\gamma t}$ everywhere it is left in Eq. (5.22),

$$\exp[-\tfrac{1}{2}x^2(m\omega_0/\hbar)e^{i\theta}]H_n(x(m\omega/\hbar)^{1/2}), \tag{5.26}$$

and remember that

$$x^2(m\omega_0/\hbar)e^{i\theta} = y^2e^{i\theta+2\gamma t},$$
$$x(m\omega/\hbar)^{1/2} = y(\cos\theta)^{1/2}e^{\gamma t}, \tag{5.27}$$

then one has exactly the expression (4.36) for the Kanai solution. When one considers the entirely different logic leading up to the two formulations of the problem, one can see that this reproduction of the Kanai solution proves our contention that it represents the problem of a particle of definite (i.e., $\Delta m = 0$) but changing mass. [Equation (5.25) was integrated at $\tau = 0$.]

We add a final word on the normalization of the solution. When the solution $\psi(x,m,t)$ of Eq. (5.22) is integrated over β to give

$$\psi(x,m,t) = (2m)^{-1/2}\sum_n A_n(\tfrac{1}{2}\ln(m/m_0)-\gamma t)$$
$$\times N_n'H_n(x(m\omega/\hbar)^{1/2})\exp(-\tfrac{1}{2}x^2\omega_0e^{i\theta}-i\epsilon_nt), \tag{5.28}$$

we can check the normalization,

$$\int dx\, dm\,|\psi(x,m,t)|^2 = 1. \tag{5.29}$$

The integral over x gives

$$\int \frac{dm}{2m}e^{\gamma t}\left(\frac{m_0}{m}\right)^{1/2}\sum_n\left|A_n\left(\frac{1}{2}\ln\left(\frac{m}{m_0}\right)-\gamma t\right)\right|^2. \tag{5.30}$$

The factor $(m_0/m)^{1/2}$ comes from the fact that N' contains the constant mass m_0, while H_n contains the variable m. If we now replace m by μ, defined by

$$\mu \equiv \tfrac{1}{2}\ln(m/m_0)-\gamma t, \quad -\infty<\mu<\infty,$$
$$d\mu = dm/2m, \tag{5.31}$$

then this integral becomes

$$\int d\mu\, e^{-\mu}\sum_n |A_n(\mu)|^2 = 1, \tag{5.32}$$

which determines the appropriate normalization of the A_n, and which is time-independent.

This solution, like that for the Kanai Hamiltonian, represents a variable mass particle. The mass in this case is a distribution (actually a separate wavepacket for each value of n), centered about $m_0e^{2\gamma t}$, and like the Kanai wavefunction, it cannot be given a constant mass reinterpretation, for exactly the same reasons, namely that the damping proceeds to $t\to\infty$.

However, as in the Kanai case, we will be modifying the Hamiltonian in order to be able to provide a constant mass reinterpretation. To that end, let us explain exactly how to interpret our wavefunction, and then how one would reinterpret it as representing a damped particle of constant mass. Again, the variable x represents the position, and ψ is normalized according to Eq. (5.29). The momentum, $p = mv \to (\hbar/i)\partial/\partial x$, represents the momentum of the packet, whose mass is centered about the value $\bar{m} = m_0e^{2\gamma t}$. The classical velocity would be given by the expectation value of the operator (p/m).

If one wanted to reinterpret the results as applying to a particle of constant mass, the first restriction one would have to apply is that the wavefunction in "mass space" be rather narrow. This is necessary, so that Δp could be interpreted as $\bar{m}\Delta v$. Otherwise contributions corresponding to terms $\bar{v}\Delta m$ would appear, and have no constant mass interpretation.

In order to apply a constant mass interpretation to the problem, one merely considers the entire concept of "changing mass" as we have introduced it, as a formal manipulation of symbols. Then our variable m can be written as

$$m = m_0\eta, \tag{5.33}$$

where η is a dimensionless dynamical variable, which we can call the "dissipation variable." As long as the system exhibits dissipation, η will be increasing. Its expectation value for the Kanai Hamiltonian will be $e^{2\gamma t}$, but for our modified Hamiltonian it increases only to a certain value and then remains constant.

Thus the mass is always m_0, and the momentum, $p_0 = m_0v$, becomes

$$p_0 = m_0 v = m_0(p/m) = p_0/\eta = \left(\frac{\hbar}{i\eta}\right)\frac{\partial}{\partial x}. \tag{5.34}$$

Expectation values of functions of the momentum, for the constant mass particle, become

$$\langle f(p_0)\rangle = \int m_0\, d\eta\, dx\, \psi^* f\left(\frac{\hbar}{i\eta}\frac{\partial}{\partial x}\right)\psi. \tag{5.35}$$

The operator $(\hbar/i)\partial/\partial m$ becomes

$$\left(\frac{\hbar}{i}\right)\frac{\partial}{\partial m} = \left(\frac{\hbar}{im_0}\right)\frac{\partial}{\partial \eta}. \tag{5.36}$$

With this interpretation, or reinterpretation, of the formalism, one need never introduce the concept of changing mass, but rather consider that one has a damped particle of constant mass.

We shall not attempt a further physical interpretation of the dissipation variable at this time, except to note that it is obviously related to entropy production. And just as the entropy itself in classical physics can be thought of as a variable which somehow contains on a macroscopic level all the integrated microscopic information relating to "order" in the system,[6] so too the same can be said for the dissipation variable.

This is the prescription for interpreting the formalism as applying to a constant mass particle. It will run into trouble for the Hamiltonian we have been treating, Eq. (3.14), because of the same uncertainty principle problem as with the Kanai Hamiltonian, namely that $\Delta p \Delta x \to 0$. However, the causes are identical as for the Kanai case, and here too we will be able to modify the Hamiltonian to provide a reasonable constant mass interpretation.

VI. MODIFICATION OF THE SOLUTION AT LOW ENERGIES

We have seen that if we want to maintain the interpretation of the solution to either the Kanai Hamiltonian or the variable mass one as relating to a particle of constant mass m_0, rather than increasing mass, then we must cope with the fact that the physical momentum, $p_k = m_0 v$, and the position, x both damp without limit, so that the uncertainty relation is violated, as in Eq. (4.41). It was pointed out in Ref. 1 that one can modify the Kanai Hamiltonian to eliminate the damping for small enough energy. The same can be done for our Hamiltonian, and we can easily write down a whole class of Hamiltonians that accomplish this.

In our problem of the Harmonic oscillator, "small enough energies" means for energies of the order of $\frac{1}{2}\hbar\omega$. Since we are working in a system where n is diagonal, we can write

$$\psi(x,m,t) = \sum_n a_n |n,m,t\rangle, \tag{6.1}$$

where the right-hand side is defined by Eq. (5.22). And since we know what the initial physical energy of the particle in state n is, at $t = 0$, namely $E_n^{(k)} \sim (n + \frac{1}{2})\hbar\omega$ [from Eq. (5.21)], then we also know how long it will take this state to decay down to energy $\epsilon_0 \sim \frac{1}{2}\hbar\omega$. Specifically, this time is given by

$$\begin{aligned} &E_n^{(k)} e^{-2\gamma t_n} \approx \epsilon_0, \\ &t_n \approx (1/2\gamma)\ln(2n+1). \end{aligned} \tag{6.2}$$

Once we know this time t_n, which is different for each state n, we can modify the Hamiltonian, which is diagonal in n,

$$\mathscr{H}\psi = \sum a_n \mathscr{H}_n |n\rangle \to \sum a_n \mathscr{H}'_n |n\rangle. \tag{6.3}$$

After we have the totality of numbers $\mathscr{H}'_n$, it may be quite difficult to reconstruct the form of the new operator H' which is valid in any representation. However we do not have to face this problem, so long as we remain in the representation $|n\rangle$. The chief requirement on $\mathscr{H}'_n$ is that

$$\begin{aligned} &\mathscr{H}'_n \to \mathscr{H}_n, \quad t \ll t_n, \\ &\mathscr{H}'_n \to \mathscr{H}^0 = p^2/2m_0 + \tfrac{1}{2}m_0\omega^2 x^2, \quad t \gg t_n, \end{aligned} \tag{6.4}$$

so that the Hamiltonian ultimately becomes that of an undamped oscillator of frequency ω (not ω_0). The conditions of Eq. (6.4) actually define an entire class of Hamiltonians with the desired properties.

The proper way to modify the Hamiltonian depends on the particular problem one is solving, and on what particular features of the actual problem one considers most realistic, and worth preserving. So without attempting to discuss the problem in any generality, we will merely outline two solutions as representative. First, the simplest thing to do is, at $t = t_n$, to abruptly change the form of $\mathscr{H}_n$ to H^0,

$$\mathscr{H}\sum a_n|n\rangle = \sum a_n \mathscr{H}_n \bar{\theta}(t - t_n)|n\rangle + \sum a_n H^0 \theta(t - t_n)|n\rangle \tag{6.5}$$

[θ is the step function, $\theta(x) = 1$, $x > 0$; $\theta(x) = 0$, $x < 0$, and $\bar{\theta} = 1 - \theta$]. Then at $t = t_n$, one will have to express the state $|n\rangle$ in terms of the eigenstates of H^0, denoted by $|l\rangle_0$,

$$|n\rangle = \sum b_l^{(n)} |l\rangle_0. \tag{6.6}$$

Actually, one is guaranteed that by time t_n, most of the contribution to this sum will come from the few lowest states of H^0, as required by consistency. One can see this first by the fact that the time has been specifically chosen so the average energy will now be close to ϵ_0, so that the lowest states must dominate the expansion. However one could also look at $\langle x^2\rangle$. In the initial state, $\langle x^2\rangle_n \sim n\langle x^2\rangle_0$, from the properties of the Hermite polynomials. But in time t_n the factor $m^{1/2}$ in the argument of H_n will effectively create the proper time dependence, $e^{-2\gamma t}$, which will bring down the width by a factor of n, from Eq. (6.2). These proofs are important in the case of high n, the semiclassical states. For small n, the states are already of the appropriate energy and width.

The second solution method we will discuss, is to alter the Hamiltonian continuously. For example, instead of the perturbation $\mathscr{H}_I = -2\gamma m\tau$, one could have

$$\mathscr{H}'_I = -2\gamma m(1 - m/m_1)\tau. \tag{6.7}$$

This replaces γ by

$$\gamma_{\text{eff}} = \gamma(1 - m/m_1). \tag{6.8}$$

At $t = 0$, one chooses m_1 such that $m \ll m_1$, so that $\gamma_{\text{eff}} \sim \gamma$. But as m approaches m_1, $\gamma \to 0$. The constant m_1 is then chosen appropriately, $m_1 = f(n)$, so that this will happen at time t_n. The classical solution for the Hamiltonian (6.7) is

$$m/m_0 = (m_1/\Delta)e^{2\gamma t}/(1 + (m_0/\Delta)e^{2\gamma t}) \tag{6.9}$$

where

$$\Delta \equiv m_1 - m_0 \approx m_1 \text{ for } n \gg 1. \tag{6.10}$$

For small and large times,

$$\begin{aligned} &m \approx m_0 e^{2\gamma t}, \quad t \ll t_0, \\ &m \approx m_1(1 - e^{-2\gamma(t - t_0)}), \quad t \gg t_0. \end{aligned} \tag{6.11}$$

The transition region is determined by the point of inflection, $\ddot{m} = 0$, given by t_0,

$$e^{2\gamma t_0} = \Delta/m_0 \tag{6.12}$$

So if

$$\Delta/m_0 = \epsilon_n/\epsilon_0 \sim 2n + 1, \tag{6.13}$$

then

$$t_0 = t_n. \tag{6.14}$$

The quantum solution to the Hamiltonian of Eq. (6.7) starts by replacing $m\tau$ by its symmetrical equivalent, Eq. (5.4), and the term $m^2\tau$, by its symmetrical equivalent,

$$\tfrac{1}{2}(m^2\tau + \tau m^2) = \left(\frac{\hbar}{i}\right)(m^2\partial/\partial m + m). \tag{6.15}$$

(One could weight in any amount of the term $m\tau m$ without altering this result!) This would replace the function M of Eq. (5.15) by

$$M = \text{const} \times \frac{m^{(i\beta - 1)/2}}{(m_1 - m)^{(i\beta + 1)/2}} \tag{6.16}$$

which produces a wavepacket centered about the solution to the classical equation (6.9), in the same way as the function M of Sec. V was related to its classical behavior.

One final remark concerns the appearance of the constant ω_0. This is the frequency of our original unperturbed Hamiltonian, $\omega_0^2 = \omega^2 + \gamma^2$. But we want our final undamped Hamiltonian to have frequency ω. So ω_0 can no longer be considered a constant but must be interpreted as a function of m,

$$\omega_0^2(m) = \omega^2 + \gamma_{\text{eff}}^2. \tag{6.17}$$

Then

$$\begin{aligned} &\omega_0(m) \to \omega_0, \quad t \ll t_n, \\ &\omega_0(m) \to \omega, \quad t \gg t_n. \end{aligned} \tag{6.18}$$

Since we are working in a representation where m is diagonal, this will still be a number, not an operator, so this offers no complications. The evolution of the variable τ is made quite complicated by altering $\mathscr{H}$, however, we are basically not interested in its behavior.

What have we accomplished by modifying the Hamiltonian in this manner? We have noted that each state takes a specific time to decay to a minimum wavepacket, at which time it approximates a particle in the ground state of the undamped oscillator. Beyond this point the particle will not decay further, because of the uncertainty principle, and so we have effectively replaced the Hamiltonian for later times by that of an undamped oscillator. Thus the wavepacket will never shrink beyond being a minimal wavepacket, for a particle of constant mass, and one can apply the constant mass reinterpretation to the modified Hamiltonian, as outlined at the end of Sec. V.

We remind the reader that by using this reinterpretation, and introducing the dissipation variable η, one need never talk about a changing mass at all. The mass of the particle is m_0 always, and the behavior of the variable η regulates the degree of dissipation present. And thus the Hamiltonians (6.4) or (6.7), where m is replaced by $m_0\eta$, will refer to systems of constant mass showing a physically reasonable behavior.

We might note that if the wavepacket is narrow in mass, as it must be if we want to maintain a constant mass interpretation, then there are certain approximations one can use in this case. For example, one could alternatively use the Kanai definition, $p_0 \approx e^{-2\gamma t}p$, for times $t < t_n$, and $p_0 \approx (m_0/m_1)p$, for times $t > t_n$, in the discontinuous case. In the continuous case, one could use $p_0 \approx e^{-2\gamma t}p$, for times $t \ll t_n$, and $p_0 \approx (m_0/m_1)p$, for times $t \gg t_n$, while for $t \sim t_n$, these approximations are not valid.

VII. THE DAMPED FREE PARTICLE

We close with a brief treatment of the force-free damped particle, in the variable mass formalism, because the solution we are interested in does not quite parallel that of the oscillator. The Kanai solution to the free damped particle, with the Hamiltonian

$$H = (p^2/2m)e^{\gamma t}, \tag{7.1}$$

is treated, among other places, in Refs. 1 and 3. The variable mass treatment of the problem starts from the Hamiltonian

$$\mathscr{H} = p^2/2m - \gamma m\tau, \tag{7.2}$$

which leads to the time independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2}\right) + \frac{\gamma\hbar}{i}\left(m\frac{\partial}{\partial m} + \frac{1}{2}\right)\psi = -\frac{\hbar}{i}\frac{\partial\psi}{\partial t}. \tag{7.3}$$

The distance scale in this problem is set by

$$\lambda = (\hbar/m\gamma)^{1/2}, \tag{7.4}$$

in terms of which the equation becomes

$$-\frac{\lambda^2}{2}\frac{\partial^2\psi}{\partial x^2} + \frac{1}{i}\left(m\frac{\partial}{\partial m} + \frac{1}{2}\right)\psi = -\frac{1}{i\gamma}\frac{\partial\psi}{\partial t}. \tag{7.5}$$

The length λ has a similar definition to that in the oscillator case, and we can use the substitution (5.8) (reading γ for ω_0), which yields the equation

$$\left[\left(-\frac{1}{2}\frac{\partial^2}{\partial y^2} + \frac{1}{2i}y\frac{\partial}{\partial y}\right) + \frac{1}{i}\left(m\frac{\partial}{\partial m} + \frac{1}{2}\right)\right]\psi$$

$$= -\frac{1}{i\gamma}\frac{\partial\psi}{\partial t}. \tag{7.6}$$

The m solutions are exactly as before, and we can write

$$\psi = M(m)\varphi(y,t)e^{-i\beta\gamma t/2}, \tag{7.7}$$

where M is given by Eqs. (5.13) and (5.15). However, in order to produce a solution directly comparable to the wavepacket formulation of the Kanai solution,[1] we proceed not by directly solving Eq. (7.6), but by first making the substitution

$$z = ye^{-\gamma t/2}. \tag{7.8}$$

Then for the variables z and t, we have

$$\frac{\partial}{\partial y} \to e^{\gamma t/2}\frac{\partial}{\partial z} = \frac{z}{y}\frac{\partial}{\partial z}, \quad \frac{\partial}{\partial t} \to -\frac{\gamma}{2}z\frac{\partial}{\partial z} + \frac{\partial}{\partial t}, \tag{7.9}$$

and Eq. (7.6) becomes

$$-\tfrac{1}{2}e^{\gamma t}\frac{\partial^2\varphi}{\partial z^2} = -\frac{1}{i\gamma}\frac{\partial\varphi}{\partial t}, \tag{7.10}$$

which is exactly the quantum Kanai equation, Ref. 1, Eq. (11). The solution is

$$\varphi(z,u) = \int dz'\, G(z-z',u)\varphi_0(z'), \tag{7.11}$$

where

$$\varphi_0(z) = \varphi(z,0), \tag{7.12}$$

the initial value of φ, and

$$G(z,u) = (2\pi iu)^{-1/2}e^{iz^2/2u}, \tag{7.13}$$

the Green's function for the problem, and the variable u is given by

$$u = 1 - e^{-\gamma t}.$$

So finally

$$\psi_1(z,m,u,t) = (4\pi m)^{-1/2}\int d\beta\,(m/m_0)^{i\beta/2}e^{-i\beta\gamma t/2} \times \int dz'\, G(z-z',u)\varphi_0^{(\beta)}(z'), \tag{7.14}$$

where one may arbitrarily choose a different function $\varphi_0^{(\beta)}$ for each value of β, and

$$\psi(x,m,t) = \psi_1[(xe^{-\gamma t/2})/\lambda(m),m,1-e^{-\gamma t},t]. \tag{7.15}$$

The β dependence of $\varphi_0^{(\beta)}(z')$ plays the same role as the function $a(\beta)$ appearing in Sec. V and the β dependence may be integrated first.

As with the oscillator, the center of each packet obeys the "classical" equation, $m = m_0e^{\gamma t}$. Also the explicit time dependence in the term $(xe^{-\gamma t/2})/\lambda$, plays the role of cancelling the time dependence due to the $m^{1/2}$ which appears in λ. So each separate mass piece of the solution behaves similarly to the Kanai solutions, but the freedom to create mass superpositions gives far greater flexibility to the total solution.

As the solution damps out, the considerations of Sec. VI are applicable here, and can be used to bring the particle to rest without further damping, and provide for a constant mass reinterpretation. However, for the free particle case the physical considerations must be very carefully taken into account in any particular problem, as pointed out in Ref. 1, Sec. IV.

VIII. SUMMARY

The standard approach to the damped harmonic oscillator, via the Kanai Hamiltonian, actually treats a particle of varying mass. The mass still plays the role of an external parameter in that theory, albeit a nonconstant one. We have gone a step beyond this and shown how one may treat the mass as a dynamical variable. Thus our solution consists of a superposition of "mass packets," the expectation value of which varies as the Kanai particle mass.

Classically, both these Hamiltonians have an alternate interpretation as a particle of constant mass subject to a damping force. However this interpretation breaks down quantum mechanically, because in both cases such an interpretation leads to an ultimate violation of the uncertainty principle. The mechanism for this is that as the particle keeps damping, its wavefunction keeps narrowing indefinitely, both in configuration and momentum space.

Nonethless, both the Kanai Hamiltonian and ours may be modified, so as to provide a viable constant mass reinterpretation. This procedure makes use of the fact that beyond a certain point, it is physically impossible to detect that the particle is still damping, and so the damping force may be safely turned off. We have also given a specific procedure for carrying out such a constant mass reinterpretation of the Hamiltonian, provided the mass wavepacket is narrow. In fact, we show that by introducing the dissipation variable it becomes possible to avoid mention of changing masses altogether. Thus, through the use of the modified Hamiltonian, the method allows one to treat the practical problem of the damped constant mass particle. In this regard, our variable mass method is more flexible than the Kanai Hamiltonian, and has the mathematical advantage of being time independent. It is also linear, of course, which gives it important advantages over the various nonlinear approaches to the problem.[7]

ACKNOWLEDGMENT

Research supported (in part) by a grant from the PSC-BHE Research Award Program of the City University of New York.

[1] D.M. Greenberger, J. Math. Phys. **20**, 762 (1979).
[2] E. Kanai, Prog. Theor. Phys. **3**, 440 (1948).
[3] The entire subject is treated in a review article by Hasse, and the reader is referred there for thorough references to work cited; R.W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[4] No details will be needed for this paper, beyond the treatment given here. However the formalism was developed in D.M. Greenberger, J. Math. Phys. **11**, 2329, 2341 (1970); **15**, 395, 406 (1974).
[5] We could have alternatively made a series of canonical transformations to "scale" the time decay out of the problem, and then diagonalize in the resulting undamped problem. See E.H. Kerner, Can. J. Phys. **36**, 371 (1958). The operators D, D^* are intimately related to the a, a^* of that problem.
[6] This approach was pioneered by L. Tisza, and is discussed in his book, L. Tisza, *Generalized Thermodynamics* (MIT Press, Cambridge, Mass., 1966), and in that of H.B. Callen, *Thermodynamics* (Wiley, New York, 1960).
[7] These are discussed in Ref. 1.

Upper bounds for the many-channel Marchenko transformation operator and its derivatives

Pierre Rochus[a)]

Physique Nucléaire Théorique, Institut de Physique, Université de Liège au Sart Tilman, B-4000 Liège 1, Belgium
(Received 4 August 1978)

We specify the class of perturbative complex matrix potentials for which the corresponding many-channel Marchenko type transformation operators are bounded and integrable. Our reference matrix potential contains Coulomb interactions, different threshold energies, and centrifugal potentials with different angular momenta. Estimates for the transformation operator and its derivatives are obtained; they enable us to improve our recent results and are necessary for the establishment of a unique solution to the "generalized Marchenko fundamental equation." From the existence of an integrable transformation operator, the analyticity of the Jost solution as a function of k_1 is deduced in the upper-half of the physical k_1 plane.

1. INTRODUCTION

The transformation operator plays a prominent part in the theory of the inverse problem of scattering: Indeed, the starting point of the method developed by Agranovich and Marchenko[1] is to specify the class of potentials for which the existence of a bounded transformation operator can be proved. Within this class of potentials, a fundamental equation for the inverse problem is then derived. In the first part of their monograph, however, they only considered coupled channels without either the centrifugal terms or Coulomb interactions and with the same threshold energy in each channel. In the second part of their book, an indirect iterative approach for the singular centrifugal part was then used via the transformations introduced by Crum and Krein.[2] We consider here a system of differential equations containing different (and even not necessarily integer) angular momenta, different threshold energies and Coulomb interactions, by a direct method using the Riemann–Green solution.[3]

Cox[4] considered a system of coupled channels with different threshold energies. He was able to apply the method of Jost and Kohn[5] and to get a generalized Gel'fand–Levitan equation. However, he did not say for which potentials his equation is valid. His Gel'fand transformation operator is not necessarily a bounded function; it could happen that the transformation operator is only defined as a distribution. In that case, it is only useful if we can show that its diagonal part is bounded so that a well-behaved potential can be obtained by differentiation. Therefore, we want to find the class of perturbative potentials for which a transformation operator exists in the sense of functions theory and not in the enlarged sense of distributions. We only consider Marchenko's type transformation operator for the reasons explained in Ref. 6, and upper bounds for this operator and its derivatives are obtained. These bounds are necessary for the establishment of a stable and unique solution to the generalized fundamental Marchenko equation.[7] The recent results obtained in Refs. 3, 6, 8 must be modified according to our new estimate of the Riemann–Green solution. The present paper is divided into five sections: The introduction forms the first section. In Sec. 2, the definition of a transformation operator and the use of the Riemann–Green functions are briefly recalled. New estimates for the Riemann–Green solution are found in Sec. 3. In Sec. 4, the results for the Riemann–Green function enable us to obtain new upper bounds for the transformation operator and to specify the class of perturbative potentials for which a bounded integrable transformation operator exists. From the existence of an integrable transformation operator, the analyticity of the Jost solution in the upper half of the physical k_1 plane is shown in Sec. 5. The paper includes five appendices. In Appendix A, a spectral representation is derived for the complete Riemann–Green solution. Upper bounds for the unperturbated Jost solution, the derivatives of the unperturbated, and the complete Riemann–Green solution are obtained in Appendices B, C, D, respectively. The derivatives of the transformation operator are estimated in Appendix E.

2. THE TRANSFORMATION OPERATOR AND THE USE OF THE RIEMANN–GREEN SOLUTION

Two systems L_0 and L of n coupled differential equations are defined by the following two equations:

$$L_0(x)\phi_0(\Lambda,x) = \left[\frac{d^2}{dx^2} I + \Lambda - V_0(x)\right]\phi_0(\Lambda,x) \tag{1}$$

and

$$L(x)\phi(\Lambda,x) = \left[\frac{d^2}{dx^2} I + \Lambda - V_0(x) - V(x)\right]\phi(\Lambda,x), \tag{2}$$

where I, Λ, V_0, V stand for the unity matrix, the diagonal matrix of different channel wave numbers k_i^2 $(i = 1,n)$, the reference matrix potential and the perturbative matrix potential respectively. The reference diagonal potential contains the usual singularities: The centrifugal potential and the Coulomb interaction, while the perturbative potential is allowed to be complex (non-Hermitian). The Jost matrix solutions ϕ_0 and ϕ satisfy the same boundary conditions:

[a)]Chercheur I.I.S.N.

 0022-2488/79/050781-12$01.00

$$\lim_{x\to\infty}[\phi(\Lambda,x)]_{ij}$$

$$= \lim_{x\to\infty}[\phi_0(\Lambda,x)]_{ij}$$

$$= \delta_{ij}\exp\{i[k_j x - l_j\pi/2 + \sigma_j - (\alpha_j/2k_j)\ln(2k_j x)]\}, \tag{3}$$

where the Coulomb phase σ_j is defined by Eq. (4):

$$e^{2i\sigma_l} = \frac{\Gamma(l_j+1+i\alpha_j/2k_j)}{\Gamma(l_j+1-i\alpha_j/2k_j)}. \tag{4}$$

ϕ_0 is the irregular Coulomb diagonal matrix defined by Eq. (5):

$$[\phi_0(\Lambda,x)]_{ij} = \delta_{ij}W_{-i\alpha_i/2k_i,l_i+1/2}(-2ik_i x)$$

$$\times\exp[-\tfrac{1}{2}i\pi(l_i+1+i\alpha_i/2k_i)+i\sigma_i], \tag{5}$$

where W denotes the Whittaker's function (see Ref. 9). From Ref. 10, we know that when the Coulomb interaction is attractive, the reference and the perturbated problems have both an infinite number of bound states.

We are looking for a possible integral representation of the form:

$$\phi(\Lambda,x) = \phi_0(\Lambda,x) + \int_x^\infty K(x,t)\phi_0(\Lambda,t)\,dt, \tag{6}$$

when Λ belongs to the spectrum of $L\phi = 0$ and where $K(x,t)$ is the transformation kernel. This kernel $K(x,t)$ is connected with the solution of the inverse problem by the equation:

$$-2\frac{d}{dx}K(x,x) = V(x). \tag{7}$$

We want to specify the class of perturbative potentials V for which such a continuous bounded kernel exists.

It is shown in Ref. 3 that the transformation matrix elements have to satisfy the partial differential system (8):

$$\left[\frac{\partial^2}{\partial x^2} + k_i^2 - \frac{l_i(l_i+1)}{x^2} - \frac{\alpha_i}{x}\right]K_{ij}(x,y)$$

$$-\sum_l V_{il}(x)K_{lj}(x,y)$$

$$= \left[\frac{\partial^2}{\partial y^2} + k_j^2 - \frac{l_j(l_j+1)}{y^2} - \frac{\alpha_j}{y}\right]K_{ij}(x,y), \tag{8a}$$

$$\lim_{y\to\infty}K_{ij}(x,y) = \lim_{y\to\infty}\frac{\partial}{\partial y}K_{ij}(x,y) = 0, \tag{8b}$$

$$K_{ij}(x,x) = \tfrac{1}{2}\int_x^\infty V_{ij}(s)\,ds, \quad i,j = 1,n. \tag{8c}$$

The partial differential system (8) is equivalent to the integral system (9):

$$K_{ij}(x,y) = \tfrac{1}{2}\int_{(x+y)/2}^\infty R_{ij}(x,y;s,s)V_{ij}(s)\,ds$$

$$+\tfrac{1}{2}\int_{\mathscr{D}}\int R_{ij}(x,y;s,u)\sum_l V_{il}(s)$$

$$\times K_{lj}(s,u)\,du\,ds, \quad i,j = 1,n. \tag{9}$$

In Eq. (9), $\mathscr{D}$ denotes the Marchenko domain represented in Fig. 1 and the $R_{ij}(x,y,s,u)$ are the Riemann–Green solutions, satisfying the partial differential equations (10):

$$\left[\frac{\partial^2}{\partial x^2} + k_i^2 - \frac{l_i(l_i+1)}{x^2} - \frac{\alpha_i}{x}\right]R_{ij}(x,y;s,u)$$

$$= \left[\frac{\partial^2}{\partial y^2} + k_j^2 - \frac{l_j(l_j+1)}{y^2} - \frac{\alpha_j}{y}\right]R_{ij}(x,y;s,u), \tag{10a}$$

$$R_{ij}(x,y;s,u) = 1 \quad \text{if} \quad |x-s| = |y-u|. \tag{10b}$$

If we use the canonical variables

$$\eta = \frac{x+y}{2}, \quad \xi = \frac{y-x}{2}, \quad \eta_0 = \frac{u+s}{2}, \quad \xi_0 = \frac{u-s}{2}, \tag{11}$$

Eq. (10) reads

$$\left[\frac{\partial}{\partial\eta\,\partial\xi} - \frac{l_j(l_j+1)}{(\eta+\xi)^2} - \frac{\alpha_j}{(\eta+\xi)} + k_j^2 + \frac{l_i(l_i+1)}{(\eta-\xi)^2} + \frac{\alpha_i}{(\eta-\xi)} - k_i^2\right]$$

$$\times R_{ij}(\eta,\xi;\eta_0,\xi_0) = 0, \tag{12a}$$

$$R_{ij}(\eta,\xi;\eta_0,\xi_0) = 1 \quad \text{if} \quad \eta = \eta_0, \quad \text{or} \quad \xi = \xi_0. \tag{12b}$$

In Appendix A, the spectral representation (A10)–(A11) is obtained for $R_{ij}(x,y;s,u)$, using the techniques of Ref. 11. However, we are not able to write this spectral representation into a close form. Simple expressions for the Riemann functions $\mathscr{R}_{ij}$ and R^0_{ij} corresponding to the cases $l_i = \alpha_i = 0$, $i = 1,n$ and $k_i = \alpha_i = 0$, $i = 1,n$ respectively, can be deduced from this spectral representation:

$$\mathscr{R}_{ij}(x,y;s,u) = \mathscr{J}_0(\{\Delta_{ij}^2[(x-s)^2 - (y-u)^2]\}^{1/2}), \tag{13}$$

where $\mathscr{J}_0$ is the Bessel function of zero order and

$$\Delta_{ij}^2 = k_i^2 - k_j^2, \tag{14}$$

and

$$R^0_{ij}(x,y;s,u) = P_l(1-2x_2)$$

$$-2x_1\int_0^1 P_l(1-2x_2+2x_2t)P_l'(1-2x_1t)\,dt, \tag{15}$$

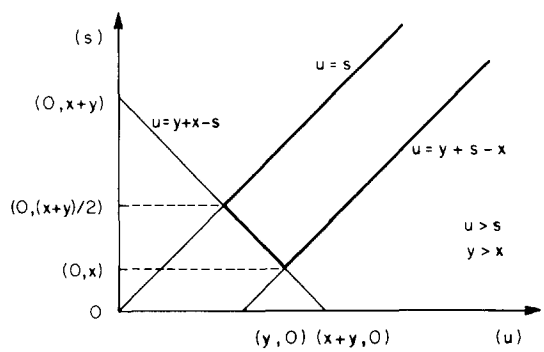

FIG. 1. Marchenko domain.

where

$$x_1 = \frac{(\eta-\eta_0)(\xi-\xi_0)}{(\eta-\xi)(\eta_0-\xi_0)}, \tag{16}$$

$$x = \frac{(\eta_0-\eta)(\xi-\xi_0)}{(\eta+\xi)(\eta_0+\xi_0)}, \tag{17}$$

and P_{l_i} is the Legendre polynomial.

This last function $R^{\,0}_{ij}$ will be used in the next section in order to evaluate the complete Riemann solution R_{ij}.

3. UPPER BOUNDS FOR THE FUNCTIONS R^{0}_{ij} AND R_{ij}

In the domain $\mathscr{D}$ of Marchenko, the following inequalities are satisfied:

$$0\leqslant\xi_0\leqslant\xi\leqslant\eta\leqslant\eta_0\leqslant\infty. \tag{18}$$

From Eqs. (16) and (17), one easily gets

$$x_1<0 \tag{19}$$

and

$$0\leqslant x_2\leqslant\frac{\eta_0\xi}{(\eta+\xi)(\eta_0+\xi_0)}\leqslant\frac{\eta_0\xi}{2\eta_0\xi}\leqslant\frac{1}{2}$$

or

$$0\leqslant 1-2x_2\leqslant 1. \tag{20}$$

Our next step is to prove the inequality (21):

$$1\leqslant 1-2x_1=1+2\frac{(\eta_0-\eta)(\xi-\xi_0)}{(\eta-\xi)(\eta_0-\xi_0)}\leqslant\frac{2(\eta+\xi)(\eta_0-\xi_0)}{(\eta-\xi)(\eta_0+\xi_0)}. \tag{21}$$

After multiplication of both sides of Eq. (21) by $(\eta-\xi)(\eta_0+\xi_0)/(\eta+\xi)(\eta_0-\xi_0)\geqslant 0$, we must show that

$$E(\eta,\xi;\eta_0,\xi_0) = \frac{(\eta-\xi)(\eta_0+\xi_0)}{(\eta+\xi)(\eta_0-\xi_0)}+2\frac{(\eta_0+\xi_0)(\eta_0-\eta)(\xi-\xi_0)}{(\eta+\xi)(\eta_0-\xi_0)^2}\leqslant 2. \tag{22}$$

The following inequalities are successively obtained:

$$E(\eta,\xi;\eta_0,\xi_0)\leqslant\frac{(\eta-\xi)(\eta_0+\xi_0)}{(\eta+\xi)(\eta_0-\xi_0)}+2\frac{(\eta_0+\xi_0)(\xi-\xi_0)}{(\eta+\xi)(\eta_0-\xi_0)} \tag{23}$$

$$\leqslant\frac{(\eta+\xi)(\eta_0+\xi_0)-2\xi_0(\eta_0+\xi_0)}{(\eta+\xi)(\eta_0-\xi_0)} \tag{24}$$

$$\leqslant\frac{(\eta+\xi)(\eta_0-\xi_0)+2\xi_0(\eta+\xi-\eta_0-\xi_0)}{(\eta+\xi)(\eta_0-\xi_0)} \tag{25}$$

$$\leqslant 1+2\frac{\xi_0(\eta+\xi-\eta_0-\xi_0)}{(\eta+\xi)(\eta_0-\xi_0)}. \tag{26}$$

As

$$\frac{\partial}{\partial\xi}\left(\frac{\eta+\xi-\eta_0-\xi_0}{\eta+\xi}\right)=\frac{\eta_0+\xi_0}{(\eta+\xi)^2}>0,$$

the right-hand side of Eq. (26) is certainly overestimated, if we replace ξ by $\eta>\xi$. Finally, we get the requested result (21):

$$E(\eta,\xi;\eta_0,\xi_0)\leqslant 1+\frac{\xi_0}{\eta}\frac{(2\eta-\eta_0-\xi_0)}{(\eta_0-\xi_0)}$$

$$\leqslant 1+\frac{\xi_0}{\eta}\frac{(\eta-\xi_0)}{(\eta_0-\xi_0)}\leqslant 2. \tag{27}$$

We are looking now for an upper bound for $|R^{\,0}_{ij}|$. Starting from its representation (15), we get

$$|R^{\,0}_{ij}|\leqslant|P_{l_j}(1-2x_2)| + \left|\int_0^1 P_{l_j}(1-2x_2+2x_2t)\,dP_{l_i}(1-2x_1t)\right|. \tag{28}$$

Since

$$0\leqslant 1-2x_2(1-t)\leqslant 1$$

and

$$1\leqslant 1-2x_1t\leqslant 2\frac{(\eta+\xi)(\eta_0-\xi_0)}{(\eta-\xi)(\eta_0+\xi_0)}$$

for $0\leqslant t\leqslant 1$, we can write

$$|R^{\,0}_{ij}(\eta,\xi;\eta_0,\xi_0)| \leqslant 1+|P_{l_i}(1-2x_1)-1| \leqslant P_{l_i}(1-2x_1)\leqslant\left[4\frac{(\eta+\xi)(\eta_0-\xi_0)}{(\eta-\xi)(\eta_0+\xi_0)}\right]^{l_i}, \tag{29}$$

where the following property has been used:

$$1\leqslant P_l(x)\leqslant(2x)^l \quad \text{for } x>1 \text{ and } l \text{ real} \geqslant 0.$$

We can construct the complete Riemann function R_{ij} from the function $R^{\,0}_{ij}$, using the composition formula

$$R_{ij}(\eta,\xi;\eta_0,\xi_0) = R^{\,0}_{ij}(\eta,\xi;\eta_0,\xi_0)+\int_{\xi_0}^{\xi}d\xi_1\int_{\eta}^{\eta_0}d\eta_1\,R^{\,0}_{ij}(\eta,\xi;\eta_1,\xi_1) \times\left(k_i^2-k_j^2-\frac{\alpha_i}{\eta_1-\xi_1}+\frac{\alpha_j}{\eta_1+\xi_1}\right)R_{ij}(\eta_1,\xi_1;\eta_0,\xi_0). \tag{30}$$

The bound (29) for $|R^{\,0}_{ij}|$ is then used to find an upper bound for $|R_{ij}|$. Setting

$$\tilde{R}^{\,0}_{ij}(\eta,\xi;\eta_0,\xi_0)=R^{\,0}_{ij}(\eta,\xi;\eta_0,\xi_0)\left[\frac{(\eta-\xi)(\eta_0+\xi_0)}{4(\eta+\xi)(\eta_0-\xi_0)}\right]^{l_i}, \tag{31}$$

$$\tilde{R}_{ij}(\eta,\xi;\eta_0,\xi_0)=R_{ij}(\eta,\xi;\eta_0,\xi_0)\left[\frac{(\eta-\xi)(\eta_0+\xi_0)}{4(\eta+\xi)(\eta_0-\xi_0)}\right]^{l_i}, \tag{32}$$

$$\mu_{ij}^2=4^{l_i}|k_i^2-k_j^2| \tag{33}$$

and

$$a_{ij}=2^{(2l_i+1)}\max(\alpha_i,\alpha_j), \tag{34}$$

Eq. (30) can be written as

$$\tilde{R}_{ij}(\eta,\xi;\eta_0,\xi_0) = \tilde{R}^{\,0}_{ij}(\eta,\xi;\eta_0,\xi_0)+\int_{\xi_0}^{\xi}d\xi_1\int_{\eta}^{\eta_0}d\eta_1\,\tilde{R}^{\,0}_{ij}(\eta,\xi;\eta_1,\xi_1) \times 4^{l_i}\left[(k_i^2-k_j^2)+\left(\frac{-\alpha_i}{\eta_1-\xi_1}+\frac{\alpha_j}{\eta_1+\xi_1}\right)\right] \times\tilde{R}_{ij}(\eta_1,\xi_1;\eta_0,\xi_0), \tag{35}$$

where

$$|\tilde{R}^{0}_{ij}(\eta,\xi;\eta_0,\xi_0)| \leqslant 1. \tag{36}$$

The successive approximations applied to the absolute value of Eq. (35) lead to

$$|\tilde{R}^{(0)}_{ij}(\eta,\xi;\eta_0,\xi_0)| \leqslant |\tilde{R}^{0}_{ij}(\eta,\xi;\eta_0,\xi_0)| \leqslant 1, \tag{37}$$

$$\begin{aligned}
&|\tilde{R}^{(1)}_{ij}|(\eta,\xi,\eta_0,\xi_0) \\
&\quad = \left| \int_{\xi_0}^{\xi} d\xi_1 \int_{\eta}^{\eta_0} d\eta_1\, \tilde{R}^{0}_{ij}(\eta,\xi;\eta_1,\xi_1) \right. \\
&\qquad \times 4^{l_i}\left(k_i^2 - k_j^2 + \frac{\alpha_j}{\eta_1+\xi_1} - \frac{\alpha_i}{\eta_1-\xi_1}\right) \\
&\qquad \left. \times \tilde{R}^{(0)}_{ij}(\eta_1,\xi_1;\eta_0,\xi_0) \right| \\
&\quad \leqslant \int_{\xi_0}^{\xi} d\xi_1 \int_{\eta}^{\eta_0} d\eta_1 \left(\mu_{ij}^2 + \frac{a_{ij}\eta_1}{\eta_1^2 - \xi_1^2}\right) \\
&\quad \leqslant \int_{\xi_0}^{\xi} d\xi_1 \int_{\eta}^{\eta_0} d\eta_1 \left(\mu_{ij}^2 + \frac{a_{ij}}{\eta_1 - \xi_1}\right) \\
&\quad \leqslant \int_{\xi_0}^{\xi} d\xi_1 \int_{\eta}^{\eta_0} d\eta_1 \frac{[\mu_{ij}^2(\eta_0-\xi_0) + a_{ij}]}{(\eta-\xi_1)^{1/2}(\eta_1-\xi)^{1/2}},
\end{aligned} \tag{38}$$

$$\begin{aligned}
&|\tilde{R}^{(p)}_{ij}(\eta,\xi;\eta_0,\xi_0)| \\
&\quad = \left| \int_{\xi_0}^{\xi} d\xi_1 \int_{\eta}^{\eta_0} d\eta_1\, \tilde{R}^{0}_{ij}(\eta,\xi;\eta_1,\xi_1) \right. \\
&\qquad 4^{l_i}\left(k_i^2 - k_j^2 + \frac{\alpha_j}{\eta_1+\xi_1} - \frac{\alpha_i}{\eta_1-\xi_1}\right) \\
&\qquad \times \int_{\xi_0}^{\xi_1} d\xi_2 \int_{\eta_1}^{\eta_0} d\eta_2\, \tilde{R}^{0}_{ij}(\eta_1,\xi_1;\eta_2,\xi_2) \\
&\qquad \times 4^{l_i}\left(k_i^2 - k_j^2 + \frac{\alpha_j}{\eta_2+\xi_2} - \frac{\alpha_i}{\eta_2-\xi_2}\right)\cdots \\
&\qquad \times \int_{\xi_0}^{\xi_{p-1}} d\xi_p \int_{\eta_{p-1}}^{\eta_0} d\eta_p\, \tilde{R}^{0}_{ij}(\eta_{p-1},\xi_{p-1};\eta_p,\xi_p) \\
&\qquad \times 4^{l_i}\left(k_i^2 - k_j^2 + \frac{\alpha_j}{\eta_p+\xi_p} - \frac{\alpha_i}{\eta_p-\xi_p}\right) \\
&\qquad \left. \tilde{R}^{(0)}_{ij}(\eta_p,\xi_p;\eta_0,\xi_0) \right| \\
&\quad \leqslant [\mu_{ij}^2(\eta_0-\xi_0) + a_{ij}]^p \int_{\xi_0}^{\xi} d\xi_1 \frac{1}{(\eta-\xi_1)^{1/2}} \\
&\qquad \times \int_{\xi_1}^{\xi_0} d\xi_1 \frac{1}{(\eta-\xi_2)^{1/2}} \cdots \int_{\xi_{p-1}}^{\xi_0} d\xi_p \frac{1}{(\eta-\xi_p)^{1/2}} \\
&\qquad \times \int_{\eta}^{\eta_0} d\eta_1 \frac{1}{(\eta_1-\xi)^{1/2}} \int_{\eta_1}^{\eta_0} d\eta_2 \frac{1}{(\eta_2-\xi)^{1/2}} \cdots \\
&\qquad \times \int_{\eta_{p-1}}^{\eta_0} d\eta_p \frac{1}{(\eta_p-\xi)^{1/2}}
\end{aligned}$$

(notice that for $p = 0$, $\xi_p = \xi$, and $\eta_p = \eta$).

$$\begin{aligned}
&|R^{(p)}_{ij}(\eta,\xi;\eta_0,\xi_0)| \\
&\quad \leqslant [\mu_{ij}^2(\eta_0-\xi_0) + a_{ij}]^p \\
&\qquad \times \frac{[\int_{\xi_0}^{\xi} d\xi_1/(\eta-\xi_1)^{1/2}]^p}{p!} \frac{[\int_{\eta_0}^{\eta} d\eta_1/(\eta_1-\xi)^{1/2}]^p}{p!}.
\end{aligned} \tag{39}$$

Since the modified Bessel function can be defined by

$$\mathcal{I}_0(z) = \sum_{p=0}^{\infty} \frac{(z^2/4)^p}{(p!)^2} = \frac{1}{\pi}\int_0^{\pi} e^{z\cos\theta}\, d\theta \leqslant e^z, \tag{40}$$

we easily get the following upper bounds:

$$\begin{aligned}
&|R_{ij}(\eta,\xi;\eta_0,\xi_0)| \\
&\quad \leqslant 4^{l_i}\left[\frac{(\eta+\xi)}{(\eta-\xi)}\frac{(\eta_0-\xi_0)}{(\eta_0+\xi_0)}\right]^{l_i} \mathcal{I}_0\left[2\left([\mu_{ij}^2(\eta_0-\xi_0) + a_{ij}] \right.\right. \\
&\qquad \left.\left. \times \int_{\xi}^{\xi_0} d\xi_1 \frac{1}{(\eta-\xi_1)^{1/2}} \int_{\eta}^{\eta_0} d\eta_1 \frac{1}{(\eta_1-\xi)^{1/2}} \right)^{1/2} \right] \\
&\quad \leqslant 4^{l_i}\left[\frac{(\eta+\xi)(\eta_0-\xi_0)}{(\eta-\xi)(\eta_0+\xi_0)}\right]^{l_i} \\
&\qquad \times \exp(4\{[\mu_{ij}^2(\eta_0-\xi_0) + a_{ij}](\eta_0-\xi_0)\}^{1/2}).
\end{aligned} \tag{41a}$$

The same method, directly applied to the case $\alpha_i = 0$, leads to

$$\begin{aligned}
&|R_{ij}(\eta,\xi;\eta_0,\xi_0)| \\
&\quad \leqslant 4^{l_i}\left[\frac{(\eta+\xi)(\eta_0-\xi_0)}{(\eta-\xi)(\eta_0+\xi_0)}\right]^{l_i} \exp[2\mu_{ij}(\eta_0-\xi_0)]
\end{aligned} \tag{42}$$

The inequality (42) differs from the inequality (41) with $a_{ij} = 0$, by a factor 2 in the exponent; this is due to the approximations we have done, in the evaluation of (41), in order to get separable integrands in Eqs. (38), (39). The estimate (41a) of R_{ij} is much more general and much easier to use than the estimate (58) of Ref. 3.

Several further approximations to Eq. (41a) can be performed:

(i) Since $\eta\xi_0 \leqslant \eta_0\xi$, we have

$$(\eta+\xi)(\eta_0-\xi_0)/(\eta-\xi)(\eta_0+\xi_0) \geqslant 1$$

and

$$\begin{aligned}
&|R_{ij}(\eta,\xi;\eta_0,\xi_0)| \\
&\quad \leqslant 4^{l_i}\left(\frac{(\eta+\xi)(\eta_0-\xi_0)}{(\eta_0+\xi_0)(\eta-\xi)}\right)^{l_{\max}} \\
&\qquad \times \exp(4\{[\mu_i^2(\eta_0-\xi_0) + a_i](\eta_0-\xi_0)\}^{1/2}),
\end{aligned} \tag{41b}$$

where

$$\mu_i = \max_j \mu_{ij}$$

and

$$a_i = \max_j a_{ij}$$

(ii) Since

$$\frac{\eta+\xi}{\eta_0+\xi_0} \leqslant \frac{\eta+\xi}{\eta_0} \leqslant \frac{\eta+\xi}{\eta} \leqslant 2,$$

we can also get from Eq. (41a)

$$|R_{ij}(\eta,\xi;\eta_0,\xi_0)|$$
$$\leqslant 8^{l_i}\left(\frac{\eta_0-\xi_0}{\eta_0+\xi_0}\right)^{l_i}$$
$$\times\exp(4\{[\mu_i^2(\eta_0-\xi_0)+a_i](\eta_0-\xi_0)\}^{1/2}). \tag{41c}$$

This inequality (41c) was the only one obtained in Ref. 2 by another method. In Sec. 4, estimates of K are obtained from the different approximations (41), and we explain why the use of inequality (41b) must be preferred.

4. AN UPPER BOUND FOR THE TRANSFORMATION OPERATOR

Using the canonical variables (11), the integral equation (9) for the transformation kernel reads

$$K_{ij}(\eta,\xi)=\tfrac{1}{2}\int_\eta^\infty d\eta_0\, R_{ij}(\eta,\xi;\eta_0,0)U_{ij}(\eta_0)$$
$$+\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\, R_{ij}(\eta,\xi;\eta_0,\xi_0)$$
$$\times\sum_l U_{il}((\eta_0-\xi_0)K_{lj}(\eta_0,\xi_0). \tag{43}$$

If we introduce, in Eq. (43), the matrices $\tilde R$, $\tilde K$, and $\tilde U$ defined by the following equations,

$$R_{ij}(\eta,\xi;\eta_0,\xi_0)$$
$$=\tilde R_{ij}(\eta,\xi;\eta_0,\xi_0)\left[\frac{(\eta+\xi)}{(\eta_0+\xi_0)}\,\frac{(\eta_0-\xi_0)}{(\eta-\xi)}\right]^{l_{\max}}$$
$$\times\exp(4\{[\mu_i^2(\eta_0-\xi_0)+a_i](\eta_0-\xi_0)\}^{1/2})4^l, \tag{44}$$

$$K_{ij}(\eta,\xi)=\tilde K_{ij}(\eta,\xi)\left(\frac{\eta+\xi}{\eta-\xi}\right)^{l_{\max}}, \tag{45}$$

$$\tilde U_{ij}(\eta)=4^l\exp(4\{[\mu_i^2(\eta_0-\xi_0)+a_i]$$
$$\times(\eta_0-\xi_0)\}^{1/2})U_{ij}(\eta), \tag{46}$$

where

$$\mu_i=\max_j(\mu_{ij}) \tag{47}$$

and

$$a_i=\max_j(a_{ij}), \tag{48}$$

we get

$$\tilde K_{ij}(\eta,\xi)=\tfrac{1}{2}\int_\eta^\infty d\eta_0\,\tilde R_{ij}(\eta,\xi;\eta_0,0)\tilde U_{ij}(\eta_0)$$
$$+\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\,\tilde R_{ij}(\eta,\xi;\eta_0,\xi_0)$$
$$\times\sum_l \tilde U_{il}(\eta_0-\xi_0)\tilde K_{lj}(\eta_0,\xi_0). \tag{49}$$

From Eq. (41b), we see that $|\tilde R_{ij}|\leqslant 1$ and successive approximations applied to the absolute value of Eq. (49) give

$$|\tilde K_{ij}^{(0)}(\eta,\xi)|\leqslant\tfrac{1}{2}\int_\eta^\infty d\eta_0|\tilde U_{ij}(\eta_0)|$$
$$\leqslant\tfrac{1}{2}\int_\eta^\infty d\eta_0\|\tilde U(\eta_0)\|=\tfrac{1}{2}\tilde\sigma^0(\eta), \tag{50}$$

where

$$\|A\|=\sup_i\sum_{j=1}^n|A_{ij}|,$$
$$|\tilde K_{ij}^{(1)}(\eta,\xi)|$$
$$\leqslant\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\sum_l|\tilde U_{il}(\eta_0-\xi_0)|\,|\tilde K_{lj}^{(0)}(\eta_0,\xi_0)|$$
$$\leqslant\tfrac{1}{2}\tilde\sigma^0(\eta)\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\|\tilde U(\eta_0-\xi_0)\|, \tag{51}$$
$$|\tilde K_{ij}^{(p)}(\eta,\xi)|$$
$$\leqslant\frac{1}{2}\frac{\tilde\sigma^0(\eta)}{p!}\left[\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\|\tilde U(\eta_0-\xi_0)\|\right]^p, \tag{52}$$

and finally

$$\|\tilde K(\eta,\xi)\|$$
$$\leqslant\frac{n}{2}\tilde\sigma^0(\eta)\exp\left[\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\|\tilde U(\eta_0-\xi_0)\|\right]. \tag{53}$$

Using the physical variables again, we get

$$\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\|\tilde U(\eta_0-\xi_0)\|$$
$$=\tfrac{1}{2}\int_x^{(x+y)/2}ds\int_{x+y-s}^{y-x+s}du\|\tilde U(s)\|$$
$$+\tfrac{1}{2}\int_{(x+y)/2}^\infty ds\int_s^{y-x+s}du\,\|\tilde U(s)\|$$
$$=\int_x^{(x+y)/2}ds\,\|\tilde U(s)\|(s-x)$$
$$+\tfrac{1}{2}\int_{(x+y)/2}^\infty ds\,\|\tilde U(s)\|(y-x)$$
$$\leqslant\int_x^\infty\|\tilde U(s)\|s\,ds=\tilde\sigma^1(x)$$

and also

$$\leqslant\tfrac{1}{2}(y-x)\tilde\sigma(x). \tag{54}$$

So we have

$$\|K(x,y)\|\leqslant\frac{n}{2}\left(\frac{y}{x}\right)^{l_{\max}}\tilde\sigma^0\left(\frac{x+y}{2}\right)$$
$$\times\exp\left[\tilde\sigma^0(x)\left(\frac{y-x}{2}\right)\right] \tag{55a}$$

and also

$$\|K(x,y)\|\leqslant\frac{n}{2}\left(\frac{y}{x}\right)^{l_{\max}}\tilde\sigma^0\left(\frac{x+y}{2}\right)\exp[\tilde\sigma^1(x)] \tag{55b}$$

$$\leqslant \frac{2^{l_{\max}-1}n}{x^{l_{\max}}}\tilde{\sigma}^{l_{\max}}\left(\frac{x+y}{2}\right)\exp[\tilde{\sigma}^1(x)] \tag{56}$$

if $\tilde{\sigma}^{l_{\max}}$ exists and where

$$\tilde{\sigma}^l\left(\frac{x+y}{2}\right) = \sup_i \sum_{j=1}^{n} \int_{(x+y)/2}^{\infty} \eta_0^l 4^{l_i} |U_{ij}(\eta_0)| \times \exp\{4[\mu_i^2(\eta_0-\xi_0)^2 + a_i(\eta_0-\xi_0)]^{1/2}\} d\eta_0, \tag{57}$$

$$\mu_i^2 = \max_j(4^{l_i}|k_i^2 - k_j^2|), \tag{58}$$

$$a_i = 2^{(2l_i+1)}\max_j(|\alpha_j|). \tag{59}$$

The same method with the use of estimate (41c) leads to the estimate

$$|K_{ij}(x,y)| \leqslant \frac{1}{2}\left(\frac{2}{x}\right)^{l_i}\tilde{\sigma}^{l_j}\left(\frac{x+y}{2}\right)\exp[\tilde{\sigma}^{(l_i-l_j+1)}(x)], \tag{60}$$

which corresponds to the one obtained in Ref. 3 and used in Refs. 6 and 8. However, from estimate (60), it is not obvious if $\lim_{x\to 0} K(x,x)$ is finite while it is so from estimate (55). From that point of view, Eq. (55) is much better. If we use estimate (41a), without any approximation, the method of successive approximations gives

$$|K_{ij}(x,y)| \leqslant \frac{1}{2}\left(\frac{y}{x}\right)^{l_i}\tilde{\sigma}^0\left(\frac{x+y}{2}\right)\exp[\|\Sigma(x,y)\|], \tag{61}$$

with

$$\Sigma_{ij}(x,y) = \tfrac{1}{2}\int_x^{(x+y)/2} ds \int_{x+y-s}^{y-x+s} du\, |\tilde{U}_{ij}(s)|\left(\frac{u}{s}\right)^{l_j-l_i} + \tfrac{1}{2}\int_{(x+y)/2}^{\infty} ds \int_s^{y-x+s} du\, |\tilde{U}_{ij}(s)|\left(\frac{u}{s}\right)^{l_j-l_i}. \tag{62}$$

If $l_i \geqslant l_j$, since $u/s > 1$, we have

$$\|\Sigma(x,y)\| \leqslant \tilde{\sigma}^1(x). \tag{63a}$$

For $l_i < l_j$, we get

$$\Sigma_{ij}(x,y) \leqslant 3^{(l_j-l_i)}\tilde{\sigma}^1\left(\frac{x+y}{2}\right) + \tfrac{1}{2}\int_x^{(x+y)/2} ds \int_{x+y-s}^{y-x+s} du\, |\tilde{U}_{ij}(s)|\left(\frac{u}{s}\right)^{l_j-l_i} \leqslant 3^{(l_j-l_i)}\tilde{\sigma}^1\left(\frac{x+y}{2}\right) + \left(\frac{y}{x}\right)^{l_j-l_i} \times\left[\tilde{\sigma}^1(x) - \tilde{\sigma}^1\left(\frac{x+y}{2}\right)\right]. \tag{63b}$$

From estimate (61), it is not obvious whether $\lim_{y\to\infty} K(x,y)$ is finite whereas it is the case from estimate (56). The most general estimate (41a) for R_{ij} will, however, be used in Appendices C, D, E in order to get estimates for the derivatives of R_{ij}^0, R_{ij}, and K_{ij}. The results of Refs. 6 and 8 may be modified according to our new estimates for K.

In order to ensure the convergence of the integral in Eq. (6), since, in Appendix B, an upper bound (B12) for the function $\phi_0(\Lambda,x)$ is found for $\mathrm{Im}k_i = \delta \geqslant 0$, $|k_i| \geqslant s > 0$, and $\alpha_i \geqslant 0$, it is sufficient to impose that

$$\int_x^{\infty} y^{l_{\max}}\tilde{\sigma}^0\left(\frac{x+y}{2}\right)e^{-\delta y}\frac{[1+2|k_i|y]^{l_i}}{(2|k_i|y)^{l_i}} < \infty \quad \text{for } x \geqslant 0$$

or

$$\int_x^{\infty} dy\, \tilde{\sigma}^{l_{\max}}\left(\frac{x+y}{2}\right) = 4^{l_{\max}}\int_x^{\infty} dy \int_{(x+y)/2}^{\infty} d\eta_0\, \eta_0^{l_{\max}}\|\tilde{U}(\eta_0)\| = 4^{l_{\max}}\int_x^{\infty} d\eta_0 \int_x^{2\eta_0-x} dy\, \eta_0^{l_{\max}}\|\tilde{U}(\eta_0)\| \leqslant 2^{(2l_{\max}+1)}\int_x^{\infty} d\eta_0\, \eta_0^{l_{\max}+1}\|\tilde{U}(\eta_0)\| = 2\tilde{\sigma}^{l_{\max}+1}(x) < \infty. \tag{64}$$

Since we suppose that $\tilde{\sigma}^{l_{\max}+1}(x)$ exists for $x \geqslant 0$, we can justify the interchange of the order of the integrals in Eq. (61) by the Tonelli–Fubini theorem. Equations (55)–(64) show that $x^{l_{\max}}K(x,y)$ is bounded and integrable if $\tilde{\sigma}^{l_{\max}+1}(0)$ remains finite. Nowhere have we made the assumption that U is a real or Hermitian matrix; our results are thus valid for non-Hermitian potentials. We see that an exponential decrease is required for the perturbation potential; the rate of this decrease is measured by $4(\mu^2x^2 + ax)^{1/2}$. In absence of Coulomb interactions, $U_{ij}(x)$ must decrease faster than $\exp[-2^{l_i+1}|k_i^2 - k_j^2|^{1/2}x]$. (The exponent obtained in Ref. 3 contains a wrong factor.) In Appendices C and D, bounds for the derivatives of R_{ij}^0 and R_{ij} are obtained. These bounds enable us to get a bound for the derivatives of the transformation operator K in Appendix E. The estimates (50), (51), and (E3) obtained for $K(x,y)$ and its derivatives are necessary if we want to prove the unicity and the stability of a suitable solution to the fundamental equation of the inverse problem the existence of which Coz and the author[7] have generalized for non-Hermitian systems of coupled channels.

5. ANALYTICITY OF THE JOST MATRIX SOLUTION

We first consider the Jost solution as a function of the n variables $(k_1,\ldots,k_n)$. Equation (6) yields

$$[\phi(\Lambda,x)]_{ij} = \delta_{ij}[\phi_0(\Lambda,x)]_{jj} + \int_x^{\infty} K_{ij}(x,y)[\phi_0(\Lambda,y)]_{jj}\, dy. \tag{65}$$

Since $[\phi_0(\Lambda,x)]_{jj}$ depends only on k_j and since the solution

$K_{ij}(x,y)$ of Eq. (8) is an entire function of

$$\Delta_{ji}^2 = k_j^2 - k_i^2 \tag{66}$$

(from a theorem demonstrated by Poincaré[12]), $[\phi(\Lambda,x)]_{ij}$ is a function of k_j and Δ_{ji}^2 only. This shows that $[\phi(\Lambda,x)]_{ij}$ is an even function of all the k_l's, except k_j as is well known from Ref. 12.

Instead of considering $\phi(\Lambda,x)$ as a function of many variables k_l ($l = 1,n$), the conservation of energy (66) between the channels may be used to eliminate $k_2,...,k_n$ in favor of the largest variable k_1 and the $n-1$ constants Δ_{1l}^2 ($l = 2,n$). By doing this, we must define, as in Refs. 13, a k_1 Riemann surface, consisting of 2^{n-1} sheets, having branch points at $k_1 = \pm \Delta_{l1}$ ($l = 2,...,n$). With each sheet, Weidenmüller[13] associates a vector τ of $n-1$ elements $\tau_l = \pm 1$, defined according the rule

$$\text{sgn Re}k_{l+1} = \tau_l \, sgn \, Rek_1, \tag{67}$$

$$\text{sgn Im}k_{l+1} = \tau_l \, \text{sgn Im}k_1. \tag{68}$$

The physical sheet is defined by $\tau_l = +1$ ($l = 1,n-1$). In order to prove the analyticity of $\Phi(k_1,x)$ with respect to k_1 in the upper half of the physical plane, the existence and the continuity of $\Phi(k_1,x)$ and its first derivative with respect to k_1 must be shown. The existence of a bounded continuous $\Phi_{ij}(k_1,x)$ has already been shown if $\tilde{\sigma}^{l_{\max}+1}(0)$ exists for $\text{Im}k_1 > 0$, hence for $\text{Im}k_j > 0$. Since the integral in Eq. (6) converges absolutely, the differentiation of Φ with respect to k_1 can be performed under the integral sign:

$$\left[\frac{d}{dk_1}\Phi(k_1,x)\right]_{ij} = \left[\frac{d}{dk_1}\Phi_0(k_1,x)\right]_{ij}\delta_{ij} + \int_x^\infty K_{ij}(x,t) \times \left(\frac{d}{dk_1}\Phi_0(k_1,t)\right)_{ij} dt. \tag{69}$$

Since $[(d/dk_1)\Phi_0(k_1,x)]$ exists and can be bounded for $\text{Im}k_1 > 0$ and since $K_{ij}(x,t)$ is absolutely integrable, $[(d/dk_1)\Phi(k_1,x)]_{ij}$ is well defined and bounded. The matrix function $\Phi(k_1,x)$ is thus analytical in the upper half of the physical k_1 plane, for all fixed $x > 0$.

6. CONCLUSION

We have found sufficient conditions that the matrix perturbative potential should satisfy in order to get a bounded and integrable transformation operator with bounded first derivatives. These conditions are, of course, dependent on the reference potential: The centrifugal part imposes that the perturbative potential has certain moments [see Eq. (56)] while the Coulomb interaction or the different threshold energies lead to an exponential decrease of the potential [see Eq. (57)]. For this class of potentials, the analyticity of the Jost solution has been shown in the upper half of the physical k_1 plane (for $\alpha_j > 0$). This property of analyticity is essential if we want to establish a fundamental equation for the inverse problem. The upper bounds (55), (56), (E3) that we have obtained for the transformation operator and its derivatives are of basic importance for two reasons:

(i) They allow us to get an upper bound for the kernel of the fundamental equation and consequently give us the conditions to be imposed to the scattering data in order to have a suitable solution to the inverse problem.

(ii) They are also necessary to ensure the stability of the inverse problem (see Ref. 14). Indeed, the experimental scattering data are only known up to a certain energy and the question naturally arises whether this is sufficient to well define the potential. Of course, this is not sufficient if we do not restrict the class of acceptable potentials (see Ref. 15). On the other hand, if we impose that the solution of the inverse problem should belong to the above-defined class of potentials, then the estimates obtained for K and its first derivatives will enable us to show that the solution is stable with respect to small changes in the phase shifts above a certain energy.

ACKNOWLEDGMENTS

The author would like to thank Professor Marcel Coz for his critical reading of the manuscript.

APPENDIX A: SPECTRAL REPRESENTATION OF THE COMPLETE RIEMANN–GREEN FUNCTION

Since the partial differential Eq. (10a) satisfied by the Riemann–Green matrix element $R_{ij}(x,y;s,u)$ is separable with respect to the variables x,y, a spectral representation can easily be obtained, by a generalization of a method developed by Riemann and described p. 328 of Ref. 11.

Setting $\Delta_{ij}^2 = k_i^2 - k_j^2$, we consider the following two differential equations:

$$\left[\frac{d^2}{dx^2} + \Delta_{ij}^2 + \lambda^2 - \frac{l_i(l_i+1)}{x^2} - \frac{\alpha_i}{x}\right] v(\lambda,x) = 0, \tag{A1}$$

$$\left[\frac{d^2}{dy^2} + \lambda^2 - \frac{l_j(l_j+1)}{y^2} - \frac{\alpha_j}{y}\right] w(\lambda,y) = 0. \tag{A2}$$

The regular $F_j(\lambda,y)$ and irregular $G_j(\lambda,y)$ real Coulomb functions, defined by the equation:

$$[\phi_0(\Lambda,y)]_{ij} = G_j(k_j,y) + iF_j(k_j,y), \tag{A3}$$

form a system of two linearly independent solutions of Eq. (A2), the Wronskian W of which is constant:

$$W = \text{Wr}[F,G] = F\frac{\partial}{\partial y}G - G\frac{\partial}{\partial y}F = -1. \tag{A4}$$

The solution v_{x_1} of Eq. (A1) is chosen in such a way that the boundary conditions (A5) and (A6) are satisfied:

$$v_{x_1}(\lambda,x) = 0 \quad \text{for } x = x_1 \tag{A5}$$

and

$$\frac{d}{dx}v_{x_1}(\lambda,x) = 1 \quad \text{for } x = x_1. \tag{A6}$$

An expression for $v_{x_1}(\lambda,x)$ in terms of F_i and G_i is easily obtained:

$$v_{x_1}(\lambda,x) = [G_i((\lambda^2 + \Delta_{ij}^2)^{1/2},x_1)F_i((\lambda^2 + \Delta_{ij}^2)^{1/2},x)$$

$$- G_i((\lambda^2 + \Delta_{ij}^2)^{1/2},x)F_i((\lambda^2 + \Delta_{ij}^2)^{1/2},x_1)]. \tag{A7}$$

We consider the solution $w(\lambda,y)$ of Eq. (A2) with which a function $\bar{w}(\lambda,y)$ can be associated, such that the following equality is verified for any function $f(x)$ belonging to $L^2(0,\infty)$:

$$f(x) = \left(\int \sum\right) d\lambda\, \bar{W}(\lambda,x) \int_0^\infty dy\, w(\lambda,y) f(y) \tag{A8}$$

where $(\int \Sigma)$ denotes the integration over the continuous spectrum, plus an infinite summation over the point spectrum when α_j is negative. From Refs. 10 and 16, we know that such functions $w(\lambda,y)$ and $\bar{w}(\lambda,y)$ exist:

$$w(\lambda,y) = \tfrac{1}{2}\pi\bar{w}(\lambda,y) = F_j(\lambda,y). \tag{A9}$$

The Riemann–Green matrix element R_{ij} can be written for $|x-s| > |y-u|$, which is always satisfied inside the Marchenko domain (and also inside the Gel'fand domain):

$$R_{ij}(x,y;s,u) = 2\,\mathrm{sgn}(s-x)\left(\int \sum\right) d\lambda\, \bar{w}(\lambda,y) w(\lambda,u) v_x(\lambda,s), \tag{A10}$$

where

$$\mathrm{sgn}(a) = a/|a|.$$

Of course, when $|x-s| = |y-u|$, we have

$$R_{ij}(x,y;s,u) = 1. \tag{A11}$$

However, we did not manage to rewrite the spectral representation (A10) for the complete Riemann–Green function into a close form. This can only be done in particular cases when $\alpha_i = l_i = 0$ or when $\alpha_i = k_i = 0$ and leads to well-known results.[11,17,18] We must acknowledge that not very much progress has been achieved since 1930: Most of the actually known Riemann-Green solutions were already discovered at that time by Darboux[11] and Chaundy,[18] in spite of the fact that new constructive methods[19,20] have been proposed.

APPENDIX B: BOUND FOR ϕ_0

Taking the absolute value of Eq. (4), we get

$$|[\phi_0(\Lambda,x)]_{ij}| = \delta_{ij}|W_{-i\alpha_i/2k_i,l_i+1/2}(-2ik_ix)|e^{\alpha_i\pi/4k_i}. \tag{B1}$$

Assuming that the Coulomb interaction is repulsive ($\alpha_i \geqslant 0$) and setting $Z = -2ik_ix$, $\kappa = -i\alpha_i/2k_i$, $\nu = l_i + \frac{1}{2}$, the Whittaker's functions W can be expressed in terms of the Kummer function U by the Eq. (B2) [see Eq. (13.1.33) of Ref. 9]

$$W_{\kappa,\nu}(Z) = e^{-Z/2}Z^{\nu+1/2}U(\tfrac{1}{2}+\nu-\kappa,1+2\nu,Z). \tag{B2}$$

If $\mathrm{Re}(1+2\nu) > \mathrm{Re}(\frac{1}{2}+\nu-\kappa) > 0$, the integral representation (13.2.5) of Ref. 9 for U can be used:

$$U(\tfrac{1}{2}+\nu-\kappa,1+2\nu,Z) = \frac{1}{\Gamma(\nu+\frac{1}{2}-\kappa)} \int_0^\infty e^{-Zt} t^{\nu-\kappa-1/2}(1+t)^{\nu+\kappa-1/2}\,dt \tag{B3}$$

if

$$\mathrm{Re}(l_i + 1 + i\alpha_i/2k_i) > 0, \tag{B4}$$

or

$$U(\tfrac{1}{2}+\nu-\kappa,1+2\nu,Z) = \frac{1}{\Gamma(\nu+\frac{1}{2}-\kappa)}\frac{1}{Z} \times \int_0^\infty e^{-t}\left(\frac{t}{Z}\right)^{\nu-\kappa-\frac{1}{2}}\left(1+\frac{t}{Z}\right)^{\nu+\kappa-\frac{1}{2}} dt. \tag{B5}$$

It readily follows from Eqs. (B2) and (B5) that

$$W_{\kappa,\nu}(Z) = \frac{e^{-Z/2}}{\Gamma(\nu+\frac{1}{2}-\kappa)Z^{\nu-1/2}} \times \int_0^\infty e^{-t}t^{2\nu-1}\left(\frac{Z}{t}+1\right)^{\nu+\kappa-1/2} dt, \tag{B6}$$

if

$$\mathrm{Im}k_i/|k_i|^2 \geqslant -(l_i+1)2/\alpha_i. \tag{B7}$$

Our Φ_0 Jost solution behaves like Z^{l_i} at the origin and has a discontinuity along the imaginary negative k_i axis [we take $-\pi < \arg Z < \pi$, $\arg(-i) = -\pi/2$, and $3\pi/2 < \arg k_i < -\pi/2$].

Setting $k_i = \gamma + i\delta$, we get for $\delta \geqslant 0$ and $|k_i| \geqslant s > 0$

$$|[\phi_0(\Lambda<x)]_{ij}| \leqslant \frac{\exp(\alpha_i\pi\delta/4|k_i|^2 - \delta x)\exp(\alpha_i\pi/2s)\delta_{ij}}{|\Gamma(l_i+1+i\alpha_i/2k_i)|(2|k_i|x)^{l_i}} \times \int_0^\infty e^{-t}t^{2l_i}\left(1+\frac{2|k_i|}{t}\right)^{l_i} dt, \tag{B8}$$

since

$$\left|\left(\frac{Z}{t}+1\right)^{(l_i - i\alpha_i/2k_i)}\right| = \exp\left\{\mathrm{Re}\left[\left(l_i - i\frac{\alpha_i}{2k_i}\right)\ln\left(\frac{Z}{t}+1\right)\right]\right\} = \left|\frac{Z}{t}+1\right|^{(l_i - \alpha_i\delta/2|k_i|^2)} \exp\left[\frac{\alpha_i\gamma}{2|k_i|^2}\arg\left(1+\frac{Z}{t}\right)\right] = \left|1+\frac{2|k_i|x}{t}\right|^{l_i} e^{\alpha_i\pi/2s}. \tag{B9}$$

The integral in (B8) can be evaluated:

$$\int_0^\infty e^{-t}t^{l_i}(t+2|k_i|x)^{l_i} = \sum_{j=0}^{l_i} C_{l_i}^j (2|k_i|x)^{l_i-j}(l_i+j)! \tag{B10}$$

$$\leqslant (2l_i)![1+2|k_i|x]^{l_i}, \tag{B11}$$

so that the final result can be written as

$$|[\phi_0(\Lambda,x)]_{ii}| \leqslant \frac{\exp(3\alpha_i\pi/4s-\delta x)}{|\Gamma(l_i+1+i\alpha_i/2k_i)|} \times (2l_i)!\,\frac{[1+2|k_i|x]^{l_i}}{[2|k_i|x]^{l_i}}, \tag{B12}$$

where

$$\left|\Gamma\left(l_i+1+\frac{i\alpha_i}{2k_i}\right)\right| = \left\{\left(l_i^2+\frac{\alpha_i^2}{4k_i^2}\right)\left[(l_i-1)^2+\frac{\alpha_i^2}{4k_i^2}\right]\cdots \times\left(1+\frac{\alpha_i^2}{4k_i^2}\right)\frac{\pi\alpha_i}{2k_i\sinh(\pi\alpha_i/2k_i)}\right\}^{1/2}. \tag{B13}$$

APPENDIX C: BOUNDS FOR $|(\partial/\partial\eta)R^0_{ij}(\eta,\xi;\eta_0,\xi_0)|$, $|(\partial/\partial\xi)R^0_{ij}(\eta,\xi;\eta_0,\xi_0)|$

The derivative of Eq. (15) with respect to η or ξ can be written as

$$R'^0_{ij} = -2x_2'P_{l_i}'(1-2x_2) - 2x_1'\int_0^1 P_{l_i}(1-2x_2t) \times P_{l_i}'(1-2x_1+2x_1t)\,dt + 2x_1 2x_2' \times\int_0^1 t\,P_{l_i}'(1-2x_2t)P_{l_i}'(1-2x_1+2x_1t)\,dt - 2x_1 2x_1'\int_0^1 (t-1)P_{l_i}(1-2x_2t) \times P_{l_i}''(1-2x_1+2x_1t)\,dt, \tag{C1}$$

where the primes denote derivatives with respect to any of the two variables η or ξ. Before proceeding further, the following bounds are recalled:

$$|P_l(x)| \leqslant 1 \quad \text{for } |x| \leqslant 1, \tag{C2}$$

$$P_l'(x) \leqslant l(l+1)/2 \quad \text{for } |x| \leqslant 1, \tag{C3}$$

$$P_l(x) \leqslant (2x)^l \quad \text{for } x \geqslant 1, \tag{C4}$$

$$P_l'(x) \leqslant l(l+1)(2x)^{l-1} \quad \text{for } x \geqslant 1. \tag{C5}$$

Setting $Z = 1 - 2x_1$ and using Eqs. (20) and (21), we obtain

$$|R'^0_{ij}| \leqslant |x_2'|l_j(l_j+1) + 4|x_1'|\,|P_{l_i}'(z)| + |x_2'|l_j(l_j+1)l_i(l_i+1)(2Z)^{l_i} \leqslant 4|x_1'|l_i(l_i+1)(2Z)^{l_i-1} + |x_2'|l_j(l_j+1)[1+l_i(l_i+1)(2Z)^{l_i}], \tag{C6}$$

$$|R'^0_{ij}| \leqslant (2Z)^{l_i}(|x_1'|+|x_2'|)l_{\max}(l_{\max}+1) \times 2[1+l_{\max}(l_{\max}+1)] \leqslant A(2Z)^{l_i}(|x_1'|+|x_2'|). \tag{C7}$$

To pursue, we have to find upper bounds for $[\,|(\partial/\partial\eta)x_1| + |(\partial/\partial\eta)x_2|\,]$ and $[\,|(\partial/\partial\xi)x_1| + |(\partial/\partial\xi)x_2|\,]$, where

$$\frac{1}{x_1}\frac{\partial x_1}{\partial\eta} = \frac{\eta_0-\xi}{(\eta-\eta_0)(\eta-\xi)} \leqslant 0,$$

$$\frac{1}{x_1}\frac{\partial x_1}{\partial\xi} = \frac{\xi_0-\eta}{(\xi-\xi_0)(\xi-\eta)} \geqslant 0,$$

$$\frac{1}{x_2}\frac{\partial x_2}{\partial\eta} = \frac{\xi+\eta_0}{(\eta-\eta_0)(\eta+\xi)} \leqslant 0,$$

$$\frac{1}{x_2}\frac{\partial x_2}{\partial\xi} = \frac{\eta+\xi_0}{(\xi-\xi_0)(\xi+\eta)} \geqslant 0.$$

$$\left|\frac{\partial x_1}{\partial\eta}\right| + \left|\frac{\partial x_2}{\partial\eta}\right| \leqslant \frac{(\xi-\xi_0)(\eta_0-\xi)}{(\eta-\xi)^2(\eta_0-\xi_0)} + \frac{(\eta_0+\xi)(\xi-\xi_0)}{(\eta+\xi)^2(\eta_0+\xi_0)} \leqslant \frac{(\eta_0-\xi)}{(\eta-\xi)^2} + \frac{(\eta_0+\xi)}{(\eta+\xi)^2} \leqslant \frac{2\eta_0}{(\eta-\xi)^2}, \tag{C8}$$

$$\left|\frac{\partial x_1}{\partial\xi}\right| + \left|\frac{\partial x_2}{\partial\xi}\right| \leqslant \frac{(\eta-\xi_0)(\eta_0-\eta)}{(\eta-\xi)^2(\eta_0-\xi_0)} + \frac{(\eta+\xi_0)(\eta_0-\eta)}{(\eta+\xi)^2(\eta_0+\xi_0)} \leqslant \frac{(\eta-\xi_0)}{(\eta-\xi)^2} + \frac{(\eta+\xi_0)}{(\eta+\xi)^2} \leqslant \frac{2\eta}{(\eta-\xi)^2}. \tag{C9}$$

The bounds for the derivatives of R^0_{ij} read then

$$\left|\frac{\partial}{\partial\eta}R^0_{ij}(\eta,\xi;\eta_0,\xi_0)\right| \leqslant \frac{2A\eta_0}{(\eta-\xi)^2}\left[4\frac{(\eta+\xi)(\eta_0-\xi_0)}{(\eta-\xi)(\eta_0+\xi_0)}\right]^{l_i} \tag{C10}$$

and

$$\left|\frac{\partial}{\partial\xi}R^0_{ij}(\eta,\xi;\eta_0,\xi_0)\right| \leqslant \frac{2A\eta}{(\eta-\xi)^2}\left[4\frac{(\eta+\xi)(\eta_0-\xi_0)}{(\eta-\xi)(\eta_0+\xi_0)}\right]^{l_i}. \tag{C11}$$

APPENDIX D: BOUNDS FOR $|(\partial/\partial\eta)R_{ij}|$ AND $|(\partial/\partial\xi)R_{ij}|$

In this appendix, the estimates (29), (41), (C10), and (C11) for R^0_{ij}, R_{ij}, and the derivatives of R^0_{ij} are used to find

bounds for the absolute values of the complete Riemann function derivatives. Taking the derivatives with respect to η and ξ of Eq. (30) and setting $T_{ij}(\eta,\xi) = k_i^2 - k_j^2 - \alpha_i/(\eta-\xi) + \alpha_j/(\eta+\xi)$, it readily follows that

$$\begin{aligned}\frac{\partial}{\partial\eta} R_{ij}(\eta,\xi;\eta_0,\xi_0) &= \frac{\partial}{\partial\eta} R^{0}_{ij}(\eta,\xi;\eta_0,\xi_0) - \int_{\xi_0}^{\xi} d\xi_1\, R^{0}_{ij}(\eta,\xi;\eta,\xi_1) \\ &\times T_{ij}(\eta,\xi_1) R_{ij}(\eta,\xi_1;\eta_0,\xi_0) \\ &+ \int_{\xi_0}^{\xi} d\xi_1 \int_{\eta}^{\eta_0} d\eta_1 \frac{\partial}{\partial\eta} R^{0}_{ij}(\eta,\xi;\eta_1,\xi_1) \\ &\times T_{ij}(\eta_1,\xi_1) R_{ij}(\eta_1,\xi_1;\eta_0,\xi_0), \end{aligned} \tag{D1}$$

$$\begin{aligned}\frac{\partial}{\partial\xi} R_{ij}(\eta,\xi;\eta_0,\xi_0) &= \frac{\partial}{\partial\xi} R^{0}_{ij}(\eta,\xi;\eta_0,\xi_0) \\ &+ \int_{\eta}^{\eta_0} d\eta_1\, R^{0}_{ij}(\eta,\xi;\eta_1,\xi) T_{ij}(\eta_1,\xi) \\ &\times R_{ij}(\eta_1,\xi;\eta_0,\xi_0) \\ &+ \int_{\xi_0}^{\xi} d\xi_1 \int_{\eta}^{\eta_0} d\eta_1 \frac{\partial}{\partial\xi} R^{0}_{ij}(\eta,\xi;\eta_1,\xi_1) \\ &\times T_{ij}(\eta_1,\xi_1) R_{ij}(\eta_1,\xi_1;\eta_0,\xi_0). \end{aligned} \tag{D2}$$

Defining

$$D_{ij}(\eta,\xi;\eta_0,\xi_0) = \left[\frac{(\eta_0+\xi_0)(\eta-\xi)}{4(\eta_0-\xi_0)(\eta+\xi)}\right]^{l_i} \frac{\partial}{\partial\eta} R_{ij}(\eta,\xi;\eta_0,\xi_0), \tag{D3}$$

$$C_{ij}(\eta,\xi;\eta_0,\xi_0) = \left[\frac{(\eta_0+\xi_0)(\eta-\xi)}{4(\eta_0-\xi_0)(\eta+\xi)}\right]^{l_i} \frac{\partial}{\partial\xi} R_{ij}(\eta,\xi;\eta_0,\xi_0), \tag{D4}$$

$$\tilde{R}^{0}_{ij}(\eta,\xi;\eta_0,\xi_0) = \left[\frac{(\eta_0+\xi_0)(\eta-\xi)}{4(\eta_0-\xi_0)(\eta+\xi)}\right]^{l_i} R^{0}_{ij}(\eta,\xi;\eta_0,\xi_0), \tag{D5}$$

$$\tilde{R}_{ij}(\eta,\xi;\eta_0,\xi_0) = \left[\frac{(\eta_0+\xi_0)(\eta-\xi)}{4(\eta_0-\xi_0)(\eta+\xi)}\right]^{l_i} R_{ij}(\eta,\xi;\eta_0,\xi_0), \tag{D6}$$

$$E(\eta_0-\xi_0) = 4^{l_i} \exp\{4[\mu_{ij}^2(\eta_0-\xi_0)^2 + a_{ij}(\eta_0-\xi_0)]^{1/2}\}, \tag{D7}$$

Eqs. (D1) and (C10) lead to

$$|D_{ij}(\eta,\xi;\eta_0,\xi_0)| \leqslant \frac{2A\eta_0}{(\eta-\xi)^2} + E(\eta_0-\xi_0)\left[\mu_{ij}^2(\xi-\xi_0) + a_{ij}\ln\frac{\eta-\xi_0}{\eta-\xi}\right]$$
$$+ \frac{2A}{(\eta-\xi)^2} E(\eta_0-\xi_0)\left[\mu_{ij}^2 \frac{(\eta_0^2-\eta^2)(\xi-\xi_0)}{2} + a_{ij}\left(\frac{(\eta_0-\eta)(\xi-\xi_0)}{2} + (\eta_0^2-\xi_0^2)\ln\frac{\eta+\xi}{\eta-\xi}\right)\right] \tag{D8}$$

and

$$|C_{ij}(\eta,\xi;\eta_0,\xi_0)| \leqslant \frac{2A\eta}{(\eta-\xi)^2} + E(\eta_0-\xi_0)\left[\mu_{ij}^2(\eta_0-\eta) + a_{ij}\ln\frac{\eta_0-\xi}{\eta-\xi}\right] + \frac{2A\eta}{(\eta-\xi)^2} E(\eta_0-\xi_0)$$
$$\times\left\{\mu_{ij}^2(\eta_0-\eta)(\xi-\xi_0) + \tfrac{1}{2}a_{ij}\left[\eta_1\ln\frac{(\eta_1+\xi)(\eta_1-\xi_0)}{(\eta_1-\xi)(\eta_1+\xi_0)}\right]_{\eta_1=\eta_0}^{\eta_1=\eta} + \tfrac{1}{2}a_{ij}\left[\xi_1\ln\frac{(\eta_0+\xi_1)(\eta_0-\xi_1)}{(\eta+\xi_1)(\eta-\xi_1)}\right]_{\xi_1=\xi_0}^{\xi_1=\xi}\right\}. \tag{D9}$$

If we use the well-known fact that $\ln x < x$ for $x > 1$, Eqs. (D7) and (D8) can be written as

$$\left|\frac{\partial}{\partial\eta} R_{ij}(\eta,\xi;\eta_0,\xi_0)\right|$$
$$\leqslant\left[\frac{4(\eta_0-\xi_0)(\eta+\xi)}{(\eta_0+\xi_0)(\eta-\xi)}\right]^{l_i}\left\{\frac{2A\eta_0}{(\eta-\xi)^2} + E(\eta_0-\xi_0)\left[\mu_{ij}^2(\xi-\xi_0) + a_{ij}\frac{\eta-\xi_0}{\eta-\xi}\right] + \frac{2A}{(\eta-\xi)^2}E(\eta_0-\xi_0)\right.$$
$$\left.\times\left[\mu_{ij}^2\frac{(\eta_0^2-\eta^2)(\xi-\xi_0)}{2} + a_{ij}\frac{(\eta_0-\eta)(\xi-\xi_0)}{2} + a_{ij}\frac{(\eta_0^2-\xi_0^2)(\eta+\xi)}{(\eta-\xi)}\right]\right\} \tag{D10}$$

and

$$\left|\frac{\partial}{\partial\xi} R_{ij}(\eta,\xi;\eta_0,\xi_0)\right|$$
$$\leqslant\left[4\frac{(\eta_0-\xi_0)(\eta+\xi)}{(\eta_0+\xi_0)(\eta-\xi)}\right]^{l_i}\left\{\frac{2A\eta}{(\eta-\xi)^2} + E(\eta_0-\xi_0)\left[\mu_{ij}^2(\eta_0-\eta) + a_{ij}\frac{\eta_0-\xi}{\eta-\xi}\right] + \frac{2A\eta}{(\eta-\xi)^2}E(\eta_0-\xi_0)\right.$$

$$\times\left[\mu_{ij}^2(\eta_0-\eta)(\xi-\xi_0)+\tfrac{1}{2}a_{ij}\eta\,\frac{(\eta+\xi)}{(\eta-\xi)}+\tfrac{1}{2}a_{ij}\xi\,\frac{\eta_0^2-\xi^2}{\eta^2-\xi^2}\right]\Bigg\}. \tag{D11}$$

APPENDIX E: UPPER BOUNDS FOR THE DERIVATIVES OF THE TRANSFORMATION OPERATOR

Performing the derivative of Eq. (43) with respect to η, we obtain

$$\begin{aligned}\frac{\partial}{\partial\eta}K_{ij}(\eta,\xi) = &-\tfrac{1}{2}R_{ij}(\eta,\xi,\eta,0)U_{ij}(\eta)+\tfrac{1}{2}\int_\eta^\infty d\eta_0\,\frac{\partial}{\partial\eta}R_{ij}(\eta,\xi,\eta_0,0)U_{ij}(\eta_0)\\ &-\sum_l\int_0^\xi d\xi_0\,R_{ij}(\eta,\xi;\eta,\xi_0)U_{il}(\eta-\xi_0)K_{lj}(\eta,\xi_0)\\ &+\sum_l\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\,\frac{\partial}{\partial\eta}R_{ij}(\eta,\xi;\eta_0,\xi_0)U_{il}(\eta_0-\xi_0)K_{lj}(\eta_0,\xi_0).\end{aligned} \tag{E1}$$

Setting $\tilde U_{ij}(\eta)=E(\eta)U_{ij}(\eta)$, we get the following upper bound:

$$\begin{aligned}\left|\frac{\partial}{\partial\eta}K_{ij}(\eta,\xi)\right|\left[\frac{\eta-\xi}{(\eta+\xi)4}\right]^{l_{\max}} \le &\tfrac{1}{2}\tilde U_{ij}(\eta)+\tfrac{1}{2}\int_\eta^\infty d\eta_0\left[\frac{2A\eta_0}{(\eta-\xi)^2}+E(\eta_0)\left(\mu_{ij}^2\xi+a_{ij}\frac{\eta}{\eta-\xi}\right)+\frac{2A}{(\eta-\xi)^2}E(\eta_0)\right.\\ &\left.\times\left(\frac{\mu_{ij}^2\eta_0^2\xi}{2}+a_{ij}\frac{3\eta\eta_0^2}{(\eta-\xi)}\right)\right]|U_{ij}(\eta_0)|+\sum_l\int_0^\xi d\xi_0\left(\frac{\eta-\xi_0}{\eta+\xi_0}\right)^{l_{\max}}|\tilde U_{il}(\eta-\xi_0)|\;|K_{lj}(\eta,\xi_0)|\\ &+\sum_l\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\left(\frac{\eta_0-\xi_0}{\eta_0+\xi_0}\right)^{l_{\max}}\left\{\frac{2A\eta_0}{(\eta-\xi)^2}+E(\eta_0-\xi_0)\left[\mu_{ij}^2(\xi-\xi_0)+a_{ij}\frac{\eta-\xi_0}{\eta-\xi}\right]\right.\\ &+\frac{2A}{(\eta-\xi)^2}E(\eta_0-\xi_0)\left[\frac{\mu_{ij}^2(\eta_0^2-\eta^2)(\xi-\xi_0)}{2}+a_{ij}\frac{(\eta_0-\eta)(\xi-\xi_0)}{2}\right.\\ &\left.\left.+a_{ij}\frac{(\eta_0^2-\xi_0^2)(\eta+\xi)}{(\eta-\xi)}\right]\right\}|U_{il}(\eta_0-\xi_0)|\;|K_{lj}(\eta_0,\xi_0)|.\end{aligned} \tag{E2}$$

Since we have

$$|K_{lj}(\eta_0,\xi_0)|\le\frac{n}{2}\left(\frac{\eta_0+\xi_0}{\eta_0-\xi_0}\right)^{l_{\max}}\tilde\sigma^0(\eta_0)\exp[\tilde\sigma^1(\eta_0-\xi_0)], \tag{55b}$$

it readily follows that

$$\begin{aligned}\left|\left|\frac{\partial}{\partial\eta}\mathbf{K}(\eta,\xi)\right|\right|\left[\frac{\eta-\xi}{4(\eta+\xi)}\right]^{l_{\max}} \le &\tfrac{1}{2}\|\tilde U(\eta)\|+\frac{A}{(\eta-\xi)^2}\sigma^1(\eta)+\tfrac{1}{2}\mu_{ij}^2\xi\left[\tilde\sigma^0(\eta)+\frac{A}{(\eta-\xi)^2}\tilde\sigma^2(\eta)\right]\\ &+a_{ij}\frac{\eta}{2(\eta-\xi)}\left[\tilde\sigma^0(\eta)+\frac{6A}{(\eta-\xi)}\tilde\sigma^2(\eta)\right]+\frac{n}{2}\tilde\sigma^0(\eta)\exp[\tilde\sigma^1(\eta-\xi)]\int_{\xi-\xi_0}^\eta\|U(s)\|\,ds\\ &+n\tilde\sigma^0(\eta)\exp[\tilde\sigma^1(\eta-\xi)]\frac{A}{(\eta-\xi)^2}\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\,\eta_0\|U(\eta_0-\xi_0)\|\\ &+\frac{n}{2}\tilde\sigma^0(\eta)\exp[\tilde\sigma^1(\eta-\xi)]\int_\eta^\infty d\eta_0\int_0^\xi d\xi_0\left\{\mu_{ij}^2\xi\left[1+\frac{A\eta_0^2}{(\eta-\xi)^2}\right]\right.\\ &\left.+\frac{a_{ij}}{(\eta-\xi)}\left[\eta+\frac{A\xi}{(\eta-\xi)}\eta_0+\frac{2A(\eta+\xi)}{(\eta-\xi)^2}\eta_0^2\right]\right\}\|\tilde U(\eta_0-\xi_0)\|.\end{aligned} \tag{E3}$$

To proceed further, we must evaluate the following integral:

$$\int_{\eta}^{\infty} d\eta_0 \int_0^{\xi} d\xi_0\, \eta_0^l f(\eta_0 - \xi_0)$$

with $l = 0$, 1, or 2 and $f(\eta_0 - \xi_0) = \|U(\eta_0 - \xi_0)\|$ or $\|\tilde{U}(\eta_0 - \xi_0)\|$. We easily obtain

$$\int_{\eta}^{\infty} d\eta_0 \int_{\eta_0 - \xi}^{\eta_0} ds\, \eta_0^l f(s) \leqslant \int_{\eta - \xi}^{\infty} ds \int_s^{s+\xi} d\eta_0\, \eta_0^l f(s)$$

$$= \int_{\eta - \xi}^{\infty} ds f(s) \frac{(s+\xi)^{l+1} - s^{l+1}}{l+1} = \frac{1}{l+1} \sum_{j=0}^{l} C_{l+1}^{j} \sum^{j} (\eta - \xi) \xi^{l+1-j}, \tag{E4}$$

where

$$\sum^{j} (\eta - \xi) = \begin{cases} \sigma^j(\eta - \xi) & \text{if } f(\eta - \xi) = \|U(\eta - \xi)\|, \\ \tilde{\sigma}^j(\eta - \xi) & \text{if } f(\eta - \xi) = \|\tilde{U}(\eta - \xi)\|. \end{cases} \tag{E5}$$

If U has the moments implied by Eqs. (E3), (E4), and (E5), the transformation operator has a bounded first derivative with respect to η. (If $l_{\max} \geqslant 3$, this does not introduce new conditions.) The same simple but tiresome considerations will prove that the ξ derivative of K is bounded and so are the x and y derivatives.

[1] L.W. Agranovich and V.A. Marchenko, *The Inverse Problem in Scattering Theory* (Gordon and Breach, New York, 1963).
[2] M.M. Crum, Quart. J. Math. **6**, 121 (1955); M.G. Krein, Dokl. Akad. Nauk. SSSR **113**, (5), 970 (1957).
[3] M. Coz and P. Rochus, J. Math. Phys. **17**, 894 (1976).
[4] J.R. Cox, Ann. Phys. (N.Y.) **39**, 237–52 (1966).
[5] R. Jost and W. Kohn, Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd. **27** (9), (1953).
[6] M. Coz and P. Rochus, J. Math. Phys. **18**, 2223 (1977).
[7] M. Coz and P. Rochus, "A fundamental equation for a non-Hermitian differential system," preprint.
[8] M. Coz and P. Rochus, J. Math. Phys. **18**, 2232 (1977).
[9] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[10] H. van Haeringen, C.V.M. van der Mee, and R. van Wageningen, J. Math. Phys. **18**, 941 (1977).
[11] E.T. Copson, Arch. Rat. Mech. Anal. **1**, 324 (1958).
[12] R.G. Newton, J. Math. Phys. **2**, 188 (1961); *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[13] (a) H.A. Weidenmüller, Ann. Phys. (N.Y.) **28**, 60 (1964); (b) J. Humblet, Nucl. Phys. **57**, 386 (1964).
[14] (a) V.A. Marchenko, Mat. Sbornik Tom **77**, 119 (1968); (b) D.Š. Lundina and V.A. Marchenko, Mat. Sbornik Tom **78**, 120 (1969).
[15] (a) G.A. Viano, Nuovo Cimento A **63**, 58 (1969); (b) S. Ciulli, C. Pomponiu, and I.S. Stefănescu, Phys. Rep. C **17**, 134 (1975).
[16] (a) E.C. Titchmarch, *Eigenfunctions Expansions I* (Clarendon, Oxford, 1962), p. 98; (b) W.O. Amrein, J.M. Jauch and K.B. Sinha, *Scattering Theory in Quantum Mechanics: Physical Principles and Mathematical Method* (Addison-Wesley, Reading, Mass., 1977).
[17] G. Darboux, *Théorie Générale des Surfaces*, 2 (Paris, 1915).
[18] T.W. Chaundy, Quart. J. Math. **10**, 237 (1939).
[19] E.A. Daggit, J. Math. Anal. Applics. **29**, 91–108 (1970).
[20] J.S. Papadakis and D.H. Wood, J. Diff. Eq. **24**, 397–411 (1977).

On conserved quantities in general relativity

Abhay Ashtekar[a)]

Enrico Fermi Institute, 5630 S. Ellis Avenue, Chicago, Illinois 60637
and Max Planck Institut, Föhringer Ring 6, 8000 München, 40, Germany

Anne Magnon-Ashtekar[b)]

Enrico Fermi Institute, 5630 S. Ellis Avenue, Chicago, Illinois 60637
(Received 7 September 1978)

Recently, definitions of total 4-momentum and angular momentum of isolated gravitating systems have been introduced in terms of the asymptotic behavior of the Weyl curvature (of the underlying space–time) at spatial infinity. Given a space–time equipped with isometries, on the other hand, one can also construct conserved quantities using the presence of the Killing fields. Thus, for example, for stationary space–times, the Komar integral can be used to define the total mass, and, the asymptotic value of the twist of the Killing field, to introduce the dipole angular momentum moment. Similarly, for axisymmetric space–times, one can obtain the ("z-component" of the) total angular momentum in terms of the Komar integral. It is shown that, in spite of their apparently distinct origin, in the presence of isometries, quantities defined at spatial infinity reduce to the ones constructed from Killing fields. This agreement reflects one of the many subtle aspects of Einstein's (vacuum) equation.

I. INTRODUCTION

Consider a self-gravitating system in the framework of general relativity. Let us suppose that matter sources are confined to a spatially bounded world tube. Let, furthermore, the space–time $(\hat{\mathbf{M}}, \hat{g}_{ab})$ describing the gravitational field of this system admit a Killing field $\hat{\xi}^a$. Then, the Komar integral[1]

$$Q = \int_{S^2} \hat{\mathcal{E}}_{abcd}\, \hat{\nabla}^c \hat{\xi}^d\, dS^{ab}$$

represents a conserved quantity; the integral is independent of the particular choice of the 2-sphere S^2 surrounding the matter sources, made in its definition. [Here, $\hat{\mathcal{E}}_{abcd}$ and $\hat{\nabla}_a$ are respectively, the alternating tensor and the derivative operator on $(\hat{\mathbf{M}}, \hat{g}_{ab})$.] If $\hat{\xi}^a$ is a stationary Killing field, Q represents the total energy of the isolated system, while if $\hat{\xi}^a$ is an axial Killing field, Q has the interpretation of the component of the total angular momentum along the corresponding axis. These conserved quantities have played an important role in the understanding of stationary, axisymmetric isolated systems.

If a given stationary space–time has the property that the manifold of orbits of the stationary Killing field is asymptotically flat at spatial infinity, one has also available the Hansen[2] multipole moments. These arise from an examination of the *asymptotic* properties of the norm and the twist of the Killing field on the manifold of its orbits. Of particular interest to the present analysis is the dipole angular momentum moment. Consider the case in which such a stationary space–time is equipped also with an axial Killing field. Then, one has available two "angular momentum like" quantities: the dipole moment constructed from the timelike Killing field and the Komar integral constructed from the axial Killing field. What is the relation between these quantities? On intuitive grounds, one might expect the dipole moment to "point along the axis," and its norm to equal the value of the Komar integral. However, detailed analysis of this issue is complicated not only because the two notions refer to two different Killing fields but also because whereas the Komar integral involves an integral on *space–time*, the dipole moment is a vector at the point at infinity on the (three-dimensional) *manifold of orbits* of the stationary Killing field.

Recently, Ashtekar and Hansen[3] have introduced the notion of the total (ADM) 4-momentum[4] and angular momentum of isolated systems in yet another way. Their definitions make no reference to isometries at all; these quantities are constructed by examining the behavior of the Weyl curvature at large spacelike separations from sources. This construction may be summarized as follows. First, a notion of asymptotic flatness at spatial infinity is introduced. In effect, a space–time is said to be asymptotically flat provided one can "attach" to it a point analogous to the Minkowskian i^0 such that the metric obtained by a suitable rescaling of the physical one admits an extension to this point which is smooth in "angular directions" and has only finite radial discontinuities in its first derivatives. (These discontinuities turn out to be a measure of the total energy–momentum of the given space–time time.) In the second step, using asymptotic conditions, the group of asymptotic symmetries at spatial infinity—called the Spi group—is obtained. In its structure, the Spi group closely resembles the BMS group[5]: it is a semidirect product of an infinite dimensional Abelian group (of Spi supertranslations) with the Lorentz group, it admits a preferred four-dimensional normal Abelian subgroup, but neither a preferred Lorentz nor a Poincaré subgroup. (However, it does differ from the BMS group in an important respect: while the BMS supertranslations correspond to free functions on a 2-sphere, the Spi supertranslations correspond to free functions on a timelike three-dimensional hyperboloid.)

a)Supported in part by the NSF contract PHY 76-81102 with the University of Chicago. Present address: Département de Physique, Université de Clermont-Fd 63170, Aubière, France.
b)Visiting Scholar. Permanent address: Département de Mathématiques, Université de Clermont-Fd. 63170 Aubière, France.

 0022-2488/79/050793-08$01.00

In the third step, asymptotic gravitational fields are introduced. These are represented by two second-rank, symmetric, tracefree tensor fields E_{ab} and B_{ab} on the hyperboloid D of unit spacelike vectors in the tangent space at i^0, "the point at spatial infinity." The information about the "$1/r^3$ part" of the Weyl curvature of the physical space–time is contained in this pair of fields on D. In the last step, one uses these asymptotic symmetries and asymptotic fields to construct conserved quantities.[6] The 4-momentum P_a may be regarded as a vector at i^0. Its definition involves Spi translations, i.e., the elements of the preferred four-dimensional normal subgroup of the Spi group and the asymptotic field E_{ab}. (The corresponding quantity constructed from B_{ab} which would represent the "magnetic analog of the gravitational charge" vanishes identically![7]) To introduce angular momentum, an *additional* condition is required. If $B_{ab}=0$, i.e., if the "$1/r^3$ part" of the physical Weyl curvature contains information only about the 4-momentum as one intuitively expects it to, and if the "$1/r^4$ contribution to the magnetic part" of the asymptotic curvature is well behaved, one can define the total angular momentum. Vanishing of B_{ab} causes a reduction of the Spi group to the Poincaré, and angular momentum then arises as a linear mapping (involving the "$1/r^4$ part") from the Lorentz Lie algebras of this Poincaré group to the reals. In the finished picture, angular momentum is represented by a skew, second rank tensor M_{ab} at i^0 which "transforms" in the familiar way under the action of Spi translations.[8] One can also construct, in the usual fashion, the spin vector S_a: $S_a := \epsilon_{abcd}P^bM^{cd}$, where ϵ_{abcd} is the alternating tensor at i^0. Like the 4-momentum P_a, the spin vector S_a is a fixed vector at i^0; it is invariant under the action of Spi translations.

The availability of these apparently distinct notions of energy–momentum and angular momentum raises a number of questions. Fix a space–time $(\hat{M}, \hat{g}_{ab})$ satisfying the asymptotic conditions at spatial infinity.[3,9] Then, one has available the 4-momentum P_a and the angular momentum M_{ab} at i^0. Let us now assume that the space–time admits a stationary Killing field $\hat{l}^a$, i.e., a Killing field which gives rise to a Spi time-translation. What then is the relation between the Komar scalar and the 4-vector P_a at i^0? Is P_a necessarily timelike? If so, is the "asymptotic rest frame" selected by the Killing field $\hat{l}^a$ the same as that selected by P_a? Consider, next, the notion of angular momentum. Does the presence of $\hat{l}^a$ automatically imply the vanishing of B_{ab} required for M_{ab} to be well defined? Does the spin vector S_a at i^0 "project down" to the Hansen dipole moment defined at the point at infinity of the manifold of orbits of $\hat{l}^a$? Let there exist, in place of $\hat{l}^a$, an axial Killing field $\hat{R}^a$. Does it then also follow that $B_{ab}=0$ at i^0? If so, what is the relation between the Komar scalar involving $\hat{R}^a$ and the spin vector S_a at i^0? The purpose of this paper is to answer these and related questions.

The plan of the paper is the following. Section II is devoted to preliminaries. Various facts about the conformal completion involving the point i^0 at spatial infinity are recalled and the asymptotic properties of Killing fields available on space–times admitting this completion are discussed. In Sec. III, asymptotically flat space–times admitting translational Killing fields are considered. It is shown that if the 4-momentum P_a is nonzero, such a field can not be spacelike at infinity. Under the assumption that it is timelike, P_a is shown to be proportional to K_a, the vector at i^0 which corresponds to the Spi translation induced by the Killing field. Finally, the proportionality factor is shown to be precisely the Komar scalar constructed from the Killing field. In Sec. IV, we examine angular momentum. Given the stationary Killing field, B_{ab} is shown to vanish and an expression is given for the spin vector S_a in terms of the asymptotic value of the twist of the Killing field. This expression is the natural "lift" to i^0 of the Hansen dipole moment, defined at the point at infinity of the manifold of orbits of the Killing field. Finally, nontranslational Killing fields are considered. Under the assumption that the 4-momentum P_a is timelike, it follows that the action of the Killing field on the tangent space at i^0 must be a spatial rotation in the 3-plane orthogonal to P_a. It is shown that B_{ab} must gain vanish and that the resulting spin vector S_a must point along the "axis vector" at i^0 defined by this rotation, its norm being equal to the Komar integral constructed from the Killing field.

II. PRELIMINARIES

Definition 1: A space–time $(\hat{M}, \hat{g}_{ab})$ will be said to be asymptotically empty and flat at spatial infinity provided there exists a space–time (M, g_{ab}) which is smooth everywhere except at a point i^0 where M is $C^{>1}$ and g_{ab} is $C^{>0}$, together with an imbedding of $\hat{M}$ in to M (with which we shall identify $\hat{M}$ and its image in M) satisfying the following conditions:

(i) $\bar{J}(i^0) = M - \hat{M}$;

(ii) There exists a function Ω on M which is C^2 at i^0, smooth elsewhere, such that on $\hat{M}$, $g_{ab} = \Omega^2 \hat{g}_{ab}$, and, at i^0, $\Omega = 0$, $\hat{\nabla}_a\Omega = 0$, and $\hat{\nabla}_a\hat{\nabla}_b\Omega = 2\hat{g}_{ab}$;

(iii) There exists a neighborhood $\mathcal{N}$ of i^0 in M such that in $\hat{M} \cap \mathcal{N}$, $\hat{g}_{ab}$ satisfies the vacuum Einstein equation.

Here, $\bar{J}(i^0)$ denotes the topological closure in M of the set of points which are causally related to i^0 and the notion of $C^{>k}$ differentiability is the same as in Refs. 3 and 9. (Thus, for example, the condition that g_{ab} be $C^{>0}$ at i^0 ensures that it is smooth in its "angular dependence" and its connection, i.e., the Christoffel symbols, have only finite "radial" discontinuities. For details, see Ref. 9.) This notion of asymptotic flatness is weaker than the one used in Ref. 3; we have omitted the conditions at null infinity since they play no role in the present analysis. Throughout this paper, we shall assume that the physical space–times under consideration satisfy these three conditions.

Given such a space–time, it is easy to show[3,9] that the Riemann tensor R_{abcd} of the rescaled metric g_{ab} is such that $\Omega^{1/2}R_{abcd}$ admits a direction-dependent limit at i^0. Set $C_{abcd}(\eta) = \lim_{\to i^0}\Omega^{1/2}C_{abcd}$, $E_{ac}(\eta) = C_{abcd}\eta^b\eta^d$ and, $B_{ac}(\eta) = {}^*C_{abcd}\eta^b\eta^d$, where η^a denotes the unit vector

tangent to the curve of approach to i^0. The fields $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$ are smooth, second rank, symmetric, tracefree tensor fields on $\mathcal{J}$, they carry information about the "$1/r^3$ part" of the physical Weyl curvature. The Bianchi identity on $\hat{C}_{abcd}$ yields the asymptotic field equations, $\mathbf{D}_{[a}\mathbf{E}_{b]c}=0$ and $\mathbf{D}_{[a}\mathbf{B}_{b]c}=0$ on $\mathcal{J}$. (The Ricci part of R_{abcd} only provides certain potentials for $\mathbf{E}_{ab}$ and $\mathbf{B}_{ab}$.) We shall refer to these asymptotic fields and equations repeatedly in the next two sections.

Next, we wish to discuss symmetries. Fix a space–time $(\hat{\mathbf{M}}, \hat{g}_{ab})$ satisfying Definition 1. Then, the group of asymptotic symmetries at spatial infinity is the Spi group.[3,9] Given a completion $(\mathbf{M}, g_{ab})$ of $(\hat{\mathbf{M}}, \hat{g}_{ab})$, every infinitesimal Spi symmetry is characterized by a pair $(\mathbf{F}_{ab}, \mathbf{K}_a(\eta))$, where $\mathbf{F}_{ab}$ is a fixed skew tensor at i^0 and $\mathbf{K}_a(\eta)$ a direction-dependent vector at i^0 satisfying $\mathbf{K}_a(\eta) = \lim \hat{\nabla}_a\chi$, where χ is a $C^{>0}$ scalar field.[10] Infinitesimal Spi symmetries with $\mathbf{F}_{ab}=0$ are the supertranslations; their action leaves not only i^0 but also the tangent space at i^0 invariant. If, in addition, $\mathbf{K}_a(\eta)$ happens to be direction *independent*, the corresponding Spi symmetry is a translation.

Fix a Killing field $\hat{\xi}^a$ on $(\hat{\mathbf{M}}, \hat{g}_{ab})$. $\hat{\xi}^a$ will be said to be a *translational* Killing field if the diffeomorphism it generates induces[11] a Spi translation at spatial infinity. (It is straightforward to verify that if the induced action of $\hat{\xi}^a$ is a Spi supertranslation, then it is necessarily a Spi translation.) $\hat{\xi}^a$ will be said to be a *rotational* Killing field if the induced[11] Spi symmetry is a spatial rotation. In this paper, we shall be concerned only with these two types of Killing fields. In either case, the pair $(\mathbf{F}_{ab}, \mathbf{K}_a(\eta))$ at i^0 arises in the following way. Since $\hat{\xi}^a$ is a Killing field on $(\hat{\mathbf{M}}, \hat{g}_{ab})$, it is a conformal Killing field on $(\hat{\mathbf{M}}, g_{ab})$. Hence, $\hat{\xi}^a$ $(\equiv \xi^a)$ is completely characterized by the quadruplet $(\xi^a, F_{ab}=\nabla_{[a}\xi_{b]}, \Phi=\frac{1}{4}\nabla_a\xi^a, K_a=\nabla_a\Phi)$ at any point of $\hat{\mathbf{M}}$, where indices are raised and lowered using g_{ab}.[12] Then, $\mathbf{F}_{ab}=\lim_{\to i^0} F_{ab}$ and $\mathbf{K}_a=\lim_{\to i^0} K_a$ characterize the Spi symmetry which arises from ξ^a. (The other two pieces of the conformal Killing data, ξ^a and Φ, vanish in the limit reflecting the fact that the isometry generated by ξ^a leaves both i^0 and the metric g_{ab} at i^0 invariant.[10,12])

Consider the case when the space–time admits a translational Killing field $\hat{t}^a$. Then, $\lim_{\to i^0}\hat{t}^a=0$, $\lim_{\to i^0}F_{ab}=0$ and $\lim_{\to i^0}\Phi=0$. Using the fact that the metric g_{ab} is $C^{>0}$ at i^0, a repeated application of the l'Hospital rule yields the following information: (i) $\lim_{\to i^0}\Omega^{-1}\hat{t}^a = -\frac{1}{2}\mathbf{K}^a + (\mathbf{K}\cdot\eta)\eta^a$; (ii) the norm $\hat{\lambda}=\hat{g}_{ab}\hat{t}^a\hat{t}^b$ of $\hat{t}^a$ is $C^{>0}$ at i^0, with $\lim_{\to i^0}\nabla_a\hat{\lambda}=\mathbf{f}\eta_a$ where $\mathbf{f}$ is a smooth function on $\mathcal{J}$, and, (iii) $\lim_{\to i^0}\Omega^{1/2}\hat{\omega}_a=0$, where, $\hat{\omega}_a = \hat{\mathcal{E}}_{abcd}\hat{t}^b\hat{\nabla}^c\hat{t}^d$ is the twist of the Killing field $\hat{t}^a$. Next, using the fact that in the vacuum region, $\hat{\omega}_a$ admits a scalar potential, $\hat{\omega}_a=\hat{\nabla}_a\omega$,[13] it follows that $\lim_{\to i^0}\Omega^{-1/2}\hat{\omega}_a$ exists as a regular, direction dependent vector at i^0.

Next, consider the case when the space–time admits a rotational Killing field $\hat{R}^a$. Then, $\lim_{\to i^0}\hat{R}^a=0$, $\lim_{\to i^0}\Phi=0$, and $\lim_{\to i^0}F_{ab}$ is direction independent. The differentiability of g_{ab} at i^0 yields, as before, the following asymptotic properties of $\hat{R}^a$: (i) $\lim_{\to i^0}\Omega^{-1/2}\hat{R}^a = \mathbf{F}^{ab}\boldsymbol{\eta}_b$; (ii) $\hat{\mu}=\hat{g}_{ab}\hat{R}^a\hat{R}^b$ is such that $\mu=\lim_{\to i^0}\Omega\hat{\mu}$ is a regular direction-dependent scalar, and (iii) $\Omega\hat{\sigma}_a$ admits a regular direction-dependent limit at i^0, where $\hat{\sigma}_a = \hat{\mathcal{E}}_{abcd}\hat{R}^b\hat{\nabla}^c\hat{R}^d$ is the twist of $\hat{R}^a$. Finally, Einstein's equation implies a stronger condition on $\hat{\sigma}_a$: $\sigma_a = \lim_{\to i^0}\Omega^{1/2}\hat{\sigma}_a$ exists as a regular direction dependent vector at i^0 with the property that $\boldsymbol{\sigma}_a\eta^a=0$.

These asymptotic properties of Killing fields will play a crucial role in the next two sections.

III. THE ENERGY-MOMENTUM

Let $(\hat{\mathbf{M}}, \hat{g}_{ab})$ be asymptotically empty and flat at spatial infinity, Let it admit a translational Killing field $\hat{t}^a$. Then, we have

Lemma 1.1: **On the hyperboloid $\mathcal{J}$ of unit spacelike vectors at i^0, the tensor field $\mathbf{E}_{ab}$ must satisfy $\tilde{\mathbf{K}}^m\mathbf{D}_m\mathbf{E}_{ab} = 3\,(\mathbf{K}.\eta)\,\mathbf{E}_{ab}$ where, $\mathbf{h}_{ab}$ is the natural metric on $\mathcal{J}$, $\mathbf{D}_a$ the derivative operator on $(\mathcal{J}, \mathbf{h}_{ab})$, $\mathbf{K}_a = \lim_{\to i^0}\frac{1}{4}\hat{\nabla}_n\hat{\nabla}_m\hat{t}^m$ is the vector at i^0 defined by the translation $\hat{t}^a$ and $\mathbf{K}^a=\mathbf{h}^{ab}\mathbf{K}_b$.**

Proof: $\mathbf{E}_{ab} := \lim_{\to i^0}\Omega^{1/2}C_{ambn}\eta^m\eta^n$

$$= \lim_{\to i^0}\Omega^{-3/2}\hat{C}_{ambn}\hat{\nabla}^m\Omega^{1/2}\hat{\nabla}^n\Omega^{1/2},$$

since $\eta^a=\lim_{\to i^0}\nabla^a\Omega^{1/2}$. Using the fact that $\hat{t}^a$ is a Killing field on $(\hat{\mathbf{M}}, \hat{g}_{ab})$ it follows that

$$\pounds_t\left[\Omega^{-3/2}\hat{C}_{ambn}(\hat{\nabla}^m\Omega^{1/2})(\hat{\nabla}^n\Omega^{1/2})\right]$$
$$= -\tfrac{1}{2}\,\Phi\Omega^{-3/2}\hat{C}_{ambn}(\hat{\nabla}^m\Omega^{1/2})\hat{\nabla}^n\Omega^{1/2} + \Omega^{-1}\hat{C}_{ambn}(\hat{\nabla}^{(m}\Phi)(\hat{\nabla}^{n)}\Omega^{1/2}).$$

The desired result in now obtained by taking the limit of this equation at i^0 if one uses the fact that $\Omega^{1/2}C_{abcd}$ admits a regular[10] direction dependent limit at i^0, that the presence of the Killing field $\hat{t}^a$ on $\hat{\mathbf{M}}$ implies $\mathbf{B}_{ab}$ must vanish at i^0 (Theorem 3, proved in Sec. IV) and the asymptotic properties of $\hat{t}^a$ listed in Sec. II.

Lemma 1.1 implies that the presence of the translational isometry constrains the asymptotic gravitational field $\mathbf{E}_{ab}$ on $\mathcal{J}$ in a nontrivial fashion. This constraint will be used to show that if $\hat{t}^a$ is a space translation i.e., if the vector $\mathbf{K}_a$ at i^0 is spacelike then $\mathbf{E}_{ab}$ must vanish identically. This would in particular imply the vanishing of the 4-momentum $\mathbf{P}_a$. Thus, as one might expect on intuitive grounds, physically interesting space–times can not admit space translations if they are asymptotically flat at spatial infinity.

Lemma 1.2: If $\mathbf{e} := \mathbf{E}_{ab}\tilde{\mathbf{K}}^a\tilde{\mathbf{K}}^b$ vanishes on $\mathcal{J}$, so must $\mathbf{E}_{ab}$, where, as before, $\tilde{\mathbf{K}}^a$ is the projection of $\mathbf{K}^a$ into $\mathcal{J}$.

Proof: Using the field equation $\mathbf{D}_{[a}\mathbf{E}_{b]c}=0$ and Lemma 1.1, it follows that $\mathbf{D}_a\mathbf{e} = (\mathbf{K}\cdot\eta)\mathbf{E}_{ab}\tilde{\mathbf{K}}^b$. Hence, if $\mathbf{e}$ vanishes on $\mathcal{J}$, so does $(\mathbf{K}\cdot\eta)\mathbf{E}_{ab}\tilde{\mathbf{K}}^b$. Since $(\mathbf{K}\cdot\eta)$ vanishes only on a set of measure zero of $\mathcal{J}$, and since both $\mathbf{E}_{ab}$ and $\tilde{\mathbf{K}}^b$ are smooth, it follows that $\mathbf{E}_{ab}\tilde{\mathbf{K}}^b=0$. Hence, $0=\mathbf{D}_b(\mathbf{E}_{am}\tilde{\mathbf{K}}^m) = 2(\mathbf{K}\cdot\eta)\,\mathbf{E}_{ab}$, whence $\mathbf{E}_{ab}=0$ on $\mathcal{J}$. □

Remarks: (i) Using the field equation on $\mathbf{E}_{ab}$ and Lemma 1.1, it follows that $\mathbf{e}$ satisfies an hyperbolic equation on $(\mathcal{J}, \mathbf{h}_{ab})$: $\mathbf{D}^m\mathbf{D}_m\,\mathbf{e} - \mathbf{e} = 0$. Hence, it follows that if $\mathbf{e}$ and its normal derivative to a 2-sphere cross section of $\mathcal{J}$ vanish on that 2-sphere, $\mathbf{e}$ must vanish everywhere.

(ii) Without reference to any isometries, one can show that if E_{ab} vanishes on a 2-sphere cross section, it must be zero everywhere on $\mathcal{J}$.

Theorem 1: If the asymptotic field E_{ab} fails to vanish identically on $\mathcal{J}$, $\hat{l}^a$ can not be a spatial translation, i.e., the vector $\mathbf{K}^a$ at i° can not be spacelike.

Proof: Using Lemma 1.1, it follows that $\tilde{K}^m D_m e = (K\cdot\eta)\, e$ on $\mathcal{J}$, where, as before, $e = E_{ab}\tilde{K}^a\tilde{K}^b$. Hence, along any integral curve of $\tilde{K}^a$, the field e is given by $e = C(K\cdot K - (K\cdot\eta)^2)^{-1/2}$ where C is a constant on the given integral curve, but can change from one curve to another. Let K_a be a spacelike vector at i°. Then, *every* integral curve of $\tilde{K}^a$ enters an arbitrarily small neighborhood of some point at which $K\cdot K - (K\cdot\eta)^2$ vanishes.[14] Since e is smooth everywhere on $\mathcal{J}$, it follows that the constant C must vanish for all integral curves, whence, $e = 0$ on $\mathcal{J}$. Lemma 1.2 now asserts that E_{ab} must also vanish identically on $\mathcal{J}$. □

Remarks: (i) As one expects to be the case, the argument is inapplicable when K^a is timelike: in this case, $K\cdot K - (K\cdot\eta)^2$ can vanish nowhere, since it is bounded above by $K\cdot K$. What is the situation when K^a is null? Now, there do exist points where $K\cdot K - (K\cdot\eta)^2 = -(K\cdot\eta)^2$ vanishes. However, it is no longer true that, given an integral curve of $\tilde{K}^a$, one can find a point on $\mathcal{J}$ at which $-(K\cdot\eta)^2$ vanishes and whose arbitrarily small neighborhoods are reached by the integral curve. Hence, it appears that one cannot rule out the possibility that space–times under consideration admit a null translation; additional asymptotic conditions at *null* infinity may be necessary for this purpose.

(ii) Note that the theorem implies, in particular, that space–times which are asymptotically empty and flat at spatial infinity cannot admit two translational Killing fields; in this case, one can always obtain a space translation by taking suitable linear combinations of these Killing fields. This result is of relevance, e.g., to the issue of uniqueness of the vacuum state in stationary space–times: If such a space–time satisfies the present asymptotic conditions, there is available a canonical vacuum state for linear quantum fields on this space–time.

From now on, we shall assume that the available translational Killing field is timelike; although the analysis involving null translations is straightforward, this case has little physical interest. The next step is the investigation of the relation between this Killing field and the ADM 4-momentum P_a. We begin by showing that the asymptotic rest frame defined by the Killing field is the same as that defined by P_a:

Lemma 2.1: The 4-momentum P_a at i° is proportional to the vector K_a defined by the Killing field $\hat{l}^a$.

Proof: Fix any vector V^a at i° satisfying $V\cdot K = 0$. We wish to show that $P\cdot V = 0$. By definition of P_a, one has $2P_aV^a = \int_S 2E_{ab}\tilde{V}^a\, dS^b$ where $\tilde{V}^a = h^a_b V^b$ is the projection of V^a into the hyperboloid $\mathcal{J}$, and where the integral is taken on any 2-sphere cross section of $\mathcal{J}$.[3,9] Choose for S^2 the 2-sphere cross section S defined by $K\cdot\eta = 0$. Then, from Lemma 1.1, it follows that on S, $\gamma^p_a\gamma^q_b D_p(E_{mq}\tilde{K}^m) = 0$, whence, $\gamma^m_a\gamma^n_b D_m(E_{pq}\tilde{K}^p\gamma^q_n) = 0$, where $\gamma_{ab} = h_{ab} + (-\tilde{K}^m\tilde{K}_m)^{-1}\tilde{K}_a\tilde{K}_b$ is the intrinsic metric on S. The last equation states that the vector field $E_{ab}\tilde{K}^a\gamma^b_n$, tangential to S, is covariantly constant thereon. Since (S, γ_{ab}) is of constant curvature, it now follows that $E_{ab}\tilde{K}^a\gamma^b_n = 0$ on S, whence $E_{ab}\tilde{K}^aV^b = 0$. Since $\tilde{K}^a$ is normal to S, one has $2P_aV^a = \int_S E_{ab}\tilde{V}^a dS^b = 0$. □

Next, we wish to reexpress the norm of P_a, i.e., the total mass associated with the space–time, in terms of the asymptotic properties of the Killing field. The expression for mass involves the asymptotic value of the Weyl curvature of the physical space–time. Since vacuum equations hold in a neighborhood of i°, we have, in this neighborhood, $\hat{\nabla}_a\hat{\nabla}_b\hat{l}_c = \hat{C}_{cba}{}^m\hat{l}_m$. *The key idea is to use this relation to re-express "mass aspect" in terms of the Killing field $\hat{l}^a$ and to use the asymptotic properties of the norm and twist discussed in* Sec. II.

Lemma 2.2: The 4-momentum P_a is given by $P_a = m(-K\cdot K)^{-1/2}K_a$ with $m = -\frac{1}{2}\int_S(\eta^p\hat{\lambda}_p)\, dS$, where $\hat{\lambda}_p = \lim_{\to i^\circ}\hat{\nabla}_p\hat{\lambda}$, S is the 2-sphere defined by $K\cdot\eta = 0$, and dS the natural volume element[15] on S. (Here, the Killing vector $\hat{l}^a$ has been so rescaled that $\hat{\lambda} = \lim_{\to i^\circ} -\hat{g}_{ab}\hat{l}^a\hat{l}^b = 1$.)

Proof: Set $m = -P_a(-K\cdot K)^{-1/2}K^a$. Then, $m = -\frac{1}{2}\int_{S2}E_{ab}\tilde{K}^b(-K\cdot K)^{-1/2}\, dS^b$. Choosing for S^2 the 2-sphere S, one obtains

$$m = -\tfrac{1}{2}\int_S E_{ab}t^at^b\, dS$$

$$\equiv -\tfrac{1}{2}\int_S \lim_{\to i^\circ}[\Omega^{1/2}C_{ambn}\eta^m\eta^n(\Omega^{-1}l^a)(\Omega^{-1}l^b)]\, dS$$

where $\eta^a = \nabla^a\Omega^{1/2}$ is a vector field whose limit at i° yields the unit tangent η^a to the curve along which the limit is taken, and C_{ambn} is the Weyl tensor of the rescaled metric g_{ab}. In terms of the physical metric $\hat{g}_{ab}$, one has $m = -\frac{1}{2}\int_S\lim_{\to i^\circ}[\Omega^{1/2}\hat{C}_{ambn}\eta^m\eta^n(\Omega^{-1}\hat{l}^a\hat{l}^b]\, dS$. We now focus on the integrand. In the neighborhood of i° where $\hat{g}_{ab}$ satisfies the vacuum equation, one has $\hat{C}_{ambn}\hat{l}^a\hat{l}^b = \hat{R}_{ambn}\hat{l}^a\hat{l}^b = \hat{l}^a\hat{\nabla}_n\hat{\nabla}_a\hat{l}_m$. Next, using the expression for $\hat{\nabla}_a\hat{l}_m$ in terms of the norm and the twist of $\hat{l}^a$, $\hat{\lambda}\hat{\nabla}_a\hat{l}_m = \hat{l}_{[m}\hat{\nabla}_{a]}\hat{\lambda} + \frac{1}{2}\hat{\mathcal{E}}_{mapq}\hat{l}^p\hat{\omega}^q$, and the asymptotic properties of $\hat{\lambda}$ and $\hat{\omega}^q$, $\hat{\lambda}$ is $C^{>0}$ at i° and $\lim_{\to i^\circ}\hat{\omega}_q$ vanishes at i°, one obtains $m = -\frac{1}{2}\int_S\eta^b\hat{\lambda}_p\, dS$. □

Finally, we wish to relate the proportionality factor m between P_a and $(-K\cdot K)^{-1/2}K_a$ to the value of the Komar integral involving the timelike Killing field $\hat{l}^a$. The integral itself is defined over a 2-sphere $\hat{S}^2$ in the vacuum region surrounding sources, the value of the integral being independent of the choice of the 2-sphere. Since the factor m has been expressed above in terms of the *asymptotic* properties of the Killing field, we shall evaluate the Komar integral also on a "2-sphere at infinity."

Lemma 2.3: The value of the Komar integral,[1,15] $m_k = \frac{1}{2}\int_{\hat{S}^2}\hat{\mathcal{E}}_{abcd}\hat{\nabla}^a\hat{l}^b\, d\hat{S}^{cd}$ equals m, the magnitude of the 4-momentum vector P_a.

Proof: Fix any three-dimensional submanifold $\mathbf{T}$ in $\mathbf{M}$ ($C^{>1}$ at i^0 and C^∞ elsewhere) which is orthogonal to $\mathbf{K}^a$ at i^0. Consider a sequence of 2-spheres $\hat{S}^2$ in $\mathbf{T}$ which converges to the point i^0. Then,

$$\lim_{\to i^0} \int_{S^2} \hat{\mathcal{E}}_{abcd} \hat{\nabla}^c \hat{l}^d \, d\hat{S}^{ab}$$

$$= \lim_{\to i^0} \int_{S^2} \hat{\mathcal{E}}_{abcd} \hat{\lambda}^{-1} (\hat{l}^d \hat{\nabla}^c \hat{\lambda} + \tfrac{1}{2} \mathcal{E}^{dcmn} \hat{l}_m \hat{\omega}_n) d\hat{S}^{ab}$$

$$= \int_{S^2} \lim_{\to i^0} [\mathcal{E}_{abcd} (\Omega^{-1} l^d) \nabla^c \hat{\lambda} - 2(\Omega^{-1} l_a) \hat{\omega}_b] dS^{ab}$$

where, S is, as before, the 2-sphere cross section of $\mathcal{D}$ defined by $\mathbf{K} \cdot \eta = 0$, and where $\mathcal{E}_{abcd}$ and ∇ are, respectively the alternating tensor and the derivative operator compatible with g_{ab}. Here, we have used the expression for $\hat{\nabla}_a \hat{l}_b$ in terms of the norm and the twist of $\hat{l}^a$ and the fact that the volume element on S can be obtained by rescaling the volume elements (induced by g_{ab}) on the sequence of 2-spheres by Ω^{-1} and then taking the limit. Finally, using the fact that $\hat{\lambda}$ is $C^{>0}$ at i^0, $\lim_{\to i^0} \Omega^{-1} l^a$ exists and $\lim_{\to i^0} \hat{\omega}_a$ vanishes, one obtains $m_k = -\frac{1}{2} \int_S \eta^a \hat{\lambda}_a \, dS$. □

Collecting the results of Lemmas 2.1, 2.2, and 2.3, we now have.

Theorem 2: The 4-momentum $\mathbf{P}_a$ is given by $\mathbf{P}_a = m_k (-\mathbf{K} \cdot \mathbf{K})^{-1/2} \mathbf{K}_a$, where, m_k is the value of the Komar integral and where $\mathbf{K}_a$ is the (timelike) vector at i^0 defined by the time translation $\hat{l}^a$ on $(\hat{\mathbf{M}}, \hat{g}_{ab})$.

Remarks: (i) Note that the results obtained above depend quite crucially on the "fall-off" properties of the norm and the twist of the timelike Killing field. The fact that these properties can be deduced directly from basic definitions at i^0 therefore lends independent support in favor of the asymptotic conditions introduced in Definition 1. In particular, although results relating the ADM 4-momentum with the Komar integral have been announced before,[16] to our knowledge, their derivation has always involved an imposition of asymptotic conditions on the norm and the twist of the Killing field by hand. Also, since a precise formulation of the notion of the "asymptotic rest frame selected by the Killing field" was not available in absence of i^0, these results were somewhat heuristic.

(ii) Using Lemma 1.1, one can easily show that in presence of a time translation, the "mass aspect" $\mathbf{E}_{ab} \tilde{\mathbf{K}}^a \tilde{\mathbf{K}}^b$ is *spherically symmetric* on the 2-sphere S defined by $\mathbf{K} \cdot \eta = 0$ (but *not* on any other 2-sphere!). Using Lemma 1.2 it then also follows that, if the mass happens to vanish, so must $\mathbf{E}_{ab}$ itself. We shall see in the next section that the presence of a time translation itself implies that $\mathbf{B}_{ab}$ must vanish on $\mathcal{D}$. Thus, at least in the stationary case, the "$1/r^3$ part" of the Weyl tensor in the physical space–time contains *"only* the mass information."

IV. ANGULAR MOMENTUM

This section is divided in to two parts. In the first, we assume that the given space–time admits a time translation isometry and examine the relation between the asymptotic properties of the twist of the translational Killing field and the spin vector $\mathbf{S}_a$ defined at i^0 in terms of the asymptotic behavior of the Weyl curvature. In the second, we assume that the space–time admits a rotational Killing field and discuss the relation between the corresponding Komar integral and the spin vector $\mathbf{S}_a$.

A. Stationary space–times

Let $(\hat{\mathbf{M}}, \hat{g}_{ab})$ be asymptotically empty and flat at spatial infinity and admit a time translation $\hat{l}^a$ as in Sec. III. Then, we have

Theorem 3: The field $\mathbf{B}_{ab}$ on $\mathcal{D}$ must vanish identically.

Proof: Using the fact[3,9] that $\mathbf{B}_{ab}$ admits a scalar potential B, with $\mathbf{B}_{ab} = \mathbf{D}_a \mathbf{D}_b \mathrm{B} + \mathrm{B} h_{ab}$, one can show that $\mathbf{B}_{ab}$ vanishes on $\mathcal{D}$ if and only if $\mathbf{B}_{ab} \mathbf{K}^a \mathbf{K}^b \equiv \lim_{\to i^0} 4\Omega^{1/2} {}^*C_{ambn} \times \hat{l}^a \hat{l}^b \eta^m \eta^n = 0$ where $\eta^a = \nabla^a \Omega^{1/2}$. [Recall that $\lim_{\to i^0} l_a = -\frac{1}{2} \mathbf{K}_a + (\mathbf{K} \cdot \eta) \eta_a$.] Using the fact that in the neighborhood of i^0 where the vacuum equation holds, one has $\hat{\nabla}_a \hat{\nabla}_b \hat{l}_c = \hat{C}_{cba}{}^m \hat{l}_m$, it follows that $\mathbf{B}_{ab} \mathbf{K}^a \tilde{\mathbf{K}}^b = 4 \lim_{\to i^0} [\Omega \ \eta^a \eta^b \}$. Proj. $[(2\lambda)^{-1} g_{ab} \omega^m \nabla_m \lambda - \nabla_b \omega_a)]$, where Proj. stands for the operation of projecting the indices of the tensor field that follows in the 3-flat orthogonal to $\hat{l}^a$, and where $\hat{\lambda}$ and $\hat{\omega}^a$ denote, as before, the norm and the twist of $\hat{l}^a$. Finally, using the asymptotic properties of $\hat{\lambda}$ and $\hat{\omega}^a$, one obtains $\mathbf{B}_{ab} \tilde{\mathbf{K}}^a \tilde{\mathbf{K}}^b = 0$. □

Thus, the presence of the Killing field $\hat{l}^a$ ensures that the "$1/r^3$ contribution to the asymptotic curvature is purely electric." As noted in the Introduction vanishing of $\mathbf{B}_{ab}$ enables one to select a canonical Poincaré subgroup of the Spi group.[3,9] (This situation is rather analogous to that at null infinity: in stationary space–times one can also select a preferred Poincaré subgroup of the BMS group.) This selection of the Poincaré subgroup is the crucial step in the introduction of an angular momentum at spatial infinity. Note, however, that vanishing of $\mathbf{B}_{ab}$ is not quite sufficient for angular momentum to be well defined: it is necessary, in addition, that the "next order," (i.e., "the $1/r^4$") contribution to the magnetic part of the asymptotic curvature be well behaved. (At least in principle, it may happen that the "magnetic part falls off as, say, $(r^3 \log r)^{-1}$.") Nonetheless, Theorem 3 does indicate that in a generic case, stationarity together with the asymptotic conditions introduced in Definition 1 will ensure that the total angular momentum is an unambiguous notion. In the rest of this subsection, we restrict ourselves to space–times for which this is the case.

Our next task is to relate the spin vector $\mathbf{S}_a$ at i^0, defined by $\mathbf{S}_a = \mathcal{E}_{abcd} \mathbf{M}^{cd} \mathbf{P}^b$, to the asymptotic behavior of the twist. We have

Theorem 4: The spin vector $\mathbf{S}_a$ satisfies[15] $\mathbf{S}_a \mathbf{V}^a = m/2 \int_S (\eta_a \omega^a)(\eta_b \mathbf{V}^b) \, dS$, where $\mathbf{V}^a$ is any vector at i^0 with $\mathbf{V} \cdot \mathbf{P} = 0$, $\omega_a = \lim_{\to i^0} \Omega^{-1/2} \hat{\omega}_a$ and S is the 2-sphere cross section of $\mathcal{D}$ defined by $\mathbf{P} \cdot \eta = 0$.

Proof: Set $\boldsymbol{\beta}_{ab} = \lim_{\to i^0} {}^*C_{ambn} \eta^m \eta^n$. Then, the tensorial angular momentum $\mathbf{M}_{ab}$ is given by[3,9] $\mathbf{M}_{ab} \mathbf{F}^{ab} = \frac{1}{2} \int_{S^2} \boldsymbol{\beta}_{ab} \boldsymbol{\zeta}^a \, d\mathbf{S}^b$, where $\mathbf{F}^{ab}$ is an arbitrary skew tensor at i^0, $\boldsymbol{\zeta} = \mathcal{E}^{abcd} \mathbf{F}_{cd} \eta_b$, and where the integral is taken on any 2-sphere cross section of $\mathcal{D}$. Using the definition of $\mathbf{S}_a$ in terms of $\mathbf{M}_{ab}$, one obtains, after some simplification, $\mathbf{S}_a \mathbf{V}^a = m \int_S (\mathbf{V} \cdot \eta) \, (\lim_{\to i^0} \Omega \hat{C}_{abcd} \hat{l}^a) \eta^b \, dS^{cd}$. Next, using the expression for second derivative of $\hat{l}^a$ in terms of the curvature tensor of $\hat{g}_{ab}$, the expression simplifies to

$$\mathbf{S}_a \mathbf{V}^a = m/2 \int_S (\mathbf{V} \cdot \eta) \times \lim_{\to i^0} (\Omega \hat{\mathcal{E}}_{cdmn} \hat{l}^m (\eta^b \hat{\nabla}_b \hat{\omega}_n)) \, dS^{cd}.$$

Finally, using the fact that $\Omega^{-1/2}\hat{\omega}_n$ admits a regular direction dependent limit ω_n at i°, one has

$$\mathbf{S}_a\mathbf{V}^a = m/2\int_S (\mathbf{V}\cdot\eta)\mathcal{E}_{cdmn}(-\mathbf{K}\cdot\mathbf{K})^{-1}\mathbf{K}^m\omega^n\,dS^{cd}$$
$$= m/2\int_S (\eta_a\omega^a)(\mathbf{V}\cdot\eta)\,dS. \quad \square$$

Remark: Hansen's[2] dipole angular momentum moment, defined under the assumption that the 3-manifold of orbits of $\hat{l}^a$ is asympotically flat at spatial infinity, also involves the asymptotic properties of the twist field $\hat{\omega}_a$. Although no theorem exists to the effect that the notion of asymptotic flatness used in the present analysis is equivalent to that used by Hansen, on intuitive grounds one does expect the equivalence to hold in an appropriate sense. One is therefore led to ask for the relation between the above expression for $\mathbf{S}_a$ in terms of ω_a and Hansen's expression for the dipole moment. We claim that there is a sense in which the two expressions are the same. Apart from an overall factor of $m/2$, Hansen's dipole moment may be regarded as the projection to the (conformally completed) manifold of orbits of $\hat{l}^a$ of the spin vector $\mathbf{S}_a$.[17]

B. Axisymmetric space–times

In this subsection, we shall assume that the given space–time $(\hat{\mathbf{M}}, \hat{\mathbf{g}}_{ab})$ is asymptotically empty and flat at spatial infinity and admits a rotational Killing field $\hat{R}^a$ in the sense of Sec. II. (Note that we do *not* assume the existence of a translational Killing field.) We shall first show that the presence of $\hat{R}^a$ again implies that $\mathbf{B}_{ab}$ must vanish on $\mathcal{I}$ and then analyse the relation between the resulting spin vector $\mathbf{S}_a$ and the Komar integral involving $\hat{R}^a$.

Lemma 5.1: If $\mathbf{R}^a$ denotes the Killing field on $(\mathcal{I}, h_{ab})$ induced by the rotational Killing field $\hat{R}^a$, $\pounds_{\mathbf{R}}\mathbf{B}_{ab} = 0$ on $\mathcal{I}$.

Proof: Set $B_{ab} = \Omega^{1/2}{}^*C_{ambn}\nabla^m\Omega^{1/2}\nabla^n\Omega^{1/2}$ on $\hat{\mathbf{M}}$ so that $\mathbf{B}_{ab} = \lim_{\to i^\circ} B_{ab}$. Then, $\pounds_{\hat{R}}B_{ab} = \frac{1}{2}\Phi B_{ab} + 4\Omega{}^*C_{ambn} \times \nabla^{(m}\Phi\nabla^{n)}\Omega^{1/2}$, where $\Phi = \frac{1}{4}\nabla_m\hat{R}^m$. The desired result follows by taking the limit of this equation and noting that $\lim_{\to i^\circ} B_{ab}$, $\lim_{\to i^\circ}\Omega^{1/2}C_{abcd}$, and $\lim_{\to i^\circ}\nabla^m\Phi$ yield regular direction dependent tensors at i° and $\lim_{\to i^\circ}\Phi$ vanishes. $\square$

Lemma 5.2: $\mathbf{B}_{ab}\mathbf{R}^a\mathbf{R}^b$ vanishes on $\mathcal{I}$ if and only if $\mathbf{B}_{ab}$ itself vanishes.

Proof: Since $\mathbf{B}_{ab}$ is symmetric and satisfies $\mathbf{D}_{[a}\mathbf{B}_{b]c} = 0$, it follows[3,9,18] that it admits a scalar potential $\mathbf{f}$ such that $\mathbf{B}_{ab} = \mathbf{D}_a\mathbf{D}_b\mathbf{f} + \mathbf{f}\mathbf{h}_{ab}$; $\mathbf{f}$ is unique up to addition of a fuction α satisfying $\mathbf{D}_a\mathbf{D}_b\alpha + \alpha\mathbf{h}_{ab} = 0$. Using this gauge freedom and Lemma 5.1, it is easy to check that one can always choose a potential $\mathbf{f}$ which satisfies $\pounds_{\mathbf{R}}\mathbf{f} = 0$. Assume that this choice has been made. Then, a direct computation yields $\mathbf{B}_{ab}\mathbf{R}^a\mathbf{R}^b = \frac{1}{2}(\mathbf{D}^b\mu)(\mathbf{D}_b\mathbf{f}) + \mathbf{f}\mu$ and $\mathbf{B}_{ab}\mathbf{R}^b = [\frac{1}{2}(\mathbf{D}^b\mu)(\mathbf{D}_b\mathbf{f}) + \mathbf{f}\mu]\mathbf{R}_a$, where $\mu = \mathbf{R}^a\mathbf{R}_a$ on $\mathcal{I}$. Hence if $\mathbf{B}_{ab}\mathbf{R}^a\mathbf{R}^b = 0$, we have $\mathbf{B}_{ab}\mathbf{R}^b = 0$. Using this result and the field equation, $\mathbf{D}_{[a}\mathbf{B}_{b]c} = 0$, on $\mathcal{I}$, it follows that $\mathbf{B}_{ab}\mathbf{D}^b\mu = 0$. Finally, since $\mathbf{B}_{ab}$ is tracefree and $\mathbf{R}^a$ and $\mathbf{D}_a\mu$ are mutually orthogonal (and vanish only on a set of measure zero), it follows that $\mathbf{B}_{ab}$ itself must vanish everywhere on $\mathcal{I}$. $\square$

We are now ready to prove the main result:

Theorem 5: The presence of an axial Killing field $\hat{R}^a$ implies that $\mathbf{B}_{ab}$ must vanish on $\mathcal{I}$.

Proof The main idea is to use the asymptotic properties of the norm $\hat{\mu}$ and the twist $\hat{\sigma}_a$ of $\hat{R}^a$ to show that $\mathbf{B}_{ab}\mathbf{R}^a\mathbf{R}^b$ must vanish on $\mathcal{I}$, vanishing of $\mathbf{B}_{ab}$ is then implied by Lemma 5.2. We have: $\mathbf{B}_{ab}\mathbf{R}^a\mathbf{R}^b = \lim_{\to i^\circ}\Omega^{3/2} \times {}^*\hat{C}_{ambn}\hat{R}^a\hat{R}^b\eta^m\eta^n = \lim_{\to i^\circ}\Omega^{3/2}\hat{\mathcal{E}}_{am}{}^{pq}\hat{R}^a\eta^m\eta^n\hat{\nabla}_n\hat{\nabla}_p\hat{R}_q$, where, as before, $\eta^m = \nabla^m\Omega^{1/2}$, and we have used the expression for the second derivative of the Killing field in terms of the curvature tensor. Expressing the derivative of $\hat{R}^a$ in terms of $\hat{\mu}$ and $\hat{\sigma}$, one obtains

$$\mathbf{B}_{ab}\mathbf{R}^a\mathbf{R}^b = \lim_{\to i^\circ}\Omega^{3/2}\eta^m\eta^n \times \mathrm{Proj.}\,((2\hat{\mu})^{-1}\hat{g}_{ab}\hat{\sigma}^m\hat{\nabla}_m\hat{\mu} - \hat{\nabla}_b\hat{\sigma}_a)$$

where "Proj." stands for the operation of projecting all indices of the tensor field that follows orthogonal to $\hat{R}^a$. Using the fact that $\lim_{\to i^\circ}\Omega\hat{\mu}$ and $\lim_{\to i^\circ}\Omega^{1/2}\hat{\sigma}_a$ exist as regular direction dependent tensors, it follows that $\mathbf{B}_{ab}\mathbf{R}^a\mathbf{R}^b = 0$ on $\mathcal{I}$. $\square$

Remarks. (i) Let us suppose that the ADM 4-momentum $\mathbf{P}_a$ is timelike. Consider an isometry in the physical space–time which induces at spatial infinity a nontranslational Spi symmetry. Since the 4-momentum is necessarily invariant under the action of isometries, it follows that the given isometry must give rise to a *spatial rotation* at i° (belonging to the rotation group acting on the 3-plane orthogonal to $\mathbf{P}_a$ in the tangent space at i°), in this case, the space-time can not admit a "boostlike" Killing field. Thus, if $\mathbf{P}_a$ is timelike, and if the space–time admits *any* nontranslational isometry, $\mathbf{B}_{ab}$ must vanish on $\mathcal{I}$.

(ii) The implications of Theorem 5 are the same as those of Theorem 3, the presence of the Killing field guarantees the "fall-off" property of the "magnetic part of the asymptotic curvature required for the reduction of the Spi group to the Poincaré, thereby indicating that the (tensorial) angular momentum at i° will probably be well defined. We now restrict ourselves to space–times for which this is the case and examine the relation between the spin vector $\mathbf{S}_a$ at i° and the Komar integral involving $\hat{R}^a$.

For simplicity, we shall now assume, in the main part of the discussion, that $\mathbf{P}_a$ is timelike and only comment at the end on other possibilities. Consider the timelike 2-plane in the tangent space at i° which is left invariant by the action of $\mathbf{R}^a$, i.e., which represents the axis of $\mathbf{R}^a$. (This 2-plane is spanned by vectors $\mathbf{V}^a$ such that $\mathbf{F}_{ab}\mathbf{V}^b = 0$, where, as in Sec. II, $\mathbf{F}_{ab} = \lim_{\to i^\circ}\nabla_a R_b$.) Let $\mathbf{A}^a$ denote the unit vector in this 2-plane which is orthogonal to $\mathbf{P}_a$. Then, we have:

Lemma 6.1: The spin vector $\mathbf{S}_a$ is proportional to $\mathbf{A}_a$.

Proof: The spin vector $\mathbf{S}_a$ must be left invariant by the action of $\mathbf{R}^a$ in the tangent space at i° and is, by definition, orthogonal to $\mathbf{P}_a$. Hence the result. $\square$

Thus, what remains is only the computation of the norm of the spin vector. We shall first express this norm in terms of the asymptotic values of the norm and

the twist of $\hat{R}^a$ and then compare the resulting expression with the Komar integral involving $\hat{R}^a$.

Lemma 6.2: S_a is given by $S_a = jA_a$ where[15] $j = -4\int_S \mu^{-1} R_a \sigma_b \, dS^{ab}$, where S is the 2-sphere cross section of $\mathcal{I}$ defined by $P \circ \eta = 0$. (Here, R^a is assumed to to be rescaled such that $\lim_{\to i^\circ} \nabla_a R_b = \epsilon_{abcd} P^c A^d$.)

Proof: $S_a A^a = \epsilon_{abcd} M^{cd} P^b A^a = \frac{1}{2}\int_S 2\beta_{ab} \xi^a d\mathbf{S}^b$ by definition[3,9] of M_{ab}, where, $\xi^a = P^{[a} A^{b]} \eta_b$ on $\mathcal{I}$ and S^2 is any 2-sphere cross section of $\mathcal{I}$. Using the definition of β_{ab} it then follows that $S_a A^a = \int_S -2(\mu)^{-1} D^b \mu \lim_{\to i^\circ} (\Omega^{3/2} {}^*C_{abcd} \hat{R}^a) \, dS^{cd}$, where $\mu = \mathbf{R} \cdot \mathbf{R}$ on $\mathcal{I}$. Next, using the fact that if $\hat{R}_{ab} = 0$, $\hat{\nabla}_a \hat{\nabla}_b \hat{R}_c = \hat{C}_{cba}{}^m \hat{R}_m$; that $\mu = \lim_{\to i^\circ} \Omega \hat{R} \circ \hat{R}$ and $\sigma_a = \lim_{\to i^\circ} \Omega^{1/2} \hat{\sigma}_a$ are well defined, and that $\sigma_a \cdot \eta^a = 0$, one obtains after considerable simplifications,

$$S_a A^a = \int_S (\mu)^{-2} R_a \left[\tfrac{1}{2} \mu^{-1} \sigma_b (D^p \mu D_p \mu) - D^p \mu D_p \sigma_b\right] dS^{ab}.$$

Finally, integrating the second term by parts and using the fact that since $\mu = R \circ R$ on $\mathcal{I}$, $D^a D_a \mu = \mu^{-1} D^a \mu D_a \mu - 4\mu$, one obtains the desired result. □

Using these results, one can now prove

Theorem 6: *The spin vector* S_a *is given by* $S_a = j_k A_a$, *where* j_k *is the value of the Komar integral involving* $\hat{R}^a$.

Proof: In the terminology of Lemma 6.2, we only have to show that $j = j_k$. Consider, as in Lemma 2.3, a spacelike three-dimensional submanifold of M, passing through i°, and a family of 2-spheres $\hat{S}^2$ on this submanifold, converging to the point i°. Then

$$j_k = \int_{\hat{S}} 2\hat{\epsilon}_{abcd} \hat{\nabla}^a R^b \, d\hat{S}^{cd}$$
$$= \int_{\hat{S}} 2\hat{\epsilon}_{abcd} (\hat{R}^b \hat{\nabla}^a \hat{\mu} + \tfrac{1}{2} \hat{\epsilon}_{abmn} \hat{R}^m \hat{\sigma}^n) \hat{\mu}^{-1} \, dS^{cd},$$

where, $\hat{S}^2$ is *any* 2-sphere in the sequence. Using the asymptotic properties of $\hat{R}^a$, $\hat{\mu}$, and $\hat{\sigma}_a$, and the fact that on $\mathcal{I}$, $\mathbf{R}^a$ is tangential to S, it follows that $j_k = -4\int_S \mu^{-1} R_a \sigma_b \, dS^{ab}$. Hence the result. □

Remark: If P_a is spacelike, Theorem 6 continues to hold although the axis vector A^a would now be timelike rather than spacelike. If P_a is null, on the other hand, some modifications are required. In this case, S_a is *proportional* to P_a; $S_a = jP_a$ where the "helicity" j is related to the Komar integral as follows: if $\hat{R}_a$ is rescaled such that F_{ab} $(= \lim_{\to i^\circ} \nabla_a R_b) = \epsilon_{abcd} P^c A^d$ for any vector A^d satisfying $P \cdot A = 1$, then j equals j_k, the Komar integral.

V. DISCUSSION

The analysis presented in the previous two sections depends quite crucially on the assumption that the physical space–time satisfies Einstein's vacuum equation in a neighborhood of spatial infinity: It is only because $\hat{R}_{ab} = 0$ near i° that we could recast the expressions for energy–momentum and angular momentum involving the asymptotic Weyl curvature into expressions involving the norms and twists of the available Killing fields. Note that each of the definitions considered has a direct but independent physical motivation. Thus, for example, the definition of mass of a stationary space–time in terms of the asymptotic behavior of the norm of the Killing field can be motivated using the fact that the square root of the norm plays the role of the Newtonian potential in the slow motion approximation, while the definition in terms of the electric part of the asymptotic Weyl curvature seems natural in the light of the geodesic deviation,[10] which of course has a purely geometric origin. Thus, each definition emphasises a particular aspect of one's intuitive understanding of the corresponding conserved quantity: each has an apparently distinct origin. The fact that these different notions actually agree in detail is a reflection of the richness of Einstein's equation.

Throughout this paper, we have focused on asymptotic flatness at spatial infinity. There are also available[5,19] definitions of energy–momentum and angular momentum in terms of asymptotic fields at null infinity. How do these compare with those at spatial infinity? Consider, first, space–times which are *stationary* and asymptotically flat at both null and spatial infinity.[9] Then, the situation is the following. The Bondi–Sachs 4-momentum, defined at null infinity, is parallel to the BMS translation induced by the stationary Killing field. (Recall that the Bondi–Sachs 4-momentum is a linear mapping from the space of BMS translations in to reals. In stationary space–times, every BMS translation, "orthogonal" to the timelike BMS translation induced by the Killing field, is mapped to zero by the 4-momentum.) Now, one can show quite generally, i.e., even in nonstationary contexts, that there is a natural isomorphism between the BMS translations and the Spi translations. In stationary space–times, the isomorphism sends the BMS translation induced by the Killing field to the induced Spi translation.[20] It therefore follows, from Lemma 2.1, that the two 4-momenta are colinear. Finally, their equality follows from the fact that the norm of each equals the Komar integral. Next, consider angular momentum. Again, using the isomorphism between BMS and Spi translations, one can show[20] that the spin vector defined at spatial infinity equals the one defined at null infinity. In nonstationary contexts, on the other hand, the situation is much more complicated. First, a notion of angular momentum which is free of "supertranslation ambiguities" is simply not available at null infinity.[21] Hence, it seems difficult to imagine a simple relation between, say, the spin vector S_a at i° and angular momentumlike quantities at null infinity. In the case of energy–momentum, the situation is somewhat better; one expects that, if the radiation "falls off" at a suitable rate as one approaches i° along, say, future null infinity, the past limit of the Bondi–Sachs 4-momentum should equal the ADM 4-momentum. However, one still does not know the precise "falloff" of the radiation field needed for this purpose.

Note added in proof:

1. Note that, unlike in the analyses involving pseudotensors [see, e.g., R. Beig, Phys. Lett. A **69**, 153 (1978)], we do not restrict ourselves to space–times which are topologically R^4. If the topology is nontrivial, a given Killing field can give rise to several distinct Komar scalars. In this event, the results in the paper (Theorems 2 and 6) refer to the Komar integrals evaluated on 2-spheres which, in the

completed space–time, can be continuously shrunk to i^0.

2. The issue of the relation between the ADM and the Bondi–Sachs 4-momenta, mentioned in Sec. V, has been recently resolved (A. Ashtekar and A. Magnon-Ashtekar; Preprint).

ACKNOWLEDGMENTS

We wish to thank Bernd Schmidt and Pong-Soo Jang for discussions. One of us (A. M. A.) thanks the Chicago Relativity Group for hospitality.

[1]A. Komar, Phys. Rev. **113**, 934 (1959).
[2]R.O. Hansen, J. Math. Phys. **15**, 46 (1974).
[3]A. Ashtekar and R.O. Hansen, J. Math. Phys. **19**, 1542 (1978).
[4]The 4-momentum was first introduced by Arnowitt, Deser, and Misner in terms of the asymptotic behavior of the initial data on a Cauchy surface. See, e.g., C.W. Misner, the article in *Gravitation, An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[5]H. Bondi, A.W.K. Metzner, and M.J.G. Van der Berg, Proc. R. Soc. (London), Ser. A **269**, 21 (1962).
[6]These conserved quantities arise only after Einstein's vacuum equation is imposed asymptotically, i.e., only after the stress–energy tensor $\hat{T}_{ab}$ (with this index structure) is required to admit a regular direction dependent limit at i^0.
[7]This situation is to be contrasted with the one in the electromagnetic case: Imposition of Maxwell's sourcefree equations in the asymptotic region does *not* restrict the total magnetic charge in any way.
[8]Note that unlike the definitions of angular momentum at null infinity, the definition at spatial infinity is free of supertranslation ambiguities.
[9]A. Ashtekar, *Asymptotic structure of the gravitational field at spatial infinity* (to appear in the Einstein birth-centenary volume, edited by P. Bergmann, J.N. Goldberg, and A. Held. Plenum).
[10]For details, See Ref. 9.
[11]We do not yet have a proof that every Killing field on a space–time satisfying Definition 1 must induce a Spi symmetry at spatial infinity. Although one certainly expects the result to be true, the proof may well be quite complicated because of the conformal singularity at i^0; one cannot, e.g., use the conformal Killing transport equations in a straightforward way.
[12]For details, see, e.g., A. Ashtekar and A. Magnon–Ashtekar, J. Math. Phys. **19**, 1567 (1978).
[13]See, e.g., A. Lichnerowicz, *Théories Relativistes de la Gravitation et de l'électromagnétisme* (Masson, Paris, 1955), or, R. Geroch, J. Math. Phys. **12**, 918 (1971), Appendix.
[14]Choose an orthonormal basis at i^0 such that $\mathbf{K}^a$ points along the z axis. Consider, e.g., the cross section of $\mathcal{D}$ defined by $\mathbf{t}\cdot\boldsymbol{\eta}=0$ where $\mathbf{t}^a$ is the timelike vector in the basis. On this 2-sphere, $\tilde{\mathbf{K}}^a = -\sin\theta(\partial/\partial\theta)$ and $\mathbf{e} = \mathbf{f}(\phi)\ (\sin\theta)^{-1}$. Since $\mathbf{e}$ is smooth, it follows that $\mathbf{f}(\phi)=0$, whence $\mathbf{e}=0$.
[15]Our convention is such that $\int_S dS = 1$.
[16]See, e.g., R. Geroch in *Asymptotic Structure of Space–time*, edited by P. Esposito and L. Witten (Plenum, New York, 1977), p. 96.
[17]Hansen's dipole moment $\tilde{\mathbf{S}}_a$ is defined by $\tilde{\mathbf{S}}_a = \frac{1}{2}\lim_{\rightarrow\Lambda}\mathrm{grad}.(\tilde{\Omega}^{-1/2}\hat{\lambda}\hat{\omega})$, where Λ is the point at infinity on the three-manifold of orbits of $\hat{t}^a$, $\tilde{\Omega}$ the conformal factor on this three-manifold, $\hat{\omega}$ the "twist potential" ($\hat{\omega}_a = \mathrm{grad}\hat{\omega}$) and grad. stands for "gradient." A simple calculation yields,

$$\tilde{\mathbf{S}}_a\mathbf{V}^a = \tfrac{1}{2}\int_{S^2}\boldsymbol{\lambda}^{-1}\mathbf{V}^a(\omega_a - \tfrac{1}{2}\omega^m\cdot\boldsymbol{\eta}_m)\,dS = \int_{S^2}(\omega^m\cdot\boldsymbol{\eta}_m)(\mathbf{V}\cdot\boldsymbol{\eta})\,dS\,,$$

where S^2 is the 2-sphere of unit vectors η at Λ, ω_a is the limiting value of $\Omega^{-1/2}\omega_a$ and V^a is an arbitrary fixed vector at Λ.
[18]P. D. Sommers, J. Math. Phys. **19**, 549 (1978).
[19]R. K. Sachs, Proc. R. Soc. (London) Ser. A **270**, 103 (1962); Phys. Rev. **128**, 2851 (1962); J. Winicour, J. Math. Phys. **9**, 861 (1968); B. D. Bramson, Proc. R. Soc. (London) Ser. A **341**, 463)1975); M. Streubel, J. Gen. Rel. Grav. **9**, 551 (1978).
[20]A. Ashtekar and M. Streubel, "On angular momentum of stationary gravitating systems" (to appear in J. Math. Phys.)
[21]At spatial infinity, the presence of a rotational Killing field in the physical space–time implies that $\mathbf{B}_{ab}$ must vanish on $\mathcal{D}$ and thus reduces the Spi group to the Poincaré. Is there any hope of a similar reduction at null infinity?

Thermodynamical properties of a class of solvable statistical models for hadronic matter

David E. Miller[a)]

Department of Theoretical Physics, [b)] *University of Bielefeld, Germany*

The exact evaluation for the thermodynamical properties of relativistic quantum gas models for excited hadronic matter using a statistical picture is carried out from the analysis of the grand partition function for a general class of excitation spectra. Some special forms of these energy spectra for hadronic matter are discussed for the physical content and the relationship to the known classical limiting cases.

I. INTRODUCTION

The basic development of the thermodynamical models in high-energy relativistic hadronic systems is briefly discussed in the context of the presented solutions.

A. Relativistic gas

The kinetic theory of the monoatomic ideal gas using the special theory of relativity was first investigated rather early in this century. In the classical work by Jüttner[1] it was found that an ideal gas within a large volume V in the rest system at an absolute temperature T (in atomic units $\hbar = c = k_b = 1$) containing N particles, each of which possess the rest mass m, the momentum $\mathbf{p}$, and the relativistic energy $(\mathbf{p}^2 + m^2)^{1/2}$, yielded, after an integration over phase space, a partition function proportional to the modified Bessel function[2] $K_2(m/T)$. Some years later the relativistic gas was reconsidered for the effects of quantum statistics,[3] whereupon it was found that this known solution was replaced by the series of such functions with the relativistic fugacity A in the form

$$\sum_{k=1}^{\infty} k^{-2} K_2(km/T) A^k.$$

In the years following, considerable further analyses and applications were done for these ideal relativistic gases[4] which provide a rather finished description for these rarified inactive systems.

B. Statistical hadronic models

Another side of relativistic statistical mechanics developed out of the statistical model[5] for high-energy interactions in the center-of-mass system. Especially important in the kinematical structure of the statistical bootstrap model[6] is the solution for the ideal quantum relativistic gases with the specified mass spectrum $\rho(m)$. Motivated by thermodynamical considerations, Touschek[7] made a careful analysis of relativistic phase space for the development of a truly covariant statistical mechanics. In further analyses[8] it was found that the difference between the invariant phase space and the ordinary separately invariant momentum space generally employed for particle physics is significant for the correct thermodynamical functions. The most immediately obvious contradiction in the use of invariant momentum space for the thermodynamics is that it yields for massless particles the wrong temperature dependence for the photon gas.

Since we are considering a variable number of particles in a given spatial volume V, we shall use structurally the previously developed approach[8] for the grand partition function as the beginning point, especially including the usual assumption[7] for the "physical" partition function.[9] This basic quantity may be found from rewriting the sum over the discrete momentum states p_α as an integral over the phase space measure $d\sigma(p,m)$ so that

$$\sum_\alpha \exp\{-\beta p_\alpha\} \to \int \exp\{-\beta p\} d\sigma(p,m). \tag{1.1}$$

This process of summing over these discrete four-momentum states p_α with an inverse temperature four-vector β means $\beta_\mu p^\mu_\alpha$, which is rewritten as $\beta p_\alpha \to \beta p$ in the continuous case. For the time being, we shall assume that this sum may always be replaced by an integral—this excludes such considerations as Bose condensation.

In a more thorough investigation in the high-energy limit the phase space measure $d\sigma(p,m)$ is usually separated into a mass spectrum $\rho(m)dm$ and a part dependent upon the phase space variables—the four-dimensional volume and momentum, which may be written together as $d\Omega(x^\mu, p^\mu)$. After a little analysis using the properties of the rest frame[8] where only the time component of β^μ comes in, we rewrite the integral in (1.1) as

$$\int \rho(m) dm \int d\Omega \exp\{-\beta E(\Omega)\}.$$

The properties of the mass spectrum $\rho(m)$ is well known from the asymptotic properties of the statistical bootstrap model,[6,10] where it has the asymptotic form

$$\rho(m) = c m^a \exp\{bm\}. \tag{1.2}$$

Furthermore, it is known that within this model the prediction of a highest or "ultimate" temperature[6] relating with the inverse of b can be proven[10] for $a = -3$. This result is assumed in the high-energy limit where $E(\Omega)$ asymptotically approaches m. It is possible to make further simplifications on the evaluation of the phase space integrals for the case of noninteracting point particles in a large volume V, whereupon the four-dimensional volume reduces[7] to the usual three-dimensional volume Vd^3p, and for $E(\Omega)$ the energy spectum $H(w)$ to a simple form dependent upon w, the single particle energies $(p^2 + m^2)^{1/2}$. We now describe the single particle energy spectrum $H(w)$ as the collective state of quasiparticles of the excited hadronic matter through a statistical picture in the rest frame of the total system.

[a)] Heinrich-Hertz Fellow 1977–78.

[b)] Now also a the Center for Interdisciplinary Research at the University of Bielefeld.

 0022-2488/79/050801-09$01.00

C. Generalized energy spectrum

The generalized energy spectra provide the structure for the total energy of the excitations of hadronic matter. The basic assumption for this quantum many-particle model is that the density matrix elements may be diagonalized in such a way that the form of the energy spectrum can be written as a expansion in all the integer powers of the single-particle energies. This situation allows us, for the general complex energy variable z within some region of the complex plane containing the positive real axis, to make a Laurent expansion for the energy spectrum in the explicit form

$$H(z) = \sum_{n=-\infty}^{\infty} a_n z^n. \tag{1.3}$$

With this assumed energy spectrum we are able to inspect the structure of the complex integral using a given measure $d\Omega(z)$,

$$\int d\Omega(z) \exp\{-\beta H(z)\},$$

for its analytical properties. This study must be carried through under the necessary physical conditions which are imposed by the related real valued integral. This case, which corresponds to the expansion in Hermitian operators, will be our main interest here.

First we would like to see what may be simply understood from the structure of (1.3) out of the evaluation of the grand partition function which is formally defined for identical noninteracting particles i as

$$\Xi(V,\beta) = \prod_i (1 + \gamma \exp\{-\beta H(z_i)\})^{\gamma}, \tag{1.4}$$

where γ is a factor which distinguishes between the various statistics ($\gamma = +1$, Fermi; $\gamma = -1$, Bose). At this point it is necessary to restrict the problem to include known special solutions.

The special cases of this general formal structure,[11] which have been previously investigated, give some specific predictions for a range of real values of particular coefficients a_n. For the sake of completeness we shall briefly state some of the necessary properties for real values of the coefficients. It is to be remarked that we find the usual grand partition function[12] for the ideal relativistic gas[3] when all of the a_n vanish except for $a_1 = 1$ and $a_0 = -(\mu + m)$, where μ is the real chemical potential as given by the Gibbs free energy per particle. It must be further stated here that a necessary requirement for the replacement of the summation over states by an integral as in (1.1) with the form of the mass spectrum in (1.2) requires that the energy spectrum for real values at high energies asymptotically approach the mass m. For this reason we redefine[11] our series and variables, especially for real values in the form $\{a_n\} \to \{c_n\}$ for $z \to my$, so that

$$H(y) = m \sum_{n=-\infty}^{\infty} c_n y^n \tag{1.5}$$

has the proper relation to the decomposition of the phase space and mass integral. For the time being we shall consider only the explicit contribution of the energy spectrum to the phase space integrals. Under the assumption that $c_0, c_1 \neq 0$ it may readily be seen that for the real values of y the convergence of the integral for $\Xi(V,\beta)$ in the thermodynamic limit may be assured by $c_n \geqslant 0$ for $n \geqslant 2$, with no restriction on c_{-n} for $n \geqslant 1$. Using the properties of the integral representation of (1.4) we find that in the thermodynamic limit

$$\lim_{V\to\infty} \frac{1}{V} \ln \Xi(V,\beta) = \frac{gm^3}{2\pi^2} \sum_{k=1}^{\infty} \frac{(-\gamma)^{k-1}}{k} I_k(\beta), \tag{1.6}$$

where g is a phase space factor for the internal quantum numbers and the phase space integral becomes

$$I_k(\beta) = \int_1^{\infty} dy y(y^2-1)^{1/2} \exp\left\{-k\beta m \sum_{n=-\infty}^{\infty} c_n y^n\right\}. \tag{1.7}$$

We see in (1.6) and (1.7) a direct analogy with a cluster expansion[13] in nonrelativistic statistical mechanics where $k = 1$ yields the simple ideal relativistic gas or single-particle term, and the following term $k = N$ contains the clusters with N particles with the forefactor $(-\gamma)^{k-1}$ which depends on the type of statistics.

D. Procedure

Finally, we make a few remarks about how we shall develop and solve this problem as formulated in Eq. (1.6) and (1.7) for hadronic matter at high energies. The fact that we are primarily interested in real values of the expansion (1.3) is similar to working with real values of the spectra of the observables in quantum mechanics once one knows that the operators have been diagonalized. This similarity means that the really strong assumption is made when it is allowed a diagonal form in terms of an expansion of single-particle energies. We shall point out later the significance of some of the powers in this expansion. In the following section we shall carry out the evaluation of $\ln\Xi(\beta)$. Directly thereafter it will be shown how some aspects of these results may be understood. The conclusions report other related work in theoretical physics.

II. EVALUATION FOR THE GENERALIZED ENERGY SPECTRUM

The properties of the generalized energy spectrum stated in Sec. C of the Introduction are further elaborated upon in order to arrive at the general solution, which is complemented by some remarks and special cases.

A. Thermodynamics

The general form of the thermodynamic potential expressed in Eq. (1.6) is similar to the cluster expansion for the classical interacting gas, for which $(-\gamma)^{k-1} I_k(\beta)/k$ yields the structure of the cluster integral $b_k(\beta) z^k$ in the expansion with z, the classical fugacity. As it was mentioned above, the ideal relativistic quantum gases involve the replacement of $I_k(\beta)$ by the form[3] $K_2(km/T)/k$. We now want to propose our extension to the expansion (1.6) under the conditions on (1.5) for the convergence of the integral $I_k(\beta)$ in (1.7). At this point it is proper to propose the restrictions on the c_n consistent with the a_n in (1.3). This proposal means that $c_1 = 1$ and

$c_0 = -(\mu + m)/m$ so that the term $c_1 y$ always dominates over all the $c_{-n} y^{-n}$ terms. For the positive integers $n > 1$ we must in general demand that $c_n \geqslant 0$, which always brings as a result faster convergence for $I_k(\beta)$ from the dominant anharmonic terms. It may be possible to relax this inequality for some $c_n < 0$ under the restriction that there must exist some $c_{n'} > 0$ for which $n' > n$. The general convergence properties of the complete "cluster" series (1.6) is a much more difficult problem, and will only be considered in certain restricted cases.

B. Evaluation of the grand potential

We shall now state in a precise form the general results of the explicit calculation of the integral $I_k(\beta)$ given in Eq. (1.7) for the evaluation of the thermodynamical functions of (1.6). These somewhat more general results will be further qualified by a series of remarks and special cases in the following sections.

Main statement: With the general energy spectrum given in (1.5), the evaluation of the thermodynamical functions from (1.6) through the calculation of (1.7) under the above proposed assumptions on $\{c_i\}$ may be stated in its most general form as

$$I_k(\beta) = (km\beta)^{-1} \sum_{n=0}^{\infty} \frac{(km\beta)^n}{n!} \sum_{j=-\infty}^{\infty} b_{nj}[\{c_i\}] \, G(nj, km\beta)(km\beta)^{nj}, \tag{2.1a}$$

$$G(nj, km\beta) = \Gamma(nj + 3, km\beta) - \sum_{l=1}^{\infty} !l\,!(km\beta)^{2l}\Gamma(nj - 2l + 3, km\beta). \tag{2.1b}$$

The coefficients $b_{nj}[\{c_i\}]$ depend directly upon the set $\{c_i\}$ of coefficients in (1.5) and the properly formed products thereof. The incomplete gamma function $\Gamma(\alpha, x)$ as well as the other notation with the double factorials are defined in Appendix A. The details for the explicit evaluation of this main statement are accordingly presented in Appendix B. With this result we shall proceed further by making some remarks aimed at deepening our appreciation of these findings. It is, however, necessary for us to be particularly strongly warned about the dangers of the limit $m\beta \to 0$, which in the form of (2.1) is not obvious how to carry out, and must be considered separately for various special cases of the given spectra in (1.5).

C. Remarks on Bessel functions

This main result as expressed in Eq. (2.1) may be enlightened by a few remarks on some interrelationships between these series of the incomplete gamma functions $\Gamma(\alpha, x)$ with the exponential integrals $E_n(x)$ and the modified Bessel functions $K_\nu(x)$ where α and ν are real numbers, n a positive integer and x a positive real variable. It is generally possible to consider many of these results through analytic continuation under certain restrictions into the complex plane. Although this process is consistent with our original expansion of $H(z)$ in (1.3), we shall, for the time being, leave such considerations.

Remark 1: Special representations for $K_2(x)$ and $K_1(x)$ in terms of the incomplete gamma functions are given as the following:

$$K_2(x) = x^{-2}\left[\Gamma(3,x) - \sum_{l=1}^{\infty} !l\,!x^{2l}\Gamma(3 - 2l, x)\right], \tag{2.2a}$$

$$K_1(x) = x^{-1}\left[\Gamma(2,x) - \sum_{l=1}^{\infty} !l\,!x^{2l}\Gamma(2 - 2l, x)\right]. \tag{2.2b}$$

These relationships will be important in identifying the terms in later work.

Remark 2: $K_0(x)$ is best represented through $E_n(x)$ as

$$K_0(x) = \sum_{l=0}^{\infty} !l\,!E_{2l+1}(x)x^{2l}. \tag{2.3}$$

A general formula may be derived from the Schläfli integral representation[2] in the given form.

Remark 3: For $\nu > -\frac{1}{2}$ we have

$$K_\nu(x) = \frac{\sqrt{\pi}}{x^\nu 2^\nu \Gamma(\nu + 1/2)}\left[\Gamma(2\nu, x) - \sum_{l=1}^{\infty} b_l(\nu)x^{2l}\Gamma(2\nu - 2l, x)\right], \tag{2.4}$$

where the coefficient $b_l(\nu)$ must be determined. Equations (2.2) and (2.3) may be found as special cases of (2.4) through the use of the correct recurrence formulas.

D. Interpretation of terms

In this section we want to physically interpret the structure of the terms $a_n z^n$ appearing in Eq. (1.3). The interpretation for the terms a_0 and $a_1 z$ in (1.3) is clear from our discussion in Sec. I.C, where they represent the terms naturally appearing in the relativistic free gas and include no information about the analytic structure of the hadronic collective excitation spectrum. We recall that a_0 can be directly related with the chemical potential μ, which disappears when there are no conserved particle numbers in the case of the photon gas of massless particles $a_0 = 0$. Because of the appearance of the inverse temperature β in the grand partition function (1.4), the logical choice for a_1 is 1. Our next consideration comes for the terms of the form $a_{-n}(p^2 + m^2)^{-n/2}$ in (1.3). Such mometum space "potentials" relate[14] directly with the Fourier transforms[15] of the Lennard–Jones type of potentials. It is known, for example, that under certain limiting conditions these Fourier transforms in coordinate space are responsible for the long-range behavior necessary for the phase transitions of the van der Waals type. We shall consider this point further in later work. The other terms with the positive powers $a_n(p^2 + m^2)^{n/2}$ for $n > 1$ can be regarded as a sort of "anharmonic" corrections in the excitation spectrum. This interpretation is particularly meaningful for the case of small m where the $a_1 z$ term relates directly with harmonic

oscillator so that the higher powers in $(p^2 + m^2)^{n/2}$ contribute directly to the anharmonicity in momentum space.

III. ANALYSES OF THE SOLUTIONS

The general solution which was given in the main statement (2.1) of the preceeding part will be further analyzed here for some special cases of physical interest.

A. Special forms

In this first section we shall list some special cases of the coefficients c_n in the expansion of the excitation spectra (1.5) under the general assumption that c_0 and c_1 are not both identically zero. It is generally possible for us to find the fitting expressions by using the proper relationships between $\Gamma(\alpha,x)$, $E_n(x)$ and $K_\nu(x)$ as remarked above. Furthermore, for simplicity we shall use $A = \exp\{m\beta c_0\}$ $= \exp\{-\beta(\mu + m)\}$, $c_1 = 1$ and $x = \beta m$.

1. The case for the ideal relativistic quantum gas[1] is readily found by including only c_0 and c_1 as

$$I_k(x) = A^k K_2 \frac{(kx)}{kx}. \tag{3.1}$$

The relationship to the main statement (2.1) is obvious from the remark (2.2a). This form (3.1) may then be replaced in (1.6) in order to find the thermodynamical functions after making the appropriate choice of γ for the statistics.

2. For the case where we have the only additional coefficient $c_{-1} = c'$, which is nonzero, we may explicitly write out the integral $I_k(x)$ in (1.7) in the form

$$I_k(x) = \frac{A^k}{kx}\Big[K_2(kx) - c'kxK_1(kx) + \sum_{n=0}^{\infty} \frac{(-c'kx)^{n+2}}{(n+2)!} kx\Big\{E_n(kx) \times \Big\{E_n(kx) - \sum_{l=1}^{\infty} l!\, l! E_{2l+n}(kx)\Big\}\Big\}\Big]. \tag{3.2}$$

We can see directly here how the additional coupling c' affects the cluster structure as it was elaborated upon in the preceeding part. This particular structure in (3.2) relates to a special limiting case, which we shall investigate afterwards.

3. The Yukawa type of expression[11] is given by a nonzero term of the form $c_{-2} = c''$ as the only additional coefficient ($c_{-1} = 0$). For this energy spectrum the evaluation of the integral yields

$$I_k(x) = \frac{A^k}{kx}\Big[K_2(kx) + \sum_{n=0}^{\infty} \frac{(-c''kx)^{n+1}}{(n+1)!} \times kx\Big\{E_{2n}(kx) - \sum_{l=1}^{\infty} l!\, l! E_{2l+2n}(kx)\Big\}\Big] \tag{3.3}$$

We notice a general similarity in form between (3.2) and (3.3). However, in (3.2) we have the additional $K_1(kx)$ term as well as the odd $E_n(kx)$. These terms turn out to be quite important to the physical structure.

4. The so-called "anharmonic" terms with $c_n \neq 0$ for some positive n can be in general written as

$$I_k(x) = \frac{A^k}{kx} \sum_{n=0}^{\infty} \frac{(-kx)^n}{n!} \sum_{j=1}^{\infty} b'_{nj}[\{c_n\}] K_{1+j}(kx), \tag{3.4}$$

where $b'_{nj}[\{c_n\}]$ depend on products and powers of the various coefficients of $\{c_i\}$. The special case of (3.4) for $i = 2$ may be readily worked out.[16]

From these special examples of the generalized energy spectra we are now in a position to see that any particular solutions relating to (2.1a,b) can be found for the c_{-n} terms as generalizations of (3.2) and (3.3), while for the c_n terms these expressions can be directly derived out of (3.4). With this knowledge we shall come to a discussion of some limiting cases.

B. Limiting cases

The study of some particular limiting cases brings out some special aspects of these models. For these limiting cases it is somewhat more convenient to work with the original expansion (1.3) since the limit $x \to 0$ can then be treated in two ways for $\beta \to 0$ (the high-temperature behavior) and $m \to 0$ (system of massless particles). The interpretation of these two limits is physically different. For a finitely massive particle system the high-temperature limit results in the classical ideal gas. However, for massless particles the thermodynamics becomes that of blackbody radiation obeying a Stefan–Boltzmann T^4 law. If, however, we would take the form (1.5), the two limits would become equivalent.

1. We now examine (1.3) for the limit $m \to 0$. It is clear that the single particle energy $w(p)$ as $(p^2 + m^2)^{1/2}$ becomes $w(p) = |\mathbf{p}|$. For the usual photon gas we have in (1.3) only $a_1 = 1$ and all the other terms identically zero including a_0 because of the masslessness. This case for $I_k(\beta)$ may be immediately solved to find the Stefan–Boltzmann form. The additional "anharmonic" terms a_n for positive n can be put in and solved under the above stated conditions, but they seem to have little physical relevance since one does not expect the photons to couple to themselves. However, a special case, which is finite for $m \to 0$, involves only $a_{-1} = u^2$ in addition to a_1. Under this assumption we may solve for the k cluster integral as $m \to 0$ to find

$$I_k(\beta) = u^k K_3(2ku\beta). \tag{3.5}$$

This form of the cluster integral shows that the coupling u has taken over the role of the mass for the massless particles. This result means that the particles which couple through a_{-1} are qualitatively different from photons, in that they possess a real clustering property. The further terms in the series for $H(y)$ of the form $a_{-n} z^{-n}$ possess the known "infrared" singularities in the small mass limit. This situation is clear from the consideration of the term $a_{-2}z^{-2}$, which for $m \to 0$ yields the usual Coulomb structure and its expected infrared divergence. The further terms, $a_{-3}z^{-3}$, $a_{-4}z^{-4}$, etc., diverge even more strongly in this limit.

2. The other limit $m \to \infty$ is very important for heavy resonances. For this case the factorized form of the excitation spectrum (1.5) is more convenient. When we assume that every order in the negative powers in (1.5) is coupled in the same way so that all the terms c_{-n} have the simple power form $(c/m)(c'/m)^n$ where c and c' are two given constants, we have the additional term in the energy spectrum of the type $(c/m)\ln[1-(c'/my)]$. Thus the k-cluster integral in (1.7) then becomes

$$I_k(\beta) = A^k \int_1^\infty dy y(y^2-1)^{1/2}\left(1-\frac{c'}{my}\right)^{-k\beta c} e^{-km\beta y}. \quad (3.6)$$

The evaluation of this integral follows the form of (2.1) which may be explicitly evaluated as

$$I_k(\beta) = \frac{A^k}{(k\beta)^3}\sum_{n=0}^{\infty} b_n(kc\beta)\left(\frac{kc'\beta}{m}\right)^n [\Gamma(3-n,kx) - \sum_{l=1}^{\infty} !l\, !x^{2l}\Gamma(3-n-2l,kx)\Big], \quad (3.7)$$

where $b_n(kc\beta)$ is in this special case of the form

$$b_n(kc\beta) = kc\beta(kc\beta+1)\cdots\frac{(kc\beta+n-1)}{n!}, \quad (3.8)$$

which comes directly from the above power form of c_{-n}.

3. We now consider another case for this large mass limit where only the positive powers c_n enter. When we assume the particular structure of the additional terms to c_0 and c_1 to be of the form c_n as $(c''m)^n/n!$, then the total expression is of the form given by

$$H(y) = c_0 m + c_1 my + c(\exp\{c''my\} - c''my - 1). \quad (3.9)$$

Again we are able to carry out the expanded integral (1.7) as

$$I_k(\beta) = A^k \exp\{k\beta c\}\sum_{n=0}^{\infty}\frac{(-kc)^n}{n!X_{k,n}}K_2(X_{k,n}), \quad (3.10)$$

where $X_{k,n}$ means $m[k\beta(1-cc'') - nc'']$. This solution is regular except in the neighborhood of zero. Because of the properties of the modified Bessel function[2,17] around the zeros, we find that at any such point the thermodynamics reduces to

$$\lim_{V\to\infty}\frac{1}{V}\ln\Xi(V,T) \sim \left(\frac{T_c - T}{T_c}\right)^{-3}, \quad (3.11)$$

where T_c takes the form $k(1-cc'')/mc''$ from the zero points of $K_2(X_{k,n})$. It is interesting to note that T_c relates with the ratio of k to n for nonzero n, which means that the "critical" temperature is determined by the additional binding or coherence n due to the additional structure in (3.9) in comparison with the cluster size k. In simple terms, this means that T_c is the relationship of the cluster size coming from the symmetry as compared to the structure from the coherence.

C. Physical interpretations

Although it would be possible at this point to do a more thorough analytical investigation of some additional special cases including the limiting forms, it is perhaps more informative to briefly look more deeply into the explicit mathematical structure of the above considered terms as typical of the others. Surprisingly enough, a numerical investigation of these integrals $I_k(x)$ turns out in most cases to be quite a lot simpler than one would think. This situation arises from the fact that the $K_n(x)$ and $E_n(x)$ both have rather workable asymptotic forms,[18] which for small values of the couplings c_{-n} usually lead to rapid convergence for the higher powers. This analysis for the Yukawa term[19] of Sec. IIIA.3 above shows that the main convergence difficulties in (3.3) for very small c'' come for x around unity, which means that the thermal energy is of the same order as the mass energy, so that neither the $x \to 0$ photon gas nor $x \to \infty$ classical ideal gas limits come into play. However, as remarked above, one must be careful in this case with $x \to 0$ because of the infrared problems. For the sake of more explicit calculations of physical interest, we mention a particular model. In order to preserve the properties necessary for the thermodynamic limit of a statistical system,[14] we propose a combination of the terms arising from c_{-1} and c_{-2} so that $c' > 0$ and $c'' < 0$. This choice can be seen to give the proper repulsive hard-core and attractive long-range behavior using the Fourier transforms of these expansion terms as derived in Appendix B. A more detailed treatment of this model with a careful discussion of the mathematical and physical properties will be carried out elsewhere.

D. Mathematical extensions

The last point of consideration in this part is the extension of these analyses to complex variables. Although the analytical structure of (1.3) is designed for an expansion of the function $H(z)$ of the complex variable z, the evaluation of the integral of the type $I_k(x)$ in (1.7) with the necessary convergence conditions upon it would be greatly complicated by the presence of the singularities and the multivalued structure found for such functions in the complex plane. It may, however, be appropriately remarked here that the term a_0 could readily be taken as complex, which would effectively make A complex as in the Yang–Lee theory of phase transitions.[18] Other terms may be included in principle as complex variables, since many of these integrals can be under the proper restrictions extended to the complex plane. However, the intention of our main result in (2.1) would then be drastically changed.

IV. GENERALIZATIONS AND CONCLUSIONS

In this concluding part of our work we want to assemble the various calculations to make some specific predictions for the thermodynamical functions. We are able to fully carry out this program only after bringing together the phase space parts, which were evaluated above, with the integration over the mass spectrum in terms of a total grand partition function $Z(V,\beta,A)$. After writing $Z(V,\beta,A)$ we shall point out its significance to some problems in hadronic matter as well as possible further work in this direction.

A. Mass spectra

Up until now we have concentrated upon the analytic properties of the phase space integral $I_k(\beta)$, treating the mass of the particles as a constant. This approach, however, is very unrealistic for high-energy physics where one very seldom has a single type of interacting stable particle, but is usually confronted with a large number of unstable particles decaying, forming resonances, and interacting to form bound states. This situation may be generally described by a mass spectrum $\rho(m)$. We shall now briefly look into the thermodynamical functions resulting from these calculations.

The origin of the mass spectrum arises primarily in the applications to relativistic statistical physics in the area of multiple-particle production for hadronic systems. For such systems one often employs the specific mass spectra $\rho_{+1}(m)$ and $\rho_{-1}(m)$ for the Fermi–Dirac and the Bose–Einstein statistics respectively. The general thermodynamical treatment of this problem was first developed by Hagedorn[6,8] for the analysis of the decaying fireballs consisting of many types of particles and resonances. We may regard the simplified system considered in detail above as possessing the simple delta function mass spectrum so that there exists only a single stable particle of mass m_0, the lightest hadron. With this idea of extending our system to include a mass spectrum we may rewrite (1.6) for the total partition function as a functional of the mass spectrum $Z[\rho(m);(V,\beta,A)]$ in the expansion depending upon the statistics from γ as

$$V^{-1}\ln Z_\gamma = \left(\frac{g}{2\pi^2}\right)\sum_k^{\infty}\sum_j \frac{A^k}{k}(-\gamma)^k \int_{m_0}^{\infty} dm\rho_\gamma(m')I_k(m',\beta), \quad (4.1)$$

where A is the relativistic correspondent of the activity or fugacity for the hadronic matter. The usual statistical bootstrap model[6,10] demands that the individual mass spectra all have the form (1.2).

The thermodynamical functions may now be calculated whenever the excitation spectrum can be specified within the above stated conditions. The pressure is immediately calculated from (4.1). For the massive particles the average particle density may be calculated in the usual way[12] through a differentiation with respect to A as $A(\partial/\partial A)\ln Z_\gamma(V,\beta,A)$, which yields a second expansion in the powers of A.

There has been a considerable amount of work using an S matrix approach for the interacting relativistic gas. It is found that for this system one can develop an S matrix formulation of the cluster expansion in statistical mechanics.[20] This approach has been further developed to study the possibility of phase transitions in large hadronic systems.

B. S-matrix approach

It is important throughout the present work not to mistake the energy spectrum $H(w(p))$ of the excitations in hadronic matter with the formal Hamiltonian operator, which itself contains all the various energy contributions including those absorbed in the mass spectrum as well as those regarded by us as interactions with the hadronic medium. Thus $H(w)$ can be interpreted as the energy "left over" from $\rho(m)$ that does not go into pure particle production, but remains as an interaction with the hadronic medium. Such an energy spectrum only has meaning as quasiparticles or excitations[13] of the interacting hadronic system, not as singly produced hadrons. However, this structure reduces to the real particle formalism in the case where the particles decouple from the medium.

During the course of this work we have made some rather strong assumptions on the excitation spectrum $H(w(p))$ in the form $mH(w(p))$, which assured the necessary structure for the statistical bootstrap model.[10] A change in this assumption or omitting it altogether could drastically alter the known critical phenomena associated with the model of hadronic matter.

C. Critical exponents

At this point we want to state briefly a few of the critical properties of some models for hadronic matter. From the calculated analytical structure of the partition function in the statistical bootstrap model[6,10] a singular structure is found relating to the parameters a and b in (1.2). When the related gas model with a mass spectrum $\rho(m)$ is considered, a set of critical exponents[21] may be associated with this singular structure in the thermodynamic limit. These ideal have been further developed in work on the quantum relativistic gas models[22] including $\rho(m)$, for which an equation of state[23] has been recently found. Furthermore, a general type of cluster model has been introduced for statistical mechanics at high energy density.[24] It has also been recently found that the presence of a coupling between the fireballs in a slightly modified statistical bootstrap model can give rise directly to a phase transition which is characteristic of this model of hadronic matter.[18]

D. Concluding remarks

Finally we want to make a few concluding remarks concerning the physical nature of our general solution given in (2.1a,b) along with the special cases mentioned in Sec. III.

1. Our above proposed model, which is characterized by an energy or excitation spectrum $H(w(p))$, has the general effect of introducing a binding together of the clusters formed by the statistics in the ideal relativistic quantum gas. This structure appears quite openly in the special cases discussed in Sec. III.A, where, in addition to the known ideal cluster terms $K_2(kx)/kx$ of the relativistic gas, we also have a series in the powers of the relevant couplings. This effect, which is particularly apparent in the examples 2 and 3, makes a direct contribution to the collective structure of the hadronic matter. In order to understand this contribution more clearly, we shall elaborate more specifically upon the structure of the model suggested at the end of Sec. III.C.

2. The general properties which are derived from a model with $c' > 0$ and $c'' < 0$ bear a resemblance to the known interaction potentials of atomic physics after performing a Fourier transformation into coordinate space. From the specific evaluations of the Fourier transforms at the end of Appendix B we may describe the corresponding interaction po-

tential $\phi(r)$ for this model in the form

$$\phi(r) = \frac{4\pi mc'}{r} K_1(mr) + \frac{2\pi^2 c''}{r} e^{-mr}. \tag{4.2}$$

We can immediately see that in the limit $m \to 0$ the repulsive part with $c' > 0$ goes over to the form $4\pi c'/r^2$, which is similar to the atomic potential arising from the centrifugal barrier, and the attractive part with $c'' < 0$ becomes $2\pi^2 c''/r$, the simple electrostatic Coulomb potential. The situation now has a basic similarity with the well-known binding structure of the diatomic molecules, for which it is possible to make a detailed analysis for the thermodynamical properties of a gas containing separate translational, rotational, and oscillatory degrees of freedom.[25] In this sense it may well be possible to envision our relativistic gas expansion as forming a binding or a coherence between the "elementary" excitations in the hadronic matter which in a special sense have a likeness with a gas of harmonic oscillators in equilibrium.[26]

3. The primary result of this work is the development of the mean field type of picture for the description of the bulk properties of hadronic matter. When we consider specifically the model suggested above, we see that the basic structure necessary for the mean field description is present in the potential (4.2), in the small mass limit with its repulsive core and its long-range attractive part. These features seem to be a direct reflection through the Fourier transform of the original expansion of the energy spectrum (1.3) in the one-particle energy $w(p)$, which arises primarily out of the fundamental relativistic invariance of $p_\mu p^\mu = m^2$.

4. Another model which seems to have a rather direct relationship to our energy spectrum arises in the description of phase transition to the spontaneous coherent state[27] as found from the Dicke laser model.[28] In fact, it is found that the form of the radiation field energy $H(\omega)$ can be written as $\hbar\omega(1 - \omega_p^2/\omega^2)$ with the extra inclusion of an additional $(\omega_p/\omega)^4$ term from the retardation effects.[29] This first expression is equivalent to our energy spectrum with an a_{-1} term in the limit $m \to 0$ as discussed in Sec. III.B. Further work in this direction has shown how to use a pseudospin model based on the Dicke model in order to describe the phase transitions in nuclear matter.[30] This model already envisions some of our future goals.

As a final remark in this work we may state that we have considered how to evaluate the equilibrium thermodynamical properties of some relativistic quantum statistical models for bulk hadronic matter.

ACKNOWLEDGMENTS

This work was started while the author was a participant in the project "Dynamical Models and Statistical Methods", in the Center for Interdisciplinary Research at the University of Bielefeld. The author would like to thank all the other participants and especially the project leader Prof. H. Satz, as well as his other colleagues in the Department of Theoretical Physics, for many discussions. He would also like to express his gratitude to the offices of the Deutsche Forschungsgemeinschaft and the Heinrich–Hertz–Stiftung for financial assistance.

APPENDIX A: DEFINITIONS OF SPECIAL FUNCTIONS

We shall state here the definitions for some quantities, which have been used throughout this work.

1. The double factorial is defined as

$$n!! = n(n-2)(n-4)\cdots, \tag{A1}$$

which for even numbers becomes

$$(2n)!! = (2n)(2n-1)\cdots(4)(2), \tag{A2}$$

and for odd numbers becomes

$$(2n-1)!! = (2n-1)(2n-3)\cdots(1), \tag{A3}$$

under the requirement that 0!! and $(-1)!!$ be unity. In our particular notation we introduce $!l!$

$$!l! = \frac{(2l-3)!!}{(2l)!!}. \tag{A4}$$

2. The incomplete gamma function [31] $\Gamma(\alpha,x)$ has the integral representation given by

$$\Gamma(\alpha,x) = \int_x^\infty dt\, t^{\alpha-1} e^{-t}. \tag{A5}$$

It relates to the ordinary complete gamma function $\Gamma(\alpha)$ for $\alpha \neq 0, -1, -2, \cdots$ by[32]

$$\Gamma(\alpha,x) = \Gamma(\alpha) - \sum_{n=0}^{\infty} \frac{(-1)^{k+1} x^k}{(\alpha+k)(k)!}, \tag{A6}$$

and for $\alpha = 0, -1, -2, \cdots, -n \cdots$ in the limit $x \to 0$ as

$$\Gamma(\alpha,x) = \frac{(-1)^{n+1}}{n!}\left[\sum_{k=0}^{\infty}(-1)^k k!\, x^{n-k-1} + x^n \ln x + 0(x^n)\right]. \tag{A7}$$

3. The exponential integral[16] $E_n(x)$ is defined by

$$E_n(x) = \int_1^\infty \frac{dt}{t^n} e^{-tx}. \tag{A8}$$

4. The modified Bessel functions[2] of the second kind $K_\nu(x)$ have the integral representation[31]

$$K_\nu(x) = \frac{(x/2)^\nu \Gamma(\frac{1}{2})}{\Gamma(\nu+\frac{1}{2})} \int_1^\infty dt e^{-xt}(t^2-1)^{\nu-1/2}, \tag{A9}$$

for $\nu > -\frac{1}{2}$.

5. The asymptotic forms for the functions used above have the following structure[16,31]:

(a) for $x \to \infty$ the exponential integral becomes

$$E_n(x) \sim \frac{e^{-x}}{x}\left[1 - \frac{n}{x} + \frac{n(n+1)}{x^2} - \cdots\right]; \tag{A10}$$

(b) for $x \to \infty$ the incomplete gamma function is given by

$$\Gamma(a,x) = \frac{e^{-x}}{x^{1-a}}\left[1 + \frac{a-1}{x} + \frac{(a-1)(a-2)}{x^2} + \cdots\right]; \tag{A11}$$

(c) the modified Bessel functions in the limit $x \to \infty$ are

$$K_n(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x}\left[1 + \frac{4n^2-1}{8x}\right.$$

$$+\frac{(4n^2-1)(4n^2-9)}{2!(8x)^2}+\cdots\Bigg]; \tag{A12}$$

(d) while in the $x\to 0$, the form is

$$K_n(x)\sim\frac{2^{n-1}(n-1)!}{x^n}\Bigg[1-\frac{x^2}{4(n-1)}+\cdots\Bigg]. \tag{A13}$$

6. An important relationship between the exponential integral and the incomplete gamma function is

$$E_n(x)=\Gamma(1-n,x)x^{n-1}. \tag{A14}$$

APPENDIX B: EXPLICIT EVALUATIONS

1. We shall now furnish the details of the evaluation of $I_k(\beta)$ in (1.7) in the form (2.1). It is to be first noted that the deviations from $\Gamma(\alpha,x)$ integral of (A5) come from the presence of $(y^2-1)^{1/2}$ and the nonlinear terms in the exponential. Because of the range of integration $(1,\infty)$ it is much more efficient to expand the form $y(1-1/y^2)^{1/2}$ in the powers $(1/y^2)^n$. This expansion we perform

$$(1-y^{-2})^{1/2}=1-\sum_{l=1}^{\infty}l!/l!y^{2l}, \tag{B1}$$

from which the form of (2.1b) is clear. The direct expansion of the exponential excluding the linear term forms the summations in (2.1a). In order to insure that these expansions are integrable we provided distinct criteria on c_n in the previous section (Sec. II.A). The coefficients $b_{j,n}$ are combinations of the coefficients c_n in the expansion (1.5) and do not depend upon the quantity $km\beta$. We separate these terms out in such a way as to find the integrals for $\Gamma(\alpha,x)$ of (2.1b).

2. The ideal relativistic gas may be found as special case which is known[1-3] to find the $K_2(kx)$. This result may be related to (2.1b) by making the above expansion (B1), which is the content of the first remark (2.2a). When the same evaluation is done for the phase space[8] we find (2.2b). The further special cases worked out in Sec. III.A may be found in a similar way with the replacement of $E_n(x)$ for $\Gamma(1-n,x)$ as stated in Appendix A. 6.

3. Finally we consider the explicit evaluations of the expressions discussed in Sec. II.D which may be directly found from the tables of Fourier transforms. From these evaluations we may directly calculate the coordinate representations of the interaction potentials $\phi(r)$ as a function of the distance r from a given point. The general form of these Fourier integrals[15,33] for $\phi(r)$ are given from real (rational)$^\nu$ as

$$\phi(r)=\left(\frac{4\pi}{r}\right)\int_0^\infty dpp[(p^2+m^2)^{1/2}]^\nu\sin pr. \tag{B2}$$

For the particular terms of interest we find that the half-integer negative powers are given by the integral[33]

$$\int_0^\infty\frac{p\sin\{pr\}}{(p^2+m^2)^{n+1/2}}=\frac{-\sqrt{\pi}}{2^n m^n\Gamma(n+1/2)}\frac{d}{dr}[r^nK_n(mr)], \tag{B3}$$

and the whole integer powers have the form[33]

$$\int_0^\infty dp\frac{p\sin\{pr\}}{(p^2+m^2)^{n+1}}=\frac{(-1)^n\pi}{n!2^{n+1}m^n}\frac{d^n}{dm^n}(e^{-mr}). \tag{B4}$$

A more general expression[33] for all $\nu>-1$, $\nu\neq 3/2, 5/2\ldots$ is

$$\int_0^\infty dpp(p^2+m^2)^{\nu-3/2}\sin\{pr\}=\frac{\sqrt{\pi}\,(2m)^\nu r^{1-\nu}}{\Gamma(\frac{3}{2}-\nu)}K_\nu(mr). \tag{B5}$$

For our particular calculations the corresponding coordinate space potentials $\phi_n(r)$ for $a_n(p^2+m^2)^{-n/2}$ with $n=1,2$, and 3 respectively are readily found to be given by

$$\phi_1(r)=\frac{4\pi m}{r}K_1(mr) \tag{B6a}$$

$$\phi_2(r)=\frac{2\pi^2}{r}e^{-mr} \tag{B6b}$$

$$\phi_3(r)=8\pi K_0(mr) \tag{B6c}$$

From these terms we can see that other similar forms may be likewise calculated from (B3)–(B5) with the exception of the positive integer powers, which fail to fulfill the necessary criteria for Fourier integrals.[15,34]

[1] F. Jüttner, Ann. Phys. (Leipzig) **34**, 856 (1911); **35**; 145 (1912).
[2] G.M. Watson, *Theory of Bessel Functions* (Cambridge, London, 1966). See particularly Chapters 3 and 6 for the properties of $K_n(x)$.
[3] F. Jüttner, Z. Phys. **47**, 542 (1928); W. Glaser, Z. Phys. **94**, 317, 677 (1935).
[4] J.L. Synge, *The Relativistic Gas* (North Holland, Amsterdam, 1957).
[5] E. Fermi, Prog. Theor. Phys. **5**, 570 (1950); I. Ya. Pomeranchuk, Doklady. Akad. Nauk SSSR, **78**, 889 (1951).
[6] R. Hagedorn, Nuovo Cimento Suppl. **3**, 147 (1965); "Thermodynamics of Strong Interactions," CERN Report 71-12 (CERN, Geneva, 1971).
[7] B. Touschek, Nuovo Cimento B, **58**, 295 (1968); see also in B. Touschek and G. Rossi, *Meccanica Statistica* (Boringhieri, Torino, 1970), especially in Capitolo 3 "Termodinamica Covariante."
[8] M. Chaichian, R. Hagedorn, and M. Hayashi, Nucl. Phys. B **92**, 445 (1975).
[9] The additional assumption that the four-vector inverse temperature β be parallel to the four-vector volume will be used in the evaluation of the "physical" partition function (see Ref. 8, p. 451).
[10] S. Frautschi, Phys. Rev. D **3**, 2821 (1971); C.J. Hamer and S. Frautschi, Phys. Rev. D **4**, 2125 (1971); W. Nahm, Nucl. Phys. B **45**, 525 (1972); R. Hagedorn and I. Montvay, Nucl. Phys. B **59**, 45 (1973).
[11] D.E. Miller, Phys. Lett. A **61**, 291 (1977); **62**, 526 (1977); Ann. Israel Phys. Soc. **2** (1978), in the "Proc. of the XIII IUPAP Conf. on Stat. Phys." Haifa, Israel, August, 1977, pp. 940–4.
[12] See any book on statistical mechanics, e.g., K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).
[13] L.D. Landau and E.M. Lifschitz, "Statistical Physics," *Course of Theoretical Physics*, (Addison–Wesley, Reading, 1958) Vol. V. See especially Chapter VI, §66–69 and Chapter VII, §77–80 for descriptions of the quasiparticle spectra as well as the discussions of the interacting Bose and Fermi gases medicated through quasiparticles, in particular after using a Bogoliubov transformation.
[14] D. Ruelle, *Statistical Mechanics, Rigorous Results* (Benjamin, New York, 1969).
[15] L. Schwartz, *Théorie des Distributions* (Hermann, Pairs, 1966) see especially Chapitre VII; I.M. Gel'fand and G.E. Shilov, Generalized Functions, Vol.I, "Properties and Operations" (Academic, New York, 1964), see especially Chapter III, §§2.7–2.10.
[16] D.E. Miller, "Invariant Phase Space Solutions for Relativistic Energy Spectra," Bielefeld Preprint, BI-TP 76/21, July, 1976.
[17] M. Abromowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[18] T. Celik and H. Satz, "Phase Transitions in Hadronic Matter," to be published in Z. Phys. C: Particles and Fields.
[19] C.N. Yang and T.D. Lee, Phys. Rev. **82**, 404 (1952).

[20]P. Fré, and L. Sertorio, Nuovo Cimento A **28**, 538 (1975); **31**, 346 (1976); P. Fré, Fortschr. Phys. **25**, 579 (1977).
[21]N. Cabibbo and G. Parisi, Phys. Lett. B **59**, 67 (1975); S. Eliezer and R.M. Weiner, Phys. Rev. D **13**, 87 (1976); D. E. Miller, Phys. Lett. B **64**, 193 (1976).
[22]I. Montvay, Nuovo Cimento A **41**, 287 (1977); I. Montvay and H. Satz, Nuovo Cimento A **39**, 425 (1977); J. Letessier, A. Tounsi, and H. Satz, Nuovo Cimento A **45**, 580 (1978).
[23]F. Karsch and D.E. Miller, "Equation of State for Hadronic Matter," Phys. Lett. B **79**, 259 (1978).
[24]L. Sertorio, "Statistical Mechanics at High Energy Density," to be published in Rev. Nuovo Cimento.
[25]see Ref 13, §47.
[26]The author is very grateful to Prof. H. Satz for his remarks and criticism concerning the use of such a model.
[27]R. Gilmore, J. Math. Phys. **18**, 17 (1977).
[28]R.H. Dicke, Phys. Rev. **93**, 99 (1954); K. Hepp and E.H. Lieb, Ann. Phys. **76**, 360 (1973); Y.K. Wang and F.T. Hioe, Phys. Rev. A **7**, 831 (1973).
[29]A.F. Sadreev and A.P. Slivinsky, Phys. Lett. A **56**, 437 (1976); Physica A **92**, 255 (1978).
[30]R. Gilmore and D.H. Feng, Nucl. Phys. A **301**, 189 (1978); Phys. Lett. B **76**, 26 (1978).
[31]"Higher Transcendental Functions," *Bateman Manuscript Project*, edited by A. Erdélyi (McGraw–Hill, New York, 1953), Vol. II.
[32]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
[33]"Tables of Integral Transforms," in *Bateman Manuscript Project*, edited by A. Erdélyi (McGraw–Hill, New York, 1954), Vol. I. See the section on Fourier-sine transforms in Chapter II.
[34]E.C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford, 1937); N. Wiener, *The Fourier Integral and Certain of its Applications* (Dover, New York, 1933).

Applications of an algebraic quantization of the electromagnetic field

A. L. Carey and C. A. Hurst

Department of Mathematical Physics, University of Adelaide, P.O. Box 498, Adelaide, S. A. 5001
(Received 3 May 1978)

We quantize the full component electromagnetic potential so that a version of the Haag–Kastler axioms applies. The resulting quasilocal theory is viewed as a gauge theory in the spirit of Doplicher, Haag, and Roberts. Interaction with a classical current is described and found to yield no surprises. Lacking a rigorous scattering theory for the electromagnetic field, we construct a rigorous analog of the asymptotic Hilbert spaces for quantum electrodynamics proposed by Kulish and Faddeev and Zwanziger. On this basis we are able to find the representations of the gauge group that are associated with each charge sector in the asymptotic space. These representations correspond, in a sector of charge q, to the subsidiary condition $\partial^{\mu}A_{\mu}(x) = -qD(x)$.

1. INTRODUCTION

In a series of previous papers[1,2,3] we endevored to establish an alternative formalism for the quantization of the electromagnetic field. This formalism suffers from defects when compared with conventional Wightman field theory,[4] nevertheless the approach has some merits. In particular, it demonstrates that Fermi's original method[5] for quantizing the electromagnetic field can be recast in rigorous form and suggests that it is not improbable that a definite metric formalism, which retains most of the features of a manifestly covariant theory, may exist for the electromagnetic field in interaction.

This paper is concerned with applying the formalism to some problems to which the indefinite metric approach[6] does not appear to be well suited. In Sec. 2 we recast the results of Refs. 1–3 in the Haag–Kastler framework of quasilocal algebras.[7] Our aim here is to present the free electromagnetic field as a gauge theory in the spirit of Doplicher, Haag, and Roberts.[8,9] This section of the paper corrects and amplifies some remarks in Ref. 3. In Sec. 3 we apply the Weyl algebra approach developed by Shale[10] (and based on Cook[11]) to the example of the quantized electromagnetic field interacting with a classical current. There are no surprises here and the results of Refs. 10 and 11 for the scalar Bose field carry over without difficulty to the full four-component theory.

Our main results are contained in Sec. 4 and 5. Here we are concerned with producing rigorous analogues of the asymptotic Hilbert spaces for quantum electrodynamics proposed by Kulish and Faddeev[12] and Zwanziger.[13] We accept the results of Refs. 12 and 13, which are based on perturbation theory, as largely correct, believing that they represent at least qualitatively the asymptotic structure of quantum electrodynamics. Our attitude here is that, because charge is a conserved quantity, and heuristically at least, gauge transformations of the second kind have generators which evolve freely, results obtained on charge sectors and gauge transformations in an asymptotic Hilbert space for quantum electrodynamics will remain true in interaction.

The conclusions we come to are essentially the same as those of Zwanziger,[13] the main point being that the Lorentz condition $\partial^{\mu}A_{\mu}(x) = 0$ cannot hold in any sense in a nonzero charge sector (a result which is reinforced by the indefinite metric arguments of Ferrari, Picasso, and Strocchi[14]). In fact, in a sector of charge q we find the subsidiary condition to be $\partial^{\mu}A_{\mu}(x) = = -qD(x)$, where D is the usual commutator function for a scalar zero mass field. This has the corollary that the algebra of quasilocal observables appears to vary depending on the charge sector in question, in contrast to what happens for global gauge groups (see Doplicher, Hagg, and Roberts[8]).

This work is intended as a preliminary step in developing a definite metric approach to quantum electrodynamics which is suited to an analysis of charge sectors and the gauge group along the lines of Doplicher, Haag, and Roberts.[8,15]

2. C^*-ALGEBRA OF QUASILOCAL OBSERVABLES FOR THE ELECTROMAGNETIC FIELD

Let $X_0^+ = \{k \in R^4 | k_0 > 0, k^2 = 0\}$. Let $\mathscr{D}$ be the space of C^∞ functions with compact support in R^3 taking values in R^4. Each pair of functions $(f,g) \in \mathscr{D} \times \mathscr{D}$ defines a function on X_0^+ by

$$\phi(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int [\omega f(\mathbf{x}) + ig(\mathbf{x})] e^{-i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{x}, \qquad (2.1)$$

where $\omega = |\mathbf{k}|$. If we define

$$\hat{\phi}(x) = \frac{1}{(2\pi)^{3/2}} \int \frac{d^3\mathbf{k}}{2\omega} (\phi(\mathbf{k}) e^{-ik\cdot x} + \overline{\phi(\mathbf{k})} e^{ik\cdot x})$$

(where $k\cdot x = \omega x_0 - \mathbf{k}\cdot\mathbf{x}$), then $\hat{\phi}$ is an R^4-valued solution of $\Box\hat{\phi} = 0$ and $\hat{\phi}(\mathbf{x},0) = f(\mathbf{x})$, $\dot{\hat{\phi}}(\mathbf{x},0) = g(\mathbf{x})$.

Let M_0 denote the real linear space of solutions $\hat{\phi}$ of the wave equation so constructed from initial data in $\mathscr{D}$. The Poincaré group acts on M_0 by $\hat{\phi} \to \hat{\phi}'$,

where

$$\hat{\phi}'(x) = \Lambda\hat{\phi}(\Lambda^{-1}(x - a)), \qquad (2.2)$$

with (a,Λ) a typical Poincaré group element. It is well known that M_0 is invariant under this group action. Furthermore M_0 has a Poincaré invariant symplectic form

 0022-2488/79/050810-10$01.00

$$B(\hat{\phi},\hat{\phi}') = \frac{1}{i}\int \frac{d^3\mathbf{k}}{2\omega}(\phi^\mu(\mathbf{k})\overline{\phi'_\mu(\mathbf{k})} - \overline{\phi^\mu(\mathbf{k})}\phi'_\mu(\mathbf{k})).$$

We introduce two Poincaré invariant subspaces of M_0:

$$N_0 = \{\hat{\phi}\in M_0 \mid \partial^\mu\hat{\phi}_\mu(x) = 0\},$$

$$T_0 = \{\hat{\phi}\in M_0 \mid \hat{\phi}_\mu(x) = \partial_\mu\chi(x) \text{ where } \chi = R^4\to R\}.$$

It is not clear *a priori* that N_0 and T_0 contain very many elements. To see that N_0 and T_0 are "sufficiently large" we introduce on M_0 a "complex structure" by setting

$$(J_F\phi)(\mathbf{k}) = -ig\phi(\mathbf{k}),$$

where $g = \mathrm{diag}(1,-1,-1,-1)$. (Note that J_F does not leave M_0 invariant.) Then M_0 becomes a complex pre-Hilbert space if we define the inner product by

$$\langle\hat{\phi},\hat{\phi}'\rangle = B(\hat{\phi},J_F\hat{\phi}') + iB(\hat{\phi},\hat{\phi}').$$

The inner product $\langle\ ,\ \rangle$ is not Poincaré invariant, however the representation (2.2) and M_0 extends to the completion M of M_0 as an indecomposable representation by bounded operators.[16]

The closed subspaces N and T of M defined by the conditions

$$k^\mu\phi_\mu(\mathbf{k}) = 0 \quad\text{and}\quad \phi_\mu(\mathbf{k}) = k_\mu\chi(\mathbf{k}),\quad \chi = X_0^+\to C,$$

respectively, are Poincaré invariant. If we denote by P_N and P_T the projections onto N and T, then $P_S = P_N - P_T$ projects onto the subspace of M consisting of radiation gauge solutions of Maxwell's equations, i.e., if $\hat{\phi}\in(P_N - P_T)M$, then $\mathbf{k}\cdot\boldsymbol{\phi}(\mathbf{k}) = 0$, $\phi_0(\mathbf{k}) = 0$.

Proposition: N_0 and T_0 are dense in N and T respectively. This result is easy to prove if one observes that elements of T_0 are constructed by taking χ: $R^4\to R$, $\Box\chi = 0$, with initial data C^∞ of compact support. Then $\partial_\mu\chi\in T_0$ and, as T_0 is Poincaré invariant and the representation of the Poincaré group in T is irreducible, T_0 is dense in T. Similarly N_0 contains T_0 as well as elements of S of helicity ± 1. The only proper closed subspace of M invariant under the Poincaré group, containing T and elements of helicity ± 1 is N (see Ref. 16). So N_0, being invariant, is dense in N.

Now we construct the algebras associated with each of these spaces. If $\hat{\phi}\in M_0$, denote by $\delta_{\hat{\phi}}$ the function on M_0 taking the value 1 on $\hat{\phi}$ and zero elsewhere. In the usual way form a *-algebra $\mathcal{A}(M_0)$ by taking the complex linear span of the $\delta_{\hat{\phi}}$'s with

$$(\Sigma_i\lambda_i\delta_{\hat{\phi}_i})^* = \Sigma_i\bar{\lambda}_i\delta_{-\hat{\phi}_i}$$

and

$$\delta_{\hat{\phi}}\delta_{\hat{\phi}'} = \exp[-(i/2)B(\phi,\phi')]\delta_{\hat{\phi}+\hat{\phi}'}.$$

Construct $\mathcal{A}_c(M_0)$, the C^*-algebra of the CCR over M_0.[1,17,18]

Local algebras are obtained by taking for each $\mathcal{O}\subseteq R^4$, open and bounded, the algebra $\mathcal{A}(\mathcal{O})$ consisting of the linear span of those $\delta_{\hat{\phi}}$ such that, for some spacelike hyperplane P, the initial data for ϕ on P has support in $P\cap\mathcal{O}$. Clearly $\mathcal{A}(M_0) = \cup_{\mathcal{O}}\mathcal{A}(\mathcal{O})$ where the union is over all bounded open sets in R^4.

The algebra $\mathcal{A}_c(\mathcal{O})$ [the completion of $\mathcal{A}(\mathcal{O})$ in $\mathcal{A}_c(M_0)$] is the local field algebra associated with $\mathcal{O}$ and we have

$$\mathcal{A}_c(M_0) = \overline{\cup_{\mathcal{O}}\mathcal{A}_c(\mathcal{O})}.$$

We may similarly associate algebras $\mathcal{A}_c(N_0)$ and $\mathcal{A}_c(T_0)$ with the subspace N_0 and T_0 of M_0, and hence define algebras $\mathcal{A}_c(N_0(\mathcal{O}))$ and $\mathcal{A}_c(T_0(\mathcal{O}))$ by

$$\mathcal{A}_c(N_0(\mathcal{O})) = \mathcal{A}_c(N_0)\cap\mathcal{A}_c(\mathcal{O}),$$

$$\mathcal{A}_c(T_0(\mathcal{O})) = \mathcal{A}_c(T_0)\cap\mathcal{A}_c(\mathcal{O}).$$

Now the elements of T_0 act as automorphisms of $\mathcal{A}_c(M_0)$ by

$$A\to\delta_\psi A\delta_{-\psi},\quad A\in\mathcal{A}_c(M_0),\quad \psi\in T_0.$$

In particular,

$$\delta_\phi\to\delta_\psi\delta_\phi\delta_{-\psi} = e^{-iB(\psi,\phi)}\delta_\phi.$$

A straightforward calculation shows that this last expression corresponds to the heuristic one, $A_\mu\to A_\mu + \partial_\mu\chi$, where A_μ is the field operator in a given representation of $\mathcal{A}_c(M_0)$ and χ is the real valued solution of $\Box\chi = 0$ such that $\psi = \partial_\mu\chi$.

The linear space T_0 is therefore represented as an additive group by inner automorphisms of $\mathcal{A}_c(M_0)$. We call T_0 the gauge group. To investigate its properties and to determine the quasilocal algebra of observables for this theory we go to a particular representation.

Let $\sigma_F(\phi) = \exp[-\frac{1}{4}B(\phi,J_F\phi)]$ be the generating functional of the Fock representation ρ_F of $\mathcal{A}_c(M_0)$. This defines a state on $\mathcal{A}_c(M_0)$ which we will also denote by σ_F. Noting that $\rho_F(\mathcal{A}_c(T_0))''$ is a commutative von Neumann algebra, we can decompose the Hilbert space $\mathcal{F}_F$ of ρ_F as a direct integral by "diagonalizing" this algebra. This may be done explicitly as follows.

Denote by $T^\perp$ the orthogonal complement of N in M. Introduce the normal weak distribution μ (i.e., cylinder measure) on $T^\perp$ with characteristic function $\psi\to\exp[-\frac{1}{4}B(\psi,J_F\psi)]$, $\psi\in T$. Define H_ξ to be the Hilbert space which carries the representation of $\mathcal{A}_c(N_0)$ given by the state σ_ξ, whose restriction to $\mathcal{A}(N_0)$ has the form, for each $\xi\in T^\perp$,

$$\sigma_\xi(\Sigma_j\lambda_j\delta_{\phi_j}) = \sum_j\lambda_j\exp[-\tfrac{1}{2}C(\phi_j,\phi_j) + iB(\xi,\phi_j)],$$

where

$$C(\phi_j,\phi_j) = -\int\overline{\phi_j^\mu(\mathbf{k})}\phi_{j\mu}(\mathbf{k})d^3\mathbf{k}/2\omega.$$

Define the Hilbert space

$$H_F = \int_{\xi\in T^\perp} H_\xi d\mu(\xi),$$

to be the completion of the space of cylinder functions F on $T^\perp$ such that $F(\xi)\in H_\xi$ (note that all the spaces H_ξ may be chosen to be identical as the states σ_ξ are obtained, one from another by the action of the automorphism α_ξ: $\delta_\phi\to e^{iB(\xi,\phi)}\delta_\phi$, $\phi\in M_0$, i.e., $\sigma_\xi = \sigma_0\circ\alpha_\xi$) in the norm

$$\|F\| = \int\|F(\xi)\|_\xi^2 d\mu(\xi),$$

where $\|\ \|_\xi$ denotes the norm in H_ξ. The Fock representation ρ_F acts by

$$\rho_F(\delta_\phi)F(\zeta) = \rho_\zeta(\delta_\phi)F(\zeta),$$

for $\phi \in N_0$ where ρ_ζ is the representation corresponding to σ_ζ. If $\phi \in T_0$, this reduces to

$$\rho_F(\delta_\phi)F(\zeta) = e^{-iB(\zeta,\phi)}F(\zeta),$$

so that $\Delta_c(T_0)$ is diagonalized. For those elements ψ lying in $T^\perp \cap M_0$ we have

$$\rho_F(\delta_\psi)F(\zeta) = \exp[-\tfrac{1}{2}B(\zeta, J_F\psi) - \tfrac{1}{4}B(\psi, J_F\psi)]F(\zeta + \psi)$$

(noting that $J_F\psi \in T$). It is straightforward to check that this is indeed equivalent to the Fock representation (cf. Ref. 19).

From this realization of the Fock representation one observes directly that σ_ζ is a pure state on $\Delta_c(N_0)$, that $\rho_F(\Delta_c(T_0))''$ is the center of $\rho_F(\Delta_c(N_0))''$, and in fact that $\rho_F(\Delta_c(N_0))' = \rho_F(\Delta_c(T_0))''$. The representation $\rho_F \upharpoonright \Delta_c(N_0)$ is therefore miltiplicity free. Furthermore we see that $\rho_F(\Delta_c(N_0(\mathcal{O}))) = \rho_F(\Delta_c(\mathcal{O})) \cap \rho_F(\Delta_c(T_0))'$.

What now is the algebra of observables for the electromagnetic field? We cannot regard $\Delta_c(N_0(\mathcal{O}))$ as the algebra of observables localized in $\mathcal{O}$ since it contains elements which create from the vacuum unphysical states in the Fock representation. In order to guarantee that the gauge degrees of freedom are eliminated, we need a representation of $\Delta_c(N_0)$ in which the elements of the gauge group T_0 act as scalars. Since $\Delta_c(T_0)$ is the center of $\Delta_c(N_0)$ this will occur wherever we have an irreducible representation of $\Delta_c(N_0)$. Note that every irreducible representation of $\Delta_c(N_0)$, on restriction to $\Delta_c(T_0)$, defines a character of this algebra. A physical interpretation of these characters has been proposed by Zwanziger.[13] To discuss it we need to make contact with more conventional notation. Let ρ be a representation of $\Delta_c(M_0)$ such that $\lambda \to \rho(\delta_{\lambda\phi})$ is continuous in $\lambda \in R$. Then $\rho(\delta_\phi) = \exp iA(\phi)$ where $A(\phi)$ may be interpreted as the electromagnetic potential smeared by ϕ. In the special case where $\phi \in T_0$, say $\phi_\mu(\mathbf{k}) = k_\mu \chi(\mathbf{k})$, then $A(\phi)$ may be seen to represent the operator $\partial^\mu A_\mu(\chi)$, i.e., the supplementary condition operator $\partial^\mu A_\mu(x)$ smeared by χ. Hence the algebra $\Delta_c(T_0)$ is essentially the algebra generated by the supplementary condition operators $\partial^\mu A_\mu(x)$ and each character of $\Delta_c(T_0)$ corresponds to a choice of supplementary condition. Before pursuing this argument further let us determine the set of all characters of $\Delta_c(T_0)$, i.e., its spectrum as a C^*-algebra.

Because ρ_F is a faithful representation of $\Delta_c(M_0)$ we can identify the elements of $T^\perp$ with a subset of the spectrum of $\Delta_c(T_0)$, dense in the weak topology. In fact we have the

Lemma: The elements of the spectrum of $\Delta_c(T_0)$ are in 1–1 correspondence with elements of the algebraic dual of T_0.

Proof: Note firstly that all elements of $T^\perp$ define characters of $\Delta_c(T_0)$. This is because the realization of ρ_F given above sets up an isomorphism of $\Delta_c(T_0)$ with a C^*-algebra of continuous functions on $T^\perp$, and hence the elements of $T^\perp$ define evaluation functions in this algebra. So in particular, the zero element of T is such a functional, call it τ_0. Given a ζ in the algebraic dual of T_0 we may extend it to M_0 by defining it to be zero on those elements of M_0 not in T_0. Then the map

$$\alpha_\zeta : \delta_\phi \to e^{i\zeta(\phi)}\delta_\phi, \quad \phi \in M_0,$$

extends to an automorphism of $\Delta_c(M_0)$[17] and hence of $\Delta_c(T_0)$. The composite character $\tau_\zeta = \tau_0 \circ \alpha_\zeta$ of $\Delta_c(T_0)$ exists and so we have a map, $\zeta \to \tau_\zeta$ from the algebraic dual of T_0 into the spectrum of $\Delta_c(T_0)$. Now suppose τ is any character of $\Delta_c(T_0)$. As the elements of $T^\perp$ are dense we can find a net $\tau_{\zeta_\alpha} \to \tau$ with $\zeta_\alpha \in T^\perp$. We may write $\tau(\delta_\phi) = \exp i\zeta(\phi)$ for $\phi \in T_0$ where $\zeta : T_0 \to R$ satisfies $\zeta(\phi + \phi') = \zeta(\phi) + \zeta(\phi')$. To show that ζ is an element of the algebraic dual of T_0 we need only show that $\zeta(\lambda\phi) = \lambda\zeta(\phi)$ for $\lambda \in R$, and $\phi \in T_0$. But $\exp i\zeta(\lambda\phi)$ is the limit of both $[\exp i\lambda\zeta_\alpha(\phi)]$ and $[\exp i\zeta_\alpha(\lambda\phi)]$ and furthermore $\exp i\lambda\zeta_\alpha(\phi) \to \exp i\lambda\zeta(\phi)$. Hence ζ is linear, completing the proof.

Returning to the general argument, Zwanziger[13] proposed that if τ_ζ is a character of $\Delta_c(T_0)$, then ζ should be related to the value of the charge (admittedly this is a restatement in our language of a nonrigorous argument in the indefinite metric formalism). In particular, if $\zeta \equiv 0$, then we are trying to impose the usual supplementary condition $\partial^\mu A_\mu = 0$ which, on the basis of Refs. 13 and 14, we would expect to hold only in the charge zero sector.

We will return to this discussion of charge sectors and characters in Sec. 5; for the moment, we will assume that we are considering the case $\zeta \equiv 0$. It then follows that if I denotes the two sided ideal of $\Delta_c(N_0)$ generated by the kernel of the character τ_0, our candidate for the quasilocal algebra of observables is the quotient $\Delta_c(N_0)/I$. It follows from Ref. 1 that this algebra is isomorphic to a subalgebra of $\Delta_c(S)$ where S has the symplectic form induced by B. Physically, therefore, $\Delta_c(N_0)/I$ has the properties of an observable algebra, since it depends only on the transverse components of the electromagnetic potential. Noting that the ideal I is the kernel of the representation ρ_0 of $\Delta_c(N_0)$ (0 denotes the zero element of $T^\perp$), we will henceforth write I as $\ker\rho_0$.

Observe from the definition of ρ_0 that the corresponding state σ_0 is Poincaré invariant. Consquently $\ker\rho_0$ is Poincaré invariant and hence the Poincaré group acts by automorphisms of $\Delta_c(N_0)/\ker\rho_0$. There is a local structure on $\Delta_c(N_0)/\ker\rho 0$ defined by setting

$$\mathfrak{A}(\mathcal{O}) = (\Delta_c(N_0(\mathcal{O})) + \ker\rho_0)/\ker\rho_0,$$

for each bounded open set $\mathcal{O} \subset R^4$. From the fact that Poincaré transformations act on $\Delta_c(M_0)$ by

$$\alpha_{a,\Lambda} : \delta_\phi \to \delta_{(a,\Lambda)\phi},$$

we have

$$\alpha_{a,\Lambda}(\Delta_c(\mathcal{O})) = \Delta_c(\Lambda\mathcal{O} + a),$$

$$\alpha_{a,\Lambda}(\Delta_c(N_0(\mathcal{O}))) = \Delta_c(N_0(\Lambda\mathcal{O} + a)),$$

and hence

$$\alpha_{a,\Lambda}(\mathfrak{A}(\mathcal{O})) = \mathfrak{A}(\Lambda\mathcal{O} + a).$$

As σ_0 is Poincaré invariant these automorphisms are implemented in the Hilbert space H_0 of the representation ρ_0. Finally we note that, as

$$\rho_F(\Delta_c(N_0(\mathcal{O}))) \subseteq \rho_F(\Delta_c(N_0(\mathcal{O}')))',$$

where $\mathcal{O}'$ is spacelike separated from $\mathcal{O}$, we have

$$[\mathfrak{A}(\mathcal{O}),\mathfrak{A}(\mathcal{O}')] = 0.$$

Putting the above together, we conclude that the algebras $\Delta_c(M_0)$ and $\Delta_c(N_0)/\ker\rho_0$ both satisfy a version of the Haag–Kastler axioms.[7]

The gauge group T_0 shares some of the properties of the global gauge groups of Refs. 8 and 15. We shall not discuss these here but refer the reader to Ref. 3.

Finally, it is not difficult to show that $\Delta_c(N_0)/\ker\rho_0$ is also isomorphic to the Weyl algebra which is normally used for the electromagnetic field,[20,21] namely that constructed over the space of functions

$$F_{\mu\nu} = \partial_\mu\hat{\phi}_\nu - \partial_\nu\hat{\phi}_\mu, \quad \hat{\phi}\in M_0.$$

Finally we remark that the above procedure suffers from a number of deficiencies. The most serious, from a technical viewpoint, is that the local C^*-algebras above depend rather critically on the choice of test functions M_0. It would be desirable to replace them by local von Neumann algebras, however we have not yet found a convenient means for doing this. To a lesser extent there is the difficulty that Lorentz boosts are not implementable in the Fock representation ρ_F of $\Delta_c(M_0)$. This means that our electromagnetic potential is not covariant. However this problem can be overcome by using a covariance algebra construction[22] as described in Refs. 1 and 2 which produces a covariant, though reducible, representation of $\Delta_c(M_0)$.

3. INTERACTION OF THE ELECTROMAGNETIC FIELD WITH A CLASSICAL CURRENT

We follow here the treatments of Shale[10] and Cook[11] of the classical source problem for the case of a scalar Bose field. We suppose that $j: R^4 \to R^4$ is a given classical current distribution, and we seek the corresponding time evolution automorphism of $\Delta_c(M_0)$ and of $\Delta_c(N)/\ker\rho_0$.

Let us suppose initially that j is sufficiently smooth and vanishes sufficiently rapidly at infinity for the definitions below to carry through. We will lift this restriction later. Denote by $t\to T(t)$ the one-parameter group

$$(T(t)\phi)(\mathbf{k}) = \exp(J_s\omega t)\phi(\mathbf{k}), \quad \hat{\phi}\in M, \tag{3.1}$$

which gives the free time evolution in the one-particle space M. (Here J_s denotes the usual complex structure of multiplication by i.) The corresponding automorphism group $t\to\alpha_0(t)$ of $\Delta_c(M_0)$ is given by

$$\alpha_0(t)\delta_\phi = \delta_{T(t)\phi}. \tag{3.2}$$

If j is constant in time [i.e., $j(x_0,\mathbf{x})\equiv j(\mathbf{x})$ for all $x_0\in R$], then following Shale[10] we can define the interacting dynamics to be given by $t\to\alpha(t) = \gamma_j\alpha_0(t)\gamma_j^{-1}$ where γ_j is the automorphism

$$\gamma_j(\delta_\phi) = \exp[-iB(\phi,A^{-1}j)]\delta_\phi$$

of $\Delta_c(M_0)$, and A is the generator of $t\to T(t)$.

All of the results of Ref. 10 now carry over to this example. We note that $\alpha(t)$ is implementable in the Fock representation of $\Delta_c(M_0)$ if and only if the Fourier transform of j,

$$\tilde{j}(\mathbf{k}) = \int e^{i\mathbf{k}\cdot\mathbf{x}}j(\mathbf{x})d^3\mathbf{x},$$

is in M. Denoting by $w(t)$ the automorphism $\alpha(t)\alpha_0(t)^{-1}$ of $\Delta_c(M_0)$ we define the Møller wave operators by

$$w_\pm = \lim_{t\to\pm\infty} w(t),$$

and observe after Ref. 10, that if σ is a generating functional for $\Delta_c(M_0)$, the generating functional $w(t)^*\sigma$ [the * denotes the dual action of automorphisms on the state space of $\Delta_c(M_0)$ or equivalently, on the generating functionals] converges as $t\to\pm\infty$ in a suitable sense to

$$w_\pm\rho = \gamma_j\rho.$$

Thus the scattering operator $S = w_+w_-$ is trivial, as is to be expected. Consider now the problem of infrared divergences. When the Fourier transform of j is too singular at the origin in momentum space to be an element of M, we find that $\alpha(t)$ is not implementable and we have the usual infrared problems. However the case of the full four-component theory differs from that of the scalar or radiation gauge theory.

To see this we impose the condition that the current j be conserved. For a time-independent source this means $j_0(\mathbf{x}) = 0$ and $\nabla\cdot\mathbf{j}(\mathbf{x}) = 0$. Then γ_j, as an automorphism of $\Delta_c(T)$, acts as the identity for

$$B(\phi,A^{-1}j) = \frac{1}{i}\int\frac{d^3\mathbf{k}}{2\omega^2}(k^\mu\chi(\mathbf{k})\overline{j_\mu(\mathbf{k})} - k^\mu\overline{\chi(\mathbf{k})}j_\mu(\mathbf{k}))$$

$$= 0 \text{ as } k^\mu j_\mu(\mathbf{k}) = 0. \tag{3.3}$$

Thus γ_j leaves the ideal $\ker\rho_0$ fixed and consequently defines an automorphism $\tilde{\gamma}_j$ of $\mathfrak{A}$ by

$$\tilde{\gamma}_j(\delta_\phi + \ker\rho_0) = (\gamma\delta_\phi) + \ker\rho_0, \quad \phi\in N_0. \tag{3.4}$$

This is just a restatement in our formalism of the well known fact that a conserved current produces a gauge invariant interaction.

It follows from this observation that it is only the transverse components of j (i.e., those in P_sM) which contribute to the evolution of the observables, and consequently the full four-component theory gives the same time evolution as the radiation gauge treatment.

The second point to note is that the full four-component theory automatically includes an infrared cutoff for the observable algebra. This is a consequence of the elementary

Lemma: If $\hat{\phi}\in N_0$, then ϕ vanishes at the origin in momentum space.

Proof: We have $\partial^\mu\hat{\phi}_\mu(x) = 0$ which implies for the initial data that

$$\dot{\hat{\phi}}(0,\mathbf{x}) = -\nabla\cdot\hat{\phi}(0,\mathbf{x}).$$

But from (3.1)

$$\phi_0(0) = \frac{-i}{(2\pi)^{3/2}}\int\hat{\phi}_0(0,\mathbf{x})d^3\mathbf{x}$$

$$= \frac{i}{(2\pi)^{3/2}} \int \nabla\cdot\hat{\boldsymbol{\phi}}(0,\mathbf{x}) d^3x$$

$$= 0,$$

as $\hat{\boldsymbol{\phi}}$ has compact support on the surface $x_0 = 0$. But $k^\mu \phi_\mu(\mathbf{k}) = 0$ so that, for example, $k_1\phi_0(k_1;0,0) - k_1\phi_1(k_1,0,0)(k_1 > 0)$. That is $\phi_0(0) = \phi_1(0) = 0$, and similarly for the other components.

By noting that the initial data for $\hat{\phi} \in M_0$ have Fourier transforms which, as function of $\mathbf{k}$ are analytic, we can expand ϕ as a Taylor series about the origin and using the Lemma and (3.1) we obtain the estimate

$$|\phi_\mu(\mathbf{k})| \leqslant \omega(1 + O(\mathbf{k})), \quad \mu = 0,1,2,3.$$

Thus provided j is at worst a tempered distribution and the Fourier transform

$$\tilde{j}(k) = \int e^{ik\cdot x} j(x) d^4x$$

of j is of the form $\sim 1/\omega$ as a function on X_0^+ near $\mathbf{k} = 0$ we have

$$|B(\phi, A^{-1}j)| \leqslant 2 \int |\phi^\mu(k) j_\mu(k)| \frac{d^3k}{2\omega^2} < \infty.$$

Hence the infrared cutoff built into elements of N_0 allows us to define $\alpha(t)$ as an automorphism of $\Delta_c(N_0)$ even for singular currents j, and hence we obtain, using (3.3) and (3.4), a one-parameter group $t \to \tilde{\alpha}(t)$ which gives the dynamics of the algebra of observables $\mathfrak{A}$.

The above analysis carries over to the case of a time dependent source. In this case, we follow Cook[11] and define the evolution in terms of a propagator,

$$\alpha(s,t) = \gamma_{j_{s,t}},$$

where $\gamma_{j_{s,t}}$ is the automorphism of $\Delta_c(M_0)$ defined as follows. First set

$$j'(t,\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int e^{-i\mathbf{k}\mathbf{x}} j(t,\mathbf{x}) d^3\mathbf{x},$$

and then define

$$j_{s,t}(k) = i \int_s^t (T_t j')(t',\mathbf{k}) dt'$$

and hence

$$\gamma_{j_{s,t}}(\delta_\phi) = \exp[iB(\phi, j_{s,t})]\delta_\phi.$$

This will be well defined whenever $j_{s,t} \in M$ and in fact (by the preceding analysis) in certain more general cases.

Now we show how the interacting fields at time t may be defined in terms of the asymptotic fields at $t = \pm\infty$. Suppose that the state of the system at time s is given by a generating functional σ_s. Then at time $t > s$ the state will be given by the functional $\alpha(t,s)^*\sigma_s$ where

$$(\alpha(t,s)^*\sigma_s)(\delta_\phi) = \sigma_s(\alpha(t,s)^{-1}\delta_\phi)$$

$$= \sigma_s(\alpha(s,t)\delta_\phi).$$

In order to define asymptotic fields we need to show that the automorphism $\alpha(-\infty, t)$, for example, exists and that for each generating functional σ, $\alpha(t,s)^*\sigma$ converges pointwise to $\alpha(t,-\infty)^*\sigma$. If j is a Schwartz space function, then it follows that

$$B(j_{s,t},\phi) = \int_s^t dx_0 \int d^3\mathbf{x} j^\mu(x)\phi_\mu(x).$$

Clearly therefore, in this case, $\lim_{s\to-\infty} \alpha(s,t) = \alpha(-\infty,t)$ exists where $\alpha(-\infty,t)$ is defined via

$$B(j_{-\infty,t},\phi) = \int_{-\infty}^t dx_0 \int d^3\mathbf{x} j^\mu(x)\phi_\mu(x). \tag{3.5}$$

More generally, the automorphisms $\alpha(s,t)$, $s,t \in R \cup \{-\infty, \infty\}$ exist whenever the function $j^\mu\phi_\mu$ is integrable on R^4 for all $\phi \in M_0$. Finally, the scattering operator is defined by $s = \alpha(-\infty, \infty)$.

Because the elements $\phi \in N_0$ vanish at $\mathbf{k} = 0$, the scattering operator and the propagators will be defined as automorphisms of $\Delta_c(N_0)$ for currents j such that $j_{s,t}$ behaves as $1/\omega$ near the origin in momentum space for all s,t. Furthermore, if j is conserved it follows from either (3.5) or

$$B(j_{s,t},\phi) = \frac{1}{i} \int (j^\mu_{s,t}(\mathbf{k})\overline{\phi_\mu(\mathbf{k})} - \overline{j^\mu_{s,t}(\mathbf{k})}\phi_\mu(\mathbf{k}) d^3\mathbf{k}$$

that $\alpha(s,t)$ acts as the identity on $\Delta_c(T_0)$ and hence defines an automorphism of the observable algebra,

$$\mathfrak{A} = \Delta_c(N_0)/\ker\rho_0.$$

Finally these automorphisms are implemented in the Fock representation of $\Delta_c(M_0)$ provided $j_{s,t} \in M$ for all $s,t \in R \cup \{-\infty, \infty\}$, in which case one can deduce easily from (3.5) that the fields at time t are related to the asymptotic field at $t = -\infty$ by the usual Yang–Feldman equation

$$A_\mu(x) = A^{\text{in}}_\mu(x) + \int j_\mu(x') D_{\text{ret}}(x - x') d^4x'.$$

4. ASYMPTOTIC HILBERT SPACE FOR QUANTUM ELECTRODYNAMICS

Recently Kulish and Faddeev[12] and Zwanziger[13] have sought to formulate the asymptotic dynamics of quantum electrodynamics so as to ensure that the S matrix is free from infrared divergences. Their work is an outgrowth of that of Kibble[21,23] and Chung.[24] Our interest in this question is rather different.

Heuristic calculations suggest that the generators of gauge transformations of the second kind evolve in interaction like a free field [these are the operators $\partial^\mu A_\mu(f)$ where f is a test function]. Thus we might expect that properties of the gauge group which hold in the asymptotic Hilbert space for quantum electrodynamics will persist in interaction. Naturally we would like to prove a theorem along these lines. However, this does not seem to be possible at present. So, in order to obtain information about charge sectors and their relationship to the gauge and the algebra of observables, we construct in this section a rigorous version of the asymptotic Hilbert space for quantum electrodynamics proposed in Refs. 12 and 13. The essential difference between

the space described here and those in Refs. 12 and 13 is that we quantize the electromagnetic field as in Sec. 1 rather than use the indefinite metric.

One question we do not attempt to answer is whether our results are consistent with a perturbation theory based on our definite metric quantization. However, recent work suggests that this should indeed be the case.[25]

We introduce firstly the Fock space for the electron/-positron field. We will follow Bongaarts.[26,27] Let H be the complex L^2 space of functions $f: R^3 \to C^4$ with inner product

$$\langle f,g\rangle = \sum_{\alpha=1}^{4} \int \overline{f_\alpha(\mathbf{x})}\, g_\alpha(\mathbf{x}) d^3\mathbf{x}.$$

Introduce the Fourier transformed space of functions on

$$X^{\pm} = \{p \in R^4 \mid p^2 = m_e^2, p_0 \lesseqgtr 0\},$$

where m_e is the electron/positron mass and

$$f(\mathbf{x}) = \int \hat{f}(\mathbf{p}) e^{i\mathbf{p}\cdot\mathbf{x}} d^3\mathbf{p}/2(\mathbf{p}^2 + m_e^2)^{1/2}.$$

Introduce the projection operators γ_+, γ_- onto the positive and negative energy solutions. They are given by multiplication by matrix valued functions $\gamma_\pm(\mathbf{p})$ on $X^+ \cup X^-$ (see for example, Ref. 28). So we may write

$$f(\mathbf{x}) = \int_{X^+} [e^{i\mathbf{p}\cdot\mathbf{x}} \gamma_+(\mathbf{p}) \hat{f}(\mathbf{p}) + e^{-i\mathbf{p}\cdot\mathbf{x}} \gamma_-(\mathbf{p}) \hat{f}(\mathbf{p})] d^3\mathbf{p}/(\mathbf{p}^2 + m_e^2)^{1/2}. \tag{4.1}$$

Using the charge conjugation operation C we may identify the space of negative energy solutions, H_- say, with the space H_+ of positive energy solutions, as $C\gamma_+ = \gamma_- C$. Let $H' = H_+ \oplus CH_-$. This is the physical one-particle space, and to distinguish the two spaces in the direct sum we write H_p for the particle space and H_a for the antiparticle space, CH_-.

Now form the antisymmetrical Fock space $\mathcal{F}_e$ over H',

$$C \oplus H' \oplus H' \otimes H' \oplus \cdots.$$

Then the annihilation operators $C(f)$, $f \in H'$ are given by extension of the operator which maps the vacuum Ω_0 to zero, $f' \mapsto \langle f, f'\rangle \Omega_0$ and

$$\sum_{\text{perm}\sigma} \text{sign}\sigma (f_{\sigma(1)} \otimes \cdots \otimes f_{\sigma(n)}) \mapsto \sqrt{n} \sum_{\sigma} \text{sign}\sigma \langle f, f_{\sigma(1)}\rangle (f_{\sigma(2)} \otimes \cdots \otimes f_{\sigma(n)}),$$

for all finite sets $f_1, \cdots f_n \in H'$. Annihilation of particles and antiparticles is given by the operators $C(\gamma_+ f)$ and $C(\gamma_+ Cf)$ respectively, $f \in H$, while creation is effected by their adjoints.

Now for each configuration of ingoing or outgoing electrons and positrons with asymptotic momenta P_a, $a = 1, \cdots n$; q_b, $b = 1, \ldots m$ respectively, we introduce the space

$$\gamma_+(p_1) C^4 \otimes \cdots \otimes \gamma_+(p_n) C^4 \otimes \gamma_+(q_1) C^4 \otimes \cdots \otimes \gamma_+(q_m) C^4 \otimes K(p_1, \ldots p_n, q_1, \ldots q_m), \tag{4.2}$$

where $K(p_1, \ldots p_n, q_1, \ldots q_m)$ may be chosen in two ways. Kulish and Faddeev define it as follows. Let $\mathcal{F}_p^{\pm}$ denote, for each $p \in X^+$, the Hilbert space carrying the representation $\rho_p^{\pm}$ of $\Delta_c(M_0)$ determined by the generating functional

$$\sigma_p^{\pm}(\phi) = \exp[\mp ieB(\eta_p, \phi)] \sigma_F(\phi),$$

where $-e$ = electron charge and $(\eta_p)_\mu(k) = ip_\mu / p\cdot k$ ($p\cdot k = p_0 k_0 - \mathbf{p}\cdot\mathbf{k}$). Let $\mathcal{F}$ denote the direct sum $\mathcal{F}_F \oplus \oplus_{p \in X^+} \mathcal{F}_p^+ \oplus \mathcal{F}_p^-$). $\mathcal{F}$ is essentially the nonseparable space introduced by Kibble to handle the problem of infrared divergences.[21,23] Then $K(p_1, \cdots, p_n, q_1, \cdots, q_m)$ is the space $\oplus_a \mathcal{F}_{p_n}^+ \oplus \oplus_b \mathcal{F}_{q_n}^-$ regarded as a subspace of $\mathcal{F}$. Alternatively, our interpretation of Zwanziger's space leads us to set $K(p_1, \cdots, p_n, q_1, \cdots, q_m)$ equal to the Hilbert space carrying the representation of $\Delta_c(M_0)$ determined by the generating functional

$$\sigma_{p_a, q_b}(\phi) = \exp[-ieB(\eta_{p_a, q_b}, \phi)] \sigma_F(\phi)$$

when η_{p_a, q_b} is the function

$$\eta^\mu_{p_a, q_b}(k) = \sum_a -\left(\frac{ep_a^\mu}{p_a\cdot k}\right) + \sum_b \frac{eq_b^\mu}{p_b\cdot k}.$$

We will adopt the second definition of $K(p_1, \cdots, p_n, q_1, \cdots, q_m)$ for the remainder of the paper, as it appears to us to be less artificial than that of Kulish and Faddeev. This choice in no way affects any of the conclusions of Kulish and Faddeev.

Denote by $H_{n,m}(p_a, q_b)$ the space defined by (4.2) with the second choice of $K(p_1, \cdots, p_n, q_1, \cdots, q_m)$. Form the direct integral, denoted $H_{n,m}$, of the spaces $H_{n,m}(p_a, q_b)$ over $X^+ \times \cdots \times X^+$ ($n + m$ factors), $H_{n,m}$ consists of equivalence classes of functions F on $X^+ \times \cdots \times X^+$ such that

(a) $F(p_1, \cdots, p_n, q_1, \cdots, q_m) \in H_{n,m}(p_a, q_b)$,

(b) $\int \|F(p_1, \cdots, p_n, q_1, \cdots, q_m)\|^2_{n,m} d\mathbf{p} d\mathbf{q} < \infty$,

where $d\mathbf{p}$ denotes

$$\prod_{a=1}^{n} \frac{d^3\mathbf{p}_a}{2(\mathbf{p}_a^2 + m_e^2)^{1/2}}, \quad d\mathbf{q} \text{ is } \prod_{b=1}^{m} \frac{d^3\mathbf{q}_b}{2(\mathbf{q}_b^2 + m_e^2)^{1/2}},$$

m_e = electron/positron mass and the norm $\| \ \|_{n,m}$ is the extension of that given on functions of the form

$$p_1, \cdots, p_n, q_1, \cdots, q_m \mapsto f_1(p_1) \otimes \cdots \otimes f_n(p_n) \otimes g_1(q_1) \otimes \cdots \otimes g_m(q_m) \otimes u(p_a, q_b)$$

by

$$\int \prod_a f_a(p_a)^* \gamma_0 f_a(p_a) \times \prod_b g_b(q_b)^* \gamma_0 g(q_b) \cdot \|u(p_a, q_b)\|^2 d\mathbf{p} d\mathbf{q}.$$

[Here $\{f_a\}$, $\{g_b\}$ are elements of H_p, H_a respectively and $u(p_a, q_b)$ is in $K(p_1, \cdots, p_n, q_1, \cdots, q_m)$.]

Finally the asymptotic Hilbert space is the direct sum

$$\mathcal{A} = \bigoplus_{n,m=0}^{\infty} H_{n,m},$$

where $H_{0,0}$ is by definition and Fock space $\mathcal{F}_F$ of the photon field. The physical interpretation of the vectors in this space

is most easily established by defining a unitary operator from $\mathscr{F}_e \otimes \mathscr{F}_F$ into $\mathscr{A}$. To do this we associate with the vector

$$C(\gamma_+ f_1)^* \cdots C(\gamma_+ f_n)^* C(\gamma_+ Cg_1)^* \cdots C(\gamma_+ Cg_m)^* \Omega_0 \otimes \sigma_F(\delta_\phi)\Omega_F$$

[where $\{f_a\}$, $\{g_b\}$ have Fourier transforms (4.1) in H, $\phi \in M_0$ and Ω_F is the vacuum in $\mathscr{F}_F$] the element

$$p_1,\cdots,p_n,q_1,\cdots,q_m \to \gamma_+(p)f_1(p_1) @ \cdots @ \gamma_+(p_n)f(p_n) @ \gamma_+(q_1)Cg_1(q_1) @ \cdots @ \gamma_+(q_m)Cg_m(q_m) \otimes \{\exp[-iB(\eta_{p_a,q_b},\phi)]\rho_{p_a,q_b}(\delta_\phi)\Omega_{p_a,q_b}\} \tag{4.3}$$

of $\mathscr{A}$. Here Ω_{p_a,q_b} is the cyclic vector in $K(p_1,\cdots,p_n,q_1,\cdots,q_m)$. It is straightforward to check that this extends to a norm preserving linear map U of $\mathscr{F}_e \otimes \mathscr{F}_F$ onto $\mathscr{A}$. It is not difficult to show that the linear span of the functions of the form (4.3) is dense in $\mathscr{A}$, so that this map U is actually an isometry from $\mathscr{F}_e \otimes \mathscr{F}_F$ onto $\mathscr{A}$. We note that U is a rigorous version of the map given in Ref. 13 by Eq. (A9). The interpretation of elements of $\mathscr{A}$ as states corresponding to a certain number of particles and antiparticles is now clear. Perhaps not too immediately apparent however, are the reasons for choosing the functions η_{p_a,q_b}. A fuller discussion of their origin is given in Refs. 12, 29, and 30. We remark only that the expression $j_\mu(k) = ep_\mu/p\cdot k$ is the Fourier transform of the current due to a point electron moving with 4- velocity $v_\mu = p_\mu/m_e$. Hence, loosely speaking, the function η_{p_a,q_b} represents the effect of the asymptotic electrons and positrons on the electromagnetic field.

One of the main results which Kulish and Faddeev wish to establish is the form of the asymptotic dynamics for the fields in $\mathscr{A}$. Contrary to the claims of Ref. 12 however, the expression for the time evolution (Eq. 10) does not define an operator on their asymptotic Hilbert space. This is essentially because the L^2 norm of $\mathbf{k} \to \eta_{p_a,q_b}(\mathbf{k})$ diverges for large $\mathbf{k}$. To be fair, However, Kulish and Faddeev point out that they expect ultraviolet divergences to be eliminated by a renormalization. In any case, since we expect this asymptotic space to be a good approximation to asymptotic QED only in its infrared behavior we could overcome this difficulty by introducing a momentum cutoff as follows.

Let f be a real function on R which is identically one on a neighborhood of zero and zero on a neighborhood of infinity. Then we introduce a new asymptotic Hilbert space $\mathscr{A}'$ defined in the same way as $\mathscr{A}$ except that we replace $\eta^\mu_{p_a,q_b}$ by the function η^μ_{f,p_a,q_b} defined by

$$\mathbf{k} \to \sum_a -e\frac{f(k\cdot p_a)p_a^\mu}{p_a\cdot k} + \sum_b e\frac{f(k\cdot q_b)q_b^\mu}{q_b\cdot k}.$$

It is then straightforward to write down an operator on $\mathscr{A}'$ corresponding to that given in Eq. 10 of Ref. 12. The difficulty with this procedure is that all the results of this and the next section are then cutoff dependent. However, see the remark at the end of Sec. 5.

5. REPRESENTATIONS OF $\Delta_c(N_0)$ AND CHARGE SECTORS

In this section our aim is to write down the representation of $\Delta_c(N_0)$ which defines the algebra of observables of the electromagnetic field. Recalling the discussion of Sec. 1, this means finding a representation of $\Delta_c(N_0)$ in which $\Delta_c(T_0)$ acts as a scalar. As with the free field, the representation ρ of $\Delta_c(M_0)$ described in Sec. 4 decomposes as a direct integral of representations of $\Delta_c(N_0)$. This decomposition is achieved by decomposing, for each asymptotic momentum distribution $p_1,\cdots,p_n,q_1,\cdots,q_m$ of electrons and positrons, the representation ρ_{p_a,q_b}:

$$\rho_{p_a,q_b}(\delta_\phi) = \exp[iB(\eta_{p_a,q_b},\phi)] \int_{T'}^{\oplus} \rho_\zeta(\delta_\phi)d\mu(\zeta).$$

We will choose for our representation of $\Delta_c(N_0)$ one of the representations occurring in this decomposition. Our only criterion here is that Lorentz transformations be implemented. (We note that for the same reason that the asymptotic dynamics cannot be defined in $\mathscr{A}$ without a momentum cutoff, so translations are not implemented in $\mathscr{A}$, again contrary to the assertions of Ref. 12. So, the closest approximation we can make to Poincaré invariance is to demand that Lorentz transformations be implemented.)

Consider therefore the generating functional

$$\tilde{\sigma}_{p_a,q_b} : \phi \to \exp[iB(\eta_{p_a,q_b},\phi) + \tfrac{1}{2}iC(\phi,\phi)], \quad \phi \in N_0,$$

which defines the representation

$$\tilde{\rho}_{p_a,q_b} : \delta_\phi \to \exp[iB(\eta_{p_a,q_b},\phi)\rho_0(\delta_\phi)] \tag{5.1}$$

of $\Delta_c(N_0)$. This occurs in the above direct integral decomposition of ρ_{p_a,q_b}.

Now we perform the same construction with the representations $\tilde{\rho}_{p_a,q_b}$ as we did with the representations ρ_{p_a,q_b}, namely, form the direct integral over $X^+ \times \overset{(n+m)}{\cdots} \times X^+$ to get a representation $\tilde{\rho}_{n,m}$, say. Denote by $\tilde{\rho}$ the direct sum $\oplus_{n,m}^{\infty} \tilde{\rho}_{n,m}$ where $\tilde{\rho}_{0,0}$ is by definition ρ_0. The representation ρ of $\Delta_c(N_0)$ acts on a space B which can be obtained by replacing in the definition of $\mathscr{A}$, the space $K(p_1,\cdots,p_n,q_1,\cdots,q_m)$ by the space $\mathscr{L}_{p_1,\cdots,p_n,q_1,\cdots,q_m}$ which carry the representations $\tilde{\rho}_{p_a,q_b}$.

We first check that the Lorentz group automorphisms of $\Delta_c(N_0)$ are implemented in B. Given the function

$$p_1,\cdots,p_n,q_1,\cdots,q_m \to \gamma_+(p_1)f_1(p_1) @ \cdots @ \gamma_+(p_n)f_n(p_n) @ \gamma_+(q_1)Cg_1(q_1) @ \cdots @ \gamma_+(q_m)Cg_m(q_m) \otimes \tilde{\rho}_{p_a,q_b}(\delta_\phi)\tilde{\Omega}_{p_a,q_b},$$

in B ($\phi \in N_0$, f_a, $g_b \in H$ and $\tilde{\Omega}_{p_a,q_b}$ is the cyclic vector for $\tilde{\rho}_{p_a,q_b}$, $a = 1,\cdots,n$, $q = 1,\cdots,m$), then a Lorentz transformation Λ takes it to

$$p_1,\cdots,p_n,q_1,\cdots,q_m \to \gamma_+(p_1)f_1'(p_1) @ \cdots @ \gamma_+(p_n)f_n'(p_n) @ \gamma_+(q_1)(g_1'(q_1) @ \cdots @ \gamma_+(q_m)Cg_m'(q_m) \otimes \tilde{\rho}_{p_a,q_b}(\alpha_\Lambda(\delta_\phi))\tilde{\Omega}_{p_a,q_b}, \tag{5.2}$$

where $f_a'(p_a) = S(\Lambda)f_a(\Lambda^{-1}p_a)$, $g_b'(q_b) = S(\Lambda)g_b(\Lambda^{-1}q_b)$,

and $\Lambda \to S(\Lambda)$ is the usual Dirac spinor representation of the Lorentz group. Observe that $\tilde{\rho}_{p_a,q_b}$ can be realized in the representation space of ρ_0 as the representation

$$\delta_\phi \to \exp[iB(\eta_{p_a,q_b},\phi)]\rho_0(\delta_\phi), \quad \phi \in N_0,$$

so that

$$\tilde{\rho}_{p_a,q_b}(\alpha_\Lambda(\delta_\phi))\tilde{\Omega}_{p_a,q_b} = \exp[iB(\eta_{\Lambda^{-1}p_a,\Lambda^{-1}q_b},\phi)]V_\Lambda\rho_0(\delta_\phi)\Omega_0,$$

where $\Lambda \to V_\Lambda$ implements α_Λ in the Hilbert space carrying ρ_0. From this it follows immediately that (5.2) extends to a unitary operator on B and implements the Lorentz group automorphisms of $\Delta_c(N_0)$.

Now recall the discussion of Sec. 1. If B is to carry a representation of an observable algebra, then the gauge group T_0 must be represented by scalars. To see that this is the case, restrict $\tilde{\rho}$ to $\Delta_c(T_0)$. Observe that for each $p_1,\cdots,p_n,q_1,\cdots,q_m$, $\tilde{\rho}_{p_a,q_b}$ restricts to the character

$$\delta_\psi \to \exp[l_{n,m}(\psi)], \quad \psi \in T_0, \tag{5.3}$$

where

$$l_{n,m}(\psi) = -(n-m)e\int[\overline{\chi(k)} - \chi(k)]d^3\mathbf{k}/2\omega, \tag{5.4}$$

and $\psi_\mu(k) = k_\mu\chi(k)$.

Thus $\tilde{\rho}$ restricted to $\Delta_c(T_0)$ is constant on the subspaces on which $(n-m)$ is constant. Now charge sectors can be introduced in the obvious way, namely, we say that a subspace of $\mathscr{A}$ or B is the charge sector of charge q if it is a maximal closed subspace on which $(n-m)e = q$. Hence $\tilde{\rho}$ restricted to $\Delta_c(T_0)$ is constant on each charge sector. When $n = m = 0$ observe that $l_{0,0} = 0$, that is, $\tilde{\rho}_{p_a,q_b} \upharpoonright \Delta_c(T_0) = \rho_0 \upharpoonright \Delta_c(T_0)$, and so on the zero charge sector $\tilde{\rho}$ defines a representation of $\Delta_c(N_0)/\ker\rho_0$. For all $n \neq m$ the kernels of the characters determined by $l_{n,m}$ are all equal, and so the kernels of the representations $\tilde{\rho}_{p_a,q_b}$ $(a = 1,\cdots,n, b = 1,\cdots,m, m \neq n)$ are all equal to some ideal, say J, of $\Delta_c(N_0)$. Thus in a nonzero charge sector $\tilde{\rho}$ defines a representation of the algebra $\Delta_c(N_0)/J$.

We interpret this to imply that the algebra of observables in a nonzero charge sector must be $\Delta_c(N_0)/J$, which differs from the zero charge sector observable algebra $\Delta_c(N_0)/\ker\rho_0$. Note that $l_{n,m}$ is invariant under the Lorentz group automorphisms of $\Delta_c(T_0)$ so that the ideal J is also Lorentz invariant. That is, the Lorentz group acts as automorphisms of $\Delta_c(N_0)/J$ in the obvious way. Finally, using (5.4), we see that in terms of the potential A_μ, this choice of observable algebra amounts to setting

$$\partial^\mu A_\mu(x) = -(n-m)eD(x),$$

which is the result we wished to establish.

We conclude with the observation that $\Delta_c(N_0)/J$ and $\Delta_c(N_0)/\ker\rho_0$ are "connected" by an automorphism of $\Delta_c(N_0)$. Namely, if we let

$$\psi'_{n,m}(k) = -i(n-m)ek^\mu/k_0^2,$$

then $B(\psi_{n,m},\psi) = l_{n,m}(\psi)$ for $\psi \in T_0$, and the automorphism $\psi_{n,m} : \delta_\phi \to \exp[iB(\psi_{n,m},\phi)]\delta_\phi$ of $\Delta_c(N_0)$ has the property that

$$\Delta_c(N_0)/J \simeq \rho_0(\gamma_{n,m}(\Delta_c(N_0)))$$

(cf. Refs. 8 and 15).

Remark: If we were to carry through the construction of this section with the cutoff function η_{f,p_a,q_b}, then it is not hard to see that (5.2) still implements Lorentz transformations. Translations are also implemented (but depend on f) so the Hilbert space carries a representation of the Poincaré group. Unfortunately the representation of $\Delta_c(N_0)$ is difficult to describe. However, suppose the cutoff is removed by choosing an increasing sequence $\{f_n\}$ of cutoff functions which are C^∞ functions of compact support, such that $f_n \to 1$ as $n \to \infty$. Then if $\mathscr{O}$ is a region of space–time such that $k \to f_n(k\cdot p) = 1$ for $k \in \mathscr{O}$ (p fixed), then it is not hard to see that the states on $\Delta_c(N_0(\mathscr{O}))$ determined by η_{f_n,p_a,q_b} restrict on $\Delta_c(T_0(\mathscr{O}))$ to the characters given by (5.4) for p_a, q_b in some bounded region of momentum space. It is therefore tempting to suggest that as $n \to \infty$, these cutoff states converge to a state on $\Delta_c(N_0)$ which restricts on $\Delta_c(T_0)$ to (5.3). Although we have not carried through this analysis we believe this suggestion to be correct, thus lending support to the relation

$$\partial^\mu A_\mu(x) = -qD(x)$$

as the appropriate subsidiary condition in a sector of charge q.

6. CONCLUSIONS

We have presented here some applications of our previously developed formalism. They suggest that a definite metric axiomatic approach to quantum electrodynamics is possible, and give some indication of how one might begin to modify the work of Doplicher, Haag, and Roberts[8] to accommodate gauge groups of the second kind. We have seen that the supplementary condition $\partial^\mu A_\mu(x) = 0$ does not hold in nonzero charge sectors. The next step in this program is to produce an axiomatic framework in which properties of the gauge group T_0 and charge sectors, analogous to those obtained here, can be proved directly for quantum electrodynamics itself.

APPENDIX

The following question was raised by the referee. If $\partial^\mu A_\mu(x) = -qD(x)$, then Maxwell's equations $\partial^\mu F_{\mu\nu} = j_\nu$ must be replaced by

$$\partial^\mu F_{\mu\nu} = j_\nu - \partial_\nu(\partial^\mu A_\mu), \tag{*}$$

and the "compensating current" $q\partial_\nu D(x)$, if due to charges "at infinity," has the peculiar property of being strongly localized at the origin. What is the explanation of this peculiarity?

The current $q\partial_\nu D(x)$, arises from the choice of the functions η_{p_a,q_b} whose localized nature is explained by Zwanziger[30] in great detail (Ref. 30, p. 3484). This point was also touched on by Feynman[31] see (Ref. 31, footnote p. 445). We

refer the reader to these papers rather than repeat the discussion here. We emphasize that one of our aims in this paper was to demonstrate that Zwanziger's arguments[13] and the possibility $\partial^\mu A_\mu \neq 0$ fit naturally into our algebraic framework, this being independent of the specific condition $\partial^\mu A_\mu(x) = -qD(x)$ which arose from the theoretical description of a particular physical process, namely the traditional scattering experiment.

The remainder of this Appendix is devoted to explaining how the analysis of this paper might be expected to carry over to QED itself in an attempt to clarify the implications of the referee's question.

To begin with, one needs a definite metric axiomatic approach to quantum electrodynamics which could accommodate our quantization of the free field and the equations (*). In such an approach we would be given a Hilbert space H on which would be defined electron/positron fields, a current j_μ, and *self-adjoint* operator-valued distributions A_μ representing the electromagnetic potential. (An indefinite metric version of what we have in mind appears in Ref. 32.) These fields would be related by $\Box A_\mu = j_\mu$, and as the operators $\partial^\mu A_\mu(f)$ are not zero we would have (*). Now construct the algebra $\mathcal{F}$ generated by the smeared electron/positron fields and the operators $\exp iA_\mu(f^\mu)$ as f^μ ranges over the test function space. This is a *field algebra* in the sense of Ref. 8.

The vectors in H do not define physical states of the system nor does the $F^{\mu\nu}$ appearing in (*) define the observable electromagnetic field. (This point can be understood in the context of the indefinite metric formalism of Ref. 32 where Maxwell's equations hold only on a subspace of the space on which the potential and current are defined.)

Following Bucholz[33] it seems probable that in this approach a scattering theory for the potential A_μ can be rigorously established. Assuming this to be the case there will exist asymptotic fields $A_\mu^{\text{out}}(x)$. Now one can define the C^*-algebra $\mathcal{T}$ generated by $\exp[i\partial^\mu A_\mu^{\text{out}}(f)]$ as f ranges over the test function space. If it could be established that $\mathcal{T}$ is Abelian and that the operators $\exp i\partial^\mu A_\mu^{\text{out}}(f)$ define gauge automorphisms of the asymptotic fields, then we would have all the structure necessary to carry through the analysis of Sec. 1.

Namely, decompose H as a direct integral over the spectrum of $\mathcal{T}$ and thereby obtain a direct integral of representations of $\mathcal{F} \cap \mathcal{T}'$. Assuming the time evolution is gauge invariant (that is, lies in $\mathcal{T}'$) this decomposition would persist in the interaction. The points of the spectrum of $\mathcal{T}$ would define superselection sectors and as before we would try to link these with the charge sectors. It is unlikely that a relation as explicit as $\partial^\mu A_\mu^{\text{out}}(x) = -qD(x)$ could be proved, nevertheless we expect the points of the spectrum of $\mathcal{T}$ to be interpretable as classical current distributions (see Zwanziger's arguments in Ref. 13). Given that a charge sector can be defined in this context we could proceed to determine the form of Maxwell's equations in each charge sector. We would expect on the basis of this paper and Ref. 13 something of the form

$$\partial^\mu F_{\mu\nu}(x) = j_\nu(x) + f_\nu^{\text{out}}(x),$$

where $f_\nu^{\text{out}}(x)$ is a c-number current. Of course one might try to include $f_\nu^{\text{out}}(x)$ in the source $j_\nu(x)$, so that Maxwell's equations are satisfied. However, perturbation theory suggests that this expedient would produce infrared divergences. The point is that scattering theory is an idealized situation, the sources specfied by the points of the spectrum of $\mathcal{T}$, being c-numbers, are never dynamic and so must relate to accumulations of sources which are asymptotic (i.e., effect but are not effected, cf. Ref. 31). As such they describe that part of the system being left out of the dynamical discussion, namely the long-range Coulomb effect of the asymptotic charged particles, which, in electrodynamics, cannot be neglected.

ACKNOWLEDGMENTS

We would like to thank A. Truman, M.A. Gerzon, and J.E. Roberts for comments and remarks which proved critical, at various stages of this work, to its eventual outcome. In particular we would like to thank J.E. Roberts for constructive criticism and suggestions. Finally we are indebted to Professor D. Kastler and Professor D.W. Robinson for an invitation for one of us (A.L.C.) to visit Marseille in June 1977 where many of the ideas discussed here had their origins.

One of us (A.L.C.) gratefully acknowledges the financial support of a Rothman's Fellowship and a Queen Elizabeth II Fellowship.

[1] A.L. Carey, J.M. Gaffney, and C.A. Hurst, J. Math. Phys. **18**, 629–40 (1977).
[2] A.L. Carey and C.A. Hurst, J. Math. Phys. **18**, 1553–61 (1977).
[3] A.L. Carey, J.M. Gaffney, and C.A. Hurst, to appear in Rep. Math. Phys.
[4] R.F. Streater and A.S. Wightman, PCT, *Spin and Statistics and All That* (Benjamin, New York, 1964).
[5] E. Fermi, Atti. Accad. Lincei **9**, 881 (1929); **12**, 431 (1930); Rev. Mod. Phys. **4**, 125 (1932).
[6] L. Garding and A.S. Wightman, Ark. Fys. **28**, 129–34 (1964).
[7] R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1969).
[8] S. Doplicher, R. Haag, and J.E. Roberts, Commun. Math. Phys. **13**, 1 (1969)
[9] R.F. Streater and I. Wilde, Nucl. Phys. B **24**, 561 (1970).
[10] D. Shale, J. Math. Phys. **2**, 33 (1961).
[11] J.M. Cook, J. Math. Phys. **3**, 915 (1962).
[12] P.P. Kulish and L.D. Faddeev, Theor, Math. Phys. **4**, 745 (1971).
[13] D. Zwanziger, Phys. Rev. D **14**, 2570 (1976).
[14] R. Ferrari, L.E. Picasso, and F. Strocchi, Nuovo Cimento A **39**, 1–8 (1977).
[15] S. Doplicher, R. Haag, and J.E. Roberts, Commun. Math. Phys. **15**, 173 (1969).
[16] A.O. Barut and R. Raczka, Ann. Inst. Henri Poincaré, **17**, 111 (1972).
[17] J. Manuceau, Ann. Inst. Henri Poincaré, **8**, 139–61 (1968).
[18] G.G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).
[19] J.M. Gaffney, Ph.D. thesis Adelaide, (1974).
[20] G. Roepstorff, Commun. Math. Phys. **19**, 301–14 (1970).
[21] T.W.B. Kibble, J. Math. Phys. **9**, 315–24 (1968).

[22]S. Doplicher, D. Kastler, and D.W. Robinson, Commun. Math. Phys. **3**, 1–28 (1966).
[23]T.W.B. Kibble, Phys. Rev. **175**, 1624 (1968).
[24]V. Chung, Phys. Rev. B **140**, 1110 (1965).
[25]J. Wright, unpublished.
[26]P.J.M. Bongaarts, in *Mathematics of Contemporary Physics*, edited by R.F. Streater (Academic, New York, 1972).
[27]P.J.M. Bongaarts, Ann. Phys. **56**, 108–39 (1970).
[28]J. Bjorken and S. Drell, *Relativistic Quantum Mechanics* (McGraw–Hill, New York, 1965).
[29]J.M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, New York, 1976), 2nd ed.
[30]D. Zwanziger, Phys. Rev. D **11**, 3481 (1975).
[31]R.P. Feynman, Phys. Rev. **80**, 440 (1950).
[32]F. Strocchi and A.S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[33]D. Bucholz, Commun. Math. Phys. **52**, 147–73 (1977).

Formal analytic continuation of Gel'fand's finite dimensional representations of gl(n,C)[a]

F. Lemire[b] and J. Patera

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada
(Received 5 September 1978)

The article contains three results: I. It is shown that among the $2^{-n}\ n!(n+1)!$ discrete series of representations of the Lie algebra gl(n,C) of complex $n\times n$ matrices described in the literature, the majority are not representations at all. Thus for $n=3$ and 4 one has respectively 12 and 45 series of representations instead of 18 and 180. II. In addition to the $p+1$ discrete unitary series of representations of u(p,q) [the Lie algebra of the group U(p,q), $p\geq q$, and $p+q=n$] there exist other discrete series of gl(n,C) which become unitary when restricted to its real subalgebra u(p,q). For $n=3$ there are four such series all corresponding to the chain u(2,1) $\supset$ u(1,1) $\supset$ u(1); for $n=4$ there exist six such series for u(3,1) and four series for u(2,2). Furthermore, some of the gl(n,C) series whose restriction to the real case do not provide unitary representations in general, do contain (infinitely many) particular representations which are unitary. Such unitary representations are contained inside of two of the four series for $n=3$ and inside of seven of the 27 series for $n=4$. III. Some properties of indecomposable representations of the Lie algebras for the groups of inhomogeneous transformations are shown using the discrete series of gl(n,C).

I. INTRODUCTION

The explicit representation theory as given by Gel'fand and Tseitlin for the groups U(n) and O(n) and later extended by Gel'fand and Graev to some representation of Gl(n,C) is undoubtedly the most suitable form of the theory for extensive computations. It is therefore important to investigate the limits of its validity. In the present paper we are concerned with this question.

In a paper by Gel'fand and Tseitlin[1] every finite dimensional irreducible representation of gl(n,C) is described by labeling the basis vectors and giving explicit formulas for the representatives of a generating set of gl(n,C). In a supplement to a later paper[2] on this subject Gel'fand and Graev present a systematic study of formal analytic continuations of both the labeling and the generating operators of these representations. The goal of this paper is threefold. First we shall show that, contrary to the claim of Gel'fand and Graev, a sizeable proportion of these analytic continuations are not representations of gl(n,C). In fact we give a necessary condition to these operators to satisfy the commutation relations of gl(n,C). Secondly, we consider the restrictions of these representations to the real forms u(p,q) of gl(n,C), determining in particular all these restrictions which are discrete unitary irreducible representations of u(p,q). Finally, we shall indicate by an example some further applications of these formal Gel'fand representations to determine interesting classes of explicitly defined representations of certain subalgebras of gl(n,C). In order to make this paper as self-contained as possible, we begin with a brief outline of the material presented in the supplement to Ref. 2.

As is well known, every finite-dimensional irreducible representation of the Lie algebra gl(n,C) of the group GL(n,C) of all nonsingular $n\times n$ complex matrices is specified by a set of n integers $m_{1n}\geqslant\cdots\geqslant m_{nn}$. The representation space H has an orthonormal basis labeled by all possible triangular arrays (patterns) of integers,

$$m=\begin{pmatrix} m_{1n} & & m_{2n} & & \cdots & & m_{nn} \\ & m_{1,n-1} & & m_{2,n-1} & & \cdots\ m_{n-1,n-1} & \\ & & & \cdots & & & \\ & & & m_{11} & & & \end{pmatrix}, \tag{1}$$

where the components m_{ij} satisfy the inequalities

$$m_{ij}\geqslant m_{i,j-1}\geqslant m_{i+1,j} \quad \text{for } i<j\leqslant n.$$

The algebra gl(n,C) consists of all complex $n\times n$ matrices and has the standard basis $\{e_{ij}|i,j=1,2,\cdots,n\}$, where e_{ij} denotes the $n\times n$ complex matrix with 1 at the intersection of the ith row and jth column and zeroes elsewhere. The commutation product of this algebra is then given by

$$[e_{ij},e_{kl}]=\delta_{jk}e_{il}-\delta_{il}e_{kj}, \tag{2}$$

[a]Work supported in part by the National Research Council of Canada and by the Ministére du l'Education de Québec.
[b]Permanent address: Department of Mathematics, University of Windsor, Windsor, Ontario, Canada.

 0022-2488/79/050820-10$01.00

where δ is the Kronecker delta function. Thus in order to specify a representation of gl(n,C) on H, it suffices to define linear operators E_{ij} (representative of e_{ij}) on H satisfying the commutation relations (2). In fact, it actually suffices to define only the representatives E_{kk}, $E_{k,k-1}$ and $E_{k-1,k}$, since the other operators can be obtained from these.

Gel'fand and Tseitlin give explicit formulas for these generating operators as follows: For any basis vector $\xi(m)$, where m is the label given in (1), we have

$$E_{kk}\xi(m) = (r_k - r_{k-1})\xi(m), \tag{3}$$

where $r_k = m_{1,k} + \cdots + m_{kk}$ for $k = 1,2,\cdots,n$ and $r_0 = 0$,

$$E_{k,k-1}\xi(m) = a^1_{k-1}\xi(m^1_{k-1}) + \cdots + a^{k-1}_{k-1}\xi(m^{k-1}_{k-1}), \tag{4}$$

where m^j_{k-1} denotes the array obtained from m replacing $m_{j,k-1}$ by $m_{j,k-1} - 1$,

$$a^j_{k-1} = \left[- \frac{\Pi^k_{i=1}(m_{ik} - m_{j,k-1} - i + j + 1)\Pi^{k-2}_{i=1}(m_{k,k-2} - m_{j,k-1} - i + j)}{\Pi_{i\neq j}(m_{i,k-1} - m_{j,k-1} - i + j + 1)(m_{i,k-1} - m_{j,k-1} - i + j)} \right]^{1/2},$$

$$E_{k-1,k}\xi(m) = b^1_{k-1}\xi(m^1_{k-1}) + \cdots + b^{k-1}_{k-1}\xi(m^{k-1}_{k-1}), \tag{5}$$

where m^j_{k-1} denotes the array obtained from m replacing $m_{j,k-1}$ by $m_{j,k-1} + 1$, and

$$b^j_{k-1} = \left[- \frac{\Pi^k_{i=1}(m_{ik} - m_{j,k-1} - i + j)\Pi^{k-2}_{i=1}(m_{i,k-2} - m_{j,k-1} - i + j - 1)}{\Pi_{i\neq j}(m_{i,k-1} - m_{j,k-1} - i + j)(m_{i,k-1} - m_{j,k-1} - i + j - 1)} \right]^{1/2}.$$

A detailed derivation of these formulas as well as an effective description of the operators of irreducible finite dimensional representations of the groups GL(n,C) and U(n) is given in the paper of Gel'fand and Graev.[2] (Another derivation of these results is given by Baird and Biedenharn.[3])

In a supplement to the Gel'fand and Graev paper the labeling patterns given in (1) are slightly altered and the generating operators E_{kk}, $E_{k,k-1}$, and $E_{k-1,k}$ defined to operate on the new basis vectors in the following way: To each $k = 1,2,\cdots,n-1$ we assign a pair of integers $\{i_k,i'_k\}$ where

$$i_k \in \{0,1,\cdots,k\}, i'_k \in \{1,2,\cdots,k+1\}, \text{and } i_k < i'_k. \tag{6}$$

For each such set of indices one defines a Hilbert space $H\{i_k,i'_k\}$ having an orthonormal basis labeled by the set of all possible triangular arrays of integers where the top row is fixed and the other components satisfy the following set of inequalities,

$$\begin{aligned}
&(1)\quad m_{jk} \geqslant m_{j+1,k}, && \text{for } j < k \leqslant n,\\
&(2)\quad m_{j-1,k+1} + 1 \geqslant m_{jk} \geqslant m_{j,k+1} + 1, && \text{for } j \leqslant i_k,\\
&(3)\quad m_{j,k+1} \geqslant m_{jk} \geqslant m_{j+1,k+1}, && \text{for } i_k < j < i'_k,\\
&(4)\quad m_{j+1,k+1} - 1 \geqslant m_{jk} \geqslant m_{j+2,k+1} - 1, && \text{for } j \geqslant i'_k
\end{aligned} \tag{7}$$

(by convention we set $m_{0,k+1} = +\infty$ and $m_{k+2,k+1} = -\infty$). The original finite dimensional space corresponds to the case where $i_k = 0$ and $i'_k = k+1$ for $k = 1,2,\ldots,n-1$. All other spaces defined above are infinite dimensional. The operators E_{kk}, $E_{k,k-1}$, and $E_{k-1,k}$ are then defined on $H\{i_k,i'_k\}$ by the same formulas (3)–(5), as in the case of the finite dimensional representation, on noting that the argument of the coefficients a^j_{k-1} and b^j_{k-1} are taken to be $(\pi/2)N$, where N is the common number of negative factors under the radial signs in the expressions for a^j_{k-1} and b^j_{k-1}. It is clear that these operators map any basis vector of $H\{i_k,i'_k\}$ into a finite linear combination of basis elements of the same space $H\{i_k,i'_k\}$. In a formal sense these operators on $H\{i_k,i'_k\}$ represent analytic continuations of the finite dimensional operators. It is claimed that in this manner, for each set of indices $\{i_k,i'_k\}$, one obtains a series of irreducible representations of the algebra gl(n,C) on the Hilbert space $H\{i_k,i'_k\}$. In Sec. II we show that this claim is false for a sizeable proportion of these sets of indices.

Leaving this problem for the moment, we now outline a second set of results in the supplement to the Gel'fand and Graev paper which we wish to expand upon. For any fixed nonnegative integers p and q with $p \geqslant q$ and $p + q = n$, we denote by U(p,q) the group of all $n \times n$ complex matrices which preserve the Hermitian form

$$|x_1|^2 + |x_2|^2 + \cdots + |x_p|^2 - |x_{p+1}|^2 - \cdots - |x_n|^2.$$

Since U(p,q) is one of the real forms of the group GL(n,C), its Lie algebra u(p,q) is a real form of gl(n,C). A representation of the algebra u(p,q) is then said to be *unitary* iff the generators of the representation are all skew-Hermitian. Gel'fand and Graev show that among the representations of gl(n,C) defined above, there exist $p+1$ series of irreducible unitary representations of u(p,q). This result follows by selecting a basis of u(p,q) consisting of

$$\begin{aligned}
&ie_{kk}, && \text{for } k = 1,2,\cdots,n,\\
&e_{kl} - e_{lk},\ i(e_{kl} + e_{lk}), && \text{for } k,l \leqslant p \text{ or } k,l > p,\\
&e_{kl} + e_{lk},\ i(e_{kl} - e_{lk}) && \text{for } k > p \text{ and } l \leqslant p,
\end{aligned} \tag{8}$$

and observing that the representatives of these basis elements on the space $H\{i_k,i'_k\}$ are skew-Hermitian iff we have

$$(1)\ E^+_{kk} = E_{kk}, \quad \text{for } k = 1,2,\cdots,n,$$

and

$$(2)\ E^{+}_{k,k-1} = \begin{cases} E_{k-1,k}, & \text{for } k \neq p, \\ -E_{p-1,p}, & \text{for } k = p. \end{cases} \tag{9}$$

These conditions are equivalent to requiring that the coefficients a^j_k are real for all $k \neq p$ and all $j = 1,2,...,k$, and that a^j_p are purely imaginary for $j = 1,2,...,p$. By inspection the only sets of indices $\{i_k,i'_k\}$ satisfying these conditions are the following:

$$\begin{aligned} &i_k = 0, \quad i'_k = k+1, \quad \text{for } k < p, \\ &i_p = l \in \{0,1,2,...,p\}, \quad i'_p = l+1, \\ &i_k = l, \quad i'_k = l+k-p+1, \quad \text{for } k > p. \end{aligned} \tag{10}$$

In each of these $p+1$ cases one observes that the labeling formalism yields an explicit description of the branching rule for the canonical chain of subalgebras $u(p,q) \supset u(p,q-1) \supset \cdots \supset u(p,0) \supset \cdots \supset u(1)$. Note that the operators E_{ij} for $i,j \leqslant \mu$ leave the top $n-\mu+1$ rows of the labels invariant, and hence the subalgebra of $u(p,q)$ consisting of linear combinations of operators from $\{E_{ij} | i,j \leqslant \mu\}$ can be viewed as operating only on the bottom $\mu - 1$ rows of the labeling arrays. It is in this sense that we associate a chain of subalgebras of $u(p,q)$ with the given basis labels.

In Sec. III we show that the restriction of other representations of $gl(n,C)$ to $u(p,q)$ also provides discrete series of unitary irreducible representations of the Lie algebra $u(p,q)$, which in some cases correspond to different chains of subalgebras.

II. A NECESSARY CONDITION FOR EXISTENCE OF A DISCRETE SERIES OF gl(*n,C*)

In this section we show that for certain sets of indices $\{i_k,i'_k\}$, allowed by condition (6), the generators E_{kk}, $E_{k,k-1}$, and $E_{k-1,k}$ defined on $H\{i_k,i'_k\}$ do *not* provide representations of the algebra $gl(n,C)$. In fact, we claim that a necessary condition for these operators to satisfy the commutation relations (2) and hence provide a representation of $gl(n,C)$ on $H\{i_k,i'_k\}$ is that for each $k = 2,3,...,n-1$ we have

$$i_{k-1},i'_{k-1} \notin [1,i_k - 1] \cup [i_k + 1,i'_k - 2] \cup [i'_k,k-1]. \tag{11}$$

Here $[a,b]$ denotes the set of integers $\{a,a+1,\cdots,b\}$; by convention $[a,b] = \emptyset$ if $b < a$.

To illustrate this condition consider the case of $n = 3$. Gel'fand and Graev claim that for each fixed set of integers $m_{13} \geqslant m_{23} \geqslant m_{33}$ their construction yields 18 inequivalent irreducible representations of $gl(3,C)$; one representation for each of the 18 different sets of indices (see Table I). Of these 18 sets of indices, six do not satisfy condition (11). To be specific, consider the particular case $(i_2,i'_2) = (0,3)$ and $(i_1,i'_1) = (0,1)$. These values are clearly within the range allowed by (6), but forbidden by (11) because $i'_1 = 1 \in [i_2 + 1, i'_2 - 2] = [1,1]$. From Table I one reads off the inequalities imposed on the elements of an array belonging to the space $H\{(0,3),(0,1)\}$. Namely, for any fixed integers $m_{13} \geqslant m_{23} \geqslant m_{33}$ one has

$$m_{13} \geqslant m_{12} \geqslant m_{23}, \quad m_{23} \geqslant m_{22} \geqslant m_{33}, \quad m_{22} - 1 \geqslant m_{11}. \tag{12}$$

Consider the vector

$$m = \begin{pmatrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{23} & \\ & & m_{23} - 1 & & \end{pmatrix} \in H\{(0,3),(0,1)\},$$

TABLE I. Gel'fand representations of $gl(3,C)$ and their restrictions.

(i_2,i'_2) \ (i_1,i'_1)	(0,2) $m_{12} \geqslant m_{11} \geqslant m_{22}$	(0,1) $m_{22} - 1 \geqslant m_{11}$	(1,2) $m_{11} \geqslant m_{12} + 1$
(0,3)≡ $m_{11} \geqslant m_{12} \geqslant m_{23}$ $m_{23} \geqslant m_{22} \geqslant m_{11}$	Unitary $su(3) \supset su(2)$	Not a representation	Not a representation
(0,2)≡ $m_{11} \geqslant m_{12} \geqslant m_{23}$ $m_{23} - 1 \geqslant m_{22}$	Unitary if $m_{11} = m_{23}$ $su(2,1) \supset su(2)$	Unitary $su(2,1) \supset su(1,1)$	Unitary $su(2,1) \supset su(1,1)$
(0,1)≡ $m_{23} - 1 \geqslant m_{12} \geqslant m_{11} - 1$ $m_{11} - 1 \geqslant m_{22}$	Unitary $su(2,1) \supset su(2)$	Not a representation	Not a representation
(1,3)≡ $m_{12} \geqslant m_{11} + 1$ $m_{23} \geqslant m_{22} \geqslant m_{11}$	Unitary if $m_{23} = m_{11}$ $su(2,1) \supset su(2)$	Unitary $su(2,1) \supset su(1,1)$	Unitary $su(2,1) \supset su(1,1)$
(1,2)≡ $m_{12} \geqslant m_{11} + 1$ $m_{11} - 1 \geqslant m_{22}$	Unitary $su(2,1) \supset su(2)$	Never unitary	Never unitary
(2,3)≡ $m_{12} \geqslant m_{11} + 1$ $m_{11} + 1 \geqslant m_{22} \geqslant m_{23} + 1$	Unitary $su(2,1) \supset su(2)$	Not a representation	Not a representation

where the elements m_{22} and m_{11} have reached their highest values compatible with (12). Then

$$E_{32}m = a_2^1 m_2^1 = a_2^1 \begin{pmatrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12}-1 & & m_{23} & \\ & & m_{23-1} & & \end{pmatrix}, \tag{13}$$

which differs from zero as long as $m_{12} > m_{23}$. If, however, $m_{12} = m_{23}$, then necessarily $a_2^1 = 0$; otherwise the array m_2^1 on the right of (13) would not satisfy the inequalities (12). Substituting the values $m_{12} = m_{23}$, $m_{22} = m_{23}$, and $m_{11} = m_{23} - 1$ into the expression for a_2^1, one verifies that both the numerator and denominator of a_2^1 contain a factor equal to zero. In order to avoid a contradiction one is forced to define that $a_2^1 = 0$ whenever the numerator contains zero, regardless of the denominator. Indeed, that convention is tacitly adopted in Ref. 2. The point we want to make here is that even then a contradiction is not avoided. It can be shown as follows:

$$(E_{22} - E_{33})\xi(m) = (2m_{12} - m_{13} - m_{33} + 1)\xi(m), \tag{14}$$

and

$$[E_{23},E_{32}]\xi(m) = [(a_2^1)^2 - (b_2^1)^2]\xi(m)$$

$$= \Big((m_{13} - m_{12} + 1)(m_{12} - m_{33} + 1) \frac{(m_{12} - m_{23})(m_{12} - m_{23} + 1)}{(m_{12} - m_{23})(m_{12} - m_{23} + 1)}$$

$$- (m_{13} - m_{12})(m_{12} - m_{33} + 2) \frac{(m_{12} - m_{23} + 1)(m_{12} - m_{23} + 2)}{(m_{12} - m_{23} + 1)(m_{12} - m_{23} + 2)} \Big) \xi(m). \tag{15}$$

Algebraic simplification of the coefficients verifies that $[E_{23},E_{32}]\xi(m) = (E_{22} - E_{33})\xi(m)$. Now, however, if we assume that $m_{12} = m_{23}$, we have

$$(E_{22} - E_{33})\xi(m) = (2m_{23} - m_{13} - m_{33} + 1)\xi(m), \tag{16}$$

and

$$[E_{23},E_{32}]\xi(m) = -(b_2^1)^2\xi(m)$$

$$= \Big(-(m_{13} - m_{23})(m_{23} - m_{33} + 2) \frac{(m_{23} - m_{23} + 1)(m_{23} - m_{23} + 2)}{(m_{23} - m_{23} + 1)(m_{23} - m_{23} + 2)} \Big) \xi(m). \tag{17}$$

[Note that the term $(a_2^1)^2$ of Eq. (15) does not occur in Eq. (17).] Equating the coefficients in (16) and (17), we obtain $(m_{23} - m_{13} - 1)(m_{23} - m_{33} + 1) = 0$, which is impossible, since $m_{13} \geqslant m_{23} \geqslant m_{33}$. We may thus conclude that the operators E_{ij} defined as above on $H\{(0,1),(0,3)\}$ do not provide a representation of the algebra gl(3,C). The problem arises here in the passage from Eq. (15) to Eq. (17) as $m_{12} \to m_{23}$. In particular the coefficient $(a_2^1)^2$ tends to $(m_{13} - m_{23} + 1)(m_{23} - m_{33} + 1) \neq 0$ as $m_{12} \to m_{23}$ whereas in the limiting case of $m_{12} = m_{23}$ this term must vanish due to the constraints on the arrays of integers belonging to $H\{(0,3),(0,1)\}$. This problem can be resolved in the context of this formalism only by insuring that whenever $m_{12} = m_{22}$ we also have $m_{11} = m_{12} = m_{22}$; this condition, translated in terms of the indices, implies that if $(i_2,i_2') = (0,3)$ we must have $(i_1,i_1') = (0,2)$.

In the general case, although the coefficients are much more complicated, we arrive at essentially the same problem. Whenever the set of indices $\{i_k,i_k'\}$ allows an array in which $m_{jk} = m_{j+1,k}$, we must also have that $m_{j-1,k-1} - 1; m_{j,k-1}$ or $m_{j+1,k-1} + 1 = m_{jk} = m_{j+1,k}$ in order to preserve the continuity of the coefficients a_k^i and b_k^j at the boundary values of the arrays belonging to $H\{i_k,i_k'\}$. This condition can easily be translated into condition (11) on the set of indices.

For $n = 3$ one can directly verify that if the set of indices $\{i_k,i_k'\}$ satisfies condition (11) the operators E_{ij} defined on $H\{i_k,i_k'\}$ do provide a representation of the algebra gl(3,C). For n arbitrary, however, we have been able to give a proof of the sufficiency of condition (11) only in certain special cases.

III. DISCRETE SERIES OF UNITARY REPRESENTATIONS OF u(p,q)

As noted earlier, Gel'fand and Graev have shown that the restrictions of certain of these series of irreducible representations of gl(n,C) to its real form u(p,q) yield discrete unitary irreducible representations with the Gel'fand bases corresponding to the chain of subalgebras u(p,q) $\supset$ u($p,q-1$) $\supset \cdots \supset$ u(p,0). In this section we show that it is possible to obtain additional unitary irreducible representations of u(p,q) corresponding to this same chain of subalgebras as well as other chains of subalgebras.

In terms of the Gel'fand basis of the space $H\{(i_k,i_k')\}$, we have $E_{ij}^{T} = E_{ji}$, and moreover, $E_{jj}^{+} = E_{jj}^{T} = E_{jj}$, where T

and + denote transposition and Hermitian conjugation, respectively. We now wish to investigate under what conditions we have $E^{+}_{\mu,\mu-1} = \pm E_{\mu-1,\mu}$. By formula (4) we have $E_{\mu,\mu-1}\xi(m) = a^{1}_{\mu-1}\xi(m^{1}_{\mu-1}) + \cdots + a^{\mu-1}_{\mu-1}\xi(m^{\mu-1}_{\mu-1})$, where the coefficient $a^{j}_{\mu-1}$ is either real or purely imaginary, depending on the set of indices $\{(i_k,i'_k)\}$. (Note this is independent of the particular array.) Thus, if $a^{j}_{\mu-1}$ is real for $j = 1,2,\ldots,\mu-1$ we have $E^{+}_{\mu,\mu-1} = E_{\mu-1,\mu}$ and if $a^{j}_{\mu-1}$ is purely imaginary for $j = 1,2,\ldots,\mu-1$ we have $E^{+}_{\mu,\mu-1} = -E_{\mu-1,\mu}$. By simply counting the number of negative factors in the expression for $a^{j}_{\mu-1}$ we can determine whether $a^{j}_{\mu-1}$ is real or purely imaginary. This determination is then displayed in the following scheme:

	$j \leqslant i_{\mu-1}$	$i_{\mu-1} < j < i'_{\mu-1}$	$i'_{\mu-1} \leqslant j$
$j \geqslant i_{\mu-2}$	real	imaginary	real
$i_{\mu-2} < j \leqslant i'_{\mu-2}$	imaginary	real	imaginary
$i'_{\mu-2} < j$	real	imaginary	real

Using this scheme we can conclude that

(1) If $(i_{\mu-2},i_{\mu-2}) = (l,l')$ and $(i_{\mu-1},i'_{\mu-1}) = (l,l'+1)$, then $E^{+}_{\mu,\mu-1} = E_{\mu-1,\mu}$.

(2) If (a) $(i_{\mu-2},i_{\mu-2}) = (0,l)$ and $(i_{\mu-1},i'_{\mu-1}) = (l,\mu)$, or (b) $(i_{\mu-2},i_{\mu-2}) = (l,\mu-1)$ and $(i_{\mu-1},i'_{\mu-1}) = (0,l+1)$, or (c) $(i_{\mu-2},i_{\mu-2}) = (0,\mu-1)$ and $(i_{\mu-1},i'_{\mu-1}) = (l,l+1)$, then $E^{+}_{\mu,\mu-1} = -E_{\mu-1,\mu}$.

In all other cases one finds that in general, the coefficients of $E_{\mu,\mu-1}\xi(m)$ contain both real and purely imaginary terms and hence $E^{+}_{\mu,\mu-1} \neq \pm E_{\mu-1,\mu}$. If, however, we place particular restrictions on the values of the defining constants $m_{1n},\cdots,m_{nn}$ of the representation it is still possible to have $E^{+}_{\mu,\mu-1} = \pm E_{\mu-1,\mu}$. Later we describe this situation for $n = 3$ and 4.

Consider now any sequence $\epsilon = \{\epsilon_1,\epsilon_2,\ldots,\epsilon_n\}$ where $\epsilon_1 = 1$ and $\epsilon_i = \pm 1$ for $i = 2,3,\ldots,n$. [This sequence can be understood to be the signature of the u(p,q) invariant form.] We shall say that a set of indices $\{(i_k,i'_k)\}$ such that $E^{+}_{\mu,\mu-1} = \epsilon_{\mu-1}\epsilon_\mu E_{\mu-1,\mu}$ for $\mu = 2,3,\ldots,n$ is *compatible* with the sequence $\{\epsilon_1,\ldots,\epsilon_n\}$. If $\epsilon_i = +1$ for all $i = 1,2,\ldots,n$ there is exactly one compatible set of indices, namely $i_k = 0$ and $i'_k = k+1$ for $k = 1,2,\ldots,n-1$. If $\nu < n$ then there are $\nu+1$ distinct sets of indices compatible with the sequence $\{\epsilon_1,\ldots,\epsilon_n\}$ where $\epsilon_1 = \cdots = \epsilon_\nu = +1$ and $\epsilon_{\nu+1} = \cdots = \epsilon_n = -1$. These are precisely the sets of indices (9) considered by Gel'fand and Graev, where one takes $p = \nu$. Finally for each other sequence there are two compatible sets of indices. Note that the sets of indices compatible with any sequence satisfy condition (11).

Choose some $p < n$ and restrict attention now to those sequences which contain either p or $q = n - p$ terms equal to $+1$. To any such sequence we associate a set of generators of a representation of the algebra u(p,q) as follow:

$$\{iE_{\mu\mu} \mid \mu = 1,2,\ldots,n\}$$
$$\cup\{i(E_{\mu,\mu-1} - E_{\mu-1,\mu}), E_{\mu,\mu-1} + E_{\mu-1,\mu} \mid \text{for all } \mu \text{ with } \epsilon_{\mu-1}\epsilon_\mu = -1\},$$
$$\cup\{i(E_{\mu,\mu-1} + E_{\mu-1,\mu}), E_{\mu,\mu-1} - E_{\mu-1,\mu} \mid \text{for all } \mu \text{ with } \epsilon_{\mu-1}\epsilon_\mu = +1\}.$$

For any set of indices $\{i_k,i'_k\}$ compatible with the given sequence each of these generators is a skew-Hermitian operator on the space $H\{(i_k,i'_k)\}$. Therefore, assuming the sufficiency of condition (11), these operators provide a discrete irreducible unitary representation of the algebra u(p,q) on the space $H\{(i_k,i'_k)\}$. If we set p_j (resp. q_j) equal to the number of + 's (resp. − 1's) in the truncated sequence $\{\epsilon_1,\cdots,\epsilon_j\}$, then the Gel'fand basis of this representation space corresponds to the chain of subalgebras $\mathrm{u}(p,q) \supset \mathrm{u}(p_{n-1},q_{n-1}) \supset \cdots \supset \mathrm{u}(1)$.

ble representations of u(p,q) the unitary condition depends solely on the set of indices $\{(i_k,i'_k)\}$ and is valid for all possible choices of the defining integral parameters $m_{1n} \geqslant \cdots \geqslant m_{nn}$, i.e., for all irreducible representations of the series. We now observe that some other sets of indices also yield unitary representations, but only for certain values of the parameters $m_{1n} \geqslant \cdots \geqslant m_{nn}$. For the cases of $n = 3$ and $n = 4$ we have tabulated these additional series of unitary representations by specifying for each such set of indices the restrictions on the values of $m_{1n} \geqslant \cdots \geqslant m_{nn}$ and the sequence to which it is compatible.

To illustrate the results of this section let us again consider the case of $n = 3$. In this case there are four possible sequences (1,1,1), (1,1, − 1), (1, − 1,1) and (1, − 1, − 1), each of which will be treated separately.

The sequence (1,1,1) gives rise to the real subalgebra u(3) of gl(3,C) and the only compatible set of indices is $\{(0,2),(0,3)\}$. Not surprisingly, this set of indices provides us with the unique finite dimensional unitary irreducible representation of u(3).

The sequence (1,1, − 1) gives rise to the real subalgebra u(2,1) of gl(3,C) and there are three compatible sets of indices, namely $\{(0,2),(0,1)\}$, $\{(0,2),(1,2)\}$, and $\{(0,2),(2,3)\}$. The associated representations of these sets of indices are precisely the three discrete unitary irreducible representations of u(2,1) described by Gel'fand and Graev. In the decomposition of each of these representations with respect to the subalgebra u(2,0) generated by $\{iE_{11},iE_{22},i(E_{12}+E_{21}),E_{12}-E_{21}\}$ we observe that the second row components (m_{12},m_{22}) label the infinite number of finite dimensional irreducible unitary u(2) subrepresentations. Thus the Gel'fand basis formalism corresponds to the chain of subalgebras $\mathrm{u}(2,1) \supset \mathrm{u}(2)$.

The sequence (1, − 1,1) also gives rise to the real subalgebra u(2,1) of gl(3,C), and there are two compatible sets of indices $\{(0,1),(1,3)\}$ and $\{(1,2),(0,2)\}$. The Gel'fand bases for the two associated representations corresponds in this case to the chain of subalgebras $\mathrm{u}(2,1) \supset \mathrm{u}(1,1)$. In fact, if we decompose these representations into their irreducible com-

ponents with respect to the subalgebra u(1,1) generated by $\{iE_{11}, iE_{22}, i(E_{12}-E_{21}), E_{12}+E_{21}\}$, we find that the components (m_{12}, m_{22}) of the second row label the infinite dimensional unitary irreducible u(1,1) subrepresentations.

Finally the sequence $(1,-1,-1)$ again gives rise to the real subalgebra u(2,1) and there are two compatible sets of indices $\{(0,1),(0,2)\}$ and $\{(1,1),(1,3)\}$. As in the previous case the Gel'fand labeling corresponds to the chain of subalgebra $u(2,1) \supset u(1,1)$.

A summary of these results for the case $n=3$ is displayed in Table I. a similar analysis of the case $n=4$ has also been carried out with the results given in Tables II(a), II(b), and II(c).

IV. EXAMPLES

In this section we illustrate the results of Secs. II and III by considering the following specific examples where for simplicity we assume that $m_{13}=m_{23}=m_{33}=0$.

A. Space in which operators do not provide a representation of gl(3,C)

The space $H\{(0,1),(0,3)\}$ has a basis consisting of

$$\left\{\xi\begin{pmatrix}0 & & 0\\ & z & \\ & & 0\end{pmatrix}\middle| z\leqslant -1\right\}.$$

Using the formulas (3)–(5) we have

$$(E_{22}-E_{33})\xi\begin{pmatrix}0 & & 0\\ & z & \\ & & 0\end{pmatrix} = -z\xi\begin{pmatrix}0 & & 0\\ & z & \\ & & 0\end{pmatrix},$$

whereas

$$[E_{23},E_{32}]\xi\begin{pmatrix}0 & & 0\\ & z & \\ & & 0\end{pmatrix} = 0.$$

Thus on this space $[E_{23},E_{32}] \neq E_{22}-E_{33}$, i.e., these operators do *not* yield a representation of gl(3,C).

B. Representation of gl(3,C) whose restriction to the real forms u(3) or u(2,1) are not unitary

The space $H\{(1,2),(1,2)\}$ has a basis consiting of

TABLE IIA. Gel'fand representations of gl(4,C) and their restrictions. Assume $(i_1,i_1')=(0,1)$, i.e., $m_{22}-1\geqslant m_{11}$.

(i_2,i_2') / (i_1,i_1')	(0,3) $m_{11}\geqslant m_{12}\geqslant m_{21}$ $m_{21}\geqslant m_{22}\geqslant m_{11}$	(0,2) $m_{11}\geqslant m_{12}\geqslant m_{21}$ $m_{11}-1\geqslant m_{22}$	(0,1) $m_{21}-1\geqslant m_{12}\geqslant m_{11}-1$ $m_{11}-1\geqslant m_{22}$	(1,3) $m_{12}\geqslant m_{11}+1$ $m_{21}\geqslant m_{22}\geqslant m_{11}$	(1,2) $m_{12}\geqslant m_{11}+1$ $m_{11}-1\geqslant m_{22}$	(2,3) $m_{12}\geqslant m_{11}+1$ $m_{11}+1\geqslant m_{22}\geqslant m_{21}+1$
(0,4) $m_{11}\geqslant m_{12}\geqslant m_{21}$ $m_{21}\geqslant m_{22}\geqslant m_{11}$ $m_{11}\geqslant m_{12}\geqslant m_{11}$	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation
(0,3) $m_{11}\geqslant m_{12}\geqslant m_{21}$ $m_{21}\geqslant m_{22}\geqslant m_{11}$ $m_{11}-1\geqslant m_{11}$	Not a representation	Unitary on su(3,1)⊃su(2,1) ⊃su(1,1)	Not a representation	Not a representation	Not a representation	Not a representation
(0,2) $m_{11}\geqslant m_{12}\geqslant m_{21}$ $m_{11}-1\geqslant m_{21}\geqslant m_{11}-1$ $m_{11}-1\geqslant m_{12}$	Not a representation	Not a representation	Not a representation	Unitary on su(2,2)⊃su(2,1) ⊃su(1,1)	Not a representation	Not a representation
(0,1) $m_{21}-1\geqslant m_{12}\geqslant m_{11}-1$ $m_{11}-1\geqslant m_{21}\geqslant m_{11}-1$ $m_{11}-1\geqslant m_{12}$	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation
(1,4) $m_{12}\geqslant m_{11}+1$ $m_{21}\geqslant m_{22}\geqslant m_{11}$ $m_{11}\geqslant m_{12}\geqslant m_{11}$	Not a representation	Not a representation	Not a representation	Unitary on su(3,1)⊃su(2,1) su(1,1)	Not a representation	Not a representation
(1,3) $m_{12}\geqslant m_{11}+1$ $m_{21}\geqslant m_{22}\geqslant m_{11}$ $m_{11}-1\geqslant m_{12}$	Not a representation	Never Unitary	Not a representation	Never unitary	Never unitary	Not a representation
(1,2) $m_{12}\geqslant m_{11}+1$ $m_{11}-1\geqslant m_{22}\geqslant m_{11}-1$ $m_{11}-1\geqslant m_{12}$	Not a representation	Not a representation	Not a representation	Never unitary	Not a representation	Not a representation
(2,4) $m_{12}\geqslant m_{11}+1$ $m_{11}+1\geqslant m_{21}\geqslant m_{21}+1$ $m_{11}\geqslant m_{12}\geqslant m_{11}$	Not a representation	Unitary on su(2,2)⊃su(2,1) ⊃su(1,1)	Not a representation	Not a representation	Not a representation	Not a representation
(2,3) $m_{12}\geqslant m_{11}+1$ $m_{11}+1\geqslant m_{21}\geqslant m_{21}+1$ $m_{11}-1\geqslant m_{12}$	Not a representation	Never unitary	Not a representation	Not a representation	Not a representation	Not a representation
(3,4) $m_{12}\geqslant m_{11}+1$ $m_{11}+1\geqslant m_{21}\geqslant m_{21}+1$ $m_{21}+1\geqslant m_{12}\geqslant m_{11}+1$	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation

TABLE IIB. Gel'fand representations of gl(4,C) and their restrictions. Assume $(i_1,i_1')=(0,2)$, i.e., $m_{12}\geqslant m_{13}\geqslant m_{22}$.

(i_2,i_2') / (i_3,i_3')	(0,3) $m_{13}\geqslant m_{12}\geqslant m_{23}$ $m_{23}\geqslant m_{24}\geqslant m_{34}$	(0,2) $m_{13}\geqslant m_{12}\geqslant m_{23}$ $m_{13}-1\geqslant m_{22}$	(0,1) $m_{23}-1\geqslant m_{12}\geqslant m_{33}-1$ $m_{13}-1\geqslant m_{22}$	(1,3) $m_{12}\geqslant m_{13}+1$ $m_{23}\geqslant m_{22}\geqslant m_{33}$	(1,2) $m_{12}\geqslant m_{13}+1$ $m_{13}-1\geqslant m_{22}$	(2,3) $m_{12}\geqslant m_{13}+1$ $m_{13}+1\geqslant m_{22}\geqslant m_{23}+1$
(0,4) $m_{14}\geqslant m_{13}\geqslant m_{24}$ $m_{24}\geqslant m_{23}\geqslant m_{34}$ $m_{14}\geqslant m_{13}\geqslant m_{44}$	Unitary on su(4)⊃su(3) ⊃su(2)	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation
(0,3) $m_{14}\geqslant m_{13}\geqslant m_{24}$ $m_{24}\geqslant m_{23}\geqslant m_{34}$ $m_{44}-1\geqslant m_{13}$	If $m_{14}=m_{24}=m_{34}$ unitary on su(3,1)⊃su(3) ⊃su(2)	If $m_{14}=m_{24}=m_{34}$ unitary on su(2,2)⊃su(2,1) ⊃su(2)	Not a representation	Not a representation	Not a representation	Unitary on su(3,1)⊃su(2,1) ⊃su(2)
(0,2) $m_{14}\geqslant m_{13}\geqslant m_{24}$ $m_{14}-1\geqslant m_{23}\geqslant m_{44}-1$ $m_{44}-1\geqslant m_{13}$	If $m_{14}=m_{24}$ unitary on su(3,1)⊃su(3) ⊃su(2)	Not a representation	Unitary on su(2,2)⊃su(2,1) ⊃su(2)	Never unitary	Not a representation	Not a representation
(0,1) $m_{24}-1\geqslant m_{13}\geqslant m_{14}-1$ $m_{14}-1\geqslant m_{23}\geqslant m_{44}-1$ $m_{44}-1\geqslant m_{13}$	Unitary on su(3,1)⊃su(3) ⊃su(2)	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation
(1,4) $m_{13}\geqslant m_{14}+1$ $m_{24}\geqslant m_{23}\geqslant m_{14}$ $m_{14}\geqslant m_{13}\geqslant m_{44}$	If $m_{24}=m_{14}=m_{44}$ unitary on su(3,1)⊃su(3) ⊃su(2)	Not a representation	unitary on su(3,1)⊃su(2,1) ⊃su(2)	If $m_{24}=m_{14}=m_{44}$ unitary on su(2,2)⊃su(2,1) ⊃su(2)	Not a representation	Not a representation
(1,3) $m_{13}\geqslant m_{14}+1$ $m_{24}\geqslant m_{23}\geqslant m_{14}$ $m_{44}-1\geqslant m_{13}$	If $m_{24}=m_{14}$ unitary on su(3,1)⊃su(3) ⊃su(2)	Never unitary	Never unitary	Never unitary	Unitary on su(2,2)⊃su(2,1) ⊃su(2)	Never unitary
(1,2) $m_{13}\geqslant m_{14}+1$ $m_{14}-1\geqslant m_{23}\geqslant m_{44}-1$ $m_{44}-1\geqslant m_{13}$	Unitary on su(3,1)⊃su(3) ⊃su(2)	Not a representation	Never unitary	Never unitary	Not a representation	Not a representation
(2,4) $m_{13}\geqslant m_{14}+1$ $m_{14}+1\geqslant m_{23}\geqslant m_{24}+1$ $m_{14}\geqslant m_{13}\geqslant m_{44}$	If $m_{14}=m_{44}$ unitary on su(3,1)⊃su(3) ⊃su(2)	Never unitary	Not a representation	Not a representation	Not a representation	Unitary on su(2,2)⊃su(2,1) ⊃su(2)
(2,3) $m_{13}\geqslant m_{14}+1$ $m_{14}+1\geqslant m_{23}\geqslant m_{24}+1$ $m_{44}-1\geqslant m_{13}$	Unitary on su(3,1)⊃su(3) ⊃su(2)	Never unitary	Not a representation	Not a representation	Not a representation	Never unitary
(3,4) $m_{13}\geqslant m_{14}+1$ $m_{14}+1\geqslant m_{23}\geqslant m_{24}+1$ $m_{24}+1\geqslant m_{13}\geqslant m_{14}+1$	Unitary on su(3,1)⊃su(3) ⊃su(2)	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation

$$\left\{\xi\begin{pmatrix} x & & y \\ & z & \end{pmatrix}\middle|\quad x\geqslant 1; y\leqslant -1 \text{ and } z\geqslant x+1\right\}.$$

By direct calculation one can verify that the operators defined by formulas (3)–(5) provide an irreducible representation of gl(3,C) on this space. However we also observe that

$$\left\langle\xi\begin{pmatrix} x+1 & & y \\ & z & \end{pmatrix}\middle|E_{23}^{+}\middle|\xi\begin{pmatrix} x & & y \\ & z & \end{pmatrix}\right\rangle = -\left(\frac{x(x+1)(x+2)(z-x-1)}{(x-y+1)(x-y+2)}\right)^{1/2} = \left\langle\xi\begin{pmatrix} x+1 & & y \\ & z & \end{pmatrix}\middle|E_{32}\middle|\xi\begin{pmatrix} x & & y \\ & z & \end{pmatrix}\right\rangle$$

and

$$\left\langle\xi\begin{pmatrix} x & & y+1 \\ & z & \end{pmatrix}\middle|E_{23}^{+}\middle|\xi\begin{pmatrix} x & & y \\ & z & \end{pmatrix}\right\rangle = -i\left(\frac{(1-y)(-y)(-y-1)(z-y)}{(x-y+1)(x-y)}\right)^{1/2} = -\left\langle\xi\begin{pmatrix} x & & y+1 \\ & z & \end{pmatrix}\middle|E_{32}\middle|\xi\begin{pmatrix} x & & y \\ & z & \end{pmatrix}\right\rangle.$$

(all factors under the radical signs have been made nonnegative). Thus $E_{32}^{+}\neq\pm E_{32}$, and consequently the restriction of this representation to either u(3) or u(2,1) does not consist solely of skew-Hermitian matrices, i.e., these restrictions are not unitary.

C. Representation of gl(3,C) whose restriction to u(2,1) is unitary but not equivalent to any of those given by Gel'fand and Graev

The space $H\{(0,1),(1,3)\}$ has a basis consisting of

$$\left\{\xi\begin{pmatrix} x & & 0 \\ & z & \end{pmatrix}\middle|\quad x\geqslant 1 \text{ and } z\leqslant -1\right\}.$$

By direct calculation one can see that the linear operators defined by formulas (3)–(5) provide an irreducible representation of gl(3,C) and moreover that the restriction of this representation to u(2,1) is unitary. This representation is however not equivalent to any of the unitary representations of u(2,1) listed by Gel'fand and Graev. This follows immedi-

TABLE IIC. Gel'fand representations of gl(4,C) and their restrictions. Assume $(i_1,i_1') = (1,2)$, i.e., $m_{11} \geqslant m_{12} + 1$.

(i_2,i_2') / (i_1,i_1')	(0,3) $m_{13} \geqslant m_{12} \geqslant m_{23}$ $m_{23} \geqslant m_{22} \geqslant m_{33}$	(0,2) $m_{13} \geqslant m_{12} \geqslant m_{23}$ $m_{23} - 1 \geqslant m_{22}$	(0,1) $m_{13} - 1 \geqslant m_{12} \geqslant m_{23} - 1$ $m_{23} - 1 \geqslant m_{22}$	(1,3) $m_{12} \geqslant m_{13} + 1$ $m_{23} \geqslant m_{22} \geqslant m_{33}$	(1,2) $m_{12} \geqslant m_{13} + 1$ $m_{23} - 1 \geqslant m_{22}$	(2,3) $m_{12} \geqslant m_{13} + 1$ $m_{23} + 1 \geqslant m_{22} \geqslant m_{33} + 1$
(0,4) $m_{14} \geqslant m_{13} \geqslant m_{24}$ $m_{24} \geqslant m_{23} \geqslant m_{34}$ $m_{34} \geqslant m_{33} \geqslant m_{44}$	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation
(0,3) $m_{14} \geqslant m_{13} \geqslant m_{24}$ $m_{24} \geqslant m_{23} \geqslant m_{34}$ $m_{44} - 1 \geqslant m_{33}$	Not a representation	Unitary on su(3,1)⊃su(2,1) ⊃su(1,1)	Not a representation	Not a representation	Not a representation	Not a representation
(0,2) $m_{14} \geqslant m_{13} \geqslant m_{24}$ $m_{34} - 1 \geqslant m_{23} \geqslant m_{44} - 1$ $m_{44} - 1 \geqslant m_{33}$	Not a representation	Not a representation	Not a representation	Unitary on su(2,2)⊃su(2,1) ⊃su(1,1)	Not a representation	Not a representation
(0,1) $m_{24} - 1 \geqslant m_{13} \geqslant m_{34} - 1$ $m_{34} - 1 \geqslant m_{23} \geqslant m_{44} - 1$ $m_{44} - 1 \geqslant m_{33}$	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation
(1,4) $m_{13} \geqslant m_{14} + 1$ $m_{24} \geqslant m_{23} \geqslant m_{34}$ $m_{34} \geqslant m_{33} \geqslant m_{44}$	Not a representation	Not a representation	Not a representation	Unitary on su(3,1)⊃su(2,1) ⊃su(1,1)	Not a representation	Not a representation
(1,3) $m_{13} \geqslant m_{14} + 1$ $m_{24} \geqslant m_{23} \geqslant m_{34}$ $m_{44} - 1 \geqslant m_{33}$	Not a representation	Never unitary	Not a representation	Never unitary	Never unitary	Not a representation
(1,2) $m_{13} \geqslant m_{14} + 1$ $m_{34} - 1 \geqslant m_{23} \geqslant m_{44} - 1$ $m_{44} - 1 \geqslant m_{33}$	Not a representation	Not a representation	Not a representation	Never unitary	Not a representation	Not a representation
(2,4) $m_{13} \geqslant m_{14} + 1$ $m_{14} + 1 \geqslant m_{23} \geqslant m_{34} + 1$ $m_{34} \geqslant m_{33} \geqslant m_{44}$	Not a representation	Unitary on su(3,1)⊃su(2,1) ⊃su(1,1)	Not a representation	Not a representation	Not a representation	Not a representation
(2,3) $m_{13} \geqslant m_{14} + 1$ $m_{14} + 1 \geqslant m_{23} \geqslant m_{24} + 1$ $m_{44} - 1 \geqslant m_{33}$	Not a representation	Never unitary	Not a representation	Not a representation	Not a representation	Not a representation
(3,4) $m_{13} \geqslant m_{14} + 1$ $m_{14} + 1 \geqslant m_{23} \geqslant m_{34} + 1$ $m_{34} + 1 \geqslant m_{33} \geqslant m_{44} + 1$	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation	Not a representation

ately on comparing the weight space decompositions of these representations.

D. Unitary representation from a series which is in general nonunitary

In general, the representations of gl(3,C) associated with the set of indices $\{(0,2),(0,2)\}$ do not restrict to unitary representations of u(3) or u(2,1), however if we take $m_{13} = m_{23} = m_{33} = 0$, then the space $H\{(0,2),(0,2)\}$ has a basis consisting of

$$\left\{\xi\begin{pmatrix} 0 & & y \\ & z & \end{pmatrix}\middle|\quad y \leqslant -1;\ z \leqslant y - 1\right\}$$

and the matrix element a_2^1 and b_2^1 are identically zero. Therefore, by writing out the linear operators on this space one can directly verify that the generators of the representation are skew-Hermitian when restricted to u(2,1) and hence the representation is unitary. Moreover, the Gel'fand basis is associated with the chain of subalgebras u(2,1)⊃u(2) and by considering the weight space decomposition, this representation is not equivalent to any of those specified in Ref. 2.

V. REPRESENTATIONS OF NONSEMISIMPLE SUBALGEBRAS OF gl(n,C)

The explicitly defined Gel'fand representations of gl(n,C) offer possible avenues to study the representations of other subalgebras of gl(n,C) in a detailed manner. As an example consider the group G of inhomogeneous transformations consisting of all 3×3 complex matrices

$$g = \begin{bmatrix} \alpha & 0 \\ z & 1 \end{bmatrix},$$

where $\alpha \in \mathrm{GL}(2,C)$ and $z \in C^2$. The elements of G can be viewed as operating on the two dimensional affine space

$\{(x,y,1) \mid x,y \in C\}$ where $(x,y,1)g = ((x,y)\alpha + z, 1)$. The Lie algebra L of G, considered as a subalgebra of gl(3,C) has a basis given by $\{e_{ij} \mid i = 1,2,3;\ j = 1,2\}$. Consider an arbitrary, finite dimensional representation (ρ, V) of L. When viewed as a representation of the subalgebra gl(2,C), the space V decomposes into a finite direct sum of irreducible representations, say $V = W_1 \oplus \cdots \oplus W_n$. Define two operators $N = \rho(e_{32})$ and $M = \rho((e_{11} - e_{22} + 1)e_{31} + e_{32}e_{21})$ on the space V and note the following properties:

(1) By an extreme vector of V we mean any vector $v \in V$ such that $\rho(e_{12})v = 0$. Since $[\rho(e_{12}), N] = [\rho(e_{12}), M] = 0$, we find that N and M map extreme vectors to extreme vectors.

(2) Denote by $\langle S \rangle$ the gl(2,C) subrepresentation of V generated by a set $S \subseteq V$. Then for any nonzero extreme vector $v \in V$ we have

$$\mathrm{Dim}\langle Nv \rangle = \mathrm{Dim}\langle v \rangle + 1 \quad \text{or } Nv = 0,$$

and

$$\mathrm{Dim}\langle Mv \rangle = \mathrm{Dim}\langle v \rangle - 1 \quad \text{or } Mv = 0.$$

(3) Combining remarks 1 and 2 with the fact that $[N,M] = 0$ we find that N and M are commuting nilpotent linear operators on V. This in turn implies that for any fixed nonzero extreme vector $v \in V$ the set of all nonzero vectors of the form $N^\mu M^\nu v$ is linearly independent. {The operators N and M are clearly simple modifications of the Nagel–Moshinsky operators [cf.(4)].}

Consider now $V_0 = \langle \{N^\mu M^\nu v_1 \mid v_1$ is an extreme vector of W_1 and μ, ν are nonnegative integers$\}\rangle$. It is clear that $W_1 \subseteq V_0$ and V_0 is an L subrepresentation of V. We shall now explicitly construct a representation of L equivalent to V_0 using the Gel'fand representations of gl(3,C).

Let $d+1$ denote the dimension of the space W_1 and let k be the smallest nonnegative integer such that $N^k v_1 \neq 0$ and $N^{k+1} v_1 = 0$. Denote by (ρ_1, U) the restriction to L of the unique finite dimensional irreducible representation of gl(3,C) labeled by the sequence $m_{13} = k + d$, $m_{23} = k$ and $m_{33} = 0$. Let U' equal to the L subrepresentation of U generated by

$$\left\{ \xi(m) \,\middle|\, m = \begin{pmatrix} k+d & & & k & & 0 \\ & k+d-\nu & & & k-\mu & \\ & & k+d-\nu & & & \end{pmatrix}, \text{where } N^\mu M^\nu v_1 = 0 \right\}.$$

We claim that V_0 and U/U' are equivalent L representations. In fact, by the properties of N and M listed above we find that as a gl(2,C) representation

$$V_0 = \sum_{\substack{\mu,\nu \text{ nonnegatice integer} \\ \text{s.t. } N^\mu M^\nu v_1 \neq 0}} \oplus \langle N^\mu M^\nu v_1 \rangle,$$

where $\langle N^\mu M^\nu v_1 \rangle$ is a $(d + 1 + \mu - \nu)$-dimensional irreducible gl(2,C) subrepresentation. The corresponding decomposition of U/U' as a gl(2,C) representation yields

$$U/U' = \sum_{\substack{\mu,\nu \text{ nonnegatice integer} \\ \text{s.t. } N^\mu M^\nu v_1 \neq 0}} \oplus \left\langle \xi \begin{pmatrix} k+d & & & k & & 0 \\ & k+d-\nu & & & k-\mu & \\ & & k+d-\nu & & & \end{pmatrix} + U' \right\rangle,$$

where

$$\left\langle \xi \begin{pmatrix} k+d & & & k & & 0 \\ & k+d-\nu & & & k-\mu & \\ & & k+d-\nu & & & \end{pmatrix} + U' \right\rangle$$

is a $(d + 1 + \mu - \nu)$-dimensional irreducible gl(2,C) subrepresentation. Thus $V_0 \cong U/U'$ as gl(2,C) representations.

To complete the verification of the equivalence of V_0 and U/U' as L representation it suffices to observe that the map ϕ, given by setting

$$\phi(N^\mu M^\nu v_1) = K_{\mu\nu} \xi \begin{pmatrix} k+d & & & k & & 0 \\ & k+d-\nu & & & k-\mu & \\ & & k+d-\nu & & & \end{pmatrix} + U',$$

where the coefficients $K_{\mu\nu}$ are constants defined by the equation

$$\rho_1(e_{32})^\mu \rho_1((e_{11} - e_{22} + 1)e_{31} + e_{32}e_{21}))^\nu \xi \begin{pmatrix} k+d & & & k & & 0 \\ & k+d & & & k & \\ & & k+d & & & \end{pmatrix} + U'$$

$$= K_{\mu\nu}\xi \begin{pmatrix} k+d & & & k & & 0 \\ & k+d-\nu & & & k-\mu & \\ & & k+d-\nu & & & \end{pmatrix} + U',$$

can be extended in a natural fashion to an L representation equivalence between V_0 and U/U'.

It is easily seen that the L representation U/U' and hence V_0 is indecomposable and one can explicitly write out the matrix elements of the representations. Thus we have a very explicit presentation for a wide class of finite dimensional representations of L.

This technique has a wide range of possible applications which would appear to be worthy of further investigation.

ACKNOWLEDGMENT

We would like to express our appreciation to R.T. Sharp for some interesting discussions, in particular relating to the su(2,1)⊃su(1,1) basis labeling problem.

[1]I.M. Gel'fand and M.L. Tseitlin, "Finite Dimensional Representations of the Group of Unimodular Matrices," Dokl. Akad. Nauk. SSSR **71**, 825 (1950), MR 12 #9.
[2]I.M. Gel'fand and M.J. Graev, "Finite Dimensional Irreducible Representations of the Unitary and the Full Linear Groups and Related Special Functions," Izdv. Akad. Nauk. SSSR Ser. Math **29**, 1329 (1965) [AMS Transl. **64**, 116 Ser. 2 (1967)] MR34 #1450.
[3]G.E. Baird and L.C. Biedenharn, "On the representations of the semisimple Lie groups. II," J. Math. Phys. **4**, 1449 (1963).
[4]J.G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); J. Patera, J. Math. Phys. **14**, 279 (1973); J. Patera, R.T. Sharp, and P. Winternitz, Rev. Fis. Mex. **23**, 81 (1974).

The scaling limit of the ϕ^2 field in the anharmonic oscillator[a)]

Dan Marchesin[b)]

Courant Institute of Mathematical Sciences, New York University, New York 10012
(Received 28 June 1978)

We prove that the rescaled and renormalized q^2 process for $H_g = [p^2 + gq^4 + (1-g)q^2]/2$ tends to the Gaussian process for the Harmonic oscillator as g tends to infinity.

I. INTRODUCTION

The double-well potential

$$V(q) = g[q^2 - (1 - 1/g)/2$$

is studied in the scaling limit $g \to \infty$. The associated stochastic process, $q(t)$, has an infinitesimal generator

$$H_g = \frac{1}{2}\left[-\frac{d^2}{dq^2} + V(q)\right],$$

and in this limit, the process decomposes into a tensor product of a Bernouilli process and a Gaussian process. The Bernouilli process labels tunneling ("instantons") between the potential wells, while the Gaussian process labels fluctuations within a well ("spin waves"). It was shown in Ref. 1 that $q(t)$, appropriately scaled, tends to the Bernouilli process, i.e., a spin-$\frac{1}{2}$ Ising model in the real line. A similar result was obtained by Ref. 2 for the limit of a q^4 lattice field theory.

In order to find the Gaussian process, we use test functions which vanish on the Ising part of the measure. Thus we are led to consider functions of q which vanish at the minima of the potential, namely $q^2 - E(q^2)$ or $q[q^2 - E(q^2)]$. These random variables converge as $g \to \infty$ to the standard Gaussian process associated with the harmonic oscillator.

We define for the former the random process at time $t \geqslant 0$

$$Q_g(t) = e^{tH_g} Q_g = e^{-tH_g}(\tfrac{1}{2}g)^{1/2}[q^2 - E(q^2)].$$

We prove in Sec. II–IV that for all $t \geqslant 0$ and N

$$\lim_{g \to \infty} E[Q_g(t_1)Q_g(t_2)\cdots Q_g(t_r)] = E[q(t_1)q(t_2)\cdots q(t_n)],$$

where $q(t) = e^{-tH_0}q$ and $H_0 = \frac{1}{2}(-d^2/dq^2 + q^2)$.

A similar result can be proved along the same lines for

$$Q_g(t) = e^{-tH_g} g^{3/4} 2^{-1/4} q[q^2 - E(q^2)].$$

Thus we see that the complement of the Ising part of the measure consists of two disjoint Gaussian processes corresponding to harmonic oscillators. It should be noted that these processes are exponentially smaller (in the sense of energy levels) than the main Bernouilli part.

The phenomena of a measure decoupling into an Ising and a Gaussian part seems to be quite general. A harder result of this type was proved by Glimm–Jaffe–Spencer for some two-dimensional q^4 quantum field theory models.[3] In that case the result was proved in low temperature ($g \to \infty$) for the field q itself—and the Ising part of the measure is a correction on the main Gaussian part. There are indications that at the critical temperature (i.e., for g_c such that the gap between the two lowest eigenvalues of H_g vanishes) the field q^2 has a Gaussian behavior. In our one-dimensional case the critical temperature and the zero temperature are indentical ($g_c = \infty$), and the Gaussian part becomes a correction on the main Ising process.

Similar results hold for the n-dimensional spherical oscillator $H_g = \frac{1}{2}[-\Delta + gr^4 + (1-g)r^2]$.[4] We omit the statements and the proofs of the corresponding theorems since they are obvious generalizations of our results and methods. (We remark only that in the proofs we use the one-dimensional momentum operators instead of the radial momentum operator.)

II. THE BEHAVIOR OF THE RENORMALIZED AND SCALED q^2 PROCESS

Consider the Hamiltonian operator defined on $\mathcal{S}(\mathbb{R})$,

$$\tfrac{1}{2}[p^2 + gq^4 + (1-g)q^2],$$

adding a constant

$$H_g = \tfrac{1}{2}\{p^2 + g[q^2 - (g-1)/2g]^2\}.$$

(The self-adjointness and other properties of H_g are studied in Ref. 1.)

The potential of this Hamiltonian has two equal minima at $(\pm 1/\sqrt{2})(g-1)/g \simeq (\pm 1/\sqrt{2})$ for large g. The behavior of the eigenvalues and eigenvectors as $g \to \infty$ is studied in Ref. 1.

We rescale it through the change of variables $x = \alpha y$ where $\alpha = [2(g-1)]^{-1/4}$.

Then the eigenvalue problem for $\Omega \in \mathcal{S}(\mathbb{R})$,

$$\frac{1}{2}\left[-\frac{d^2}{dx^2} + g\left(x^2 - \frac{g-1}{2g}\right)^2\right]\Omega(x) = \varepsilon\Omega(x), \quad x \in \mathbb{R},$$

is transformed into

$$\frac{1}{2}\left[-\frac{d^2}{dy^2} + \nu\left(y^2 - \frac{1}{4\nu}\right)^2\right]\Omega(\alpha y) = [2(g-1)]^{1/2}\varepsilon\Omega(\alpha y),$$

[a)]Supported in part by the National Science Foundation under grant PHY76-17191, by the National Research Council of Brazil (CNPq), and by NASA–Goddard Space Flight Center grant No. NSG 5034.

[b)]New York University and Pontificia Universidade Catolica do Rio de Janeiro.

 0022-2488/79/050830-07$01.00

where

$$\nu = g/[2(g-1)]^{3/2}, \quad \nu \to 0 \quad \text{as } g \to \infty.$$

Let us establish the notation

$$H_\nu = \tfrac{1}{2}[p^2 + \nu(q^2 - 1/4\nu)^2] \quad \text{on } \mathscr{S}(\mathbb{R}).$$

We will denote its eigenvalues by $E^0_\nu < E^1_\nu < E^2_\nu \cdots$ and corresponding eigenstates by $\Omega^0_\nu, \Omega^1_\nu, \Omega^2_\nu, \ldots$.

Notice that close to its minima (given by $y_\pm = \pm 1/2\sqrt{\nu}$) the potential behaves like ($y \cong 1/2\sqrt{\nu}$)

$$\nu\left(y^2 - \frac{1}{4\nu}\right)^2 = \nu\left(y - \frac{1}{2\sqrt{\nu}}\right)^2\left(y + \frac{1}{2\sqrt{\nu}}\right)^2 \cong \left(y - \frac{1}{2\sqrt{\nu}}\right)^2.$$

Thus in the translated variables $\tilde{y} = y - 1/2\sqrt{\nu}$ the eigenvalue problem $H_\nu \Psi(\tilde{y}) = E\Psi(\tilde{y})$ becomes asymptotically

$$\tfrac{1}{2}(p^2 + \tilde{y}^2)\Psi = E\Psi.$$

Since the eigenvalues of this harmonic oscillator are given by $n + \frac{1}{2}$, $n = 0,1,2\ldots$ the results of 1 become in this scaling

$$\lim_{\nu\to 0} E^{2j}_\nu = \lim_{\nu\to 0} E^{2j+1}_\nu = \lim_{\nu\to 0} \lambda^j_\nu = j + \tfrac{1}{2}, \quad j = 0,1,2,\ldots, \quad \text{(II.1a)}$$

$$\lim_{\nu\to 0} \|\Omega^{2j}_\nu - \Psi^{j,e}_\nu\|_2 = \lim_{\nu\to 0} \|\Omega^{2j+1}_\nu - \Psi^{j,o}_\nu\| = 0, \quad \text{(II.1b)}$$

where $\Psi^{j,e}_\nu$ and $\Psi^{j,o}_\nu$ are given in (III.27). The eigenfunctions of $H_0 \equiv \frac{1}{2}(p^2 + q^2)$, the harmonic oscillator, satisfying $H_0\Omega^j_0 = E^j_0\Omega^j_0$, $j = 0,1,2,\ldots$ ($E^j_0 \equiv j + \frac{1}{2}$), will be denoted by $\Omega^0_0, \Omega^1_0, \ldots$.

Our problem stems essentially from the desire to understant the behavior of

$$q^2_{\text{ren},\nu} \equiv q^2 - \langle \Omega^0_\nu q^2 \Omega^0_\nu \rangle.$$

Now, it is proved in Ref. 1 that, for our H_ν,

$$\lim_{\nu\to 0} (4\nu)\langle \Omega^0_\nu q^2 \Omega^0_\nu \rangle = 1.$$

So we see that $q^2_{\text{ren},\nu} \simeq q^2 - (4\nu)^{-1}$. Notice that at the "bottom" of the potential wells (say, at the rightmost one) $q \simeq 1/2\sqrt{\nu} + \Delta q$,

$$q^2_{\text{ren},\nu} \simeq q^2 - \frac{1}{4\nu} = \left(q - \frac{1}{2\sqrt{\nu}}\right)\left(q + \frac{1}{2\sqrt{\nu}}\right) \simeq \frac{1}{\sqrt{\nu}}\left(q - \frac{1}{2\sqrt{\nu}}\right).$$

Since the two "wells" become disjoint as $\nu \to 0$ it becomes obvious that (after a rescaling) $q^2_{\text{ren},\nu}$ behaves like a q operator for the harmonic oscillator. This is actually the contents of Theorems 1–3 which follow. We recall that in the original scaling the weak operator limit of $q^2_{\text{ren},\nu}$ as $\nu \to 0$ is zero. Loosely speaking, this is due to the fact that $q^2_{\text{ren},\nu}$ is zero close to the region of highest probability density, $q = \pm 1/2\sqrt{\nu}$.

Through a Legendre transformation, we introduce $Q_\nu = \nu^{1/2} q^2_{\text{ren},\nu}$ and its conjugate momentum P_ν, where $[H_\nu, Q_\nu] = -iP_\nu$.

Theorem 1:

$$\lim_{\nu\to 0} \langle \Omega^{2k}_\nu Q^n_\nu \Omega^{2l}_\nu \rangle = \langle \Omega^k_0 q^n \Omega^l_0 \rangle, \quad \text{for} \quad k,l,n = 0,1,2,\ldots.$$

Theorem 2: For $s_i \geqslant 0$, $i = 1,2,\ldots,m-1$, $m = 1,2,3,\ldots$,

$$\lim_{\nu\to 0} \langle \Omega^0_\nu Q_\nu e^{-s_1 H_\nu} Q_\nu e^{-s_2 H_\nu} \cdots Q_\nu e^{-s_m H_\nu} Q_\nu \Omega^0_\nu \rangle = \langle \Omega^0_0 q e^{-s_1 H_0} q e^{-s_2 H_0} \cdots q \Omega^0_0 \rangle$$

Theorem 3: For $s_i \geqslant 0$ real numbers, $j_i \in \mathbb{N}$, $i = 0,\ldots,m-1$, $m \in \mathbb{N}^*$,

$$\lim_{\nu\to 0} \frac{\partial^{j_1}}{\partial s_1^{j_1}} \frac{\partial^{j_2}}{\partial s_2^{j_2}} \cdots \frac{\partial^{j_m}}{\partial s_m^{j_m}} \langle \Omega^0_\nu Q_\nu e^{-s_1 H_\nu} \times Q_\nu e^{-s_2 H_\nu} \cdots Q_\nu e^{-s_m H_\nu} Q_\nu \Omega^0_\nu \rangle = \frac{\partial^{j_1}}{\partial s_1^{j_1}} \frac{\partial^{j_2}}{\partial s_2^{j_2}} \cdots \frac{\partial^{j_m}}{\partial s_m^{j_m}} \langle \Omega^0_0 q e^{-s_1 H_0} q e^{-s_2 H_0} \cdots q \Omega^0_0 \rangle.$$

For the sake of completeness, we return to the original parameter g. Then $q^2_{\text{ren},g} \equiv q^2 - \langle \Omega^0_g q^2 \Omega^0_g \rangle$, $\lim_{g\to\infty} \langle \Omega^0_g q^2 \Omega^0_g \rangle = \frac{1}{2}$. $\alpha \equiv \alpha(g) = [2(g-1)]^{-1/4}$ has the property $\lim_{g\to\infty} \alpha = 0$. By means of a simple rescaling the theorems above become, for $\hat{H}_g \equiv \alpha(g)^2 H_g$.

Corollary 1:

$$\lim_{g\to\infty} \langle \Omega^{2k}_g [(\tfrac{1}{2}g)^{1/2} q^2_{\text{ren},g}] \Omega^{2l}_g \rangle = \langle \Omega^k_0 q^m \Omega^l_0 \rangle$$

for $k,l,m = 0,1,2,\cdots$.

Corollary 2: For $s_i \geqslant 0$, $i = 1,2,\cdots, m-1$

$$\lim_{g\to\infty} \langle \Omega^0_g (\tfrac{1}{2}g)^{1/2} q^2_{\text{ren},g} e^{-s_1 \hat{H}_g} (\tfrac{1}{2}g)^{1/2} \times q^2_{\text{ren},g} \cdots e^{-s_m \hat{H}_g} (\tfrac{1}{2}g)^{1/2} q^2_{\text{ren},g} \Omega^0_g \rangle = \langle \Omega^0_0 q e^{-s_1 H_0} q e^{-s_2 H_0} \cdots e^{-s_m H_0} q \Omega^0_0 \rangle.$$

Corollary 3 is analogous to Corollary 2 above.

Up to now we were working within the subspace of an even function on $\mathbb{R}$, or equivalently on $[0,\infty)$ with Neumann boundary conditions (reflective barrier) at 0. If we work in the space of odd functions on $\mathbb{R}$, or equivalently on $[0,\infty)$ with Dirichlet boundary conditions at 0, all the results above have similar counterparts, obtained by replacing $\nu^{1/2} q^2_{\text{ren},\nu}$ by $\frac{1}{2} q q^2_{\text{ren},\nu}$.

Finally, we can add a weak "external field" to H_ν, i.e., consider

$$\hat{H}_\nu = H_\nu + c(\nu)x.$$

If $c(\nu) = o(\nu)$, all the results above still hold. An explicit computation using parabolic cylinder functions[1] shows that if $c(\nu) = O(1)$ the results are not valid any more

III. PROOF OF THEOREM 1

Lemma 1: Let

$$H_\nu = \tfrac{1}{2}[p^2 + \nu(q^2 - 1/4\nu)^2], \tag{III.1}$$

$$Q_\nu = \nu^{1/2}(q^2 - \langle \Omega^0_\nu q^2 \Omega^0_\nu\rangle), \tag{III.2}$$

$$P_\nu = \nu^{1/2}(pq + qp). \tag{III.3}$$

Then for all $n\in\mathbb{N}$ there exist polynomials in H_ν, $A_{\nu,n}(H_\nu)$, and $B_{\nu,n}(H_\nu)$, with nonnegative coefficients which are bounded as ν tends to zero, such that

$$\text{(i)}\ \|H_\nu Q^n_\nu v\| \leqslant \|A_{\nu,n}(H_\nu)v\|, \tag{III.4}$$

$$\text{(ii)}\|Q^n_\nu v\| \leqslant \|B_{\nu,n}(H_\nu)v\|, \tag{III.5}$$

where $v\in\mathscr{S}(\mathbb{R})$ may depend on ν.

Proof: The proof is done by induction on n. When $n = 0$ there is nothing to be proved. We assume that the result is true up to $n - 1$. In order to prove that it is true for n, we need the two commutator equalities

$$[H_\nu, Q_\nu] = -iP_\nu, \tag{III.6}$$

$$[P_\nu, Q_\nu] = -i4\nu^{1/2}Q_\nu - i4\nu\langle \Omega^0_\nu q^2 \Omega^0_\nu\rangle. \tag{III.7}$$

We use (III.6) and (III.7) to establish the following equality:

$$H_\nu Q^n_\nu = 2(n-1)n(\nu^{1/2}Q^{n-1}_\nu + \nu\langle \Omega^0_\nu q^2 \Omega^0_\nu\rangle Q^{n-2}_\nu) - inP_\nu Q^{n-1}_\nu + Q^n_\nu H_\nu. \tag{III.8}$$

In fact,

$$H_\nu Q^n_\nu = \sum_{k=0}^{n-1} Q^k_\nu [H_\nu, Q_\nu] Q^{n-k-1}_\nu + Q^n_\nu H_\nu$$

$$= -i\sum_{k=0}^{n-1} Q^k_\nu P_\nu Q^{n-k-1}_\nu + Q^n_\nu H_\nu. \tag{III.9}$$

Now,

$$Q^k_\nu P_\nu Q^{n-k-1}_\nu = -\sum_{l=0}^{k-1} Q^{k-l-1}_\nu [P_\nu, Q_\nu] Q^{n-k-1+l}_\nu + P_\nu Q^{n-1}_\nu$$

$$= 4ki(\nu^{1/2}Q^{n-1}_\nu + \nu\langle \Omega^0_\nu q^2 \Omega^0_\nu\rangle Q^{n-2}_\nu) + P_\nu Q^{n-1}_\nu. \tag{III.10}$$

Substituting (III.10) in (III.9) we establish (III.8).

Let us proceed with the induction for (ii) first;

$$\|Q^n_\nu v\|^2 = \langle Q^{n-1}_\nu v, Q^2_\nu Q^{n-1}_\nu v\rangle. \tag{III.11}$$

Using (III.2) and $(a + b)^2 \leqslant 2(a^2 + b^2)$, we obtain

$$Q^2_\nu = \nu\left[q^2 - \frac{1}{4\nu} - \langle \Omega^0_\nu\left(q^2 - \frac{1}{4\nu}\right)\Omega^0_\nu\rangle\right]^2$$

$$\leqslant 2\left[\nu\left(q^2 - \frac{1}{4\nu}\right)^2 + \langle \Omega^0_\nu \nu^{1/2}\left(q^2 - \frac{1}{4\nu}\right)\Omega^0_\nu\rangle^2\right]. \tag{III.12}$$

The last term is estimated using the Schwarz inequality, and so

$$Q^2_\nu \leqslant 2\left[2H_\nu + \langle \Omega^0_\nu \nu\left(q^2 - \frac{1}{4\nu}\right)^2 \Omega^0_\nu\rangle\langle \Omega^0_\nu \Omega^0_\nu\rangle\right]$$

$$\leqslant 4(H_\nu + E^0_\nu)]. \tag{III.13}$$

Since $\lim_{\nu\to 0} E^0_\nu = \frac{1}{2}$ it is easy to use (III.13) and part (i) of the induction hypothesis in (III.11) to obtain (III.5) for n.

To prove (i) we use (III.8) and obtain

$$\|H_\nu Q^n_\nu v\| \leqslant 2(n-1)n(\nu^{1/2}\|Q^{n-1}_\nu v\| + |\langle \Omega^0_\nu \nu q^2 \Omega^0_\nu\rangle|\|Q^{n-2}_\nu v\|) + \|Q^n_\nu H_\nu v\| + n\|P_\nu Q^{n-1}_\nu v\|. \tag{III.14}$$

Since the Schwarz inequality yields

$$\langle \Omega^0_\nu \nu q^2 \Omega^0_\nu\rangle = \langle \Omega^0_\nu \nu(q^2 - 1/4\nu)\Omega^0_\nu\rangle + \tfrac{1}{4} \leqslant \nu^{1/2}(2E^0_\nu)^{1/2} + \tfrac{1}{4} \tag{III.15}$$

and $\lim_{\nu\to 0} E^0_\nu = \frac{1}{2}$, we can use (III.4) of the induction hypothesis to bound each of the first three terms of the right-hand side of (III.14) in terms of $\|C_{\nu,m}(H_\nu)v\|$ with $m = n - 1, n - 2, n + 1$. We estimate the last term of (III.14) as follows,

$$P^2_\nu = \nu(2qp - i)(2pq + i) = \nu(4qp^2q - 1). \tag{III.16}$$

It is easily verified that

$$2qp^2q = qp^2q + (qp^2q)^* = (q^2p^2 - i2qp) + (p^2q^2 + i2pq). \tag{III.17}$$

This gives us

$$P^2_\nu = 2\nu(p^2q^2 + q^2p^2) + 3\nu. \tag{III.18}$$

now

$$\nu(p^2q^2 + q^2p^2) \leqslant \nu\left\{\left[p^2 + \nu\left(q^2 - \frac{1}{4\nu}\right)^2\right]q^2 + q^2\left[p^2 + \nu\left(q^2 - \frac{1}{4\nu}\right)^2\right]\right\}$$

$$= 2\nu\left[H_\nu\left(q^2 - \frac{1}{4\nu}\right) + \left(q^2 - \frac{1}{4\nu}\right)H_\nu\right] + 2\nu\left(\frac{H_\nu}{4\nu} + \frac{H_\nu}{4\nu}\right)$$

$$= 2\nu^{1/2}\left[H_\nu \nu^{1/2}\left(q^2 - \frac{1}{4\nu}\right) + \nu^{1/2}\left(q^2 - \frac{1}{4\nu}\right)H_\nu\right] + H_\nu. \tag{III.19}$$

Since for self-adjoint operators H and $D = \nu^{1/2}(q^2 - 1/4\nu)$ the inequality $(H - D)^2 \geqslant 0$ implies $HD + DH \leqslant H^2 + D^2$, we have

$$\nu(p^2q^2 + q^2p^2) \leqslant 2\nu^{1/2}[H^2_\nu + \nu(q^2 - 1/4\nu)^2] + H_\nu$$

$$\leqslant 2\nu^{1/2}(H^2_\nu + 2H_\nu) + H_\nu$$

$$\leqslant (4\nu^{1/2} + \tfrac{1}{2})(H^2_\nu + 1). \tag{III.20}$$

Substituting (III.20) in (III.18), we obtain

$$\|P_\nu Q_\nu^{n-1}v\|^2 \leqslant 8(\nu+\nu^{1/2}+1)(\|Q_\nu^{n-1}v\|^2+\|H_\nu Q_\nu^{n-1}v\|^2). \tag{III.21}$$

Each term of the right-hand side of (III.21) can be estimated by $\|C_\nu(H_\nu)v\|^2$ using the induction hypotheses (i) and (ii). Using the equivalence of norms in $\mathbb{R}^m$ it is easy to combine all the estimates in (III.14) in a single inequality of the form

$$\|H_\nu Q_\nu^n v\| \leqslant \|A_{\nu,n}(H_\nu)v\|, \tag{III.22}$$

where $A_{\nu,n}$ is a polynomial with nonnegative coefficients which are bounded as ν tends to zero. The proof of the lemma is complete.

Lemma 2: For all $m,l\in\mathbb{N}$,

$$\langle \Omega_\nu^{2l}, [\nu(q^2-1/4\nu)^2]^m \Omega_\nu^{2l}\rangle$$

is bounded as ν tends to zero

Proof:

$$\nu^{1/2}(q^2-1/4\nu) = Q_\nu + \langle \Omega_\nu^0, \nu^{1/2}(q^2-1/4\nu)\Omega_\nu^0\rangle, \tag{III.23}$$

$$\left\langle \Omega_\nu^{2l}\left[\nu\left(q^2-\frac{1}{4\nu}\right)^2\right]^m \Omega_\nu^{2l}\right\rangle = \sum_{j=0}^{2m} C_{2m}^j \langle \Omega_\nu^{2l}, Q_\nu^{2m-j}\Omega_\nu^{2l}\rangle \times \left\langle \Omega_\nu^0 \nu^{1/2}\left(q^2-\frac{1}{4\nu}\right)\Omega_\nu^0\right\rangle^j. \tag{III.24}$$

The result follows from Schwarz's inequality and Lemma 1,-part (ii).

Lemma 3: The generalized Laguerre functions satisfy

$$\lim_{\nu\to 0}\langle \psi_\nu^{k,e}, Q_\nu^m \psi_\nu^{l,e}\rangle = \langle \Omega_0^k q^m \Omega_0^l\rangle, \quad \forall\ k,l,m\in\mathbb{N}.$$

Remark: The generalized Laguerre functions that we use are a complete orthonormal set of eigenfunctions of the eigenvalue equation

$$\tfrac{1}{2}[p^2+\tfrac{1}{4}(x-1/4\nu x)^2]\psi(x) = \lambda\psi(x). \tag{III.25}$$

The eigenvalues are $\lambda_\nu^n = n+\tfrac{1}{2}+\tfrac{1}{2}a-(16\nu)^{-1}$, $n\in\mathbb{N}$, where

$$a = \tfrac{1}{2}(1+1/16\nu^2)^{1/2}. \tag{III.26}$$

The corresponding eigenvectors are

$$\psi_\nu^{n,e}(x) = (-1)^n 2^{-1/4}[\Gamma(n+1)/\Gamma(a+n+1)]^{1/2} \times (\tfrac{1}{2}x^2)^{a/2+1} e^{-x^2/4} L_n^a(\tfrac{1}{2}x^2) \tag{III.27}$$

and

$$\psi_\nu^{n,o}(x) = \theta(x)\psi_\nu^{n,e}(x) - \theta(-x)\psi_\nu^{n,e}(-x),$$

where $\theta(x)$ is the Heaviside function and L_n^a are generalized Laguerre polynomials. Notice that $\lim_{\nu\to 0}\lambda_\nu^n = n+\tfrac{1}{2}$. For a more complete discussion see Ref. 1.

Proof of Lemma 3: Due to parity consideration it suffices to prove that $\forall k,l,m\in\mathbb{N}$,

$$\lim_{\nu\to 0} 2\int_0^\infty \psi_\nu^k(x) Q_\nu^m(x)\psi_\nu^l(x)dx = \int_{-\infty}^{\infty} \Omega_0^k(x)x^m\Omega_0^l(x)dx, \tag{III.28}$$

where we used the fact that $\psi_\nu^{k,e}(x) = \psi_\nu^{k,o}(x)$ for $x>0$ to drop the superscripts e and o (even and odd). The proof is divided in three parts.

Part 1: We claim that $\forall\ k,l,m\in\mathbb{N}$,

$$\lim_{\nu\to 0} 2\int_0^\infty \psi_\nu^k(x)\left[\frac{1}{8a}(x^2-2a)^2\right]^{m/2}\psi_\nu^l(x)dx = \int_{-\infty}^{+\infty} \Omega_0^k(x)x^m\Omega_0^l(x)dx, \tag{III.29}$$

where $a\equiv a(\nu)$ is given by (III.26).

In fact, for $I \equiv 2\int_0^\infty \psi_\nu^k(x)[(1/8a)(x^2-2a)^2]^{m/2}\psi_\nu^l(x)dx$, using (III.27) and performing the change of variables $x^2/2 = y$, we obtain

$$I = (-1)^{l+k}\{k!l!/[\Gamma(a+k+1)\Gamma(a+l+1)]\}^{1/2} \times \int_0^\infty y^a e^{-y} L_k^a(y) L_l^a(y)\left[\frac{1}{2a}(y-a)^2\right]^{m/2} dy. \tag{III.30}$$

Essentially, the proof uses the fact that $y^a e^{-y} \cong \exp[-(y-a)^2/2a]$ as a tends to infinity. We change variables again: $(y-a)/\sqrt{a} = x$.

$$I = (-1)^{l+k}\{k!l!/[\Gamma(a+k+1)\Gamma(a+l+1)]\}^{1/2} 2^{-m/2} e^{a\ln a - a}\int_{-\sqrt{a}}^{\infty} e^{-\sqrt{a}x + a\ln(1+x/\sqrt{a})} L_k^a(\sqrt{a}x+a) \times L_l^a(\sqrt{a}x+a)x^m dx. \tag{III.31}$$

We estimate the exponent

$f_a(a) \equiv -\sqrt{a}x + a\ln(1+x/\sqrt{a})$. Since for $0\leqslant z<1$, $\ln(1-z)\leqslant -z-z^2/2$, we have $f_a(x) \geqslant -\sqrt{a}x + a(x/\sqrt{a} - x^2/2a) = -x^2/2$ for $-\sqrt{a}<x\leqslant 0$.
For $x\geqslant 0$ and $a\geqslant a'$,

$$f_a(x) = \int_0^x f_a'(t)dt = -\int_0^x t\left(1+\frac{t}{\sqrt{a}}\right)^{-1} \leqslant -\int_0^x t\left(1+\frac{t}{\sqrt{a'}}\right)^{-1} dt = f_a(x)$$

It is also clear that $\lim_{a\to\infty} e^{f_a(x)} = e^{-x^2/2}, \forall x\in(-\sqrt{a},\infty)$ and

$$\chi_{(-\sqrt{a},\infty)}(x) e^{-\sqrt{a}x + a\ln(1+x/\sqrt{a})}$$

$$\leqslant \begin{cases} e^{-x^2/2} & x \leqslant 0, \\ e^{-\sqrt{a'}x + a'\ln(1 + x/\sqrt{a'})}, & x \geqslant 0, a \geqslant a'. \end{cases} \tag{III.32}$$

Now Eq. (5.5.4) of Ref. 5 states that

$$\lim_{a\to\infty} a^{-n/2} L_n^a(\sqrt{a}x + a) = [(-1)^n/n!] He_n(x). \tag{III.33}$$

This implies that all the coefficients of the polynomial in x, $a^{-n/2}L_n^a(\sqrt{a}x + a)$, tend to the corresponding coefficients of $[(-1)^n/n!]He_n(x)$. In particular, they are bounded functions of a in any interval $a \in [a_0, \infty)$. Fix a_0, and take $a > a_0$. The remarks above imply that we can find an integrable $g(x) \geqslant 0$ such that the absolute value of the integrand is estimated by this $g(x)$. Using the Lebesgue dominated convergence theorem,

$$\lim_{a\to\infty} I = \lim_{a\to\infty} \{a^{k+l}/[k!\,l!\,\Gamma(a+k+1)\Gamma(a+l+1)]\}^{1/2} \times e^{a\ln a - a} 2^{-m/2} \int_{-\infty}^{+\infty} e^{-x^2/2} He_k(x) He_l(x) x^m dx. \tag{III.34}$$

Changing variables ($x/\sqrt{2} = y$) and using the definition of $\Omega_0^n(x)$, we obtain

$$I = \lim_{a\to\infty} \{2\pi a^{k+l}/[\Gamma(a+k+1)\Gamma(a+l+1)]\}^{1/2} e^{a\ln a - a} \times \int_{-\infty}^{+\infty} \Omega_0^k(x) x^m \Omega_0^l(x) dx. \tag{III.35}$$

An application of Stirling's formula shows that the limit in (III.35) is 1, so (III.29) is proved.

Part 2: We claim that

$$\lim_{\nu\to 0}(\langle \psi_\nu^{0,e} \nu^{1/2} q^2 \psi_\nu^{0,e}\rangle - \langle \Omega_\nu^0 \nu^{1/2} q^2 \Omega_\nu^0\rangle) = 0. \tag{III.36}$$

In fact,

$$\left|\left\langle \psi_\nu^{0,e}\nu^{1/2}\left(q^2 - \frac{1}{4\nu}\right)\psi_\nu^{0,e}\right\rangle - \left\langle \Omega_\nu^0 \nu^{1/2}\left(q^2 - \frac{1}{4\nu}\right)\Omega_\nu^0\right\rangle\right|$$

$$= \left|\left\{(\psi_\nu^{0,e}, -\Omega_\nu^0)\nu^{1/2}\left(q^2 - \frac{1}{4\nu}\right)\psi_\nu^{0,e}\right\} + \left\langle \Omega_\nu^0 \nu^{1/2}\left(q^2 - \frac{1}{4\nu}\right)(\psi_\nu^{0,e} - \Omega_\nu^0)\right\rangle\right|$$

$$\leqslant \|\psi_\nu^{0,e} - \Omega_\nu^0\|^{1/2}\left(\left\langle \psi_\nu^{0,e}\nu\left(q^2 - \frac{1}{4\nu}\right)^2 \psi_\nu^{0,e}\right\rangle^{1/2} + \left\langle \Omega_\nu^0 \nu\left(q^2 - \frac{1}{4\nu}\right)^2 \Omega_\nu^0\right\rangle^{1/2}\right). \tag{III.37}$$

Since[6] $\lim_{\nu\to 0}\|\psi_\nu^{0,l} - \Omega_\nu^0\| = 0$ and the second expectation in (III.37) is bounded by $(E_\nu^0)^{1/2}$, which tends to $1/\sqrt{2}$ as ν tends to zero, (III.36) will be established as soon as we prove that the first expectation in (III.37) is bounded as ν tends to zero.

Using the inequality $(x+y)^2 \leqslant 2(x^2+y^2)$,

$$\left\langle \psi_\nu^{0,e}\nu\left(q^2 - \frac{1}{4\nu}\right)^2 \psi_\nu^{0,e}\right\rangle = (8a\nu)\left\langle \psi_\nu^{0,e}\frac{1}{8a}\left[q^2 - 2a + \left(2a - \frac{1}{4\nu}\right)\right]^2 \psi_\nu^{0,e}\right\rangle \leqslant 2(8a\nu)\left[\left\langle \psi_\nu^{0,e}\frac{1}{8a}(q^2 - 2a)^2 \psi_\nu^{0,e}\right\rangle + \left(2a - \frac{1}{4\nu}\right)^2\right]. \tag{III.38}$$

Using (III.27) we see that $\lim_{\nu\to 0} 8a\nu = 1$ and $\lim_{\nu\to 0}(2a - 1/4\nu)^2 = 0$. Finally (III.29) for $k = l = 0$ and $m = 1$ yields the boundedness of (III.38), which establishes (III.36).

The equality below can be proved by means of essentially the same arguments used for (III.36):

$$\lim_{\nu\to 0}(\langle \psi^{0,e}\nu^{1/2} q^2 \psi_\nu^{0,e}\rangle - 2a\nu^{1/2}) = 0. \tag{III.39}$$

Finally, we combine (III.39) and (III.36) to obtain

$$\lim_{\nu\to 0}(2a\nu^{1/2} - \nu^{1/2}\langle \Omega_\nu^0 q^2 \Omega_\nu^0\rangle) = 0. \tag{III.40}$$

Part 3:

$$\langle \psi_\nu^{k,e} Q_\nu^m \psi_\nu^{l,e}\rangle = \langle \psi_\nu^{k,e}[\nu^{1/2}(q^2 - 2a) + \nu^{1/2}(2a - \langle \Omega_\nu^0 q^2 \Omega_\nu^0\rangle)]^m \psi_\nu^{l,e}\rangle. \tag{III.41}$$

We expand (III.41) in powers of m. From (III.40) and $\lim_{\nu\to 0}(\nu/8a)^{1/2} = 1$ we get

$$\lim_{\nu\to 0}\langle \psi_\nu^{k,e} Q_\nu^m \psi_\nu^{l,e}\rangle = \lim_{\nu\to 0}\langle \psi_\nu^{k,e}[\nu^{1/2}(q^2 - 2a)]^m \psi_\nu^{l,e}\rangle = \lim_{\nu\to 0}\langle \psi_\nu^{k,e}[(8a)^{-1/2}(q^2 - 2a)]^m \psi_\nu^{l,e}\rangle. \tag{III.42}$$

The proof of the lemma is completed using (III.29) in (III.42).

Theorem 1:

$$\lim_{\nu\to 0}\langle \Omega_\nu^{2k} Q_\nu^n \Omega_\nu^{2l}\rangle = \langle \Omega_0^k q^n \Omega_0^l\rangle, \qquad \forall\ k,l,n \in \mathbb{N}.$$

Proof:

$$|\langle \Omega_\nu^{2k} Q_\nu^n \Omega_\nu^{2l}\rangle - \langle \psi_\nu^{k,e} Q_\nu^n \psi_\nu^{l,e}\rangle| \leqslant |\langle (\Omega_\nu^{2k} - \psi_\nu^{k,e}), Q_\nu^n \Omega_\nu^{2l}\rangle| + |\langle \psi_\nu^{k,e}, Q_\nu^n(\Omega_\nu^{2l} - \psi_\nu^{l,e})\rangle| \leqslant \|\Omega_\nu^{2k} - \psi_\nu^{k,e}\|\ \|Q_\nu^n \Omega_\nu^{2l}\| + \|Q_\nu^n \psi_\nu^{k,e}\|\ \|\Omega_\nu^{2l} - \psi_\nu^{l,e}\|.$$

It is proved in Ref. 1 that $\|\Omega_\nu^{2k} - \psi_\nu^{k,e}\|$ and $\|\Omega_\nu^{2l} - \psi_\nu^{l,e}\|$ tend to zero as ν tends to zero. The term $\|Q^n \Omega_\nu^{2l}\|$ stays bounded by virtue of Lemma 1 and the fact that

$\lim_{\nu\to 0} E_\nu^{2l} = l + \frac{1}{2}$ (also proved in Ref. 1. Lemma 3 guarantees the boundedness of $\|Q_\nu^n \psi_\nu^{k,e}\|$ as ν tends to zero.

Theorem 1 follows from the above estimate and from another application of Lemma 3.

IV. PROOF OF THEOREMS 2 AND 3

Lemma 4: Let $P_{\nu,n}$ be the orthogonal projection on $\{\Omega_\nu^0, \Omega_\nu^1, \ldots, \Omega_\nu^{n-1}\}^\perp$. (These vectors were defined in Sec. II as the eigenvectors of H_ν with corresponding eigenvalues $E_\nu^0 \leqslant E_\nu^1 \leqslant E_\nu^2 \cdots$.) Then the following inequalities hold:

(i)

$$e^{-tH_\nu} P_{\nu,n} \leqslant e^{-tE_\nu^n}, \quad \forall\ n \in \mathbb{N}, \quad t \geqslant 0, \tag{IV.1}$$

(ii)

$$H_\nu^m e^{-tH_\nu} \leqslant (m/t)^m e^{-m}, \quad \forall m \in \mathbb{N}, \quad t > 0. \tag{IV.2}$$

Proof: Using the spectral theorem, (IV.1) is obvious, and (IV.2) follows from the inequalities $H_\nu \geqslant 0$ and $\sup_{x\geqslant 0} x^m e^{-tx} \leqslant (m/t)^m e^{-m}$.

Lemma 5: Let $k \in \mathbb{N}$ and $t_i > 0$, $i = 1,\ldots,k$, and $m_1, m_2, \ldots, m_k \in \mathbb{N}^+$. Let

$$e_k \equiv \|e^{-t_k H_\nu} P^\# Q_\nu^{m_k} e^{-t_{k-1} H_\nu} \times P^\# Q_\nu^{m_{k-1}} \cdots e^{-t_1 H_\nu} P^\# Q_\nu^{m_1} \Omega_\nu^0\|, \tag{IV.3}$$

where each $P^\#$ stands for either $P_{\nu,n}$ or $P'_{\nu,n} \equiv I - P_{\nu,n}$.

(i) There exist $c = c(t_1,\ldots,t_k, m_1,\ldots,m_k)$ such that $e_k \leqslant c$.

(ii) If the leftmost $P^\#$ is $P_{\nu,n}$ (i.e., $e^{-t_k H_\nu} P^\# \equiv e^{-t_k H_\nu} P_{\nu,n}$), then for any given $\epsilon > 0$, there exist $\delta = \delta(t_1,\ldots,t_k, m_1,\ldots,m_k, \epsilon)$ and $N = N(t_1,\ldots,t_k, m_1,\ldots,m_k, \epsilon)$ such that for any ν satisfying $0 < \nu \leqslant \delta$ and any $n \geqslant N$ we have $e_k \leqslant \epsilon$.

Proof: We prove (i) by induction on k. For $k = 0$ there is nothing to be proved. Let us assume that (i) is valid for $k - 1$,

$$e_k \leqslant \|e^{-t_k H_\nu}\|\ \|P^\#\| \times \|Q_\nu^{m_k} e^{-(t_{k-1}) H_\nu} P^\# Q_\nu^{m_{k-1}} \cdots e^{-t_1 H_\nu} P^\# Q_\nu^{m_1} \Omega_\nu^0\|.$$

Using $H_\nu \geqslant 0$ and (ii) of Lemma 1, we have

$$e_k \leqslant \|B_{\nu,m_k}(H_\nu) e^{-(t_{k-1}) H_\nu/2} e^{-(t_{k-1}) H_\nu/2} \times P^\# Q_\nu^{m_{k-1}} \cdots e^{-t_1 H_\nu} P^\# Q_\nu^{m_1} \Omega_\nu^0\|.$$

We now use the triangle inequality and Lemma 4, part (ii), to obtain

$$e_k \leqslant d_k \|e^{-(t_{k-1}) H_\nu/2} P^\# Q_\nu^{m_{k-1}} \cdots e^{-t_1 H_\nu} P^\# Q_\nu^{m_1} \Omega_\nu^0\|.$$

Finally, we use the induction hypothesis, and part (i) is proved.

We prove (ii) now, using Lemma 4, part (i),

$$e_k = \|e^{-t_k H_\nu/2} P_{\nu,n} e^{-t_k H_\nu/2} P^\# Q_\nu^{m_k} e^{-t_{k-1} H_\nu} \times P^\# Q_\nu^{m_{k-1}} \cdots e^{-t_1 H_\nu} P^\# Q_\nu^{m_1} \Omega_\nu^0\|$$

$$\leqslant e^{-t_k E_\nu^n/2} \|e^{-t_k H_\nu/2} P^\# \cdots Q_\nu^{m_1} \Omega_\nu^0\|.$$

Since $E_\nu^m \leqslant E_\nu^{m+1}$, $\forall m \in \mathbb{N}$, and $\lim_{\nu\to 0} E_\nu^{2n} = \lim_{\nu\to 0} E^{2n+1} = n + \frac{1}{2}$, and application of (i) yields (ii).

Theorem 2: For $s_i \geqslant 0$, $i = 1,2,\ldots,m$, $m \in \mathbb{N}$,

$$\lim_{\nu\to 0} \langle \Omega_\nu^0 Q_\nu e^{-s_1 H_\nu} Q_\nu e^{-s_2 H_\nu} \cdots Q_\nu e^{-s_m H_\nu} Q_\nu \Omega_\nu^0 \rangle = \langle \Omega_0^0 q e^{-s_1 H_0} q e^{-s_2 H_0} \cdots q \Omega_0^0 \rangle.$$

Proof: Consider the correlation function ($s_i \geqslant 0$, $i = 1,\ldots,l$),

$$S = \langle \Omega_0^\nu Q_\nu e^{-s_1 H_\nu} Q_\nu e^{-s_2 H_\nu} \cdots e^{-s_l H_\nu} Q_\nu \Omega_0^\nu \rangle.$$

If all s_i are zero, $i = 1,\ldots,l$, then the theorem is just a particular case of Theorem 1. Thus we may assume that at least some s_i is positive, and we can rewrite S as follows, for $t_i > 0$, $i = 1,\ldots,k$,

$$S = \langle \Omega_\nu^0 Q_\nu^{m_1} e^{-t_1 H_\nu} Q_\nu^{m_2} e^{-t_2 H_\nu} \cdots Q_\nu^{m_k} \Omega_\nu^0 \rangle. \tag{IV.4}$$

We make the substitution $e^{-t_i H_\nu} = e^{-t_i H_\nu}(P_{\nu,n} + P'_{\nu,n})$ (see Lemma 4) for n to be specified later, and expand S into a sum of several terms. There will be one term where all the $e^{-t_i H_\nu}$ are multiplied by $P_{\nu,n}$, i.e.,

$$S' = \langle \Omega_\nu^0 Q_\nu^{m_1} e^{-t_1 H_\nu} P_{\nu,n} Q_\nu^{m_2} e^{-t_2 H_\nu} P_{\nu,n} \cdots Q_\nu^{m_k} \Omega_\nu^0 \rangle, \tag{IV.5}$$

and a sum of terms of the form

$$S^\# = \langle \Omega_\nu^0 P^\# Q_\nu^{m_1} e^{-t_1 H_\nu} P^\# Q_\nu^{m_2} e^{-t_2 H_\nu} P^\# \cdots Q_\nu^{m_k} \Omega_\nu^0 \rangle, \tag{IV.6}$$

where $P^\#$ are $P_{\nu,n}$ and $P'_{\nu,n}$ and at least one $P^\#$ stands for a $P_{\nu,n}$. Using Schwarz's inequality and Lemma 5, (i) and (ii), it is easy to show that given $\epsilon > 0$ there exist $\delta > 0$ and $N \in \mathbb{N}$ such that for all $n \geqslant N$ and $0 < \nu \leqslant \delta$, we have $|S^\#| \leqslant \epsilon$ for all such terms $S^\#$.

As to (IV.5) we write the full eigenfunction expansion,

$$S' = \sum_{i_1,i_2,\ldots,i_k}^{[(n-1)/2]} \langle \Omega_\nu^0 Q_\nu^{m_1} \Omega_\nu^{2i_1} \rangle \exp(-t_1 E_\nu^{2i_1}) \times \langle \Omega_\nu^{2i_1} Q_\nu^{m_2} \Omega_\nu^{2i_2} \rangle \exp(-t_2 E_\nu^{2i_2}) \cdots \langle \Omega_\nu^{2i_k} Q_\nu^{m_k} \Omega_\nu^0 \rangle. \tag{IV.7}$$

Using Theorem 1 and $\lim_{\nu\to 0} E_\nu^{2i} = E_0^i$, we obtain

$$\lim_{\nu\to 0} S' = \sum_{i_1,i_2,\ldots,i_k}^{[(n-1)/2]} \langle \Omega_0^0 q^{m_1} \Omega_0^{i_1} \rangle \exp(-t_1 E_0^{i_1}) \times \langle \Omega_0^{i_1} q^{m_2} \Omega_0^{i_2} \rangle \exp(-t_2 E_0^{i_2}) \cdots \langle \Omega_0^{i_k} q^{m_k} \Omega_0^0 \rangle. \tag{IV.8}$$

Thus, given $\epsilon > 0$, choose N so large that the right-hand side of (IV.8) differs from $\langle \Omega_0^0 q^{m_1} e^{-t_1 H_0} q^{m_2} e^{-t_2 H_0} \cdots q^{m_k} \Omega_0^0 \rangle$ by less than $\epsilon/3$ in absolute value. Then choose N, δ so small that for $0 < \nu < \delta$ the sum of the terms (IV.6) is less than $\epsilon/3$. Finally, reduce δ even more so that S' differs from (IV.8) by less than $\epsilon/3$ in absolute value. The theorem is proved.

Corollary 4: Let u_n be the Hermite functions, eigenvec-

tors of $\frac{1}{2}(p^2+q^2)$. For $n\in\mathbb{N}$, let

$$U_\nu^{n,e}(x)=\frac{1}{\sqrt{2}}\left[u_n\left(x-\frac{1}{2\sqrt{\nu}}\right)+u_n\left(x+\frac{1}{2\sqrt{\nu}}\right)\right],$$

$$U_\nu^{n,o}(x)=\frac{1}{\sqrt{2}}\left[u_n\left(x-\frac{1}{2\sqrt{\nu}}\right)-u_n\left(x+\frac{1}{2\sqrt{\nu}}\right)\right].$$

Then

$$\lim_{\nu\to 0}\|\Omega_\nu^{2n}-U_\nu^{n,e}\|=\lim_{\nu\to 0}\|\Omega_\nu^{2n+1}-U_\nu^{n,o}\|=0.$$

Proof: We claim that

$$\lim_{\nu\to 0}\|\psi_\nu^{n,e}-U_\nu^{n,e}\|^2=\lim_{\nu\to 0}\|\psi_\nu^{n,o}-U_\nu^{n,o}\|^2=0.$$

(See Lemma 3.)

To verify this claim we expand the norm and compute each of the scalar products as in the proof of Lemma 3 for $m=0$. Now we use the Theorem 4.1 of Ref. 1, namely

$$\lim_{\nu\to 0}\|\Omega_\nu^{2n}-\psi_\nu^{n,e}\|=\lim_{\nu\to 0}\|\Omega_\nu^{2n+1}-\psi_\nu^{n,o}\|=O,$$

and the proof is complete. This corollary is essentially contained in Ref. 1.

ACKNOWLEDGMENT

I wish to express may gratitude to Professor J. Glimm.

[1] D. Isaacson, "Singular Perturbation and Asymptotic Eigenvalue Degeneracy," Commun. Pure Appl. Math. **29**, 531–51 (1976).
[2] J. Rosen, "The Ising model limit of ϕ^4 lattice fields," Rockefeller U. preprint New York (1976).
[3] J. Glimm, A. Jaffe, and T. Spencer, "A Convergent Expansion about Mean Field Theory, Parts I and II," Ann. Phys. (N.Y.) **101**, 610–69 (1976).
[4] D. Isaacson and D. Marchesin, "The Eigenvalues and Eigenfunctions of a Spherically Symmetric Anharmonic Oscillator," Commun. Pure Appl. Math. **31**, 659–70 (1978).
[5] W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer–Verlag, New York, 1966), p. 247.
[6] D. Isaacson, "The Critical Behavior of ϕ^4_1," Commun. Math. Phys. **53**, 257–75 (1977).

Radiating Kerr–Newman metric[a)]

C. González, L. Herrera, and J. Jiménez

Departamento de Física, Facultad de Ciencias, U.C.V. Caracas, Venezuela
(Received 7 June 1978)

A complete generalization of the Kerr–Newman solution to the nonstationary case is given. The possibility of associating the energy–momentum tensor with the electromagnetic field is discussed.

I. INTRODUCTION

In the past, generalizations of differents static or stationary metrics to the nonstationary situation[1-3] have been considered, especially, in relation to certain astrophysical problems, namely, when the energy density of the radiation emitted by the source cannot be neglected (e.g., supernovae, quasistellar radio sources.)[4,5]

The procedure for obtaining such generalizations is very simple. The parameters of the metric are replaced by arbitrary functions of the timelike coordinate, obtaining in this way a time dependent metric which is a solution of the Einstein equations with an energy–momentum tensor describing, in principle, the material content outside the star.

In this paper we present a nonstationary generalization of the Kerr–Newman[6] metric, allowing the three parameters a, m, and e to be arbitrary functions of the timelike coordinate. The resulting energy–momentum tensor is then analyzed in different cases. In the most general situation (a,m,e variables) the energy–momentum tensor has a nonzero trace, which could suggest the possible existence of an scalar field. It can be proved very easily that this is not the case.

In the third section we consider the case a = const, m and e variables. It is shown that no term of the energy–momentum tensor can be identified with the electromagnetic field, except when spacelike currents are accepted. In the more restricted case a,e = const, m variable, there is the electromagnetic contribution of the Kerr–Newman source, but the remaining terms are not of an electromagnetic nature. For the case $a = 0$, m and e variables, we recover the Bonnor–Vaidya solution for null currents.[7]

A brief discussion of the results is given in the conclusion.

II. THE SOLUTION

The Newman–Penrose formalism is used throughout the paper. For details we refer the reader to the original paper,[8] hereafter referred to as NP.

In the null coordinate system $x^0 \equiv u$, $x^1 \equiv r$, $x^2 \equiv \theta$, $x^3 \equiv \varphi$, the metric we are considering is

$$ds^2 = g_{\alpha\beta}\,dx^\alpha dx^\beta = \left[1 - \frac{2mr - e^2}{r^2 + a^2\cos^2\theta}\right] du^2 + 2\,du\,dr + \frac{2a\sin^2\theta}{r^2 + a^2\cos^2\theta}(2mr - e^2)\,du\,d\varphi - 2a\sin^2\theta\,dr\,d\varphi$$

$$- (r^2 + a^2\cos^2\theta)\,d\theta^2 - \sin^2\theta\left\{r^2 + a^2 + \frac{a^2\sin^2\theta}{r^2 + a^2\cos^2\theta}(2mr - e^2)\right\} d\varphi^2, \tag{1}$$

where, a, e, and m are arbitrary functions of the timelike coordinate u.

The corresponding null tetrad Z^μ_m can be chosen to be[9]

$$Z^\mu_0 \equiv l^\mu = \delta^\mu_1 \quad Z^\mu_3 \equiv n^\mu = \frac{1}{\Sigma}\left(\Omega\delta^\mu_0 - \frac{\Upsilon}{2}\delta^\mu_1 + a\delta^\mu_3\right), \quad Z^\mu_2 \equiv m^\mu = \frac{1}{\sqrt{2}(r + ia\cos\theta)}(ia\sin\theta\delta^\mu_0 + \delta^\mu_2 + i\cot\theta\delta^\mu_3), \tag{2}$$

whereas the covariant components are given by

$$l_\mu = \delta^0_\mu - a\sin^2\theta\delta^3_\mu, \quad n_\mu = \frac{1}{\Sigma}\left(\frac{\Upsilon}{2}\delta^0_\mu + \Sigma\delta^1_\mu - a\sin^2\theta\frac{\Upsilon}{2}\delta^3_\mu\right), \quad m_\mu = \frac{1}{\sqrt{2}(r + ia\cos\theta)}(ia\sin\theta\delta^0_\mu - \Sigma\delta^2_\mu - i\Omega\sin\theta\delta^3_\mu); \tag{3}$$

in the above

$$\Omega \equiv r^2 + a^2, \quad \Sigma \equiv r^2 + a^2\cos^2\theta, \quad \Upsilon \equiv r^2 + a^2 + e^2 - 2mr.$$

The spin coefficients are

$$\kappa = \sigma = 0, \quad \rho = -(r - ia\cos\theta)^{-1}, \quad \tau = -\frac{ia\sin\theta}{\sqrt{2}}\rho\bar{\rho} + \frac{i\dot{a}\sin\theta}{2\sqrt{2}}\bar{\rho}, \quad \alpha = \frac{i\dot{a}}{4\sqrt{2}}\bar{\rho}\sin\theta + \frac{ia\sin\theta}{\sqrt{2}}\rho^2 + \frac{\cot\theta}{2\sqrt{2}}\bar{\rho},$$

[a)]Partially supported by CONICIT.

 0022-2488/79/050837-07$01.00

$$\beta = -\frac{\cot\theta}{2\sqrt{2}}\bar{\rho} - \frac{\dot{a}\sin\theta}{2\sqrt{2}}\bar{\rho}\left(\frac{i}{2} - a\cos\theta\rho\right), \quad \gamma = \frac{\Upsilon}{2}\rho^2\bar{\rho} + \frac{(r-m)}{2}\rho\bar{\rho} - \frac{i\dot{a}}{2}\rho^2\bar{\rho}^2\Omega r\cos\theta,$$

$$\epsilon = 0, \quad \nu = \frac{i\dot{a}\sin\theta}{\sqrt{2}}\rho^2\bar{\rho}\left(\frac{\Upsilon}{2} + a^2 - a^2\Upsilon\rho\bar{\rho}\cos^2\theta\right) + \frac{i(e\dot{e} - \dot{m}r)a\sin\theta}{\sqrt{2}}\rho^2\bar{\rho}, \tag{4}$$

$$\mu = \frac{\Upsilon}{2}\rho^2\bar{\rho} + a\dot{a}\rho\bar{\rho}, \quad \lambda = -a\dot{a}\rho^3\bar{\rho}r^2\sin^2\theta, \quad \pi = \frac{i\dot{a}\sin\theta}{2\sqrt{2}}\rho + \frac{ia\sin\theta}{\sqrt{2}}\rho^2.$$

To calculate the tetrad components of the trace free Ricci tensor, the tetrad components of the Weyl tensor and the Ricci scalar, we use Eqs. (4.2a) to (4.2r) in NP.

A straightforward but lengthy calculation gives

$$\phi_{00} = 0, \tag{5}$$

$$\phi_{10} = \frac{\dot{a}\sin\theta}{2\sqrt{2}}(3ar\cos\theta - ir^2)\rho^2\bar{\rho}^2, \tag{6}$$

$$\phi_{20} = \frac{\ddot{a}a\sin\theta}{4}\rho^2 + \frac{\dot{a}^2}{2}\sin^2\theta\rho^2(\tfrac{1}{4} - a^2\cos^2\theta\rho\bar{\rho}) + a\dot{a}\sin\theta\rho^3\left(1 - 2a^2r\rho\bar{\rho}^2\cos^2\theta + \rho\bar{\rho}r^2 + \frac{ia\cos\theta}{2}(\bar{\rho} - 2\rho)\right), \tag{7}$$

$$\phi_{11} = \frac{e^2\rho^2\bar{\rho}^2}{2} - \frac{\ddot{a}a^3\sin^2\theta\cos^2\theta}{4}\rho^2\bar{\rho}^2 + \frac{\dot{a}^2\sin^2\theta\rho\bar{\rho}}{32}[1 + \rho^2\bar{\rho}^2(r^4 + a^4\cos^4\theta - 14r^2a^2\cos^2\theta)]$$
$$+ \frac{\dot{a}\rho^3\bar{\rho}^3ar\sin^2\theta}{8}(13r^2 + 5a^2\cos^2\theta), \tag{8}$$

$$\phi_{12} = -\frac{ia\sin\theta}{2\sqrt{2}}\left(2e\dot{e} + \frac{\dot{m}}{\bar{\rho}}\right)\rho^2\bar{\rho}^2 + \frac{i\ddot{a}a\sin\theta\Omega}{4\sqrt{2}}\rho\bar{\rho}^2(1 + 2a^2\cos^2\theta\rho\bar{\rho})$$
$$+ \dot{a}^2\frac{\sin\theta\cos\theta}{2\sqrt{2}}a\rho^3\bar{\rho}^4(-ia^4\cos^3\theta + ir^4\cos\theta - ia^2r^2\cos^3\theta$$
$$- 2ar^3\sin^2\theta) + \frac{\dot{a}\sin\theta\rho^3\bar{\rho}^4}{4\sqrt{2}}[10a^5\cos^3\theta - ir^5 - 2ia^4m\cos^4\theta + a^4r(3i\cos\theta + 2i\cos^4\theta) - 2a^3(2mr - e^2)\cos^3\theta$$
$$+ a^3r^2(10\cos\theta - 2\sin^2\theta\cos\theta) + 2ir^3(e^2 - 2mr) + 3ia^2r\cos^2\theta(2mr - e^2) + a^2r^3(8i - 15i\sin^2\theta)$$
$$+ 4ar^2\cos\theta(2mr - e^2)\Big], \tag{9}$$

$$\phi_{22} = \frac{(\dot{e}^2 + e\ddot{e} - \ddot{m}r)a^2\sin^2\theta}{2}\rho^2\bar{\rho}^2 + (e\dot{e} - \dot{m}r)r\rho^2\bar{\rho}^2 + \ddot{a}a\rho^2\bar{\rho}^2\left[\frac{\sin^2\theta}{2}\left(a^2 + \frac{\Upsilon}{2} - a^2\Upsilon\cos^2\theta\rho\bar{\rho}\right) - \Omega\right] + \dot{a}^2\rho^2\bar{\rho}^2$$
$$\times\left(-\tfrac{9}{4}a^2\Upsilon\sin^2\theta\cos^2\theta\rho\bar{\rho} - \tfrac{3}{2}a^2\sin^2\theta\cos^2\theta\rho\bar{\rho} + \tfrac{5}{2}a^2\sin^2\theta + 2a^4\Upsilon\sin^2\theta\cos^4\theta\rho^2\bar{\rho}^2 - \Omega + 2a^2\Omega\cos^2\theta\rho\bar{\rho} + a^2 + \frac{\Upsilon}{2}\sin^2\theta\right.$$
$$\left. - a^2r^4\sin^4\theta\rho^2\bar{\rho}^2\right) + \dot{a}\frac{a\rho^4\bar{\rho}^4}{2}\left[r^5\left[\tfrac{1}{2} - \tfrac{5}{2}\cos^2\theta - 4\sin^2\theta\left(\dot{m} - \frac{e\dot{e}}{r}\right) + \frac{r^3}{2}(2mr - e^2)(1 + 5\cos^2\theta)\right]\right.$$
$$+ a^2r^3\left[-\tfrac{5}{2}\sin^2\theta + 4 - 4\sin^2\theta\cos^2\theta\left(\dot{m} - \frac{e\dot{e}}{r}\right) - 10\cos^2\theta + \tfrac{3}{2}\sin^2\theta\cos^2\theta\right]$$
$$\left. + \frac{a^2r^2}{2}(2mr - e^2)(5\cos^2\theta + 3\cos^4\theta) - 2a^4r\cos^2\theta\right], \tag{10}$$

$$\psi_0 = 0, \tag{11}$$

$$\psi_1 = \frac{1}{2\sqrt{2}}\dot{a}\sin\theta\rho^2\bar{\rho}^2(2ia^2\cos^2\theta - ar\cos\theta), \tag{12}$$

$$\psi_2 = \rho^3(m + e^2\bar{\rho}) - 2\Lambda - \frac{\ddot{a}a}{4}\sin^2\theta\rho\bar{\rho} - \frac{\dot{a}\sin^2\theta}{8}\rho\bar{\rho}$$

$$+ \frac{\dot{a}}{4}\rho\bar{\rho}(2i\cos\theta - 8ar\rho\bar{\rho} - 3a\sin^2\theta\rho + 2ia^2\sin^2\theta\cos\theta\rho^2 - 4a\rho + 2ar\sin^2\theta\rho\bar{\rho} + a\sin^2\theta\bar{\rho}), \tag{13}$$

$$\psi_3 = -i\left(\dot{m} - \frac{e\dot{e}}{r}\right)\frac{a\sin\theta}{\sqrt{2}}\rho^2\bar{\rho}(\tfrac{1}{2} + 2r\rho) + \ddot{a}\frac{\sin\theta\rho\bar{\rho}^2\Omega}{4\sqrt{2}}(i - 2ar\cos\theta\rho^2) + \frac{\dot{a}^2\sin\theta}{8\sqrt{2}}\rho\bar{\rho}(-2\Omega\cos\theta\bar{\rho}^2 + 8i\cos^2\theta\Omega r\rho^3\bar{\rho}^2$$

$$- 4ia\cos^2\theta\Omega r\rho^2\bar{\rho}^2 - 8a^2r\cos\theta\rho^2\bar{\rho}^2 - 2ar^2\sin^2\theta\rho^2\bar{\rho}^2 - 4a^2\sin^2\theta\cos\theta r^2\bar{\rho}\rho^3 + 2ia\bar{\rho} - 2\Omega r\cos\theta\rho^2\bar{\rho})$$

$$+ \frac{\dot{a}\rho\bar{\rho}}{8\sqrt{2}}\Big(-2ia\sin\theta\frac{\Upsilon}{2}\bar{\rho}^2 + 8i\Omega\sin\theta\rho^2 + 16a\Omega\sin\theta\cos\theta\rho^3$$

$$- 4i\Omega\frac{\cos^2\theta}{\sin\theta}\rho^2 + 16a^3\sin\theta\cos\theta\rho^2\bar{\rho} - 8a\sin\theta\cos\theta\frac{\Upsilon}{2}\rho\bar{\rho}^2$$

$$+ 4a(r - m)\sin\theta\cos\theta\bar{\rho}^2 - 4ar\Omega\sin\theta\cos\theta\rho^2\bar{\rho}^2 + 4ir\Omega\sin\theta\rho^2\bar{\rho} - 8i\sin\theta\rho^2(\Upsilon/2 + a^2 - a^2\Upsilon\cos^2\theta\rho\bar{\rho})$$

$$+ 4ar^2\sin\theta\cos\theta\rho^2\bar{\rho} + 12a\Omega r\sin\theta\cos\theta\rho^3\bar{\rho} - 2i\Omega r\cos\theta\cot\theta\rho^2\bar{\rho} - i\sin\theta(r - m)\bar{\rho} + 4a\cot\theta\rho$$

$$- 6i\frac{\Upsilon}{2}\sin\theta\rho^2 - 3i\sin\theta(r - m)\rho - 10a\Upsilon\sin\theta\cos\theta\rho^2\bar{\rho} - 6a(r - m)\sin\theta\cos\theta\rho\bar{\rho}, \tag{14}$$

$$\psi_4 = ra^2\sin^2\theta\rho^3\bar{\rho}\left[\rho\left(\dot{m} - \frac{e\dot{e}}{r}\right) - \frac{1}{2}\left(\ddot{m} - \frac{\dot{e}^2 + e\ddot{e}}{r}\right)\right] + \ddot{a}ar^2\Omega\sin^2\theta\rho^4\bar{\rho}^2 + \dot{a}^2r^2\Omega\sin^2\theta\rho^4\bar{\rho}^2$$

$$- 3i\dot{a}^2ar^2\Omega\sin^2\theta\cos\theta\rho^5\bar{\rho}^2 + i\dot{a}^2ar^2\Omega\sin^2\theta\rho^4\bar{\rho}^3 - a\dot{a}r\Upsilon\sin^2\theta\rho^4\bar{\rho}^2 - \tfrac{3}{2}a\dot{a}r^2\Upsilon\sin^2\theta - \frac{a\dot{a}}{2}\Upsilon r^2\sin^2\theta\rho^4\bar{\rho}^3$$

$$- \frac{a\sin^2\theta}{2}\ddot{a}\rho^3\bar{\rho}\xi + ia\dot{a}^2\sin^2\theta\cos\theta\rho^4\bar{\rho}\xi - ia\dot{a}\frac{\sin^2\theta\cos\theta}{2}\rho^3\bar{\rho}^2\xi - \frac{a^2\dot{a}^2\sin^2\theta}{2}\rho^3\bar{\rho} - ar\frac{\dot{m}\dot{a}\sin^2\theta}{2}\rho^3\bar{\rho}$$

$$+ a^2\dot{a}^2\sin^2\theta\rho^3\bar{\rho} + a^4\dot{a}^2\sin^2\theta\cos^2\theta\rho^4\bar{\rho}^2 - a^3r\dot{m}\dot{a}\sin^2\theta\cos^2\theta\rho^4\bar{\rho}^2$$

$$+ 2a^2\dot{a}^2\frac{\Upsilon}{2}\sin^2\theta\cos^2\theta\rho^4\bar{\rho}^2 + ia^3\Upsilon\dot{a}^2\sin^2\theta\cos^3\theta\rho^5\bar{\rho}^2 + i\Upsilon a^3\dot{a}^2\sin^2\theta\cos^3\theta\rho^4\bar{\rho}^3 + \frac{ar\dot{m}\dot{a}\sin^2\theta\rho^3\bar{\rho}}{2}$$

$$- ia^2r\dot{a}\dot{m}\sin^2\theta\cos\theta\rho^4\bar{\rho} + \frac{ira^2\dot{a}\dot{m}\sin^2\theta\cos\theta}{2}\rho^3\bar{\rho}^2 - \frac{i\dot{a}\rho}{2}\{\xi(\cos\theta\rho^2\bar{\rho} + 2ia\sin\theta\rho^3\bar{\rho} - ia\sin^2\theta\rho^2\bar{\rho}^2)$$

$$- a^2\Upsilon\sin\theta[(-2\rho\bar{\rho}\sin\theta\cos\theta - ia\sin\theta\cos^2\theta\rho\bar{\rho}^2) + ia\sin\theta\cos^2\theta\rho^2\bar{\rho}]\} - 2ia\dot{a}^2\Omega r^3\sin^2\theta\cos\theta\rho^5\bar{\rho}^3$$

$$+ a\dot{a}r^2\Upsilon\sin^2\theta\rho^4\bar{\rho}^2(3\rho + \bar{\rho}) - a\dot{a}r^2(r - m)\sin^2\theta\rho^4\bar{\rho}^2 + 2a^2\dot{a}^2r^2\sin^2\theta\rho^4\bar{\rho}^2 + a\dot{a}r^3\Upsilon\sin^2\theta\rho^5\bar{\rho}^3$$

$$+ \left(\frac{ia\sin\theta}{\sqrt{2}}(4\rho - \bar{\rho}) - \frac{3ir\dot{a}\sin\theta}{2}\theta\rho\bar{\rho}\right) + \frac{\cot\theta\rho^2}{\sqrt{2}} + a\dot{a}\sin\theta\cos\theta\frac{\rho\bar{\rho}}{2\sqrt{2}} + \frac{i\dot{a}\sin\theta\rho}{2\sqrt{2}}\Big),$$

$$\times\left(i\dot{a}\,\frac{\sin\theta\rho^2\bar{\rho}}{\sqrt{2}}\xi - i\dot{m}\,\frac{ra\sin\theta}{\sqrt{2}}\rho^2\bar{\rho}\right), \tag{15}$$

$$\xi \equiv \frac{\Upsilon}{2} + a^2 - a^2\Upsilon\cos^2\theta\rho\bar{\rho}.$$

For the scalar $\Lambda = -R/24$, we get

$$\Lambda = -\ddot{a}\,\frac{a\sin^2\theta\rho\bar{\rho}}{12}(1 + a^2\cos^2\theta\rho\bar{\rho}) + \frac{\dot{a}^2}{3}\rho\bar{\rho}\sin^2\theta\left(-\tfrac{3}{8}a^2\cos^2\theta\rho\bar{\rho} - \frac{7}{32} + \frac{\rho^2\bar{\rho}^2}{32}(r^4 + 13a^4\cos^4\theta - 2a^2r^2\cos^2\theta)\right)$$

$$+ \frac{\dot{a}r}{3}\rho^2\bar{\rho}^2a(4 - \tfrac{11}{8}r^2\sin^2\theta\rho\bar{\rho} - \tfrac{5}{8}a^2\sin^2\theta\cos^2\theta\rho\bar{\rho}). \tag{16}$$

From the Einstein equations if follows that the energy–momentum tensor, when it is written in terms of the scalars defined above, is given by the expression

$$T_{\mu\nu} = 2\phi_{22}l_\mu l_\nu + 2\phi_{00}n_\mu n_\nu + 2\phi_{20}m_\mu m_\nu + 2\bar{\phi}_{20}\bar{m}_\mu\bar{m}_\nu + 4\phi_{11}[l_{(\mu}n_{\nu)} + m_{(\mu}\bar{m}_{\nu)}]$$

$$- 4\phi_{21}l_{(\mu}m_{\nu)} - 4\bar{\phi}_{21}l_{(\mu}\bar{m}_{\nu)} - 4\phi_{10}n_{(\mu}m_{\nu)} - 4\bar{\phi}_{10}n_{(\mu}\bar{m}_{\nu)} + 6\Lambda\,[2l_{(\mu}n_{\nu)} - 2m_{(\mu}\bar{m}_{\nu)}]. \tag{17}$$

Using (5)–(10) and (16), Eq. (17) becomes

$$T_{\mu\nu} = 2\Bigg\{\frac{\dot{e}^2 + e\ddot{e} - \ddot{m}r)a^2\sin^2\theta}{2}\rho^2\bar{\rho}^2 + (e\dot{e} - \dot{m}r)r\rho^2\bar{\rho}^2 + a\ddot{a}\rho^2\bar{\rho}^2\left[\frac{\sin^2\theta}{2}\left(a^2 + \frac{\Upsilon}{2} - a^2\Upsilon\cos^2\theta\rho\bar{\rho}\right) - \Omega\right] + \dot{a}^2\rho^2\bar{\rho}^2$$

$$\times\left(-\tfrac{9}{4}a^2\Upsilon\sin^2\theta\cos^2\theta\rho\bar{\rho} + \tfrac{5}{2}a^2\sin^2\theta - \tfrac{3}{2}a^2\sin^2\theta\cos^2\theta\rho\bar{\rho} + 2a^4\Upsilon\sin^2\theta\cos^4\theta\rho^2\bar{\rho}^2 - \Omega + 2a^2\Omega\cos^2\theta\rho\bar{\rho} + a^2\right.$$

$$\left. - a^2r^4\sin^2\theta\rho^2\bar{\rho}^2 + \frac{\Upsilon}{2}\sin^2\theta\right) + \frac{a\dot{a}\rho^4\bar{\rho}^4}{2}\left[r^5\left(\tfrac{1}{2} - \tfrac{5}{2}\cos^2\theta\right) - 4\sin^2\theta r^4(\dot{m}r - e\dot{e}) + \frac{r^3}{2}(2mr - e^2)(1 + 5\cos^2\theta)\right.$$

$$\left. + a^2r^3(4 - \tfrac{5}{2}\sin^2\theta - 10\cos^2\theta) - 4\sin^2\theta\cos^2\theta\left(\dot{m} - \frac{e\dot{e}}{r}\right) + \tfrac{3}{2}\sin^2\theta\cos^2\theta + \frac{a^2r}{2}(2mr - e^2)(5\cos^2\theta\right.$$

$$\left. + 3\cos^4\theta - 2a^4r\cos^2\theta)\right]\Bigg\}l_\mu l_\nu + 4\,\mathrm{Re}\Bigg\{\Bigg[\Bigg[\ddot{a}\,\frac{a\sin^2\theta}{4}\rho^2 + \dot{a}^2\left(\frac{\sin^2\theta}{8}\rho^2 - a^2\,\frac{\sin^2\theta\cos^2\theta}{2}\rho^3\bar{\rho}\right) + a\dot{a}\sin^2\theta\rho^3$$

$$\times\left(1 + \rho\bar{\rho}r^2 - 2a^2r\cos^2\theta\rho\bar{\rho}^2 + \frac{ia\cos\theta}{2}(\bar{\rho} - 2\rho)\right)\Bigg]m_\mu m_\nu\Bigg\}$$

$$+ 4\Bigg\{\frac{e^2}{2}\rho\bar{\rho} - \ddot{a}\,\frac{a^3\sin^2\theta\cos^2\theta}{4}\rho^2\bar{\rho}^2 + \frac{\dot{a}\sin^2\theta\rho\bar{\rho}}{32}[1 + \rho^2\bar{\rho}^2(r^4 + a^4\cos^4\theta - 14r^2a^2\cos^2\theta)]$$

$$+ \frac{\dot{a}\rho^3\bar{\rho}^3ar\sin^2\theta}{8}(13r^2 + 5a^2\cos^2\theta)\Bigg\}(l_{(\mu}n_{\nu)} + m_{(\mu}\bar{m}_{\nu)})$$

$$- 8\,\mathrm{Re}\Bigg\{\Bigg[-\frac{ia\sin\theta}{2\sqrt{2}}\left(2e\dot{e} + \frac{\dot{m}}{\bar{\rho}}\right)\rho^2\bar{\rho}^2 + i\ddot{a}\,\frac{a\sin\theta}{4\sqrt{2}}\Omega\rho\bar{\rho}^2(1 + 2a^2\cos^2\theta\rho\bar{\rho}) + \dot{a}^2\,\frac{\sin\theta\cos\theta}{2\sqrt{2}}$$

$$\times a\rho^3\bar{\rho}^4(-ia^4\cos^3\theta - ia^2r^2\cos^3\theta + ir^4\cos\theta - 2ar^3\sin^2\theta) + \frac{\dot{a}\sin\theta}{4\sqrt{2}}\rho^3\bar{\rho}^4(10a^5\cos^3\theta - ir^5 - 2ia^4m\cos^4\theta)$$

$$+ ra^4(3i\cos\theta + 2i\cos^4\theta) - 2a^3(2mr - e^2)\cos^3\theta + a^3r^2(10\cos\theta - 2\sin^2\theta\cos\theta)$$

$$+ 6ia^2r^2m\cos^2\theta + a^2r^3(8i - 15i\sin^2\theta) + 4ar^2\cos\theta(2mr - e^2) - 2ir^3(2mr - e^2)]\, l_{(\mu}\bar{m}_{\nu)}\}$$

$$- 8\,\mathrm{Re}\left(\frac{\dot{a}}{2\sqrt{2}}\sin\theta(3ar\cos\theta - ir^2)\rho^2\bar{\rho}^2\, n_{(\mu}m_{\nu)}\right) + 6\Lambda\,(2l_{(\mu}n_{\nu)} - 2m_{(\mu}\bar{m}_{\nu)}). \tag{18}$$

III. EINSTEIN–MAXWELL FIELDS

In this sintion we shall assume that $a = \text{const}$, in which case we have, from (5)–(17):

$$\begin{aligned}
&\kappa = \epsilon = \sigma = \lambda = 0, && \pi = ia\sin\theta\rho^2/\sqrt{2}\\
&\rho = -(r - ia\cos\theta)^{-1}, && \beta = -\cot\theta\bar{\rho}/2\sqrt{2},\\
&\tau = -ia\sin\theta\rho\bar{\rho}/\sqrt{2}, && \alpha = \pi - \bar{\beta},\\
&\nu = i(e\dot{e} - \dot{m}r)a\sin\theta\rho^2\bar{\rho}/\sqrt{2}, && \gamma = \frac{(r-m)}{2}\rho\bar{\rho} + \mu,
\end{aligned} \tag{19}$$

$$\mu = \Upsilon\rho^2\bar{\rho}/2,$$

$$\phi_{00} = \phi_{10} = \phi_{20} = 0,$$

$$\phi_{11} = \frac{e^2}{2}\rho^2\bar{\rho}^2, \quad \phi_{12} = -\frac{ia\sin\theta}{2\sqrt{2}}(2e\dot{e} + \dot{m}/\bar{\rho})\rho^2\bar{\rho}^2, \tag{20}$$

$$\phi_{22} = \frac{(\dot{e}^2 + e\ddot{e} - \ddot{m}r)}{2}a^2\sin^2\theta\rho^2\bar{\rho}^2 + (e\dot{e} - \dot{m}r)r\rho^2\bar{\rho}^2,$$

$$\begin{aligned}
T_{\mu\nu} = {}& [(\dot{e}^2 + e\ddot{e} - \ddot{m}r)a^2\sin^2\theta + 2(e\dot{e} - \dot{m}r)r]\rho^2\bar{\rho}^2 l_\mu l_\nu\\
&+ 2e^2\rho^2\bar{\rho}^2(l_{(\mu}n_{\nu)} + m_{(\mu}\bar{m}_{\nu)})\\
&- i\sqrt{2}a\sin\theta(2e\dot{e} + \dot{m}/\bar{\rho}^2\bar{\rho}^2 l_{(\mu}\bar{m}_{\nu)}\\
&+ i\sqrt{2}a\sin\theta(2e\dot{e} + \dot{m}/\rho)\rho^2\bar{\rho}^2 l_\mu m_{\nu}.
\end{aligned} \tag{21}$$

In order to identify any term in (21) with an electromagnetic field it is necessary that it can be written as

$$_M T_{\mu\nu} = -F_{\mu\lambda}F_\nu{}^\lambda + \tfrac{1}{4}(F_{\alpha\beta}F^{\alpha\beta})g_{\mu\nu}, \tag{22}$$

where the Maxwell tensor $F^{\alpha\beta}$ should satisfy the Maxwell equations

$$F^{\alpha\beta}{}_{;\beta} = 4\pi j^\alpha, \quad F_{[\alpha\beta;\sigma]} = 0. \tag{23}$$

In the tetrad notation, the equations (22) and (23) read respectively:

$$\begin{aligned}
&{}_M T_{\mu\nu}\\
&= 2[|\phi_2|^2 l_\mu l_\nu + |\phi_0|^2 n_\mu n_\nu + \bar{\phi}_0\phi_2 m_\mu m_\nu + \phi_0\bar{\phi}_2\bar{m}_\mu\bar{m}_\nu]\\
&\quad + 4|\phi_1|^2[l_{(\mu}n_{\nu)} + m_{(\mu}\bar{m}_{\nu)}] - 4\bar{\phi}_1\phi_2 l_{(\mu}m_{\nu)} - 4\phi_1\phi_2 l_{(\mu}\bar{m}_{\nu)}\\
&\quad + 4\bar{\phi}_0\phi_1 n_{(\mu}m_{\nu)} - 4\phi_0\bar{\phi}_1 n_{(\mu}\bar{m}_{\nu)},
\end{aligned} \tag{24}$$

$$\begin{aligned}
D\phi_1 - \bar{\delta}\phi_0 &= (\pi - 2\alpha)\phi_0 + 2\rho\phi_1 - \kappa\phi_2 + 2\pi J_0,\\
\delta\phi_1 - \Delta\phi_0 &= (\mu - 2\gamma)\phi_0 + 2\tau\phi_1 - \sigma\phi_2 + 2\pi J_2,
\end{aligned} \tag{25}$$

$$\begin{aligned}
D\phi_2 - \bar{\delta}\phi_1 &= -\lambda\phi_0 + 2\pi\phi_1 + (\rho - 2\epsilon)\phi_2 + 2\pi J_3,\\
\delta\phi_2 - \Delta\phi_1 &= -\nu\phi_0 + 2\mu\phi_1 + (\tau - 2\beta)\phi_2 + 2\pi J_1,
\end{aligned}$$

where

$$\begin{aligned}
&\phi_0 \equiv F_{\mu\nu}l^\mu n^\nu, \quad \phi_2 \equiv F_{\mu\nu}\bar{m}^\mu n^\nu,\\
&\phi_1 \equiv \tfrac{1}{2}F_{\mu\nu}(l^\mu n^\nu + \bar{m}^\mu m^\nu),
\end{aligned} \tag{26}$$

and $J_m \equiv j^\alpha Z_{\alpha m}$ [note that the π appearing in the last terms of (25) is the usual constant and not the spin coefficient].

It is easy to see that only two terms in (21) can be written down as (22), namely, the term

$$[(\dot{e}^2 + e\ddot{e} - \ddot{m}r)a^2\sin^2\theta + (2e\dot{e} - \dot{m}r)r]\rho^2\bar{\rho}^2 l_\mu l_\nu \tag{27}$$

and the term

$$2e^2\rho^2\bar{\rho}^2(l_{(\mu}n_{\nu)} + m_{(\mu}\bar{m}_{\nu)}). \tag{28}$$

For (27) we have

$$\phi_0 = \phi_1 = 0$$

$$\phi_2 = [(\dot{e}^2 + e\ddot{e} - \ddot{m}r)a^2\sin^2\theta + (e\dot{e} - \dot{m}r)2r]^{1/2}\rho\bar{\rho}\exp(i\psi). \tag{29}$$

Yet, it is impossible to satisfy the Maxwell equations with (29) if $\dot{e} = \dot{m} = 0$.

Considering next the term given by (28) we put

$$\phi_0 = \phi_2 = 0, \quad \phi_1 = \sqrt{e^2/2}\rho\bar{\rho}\exp(i\psi) \tag{30}$$

(where ψ is an arbitrary real function).

We shall see that Maxwell equations are satisfied only if $\dot{e} = 0$ or $\dot{e} \neq 0$ and spacelike currents are admitted.

In fact, assuming $\dot{e} = 0$ in which case $J_m = 0$, the Maxwell equations become:

$$\begin{aligned}
&D\phi_1 = 2\rho\phi_1, \quad \delta\phi_1 = 2\tau\phi_1,\\
&-\bar{\delta}\phi_1 = 2\pi\phi_1, \quad -\Delta\phi_1 = 2\mu\phi_1,
\end{aligned} \tag{31}$$

and a straightforward calculation shows that if we choose $\psi = -2\tan^{-1}(r/a\cos\theta)$ in (30), Eq. (31) are automatically satisfied.

Now, the Maxwell scalars, with the ψ chosen as above, become

$$\phi_0 = \phi_2 = 0, \quad \phi_1 = \sqrt{e^2/2}/(r - ia\cos\theta) \tag{32}$$

and the Maxwell tensor is given by

$$F_{\mu\nu} = \frac{2(2e^2)^{1/2}}{(r^2 + a^2\cos^2\theta)^2}[(r^2 - a^2\cos^2\theta)(\delta^0_{[\mu}\delta^1_{\nu]}$$

$$- a \sin^2\theta \delta^3_{[\mu}\delta^1_{\nu]}$$

$$- 2a \sin\theta \cos\theta (a\delta^0_{[\mu}\delta^2_{\nu]} + \Omega\delta^2_{[\mu}\delta^3_{\nu]}]. \tag{33}$$

This is an expected result, since (33) describes the electromagnetic field outside a usual (stationary) Kerr–Newman source.

Let us now consider the case $\dot{e}\neq 0$. The Maxwell equations to be satisfied are:

$$D\phi_1 = 2\pi\rho\phi_1 + 2\pi J_0, \quad \delta\phi_1 = 2\tau\phi_1 + 2\pi J_2,$$
$$- \bar{\delta}\phi_1 = 2\pi\phi_1 + 2\pi J_3, \quad - \Delta\phi_1 = 2\mu\phi_1 + 2\pi J_1, \tag{34}$$

where, as before

$$\phi_1 = (e^2/2)^{1/2}\rho\bar{\rho}\exp(i\psi),$$

but now $e = e(u)$. Instead of (34) we shall consider the equivalent system:

$$\frac{\partial}{\partial u}[(e^2)^{1/2}\exp(i\psi)] = \pi\sqrt{2}\left(\Upsilon J_0 - \frac{2}{\rho\bar{\rho}}J_1\right) - 4i\pi a \sin\theta \operatorname{Re}(J_3/\rho).$$

$$\frac{\partial}{\partial r}[(e^2)^{1/2}\exp(i\psi)] = \frac{2\pi\sqrt{2}}{\rho\bar{\rho}}J_0 - 2ia\cos\theta\rho\bar{\rho}[(e^2)^{1/2}\exp(i\psi)].$$

$$\frac{\partial}{\partial\theta}[(e^2)^{1/2}\exp(i\psi)] = \frac{4i\pi}{\rho\bar{\rho}}\operatorname{Im}(J_3/\rho) - 2iar\sin\theta\rho\bar{\rho}[(e^2)^{1/2}\exp(i\psi)].$$

$$\frac{\partial}{\partial\phi}[(e^2)^{1/2}\exp(i\psi)] = 4i\pi\Omega\sin\theta\operatorname{Re}(J_3/\rho) - \pi\sqrt{2}a\sin^2\theta\left(\Upsilon J_0 - \frac{2}{\rho\bar{\rho}}J_1\right). \tag{35}$$

From the sinond equation of (35) we get

$$J_0 = 0 \quad \text{or} \quad \psi = (2n+1)\frac{\pi}{2}.$$

The sinond possibility is excluded by the third equation in (35), and thus

$$J_0 \equiv j^\alpha l_\alpha = 0, \quad \frac{\partial\psi}{\partial r} = -2a\cos\theta\rho\bar{\rho}. \tag{36}$$

Again, from the third equation in (34) and (36) it follows that

$$\operatorname{Im}(J_3/\rho) = 0, \tag{37}$$

$$\frac{\partial\psi}{\partial\theta} = -2are\sin\theta. \tag{38}$$

Feeding back (37) and (38) in the last equation of (34) we get:

$$\sin\psi\frac{\partial\psi}{\partial\varphi} = -\frac{2\pi(2/e^2)^{1/2}}{\rho\bar{\rho}}a\sin^2\theta J_1, \tag{39}$$

$$\cos\psi\frac{\partial\psi}{\partial\varphi} = \frac{4\pi\Omega\sin\theta}{(e^2)^{1/2}}\frac{J_3}{\rho}. \tag{40}$$

It follows from (39) and (49) that if $J_3 = 0$, then $J_1 = 0$. Since the current is not zero, we should have $J_3\neq 0$ and thus

$$j^\alpha j_\alpha = 2J_0J_1 - 2|J_3|^2 = -2|J_3|^2 < 0. \tag{41}$$

Let us now consider the special case $a = 0$, m, e variable. The system (34) is transformed into:

$$\frac{\partial}{\partial u}[(e^2)^{1/2}\exp(i\psi)] = \pi\sqrt{2}\left(\Upsilon J_0 - \frac{2}{\rho\bar{\rho}}J_1\right),$$
$$\frac{\partial}{\partial r}[(e^2)^{1/2}\exp(i\psi)] = 2\pi\sqrt{2}J_0/\rho\bar{\rho},$$
$$\frac{\partial}{\partial\theta}[(e^2)^{1/2}\exp(i\psi)] = \frac{4i\pi}{\rho\bar{\rho}}\operatorname{Im}(J_3/\rho),$$
$$\frac{\partial}{\varphi}[(e^2)^{1/2}\exp(i\psi)] = 4i\pi\Omega\sin\theta\operatorname{Re}(J_3/\rho), \tag{42}$$

where $\rho \equiv -1/r$, $\Upsilon \equiv r^2 + e^2 - 2mr$, and $\Omega \equiv r^2$.

From the second equation we get

$$J_0 = 0, \quad \text{or} \quad \psi = (2n+1)\frac{\pi}{2}.$$

Taking $\psi = (2n+1)\pi/2$ one gets, from the last three equations of (42),

$$J_0 = J_3 = 0,$$

and so

$$j^\alpha j_\alpha = 0. \tag{43}$$

the election $J_0 = 0$ gives the same result by virtue of the two last equations of (42).

Thus, the physical system described by (30) consists of a flux of charged particles traveling with the speed of light.

IV. CONCLUSIONS

We have extended the Kerr–Newman metric assuming that the three parameters of the solution are arbitrary functions of the timelike coordinate (u).

In this case the energy–momentum has a nonzero trace but cannot be associated with a scalar field. Since the trace of the energy–momentum tensor depends only on $\dot{a}$ and $\ddot{a}$, in the case m, e constant, a variables, changes in the angular momentum of the source occurs, not by emission of radiation, but by the interaction with the "physical" entity, described by (21), which we could not identify.

In the case a = const, e, m variable, identification of different terms in the energy–momentum tensor as Maxwell fields yields negative results, except in the subcase e = const, in which case the usual electromagnetic contribution of the Kerr–Newman source is recovered.

In the case e variable, solutions of the Maxwell equations can be obtained only if the spacelike currents are accepted (charged tachyons). This result agrees with the point of view that the charge in the Kerr–Newman metric[10,11] is hidden by the horizon, so for the charges to get out from the

singularity to a region outside the horizon, the corresponding current vector should be spacelike.

In the more restricted case, e, m variable, $a = 0$, the situation already described in the literature[7] is recovered.

It should be stressed that, since the propagation of electromagnetic radiation in a curved space–time is always accompanied by a backscattering radiation,[12–14] metrics of radiating sources should allow tail terms in the Maxwell scalars. This is not the case in our metric nor in the Vaidya's or in the Carmeli–Kaye's metric.

In fact, when Carmeli and Kaye state that in the case $a = \text{const}$, m variable, $e = 0$, one of the terms of the energy–momentum tensor asymptotically may be interpreted in terms of the Maxwell field; this should be taken to mean that this part is a Maxwell term up to tail terms.

In other words, our feeling is that to just take the parameter entering the metric as functions of the timelike coordinate, does not generalize the solutions enough as to describe radiation process.

[1]P.C. Vaidya, Curr. Sci. **12**, 183 (1943).
[2]M. Carmeli and M. Kaye, Phys. Lett. A **53**, 11 (1975).
[3]L. Herrera, Nuovo Cimento B **43**, 283 (1978).
[4]C.W. Misner, Phys. Rev. **137**, B1360 (1965).
[5]C.W. Misner, R. Lindquist, and R. Schwarz, Phys. Rev. **137**, B1364 (1965).
[6]E.T. Newman, E. Couch, K. Chinapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[7]W. Bonnor and P. Vaidya, Gen. Rel. Grav. **1**, 127 (1970).
[8]E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[9]S. Bose, J. Math. Phys. **16**, 772 (1975).
[10]E. Newman and A. Janis, J. Math. Phys. **6**, 915 (1965).
[11]W. Israel, Phys. Rev. D **2**, 641 (1970).
[12]W. Couch and W. Hallid, J. Math. Phys. **12** 2170 (1971).
[13]A. Papapetrou, J. Phys. A **8**, 313 (1975).
[14]W. Kundt and E. Newman, J. Math. Phys. **9**, 2193 (1968).

The semiclassical expansion of the anharmonic-oscillator propagator[a)]

Maurice M. Mizrahi

Center for Naval Analyses of the University of Rochester, 1401 Wilson Blvd., Arlington, Virginia 22209
(Received 5 September 1978)

This paper shows how to calculate the terms of a semiclassical (WKB) expansion of the quantum-mechanical propagator corresponding to the quartic anharmonic-oscillator potential, $V = m\omega^2 q^2/2 + \lambda q^4/4$. This nonperturbative treatment expresses each term in the series as a path integral, which is then evaluated in the framework of a formalism, introduced by C. DeWitt-Morette, which does not entail the usual time-slicing operation followed by a limiting procedure. The Gaussian measure used absorbs all the quadratic terms in the expansion of the action functional about a classical path. The covariance of this Gaussian measure is Feynman's Green function for the small-disturbance operator of the system. This function can be obtained by varying the constants of integration in the classical solution, and therefore the coefficients of the expansion depend only on this classical solution. If the latter is chosen to be the one which tends to its harmonic counterpart when $\lambda \to 0$, then it is seen that the propagator also tends to its harmonic counterpart when $\lambda \to 0$.

I. INTRODUCTION

The one-dimensional quartic anharmonic oscillator is a particle of mass m in a potential given by

$$V(q) = \frac{m\omega^2 q^2}{2} + \frac{\lambda q^4}{4}. \tag{1}$$

It is an important model in physics as a prototype nonlinear field theory and has generated a great deal of activity in recent years for several reasons. First, it is a simple example of a perturbation which causes the associated quantum-mechanical quantities to be nonanalytic in the coupling constant λ. Therefore, the usual perturbation series in powers of the coupling constant are divergent, although it has been shown[1] that the Padé approximants of the Rayleigh–Schrödinger series for the energy levels converge to the correct eigenvalues of the Hamiltonian, which has a positive-definite spectrum for $\lambda > 0$. The anharmonic oscillator is also the simplest nonlinear interaction which still yields plane-wave periodic solutions in the associated $\lambda\varphi^4$ field theory, and even admits of a restricted superposition principle.[2]

While the energy spectrum has been studied rather extensively,[1,3,4] the propagator $K \equiv \langle q_b, t_b | q_a, t_a \rangle$, or probability amplitude that a particle at q_a at time t_a will be at q_b at time t_b, has not. The purpose of this paper is to show how to calculate the terms of a semiclassical (WKB) expansion of this propagator (in powers of $\hbar$). This treatment, of necessity nonperturbative since it does not hinge on any expansion in powers of λ, expresses each term in the series as a path integral. The latter is then evaluated in the framework of a formalism where the usual approach of time-slicing followed by a limiting procedure is replaced by a more tractable definition, introduced by C. DeWitt-Morette,[5] which greatly simplifies calculations. This approach enabled us to systematically generate all the terms in the semiclassical expansion, which represents some progress over previous studies of approximating the anharmonic oscillator propagator by path-integral techniques (Lam,[6] Sarkar,[7] Mathews and Seshadri[8]).

First, the classical system is studied: The classical paths joining two fixed end points are calculated and the limit of zero coupling constant is discussed. Then, the classical action and other elements of the WKB expansion (Jacobi commutator, Van Vleck–Morette function, Feynman's Green function) are derived explicitly, and their connection with the small-disturbance equation investigated. Finally, the path integrals constituting the terms of the WKB expansion are exhibited and reduced to definite integrals over known functions, first for an arbitrary potential, then for the anharmonic oscillator.

II. THE CLASSICAL SYSTEM

The potential

The potential, given in (1), is sketched below for $\lambda > 0$ and $\lambda < 0$ (see Fig. 1).

The potential well is always present for $\lambda < 0$, so there will always be harmonic motion in some neighborhood of the origin. As $|\lambda|$ decreases, the well gets deeper and deeper, the maxima go higher and higher, and the points where the potential crosses the horizontal axis are rejected farther and farther. The drastic change in the shape of V as λ changes sign is the cause for the nonanalyticity in λ. For $\lambda > 0$, there will always be a stable ground state, whereas for $\lambda < 0$, the ground state is unstable, as there is a finite probability for the particle to "leak out" of the well. The failure of perturbation theory is due to the fact that at large distances the q^4 term will always dominate the q^2 term, regardless of how small λ is.

Dynamical equation

The dynamical equation for the classical path $q_c(t)$ is

$$\ddot{q}_c(t) + \omega^2 q_c(t) + \frac{\lambda}{m} q_c^3(t) = 0. \tag{2}$$

It can be solved in terms of the (biperiodic) elliptic functions. Our source for the latter is Byrd and Friedman's handbook.[9]

[a)]This paper is based in part on the author's Ph.D. dissertation, "An Investigation of the Feynman Path Integral Formulation of Quantum Mechanics", The University of Texas at Austin, 1975.

 0022-2488/79/050844-12$01.00

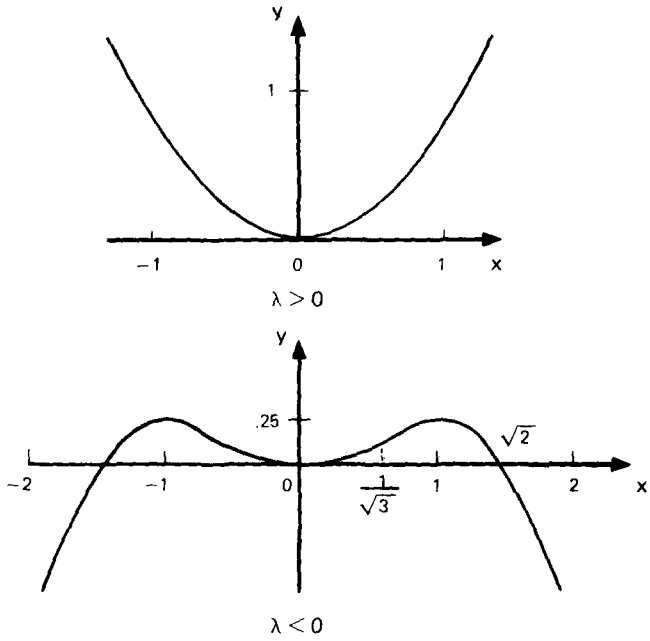

FIG. 1. The anharmonic oscillator potential $V(q) = m\omega^2 q^2/2 + \lambda q^4/4$ [$y = x^2/2 \pm x^4/4$; $y \equiv V(q)|\lambda|/m^2\omega^2$, $x \equiv q|\lambda|^{1/2}/\omega m^{1/2}$].

We choose the following form for the solution of (2):

$$q_c(t) = q_m \,\mathrm{cn}[\Omega(t - t_0), k], \tag{3}$$

where

$$\Omega^2 \equiv \omega^2 + \lambda q_m^2, \quad k^2 \equiv \frac{\lambda}{2}\left(\frac{q_m}{\Omega}\right)^2.$$

This corresponds to the case where the particle is released at q_m at time $t = t_0$ with no initial velocity. (For simplicity, we take the mass m equal to 1; it can always be restored by replacing λ by λ/m.) The form (3) assumes $\lambda > 0$. If $\lambda < 0$, the form $q_c(t) = q_m \,\mathrm{sn}[\Omega(t - t_0), k]$, with $\Omega^2 \equiv \omega^2 + \lambda q_m^2/2$ and $k^2 \equiv -\lambda q_m^2/2\Omega^2$, must be used. We assume $\lambda > 0$ here for definiteness. Note that the modulus k lies always between 0 and $1/\sqrt{2}$ ($= 0.707...$). If we take the modulus k and the phase t_0 to be our constants of integration, we get

$$q_c(t) = \left[\frac{2k^2\omega^2}{\lambda(1 - 2k^2)}\right]^{1/2} \mathrm{cn}\left[\frac{\omega(t - t_0)}{(1 - 2k^2)^{1/2}}, k\right]. \tag{4}$$

Classical paths

The classical paths of interest for the calculation of the propagator are those for which the initial and final positions are specified:

$$q_c(t_a) = q_a, \quad q_c(t_b) = q_b.$$

Substituting these conditions in (4) yields the relationship between the set (k, t_0) and the set (q_a, q_b):

$$\frac{\omega(t_a - t_0)}{(1 - 2k^2)^{1/2}} = \pm \,\mathrm{cn}^{-1}\left\{\frac{q_a}{q_m}, k\right\} + 4nK(k), \tag{5a}$$

$$\frac{\omega(t_b - t_0)}{(1 - 2k^2)^{1/2}} = \pm \,\mathrm{cn}^{-1}\left\{\frac{q_b}{q_m}, k\right\} + 4n'K(k), \tag{5b}$$

where n and n' are integers,

$$q_m \equiv [2k^2\omega^2/\lambda(1 - 2k^2)]^{1/2}, \tag{6}$$

and $K(k)$ is the quarter period of the cn function. Subtracting (5a) from (5b) yields the final transcendental equations giving k in terms of q_a and q_b:

$$\pm \frac{\omega T}{(1 - 2k^2)^{1/2}} = \varphi_\pm(k^2) + 4NK(k), \tag{7}$$

where any combination of signs is permitted, $T \equiv t_b - t_a$, N is an integer, and

$$\varphi_\pm(k^2) = \mathrm{cn}^{-1}(q_a/q_m) \pm \mathrm{cn}^{-1}(q_b/q_m).$$

Equation (7) must be solved graphically for k [t_0 is then determined, for example, by Eq. (5a)]. Since $\mathrm{cn}^{-1}u$ is defined only for $u \in [-1, 1]$, we must have $|q_a| \leqslant q_m$ and $|q_b| \leqslant q_m$. Thus, in addition to the upper bound $\frac{1}{2}$ on k^2, we have a lower bound:

$$k^2_{\min} \leqslant k^2 \leqslant \tfrac{1}{2},$$

where

$$k^2_{\min} \equiv \frac{\lambda}{2(\lambda + \omega^2\alpha^2)}, \quad \alpha \equiv \frac{1}{\max(|q_a|, |q_b|)}.$$

Note that cn^{-1} is always positive. It monotonically decreases from $\mathrm{cn}^{-1}(-1) = 2K(k)$ to $\mathrm{cn}^{-1}(1) = 0$, with an inflexion point at $(0, K(k))$. When $0 \leqslant k^2 \leqslant \frac{1}{2}$, $K(k)$ monotonically increases from $K(0) \simeq 1.58$ to $K(1/\sqrt{2}) \simeq 1.85$.

A sample graphical solution of (7) is shown in Fig. 2, for $\omega = T = q_a = q_b = 1$. The cases $\lambda = 0.001$, 0.5, and 1 are shown. The curve $\omega T/(1 - 2k^2)^{1/2}$ intersects $\varphi_+(k^2)$ once, twice, or not at all. Each intersection gives the modulus k for a possible classical path such that $q(t_a) = q_a$ and $q(t_b) = q_b$. As N increases beyond a certain point N_0, each of the curves $\varphi_+(k^2) + 4NK(k)$ (one for each N) intersects $\omega T/(1 - 2k^2)^{1/2}$ twice for each $N > N_0$. Therefore, there is always a countably infinite number of paths, with a cluster point at $k^2 = \frac{1}{2}$. Another set, independent of λ, is obtained from the intersection of the curve $\omega T/(1 - 2k^2)^{1/2}$ with the curves $4NK(k)$. The higher the k, the higher the amplitude of the corresponding path [as revealed by Eq. (4)].

Behavior as $\lambda \to 0$: We shall be particularly concerned with the behavior of our expressions as λ approaches 0. What happens to the classical solution as $\lambda \to 0$? For initial boundary conditions, it appears, according to (3), that we retrieve harmonic motion: indeed, as $\lambda \to 0$, $k \to 0$, $\Omega \to \omega$, and cn $\to$ cos. However, for other boundary conditions, it appears, according to (4), that we have a $1/\sqrt{\lambda}$ singularity as $\lambda \to 0$: Indeed, for arbitrary values of the constants of integration (say $k^2 = 0.3$ and $t_0 = 2$ s), (4) indicates that $q_c(t) \sim 1/\sqrt{\lambda}$ as $\lambda \to 0$. Is harmonic motion irretrievable then as a limiting case?

The answer is no. The reason is that only *physical* boundary conditions (such as position and velocity at certain times) are acceptable.[10] $k^2 = 0.3$ is *not* a physical boundary condition. When the latter are inserted, *k will depend on* λ in such a manner as to make at least one classical path $q_c(t)$ reduce to harmonic motion when $\lambda \to 0$.

In the case of endpoint boundary conditions, (4) shows

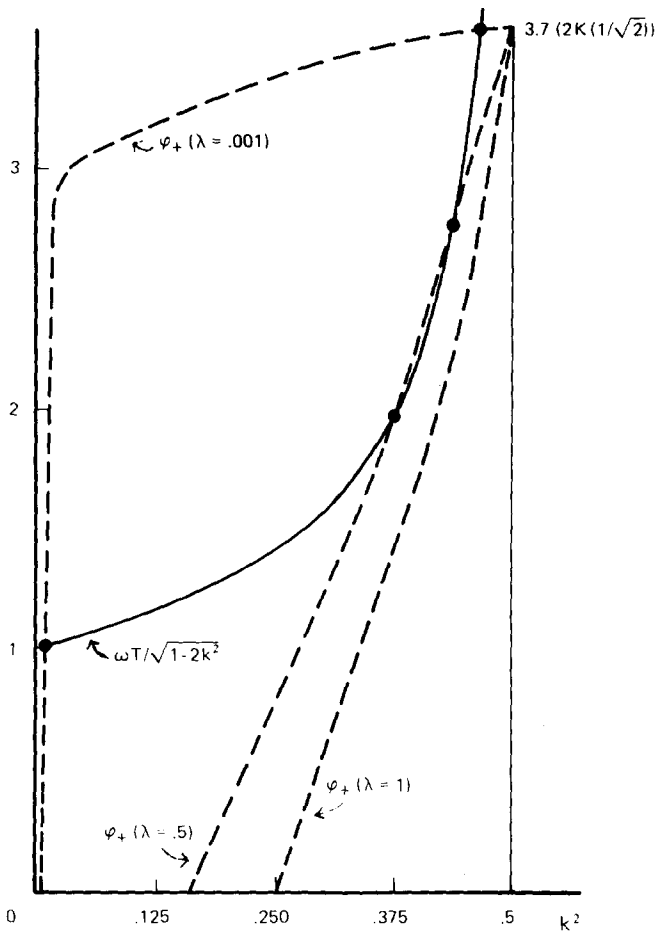

FIG. 2. Classical paths for the anharmonic oscillator. Each intersection (black dots) gives a value of k which corresponds to a classical solution of the dynamical equation for fixed-end-point boundary conditions.

that the only way that $q_c(t)$ can retain its constant, preassigned values at t_a and t_b is if k^2 goes to 0 as fast as λ. The ratio k^2/λ is then an arbitrary constant A, which may be dependent on ω, and (4) becomes $q_c(t) = A\cos\omega(t - t_0)$, which is harmonic motion. Figure 2 shows that as λ approaches 0 there is always one solution k^2 which also approaches 0. This solution, which we call $q_{c0}(t)$, is the lowest-amplitude (or lowest-energy) path, and coincides, when $\lambda = 0$, with the (generally) unique harmonic-oscillator path between the two fixed endpoints. The other paths correspond to values of k which do not go to 0 with λ, and hence their amplitudes increase without bound as $\lambda \to 0$. Their graph becomes, in the limit, a set of parallel lines perpendicular to the t axis, one of which goes through t_a and another through t_b.

Our semiclassical expansion of the propagator will be about this regular path $q_{c0}(t)$. Since all the coefficients will depend, directly or indirectly, on q_{c0} alone, the anharmonic propagator will tend toward the harmonic propagator as the coupling constant tends to 0.

Classical action

The classical action (or action functional evaluated at a classical path) for the anharmonic oscillator is needed for the WKB approximation. It is given by

$$S_c \equiv \int_{t_a}^{t_b} L(\dot{q}_c(t), q_c(t), t)\,dt$$

$$= \int_{t_a}^{t_b} [\tfrac{1}{2}\dot{q}_c^2(t) - \tfrac{1}{2}\omega^2 q_c^2(t) - \tfrac{1}{4}\lambda q_c^4(t)]\,dt.$$

Using the integrals 312.02 (p. 193), 361.02 (p. 212), 312.04 (p. 193) of Ref. 9 and the formula

$$E(u') - E(u) = E(u' - u) - k^2\,\mathrm{sn}u\cdot\mathrm{sn}u'\cdot\mathrm{sn}(u' - u), \quad (8)$$

[derived from formulas 116.01 (p.13) and 123.01 (p.23) of Ref. 9] we obtain the answer:

$$S_c = \frac{-2\omega^3}{3\lambda(1 - 2k^2)^{1/2}} E\left(\frac{\omega T}{(1 - 2k^2)^{1/2}}\right)$$

$$+ \frac{2\omega^3 k^2}{3\lambda(1 - 2k^2)^{1/2}}\left\{\mathrm{sn}(u_a)\,\mathrm{sn}(u_b)\,\mathrm{sn}\frac{\omega T}{(1 - 2k^2)^{1/2}}\right.$$

$$+ \frac{1}{1 - 2k^2}[\mathrm{sn}(u_a)\,\mathrm{cn}(u_a)\,\mathrm{dn}(u_a) - \mathrm{sn}(u_b)$$

$$\left.\times\mathrm{cn}(u_b)\,\mathrm{dn}(u_b)]\right\} + \frac{\omega^4(1 - k^2)(2 - 3k^2)T}{3\lambda(1 - 2k^2)^2}, \quad (9)$$

where $u_{a/b} \equiv \omega(t_{a/b} - t_0)/(1 - 2k^2)^{1/2}$.

Behavior when $\lambda \to 0$: Let us look at the behavior of S_c as $\lambda \to 0$ along the path q_{c0}, where $k^2 \to 0$ as $\lambda \to 0$ such that k^2/λ is a constant. Using the fact that $E(u) = u + O(k^2)$, we easily see that S_c is regular at $\lambda = 0$, and reduces to the classical action for the harmonic oscillator.

III. THE QUANTUM SYSTEM

The small-disturbance equation

Just as the classical system is dominated by the dynamical (or Euler–Lagrange) equation, the quantum system is dominated by the small-disturbance (or Jacobi) equation. The latter is the equation satisfied by a small variation in the classical path, obtained, for example, by a small change in a constant of integration, such as the total energy or an end point. The small-disturbance equation is studied in more detail in Appendix A and in Refs. 11 and 5(c). For the anharmonic oscillator, it is

$$\left[-\frac{d^2}{dt^2} - \omega^2 - 3\lambda q_c^2(t)\right] f(t) = 0. \quad (10)$$

Solutions of the small-disturbance equation

Solutions of the small-disturbance equation can always be generated by differentiating the classical solution with respect to a constant of integration. This simple procedure was known to Jacobi,[12] but it seems to be sometimes forgotten today, as one still finds attempts at solving the equation

directly; for example, Sarkar[7] has undertaken this very difficult task for the anharmonic oscillator [Eq. (10)].

The functions we will need for the path-integral treatment of the propagator are the Jacobi commutator $J(t,t')$, the Van Vleck–Morette (VVM) function $M(t_a,t_b)$, and the Feynman Green function $G(t,t')$. Their expressions are given below, followed by their definition and derivation.

Jacobi commutator:

$$J(t,t') = \frac{(1-2k^2)^{3/2}}{\omega}\,\mathrm{sn}u\,\mathrm{sn}u'\,\mathrm{dn}u\,\mathrm{dn}u'$$
$$\times\left[\frac{-1}{1-2k^2}\left(\frac{\mathrm{cn}u'}{\mathrm{sn}u'\,\mathrm{dn}u'}-\frac{\mathrm{cn}u}{\mathrm{sn}u\,\mathrm{dn}u}\right)+\frac{u'-u}{1-2k^2}\right.$$
$$-\frac{E(u'-u)}{k'^2}+\frac{k^2}{k'^2}\left(\frac{\mathrm{sn}u'\mathrm{cn}u'}{\mathrm{dn}u'}-\frac{\mathrm{sn}u\,\mathrm{cn}u}{\mathrm{dn}u}\right.$$
$$\left.\left.+\mathrm{sn}u\,\mathrm{sn}u'\,\mathrm{sn}(u'-u)\right)\right], \tag{11}$$

where

$$u\equiv\frac{\omega(t-t_0)}{(1-2k^2)^{1/2}},\quad u'\equiv\frac{\omega(t'-t_0)}{(1-2k^2)^{1/2}},\quad k'^2\equiv 1-k^2;$$

VVM function:

$$M(t_a,t_b) = [J(t_a,t_b)]^{-1}; \tag{12}$$

Feynman's Green function:

$$G(t,t') = \frac{J(t',t_a)J(t_b,t)Y(t-t')+J(t,t_a)J(t_b,t')Y(t'-t)}{J(t_a,t_b)}. \tag{13}$$

Definitions and derivations

The Jacobi commutator: This function $J(t,t')$ of two variables can be defined as follows: The unique, retarded Green function of the small-disturbance operator, satisfying

$$\left[-\frac{d^2}{dt^2}-\omega^2-3\lambda q_c^2(t)\right]G(t,t') = \delta(t-t'), \tag{14}$$

is $G^{-}(t,t') = J(t,t')Y(t-t')$, where $Y(x)\equiv 1$ for $x>0$ and 0 otherwise. $J(t,t')$ is antisymmetric and satisfies the small-disturbance equation in both t and t'. It is called the commutator because, as shown in Appendix A, it can be written as a Poisson bracket of position at different times with respect to initial (or final) position and momentum; when the system is quantized, this expression becomes the commutator. For example, for initial boundary conditions, we have

$$J(t,t') = \frac{\partial q_c(t)}{\partial q_a}\frac{\partial q_c(t')}{\partial p_a}-\frac{\partial q_c(t')}{\partial q_a}\frac{\partial q_c(t)}{\partial p_a}$$
$$\equiv\{q_c(t),q_c(t')\}_{(q_a,p_a)}\to\frac{1}{i\hbar}[\mathbf{Q}(t),\mathbf{Q}(t')]. \tag{15}$$

For any two convenient constants of integration α_1 and α_2 we can write the commutator as (see proof in Appendix A):

$$J(t,t') = \left[\frac{\partial q_c(t)}{\partial\alpha_1}\frac{\partial q_c(t')}{\partial\alpha_2}-\frac{\partial q_c(t')}{\partial\alpha_1}\frac{\partial q_c(t)}{\partial\alpha_2}\right]$$
$$\times\left[\frac{\partial q_c(t_b)}{\partial\alpha_1}\frac{\partial p_c(t_b)}{\partial\alpha_2}-\frac{\partial p_c(t_b)}{\partial\alpha_1}\frac{\partial q_c(t_b)}{\partial\alpha_2}\right]^{-1} \tag{16}$$

(or a similar expression with t_b replaced by t_a), where $p_c(t)$ is the classical momentum [equal to $\dot q_c(t)$ for the anharmonic oscillator]. We will use this formula with $q_c(t)$ given by (4) and $\alpha_1 = k$, $\alpha_2 = t_0$. The velocity is given by

$$\dot q_c(t) = \frac{-k\omega^2}{1-2k^2}\sqrt{\frac{2}{\lambda}}\,\mathrm{sn}u\cdot\mathrm{dn}u = -\frac{\partial q_c(t)}{\partial t_0},$$

where u is defined in (11). The formulas for differentiating the elliptic functions with respect to the modulus k are found in Ref. 9 (710.51-3, p. 283). Since the argument of the elliptic functions also depends on k, the chain rule must be used to evaluate $\partial q_c(t)/\partial k$ and $\partial\dot q_c(t_b)/\partial k$. We obtain

$$\frac{\partial q_c(t)}{\partial k} = \omega\left(\frac{2}{\lambda}\right)^{1/2}\frac{\mathrm{cn}u}{(1-2k^2)^{3/2}}-\left(\frac{2k^2\omega^2}{\lambda(1-2k^2)}\right)^{1/2}$$
$$\times\mathrm{sn}u\cdot\mathrm{dn}u\cdot\left\{\frac{2k\omega(t-t_0)}{(1-2k^2)^{3/2}}+\frac{1}{kk'^2}[-E(u)\right.$$
$$\left.+k'^2u+k^2\,\mathrm{sn}u\cdot\mathrm{cn}u/\mathrm{dn}u]\right\}.$$

The denominator in (16) is calculated to be:

$$\frac{\partial q_b}{\partial k}\frac{\partial\dot q_c(t_b)}{\partial t_0}-\frac{\partial\dot q_c(t_b)}{\partial k}\frac{\partial q_b}{\partial t_0} = \frac{2\omega^4 k}{\lambda(1-2k^2)^3}.$$

These formulas, along with (8), lead us to the stated expression (11) for $J(t,t')$. We see that for $q_c = q_{c0}$, i.e., when k^2 goes to 0 with λ, we have $E(u)\to u$, $\mathrm{dn}u\to 1$, $\mathrm{sn}u\to\sin u$, $\mathrm{cn}u\to\cos u$, $u\to\omega(t-t_0)$, and $J(t,t')\to\omega^{-1}\sin\omega(t'-t)$, which is the harmonic-oscillator commutator function.

The VVM function: The WKB approximation to the propagator is given by the well-known formula

$$K_{\mathrm{WKB}} = (M/2\pi i\hbar)^{1/2}\exp(iS_c/\hbar), \tag{17}$$

where

$$M\equiv-\frac{\partial^2 S_c}{\partial q_a\partial q_b} = -\frac{\partial\dot q_c(t_b)}{\partial q_a} \tag{18}$$

is the Van Vleck–Morette function. The second expression for M, which will be used in the evaluation, uses the fact that $\partial S_c/\partial q_b + p_c(t_b) = \dot q_c(t_b)$.[13] Therefore, to get M in terms of k and t_0, we must use the chain rule:

$$M = -\frac{\partial\dot q_c(t_b)}{\partial k}\frac{\partial k}{\partial q_a}-\frac{\partial\dot q_c(t_b)}{\partial t_0}\frac{\partial t_0}{\partial q_a}. \tag{19}$$

In order to calculate M, we must express $\partial k/\partial q_a$ and $\partial t_0/\partial q_a$ in terms of $\partial q_a/\partial k$, $\partial q_a/\partial t_0$, etc. Since we must have

$$\begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix}\begin{pmatrix} u_1 & u_2 \\ u_3 & u_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

where

$$k_1 \equiv \frac{\partial q_a}{\partial k}, \quad k_2 \equiv \frac{\partial q_a}{\partial t_0}, \quad u_1 \equiv \frac{\partial k}{\partial q_a}, \quad u_2 \equiv \frac{\partial k}{\partial q_b},$$

$$k_3 \equiv \frac{\partial q_b}{\partial k}, \quad k_4 \equiv \frac{\partial q_b}{\partial t_0}, \quad u_3 \equiv \frac{\partial t_0}{\partial q_a}, \quad u_4 \equiv \frac{\partial t_0}{\partial q_b},$$

we can easily solve for the u's in terms of the k's, to get

$$u_1 = \frac{k_4}{\mathscr{D}}, \quad u_2 = -\frac{k_2}{\mathscr{D}}, \quad u_3 = -\frac{k_3}{\mathscr{D}}, \quad u_4 = \frac{k_1}{\mathscr{D}},$$

where $\mathscr{D} \equiv k_1k_4 - k_2k_3$. Substituting this result in (19), and comparing with (16), we see that we get the value of M stated in (12), namely $M = [J(t_a,t_b)]^{-1}$.

Feynman's Green function: Feynman's Green function $G(t,t')$, satisfying (14), is the unique Green function of the small-disturbance operator which vanishes at both end-points. It is important for our treatment because it is the covariance of the Gaussian measure used to express the propagator as a path integral. As was stated before (and proved in Appendix A), $G^{-}(t,t') \equiv J(t,t')Y(t-t')$, with J as in (16), satisfies (14). The function

$$G(t,t') \equiv J(t,t')Y(t-t') + \frac{J(t,t_a)J(t_b,t')}{J(t_a,t_b)}$$

is also a Green function, since the addition to $G^{-}(t,t')$ is a homogeneous solution (of the small-disturbance equation) in t and t'. Further, $G(t_a,t') = G(t_b,t') = 0$. Therefore, $G(t,t')$ is Feynman's Green function. To put it in the form given in (13) requires use of the identity

$$J(t,t') \equiv \frac{J(t',t_a)J(t_b,t) - J(t,t_a)J(t_b,t')}{J(t_a,t_b)},$$

easily proved by using (16).

IV. WKB EXPANSION OF THE PROPAGATOR BY PATH INTEGRALS

Arbitrary potential

The framework for a WKB expansion of the propagator by phase-space path integrals without limiting procedure was set in an earlier paper[14] and will be only briefly summarized here. For a simple Hamiltonian of the form $p^2/2m + V(q,t)$ considered here, the phase-space path integral becomes a configuration-space path integral, since the momentum-dependent terms are rolled into the measure and only position-dependent terms remain to be path-integrated. The first step is to expand the classical action functional about the classical path $q_c(t)$:

$$S[q] \equiv S[q_c + x] = S_c + \tfrac{1}{2}\int_T [\dot{x}^2(t) - V''(t)x^2(t)]dt$$

$$- \sum_{n=3}^{\infty} \frac{1}{n!}\int_T V^{(n)}(t)x^n(t)dt,$$

where

$$V^{(n)}(t) \equiv [\partial^n V(q,t)/\partial q^n]_{q=q_c},$$

$T \equiv [t_a,t_b]$, and $x \in \mathscr{C}_0$, which is the space of paths such that $x(t_a) = x(t_b) = 0$. The classical action S_c becomes part of the WKB approximation, K_{WKB}, and the quadratic terms are rolled into the Gaussian measure, leaving the sum term for path integration. The result is

$$K = K_{\mathrm{WKB}} \int_{\mathscr{C}_0} dw_0(x)$$

$$\times \exp\left[\frac{-i}{\hbar}\sum_{n=3}^{\infty}\int_T V^{(n)}(t)\frac{x^n(t)}{n!}dt\right], \tag{20}$$

where the measure w_0 is defined by its Fourier transform:

$$\mathscr{F}w_0(\mu) \equiv \exp\left[-\frac{i\hbar}{2}\int_T\int_T G(t,t')d\mu(t)d\mu(t')\right],$$

$G(t,t')$ being Feynman's Green function defined earlier and μ being a bounded measure on the time-interval T. K_{WKB} is given by (17). The exponential in (20) can be expanded to yield

$$K = K_{\mathrm{WKB}}\left[1 + \sum_{j=1}^{\infty}\frac{1}{j!}\left(\frac{-i}{\hbar}\right)^j \sum_{n_1=3}^{\infty}\cdots\sum_{n_j=3}^{\infty}\right.$$

$$\times \int_{T^j}\frac{dt_1\cdots dt_j}{n_1!\cdots n_j!}V^{(n_1)}(t_1)\cdots V^{(n_j)}(t_j)$$

$$\left.\times \int_{\mathscr{C}_0} x^{n_1}(t_1)\cdots x^{n_j}(t_j)dw_0(x)\right]. \tag{21}$$

To evaluate the path integral, we need the moments formula (see, e.g., Ref. 15)

$$\int_{\mathscr{C}_0} x(t_1)x(t_2)\cdots x(t_n)dw_0(x)$$

$$= \begin{cases} 0 & \text{if } n \text{ is odd,} \\ (i\hbar)^m \Sigma' G(t_{i_1},t_{i_2})G(t_{i_3},t_{i_4}) & \\ \cdots G(t_{i_{2m-1}},t_{i_{2m}}) & \text{if } n = 2m \text{ is even,} \end{cases} \tag{22}$$

where Σ' denotes the sum over all different combinations of different indices i_j with $\{i_1,i_2,\ldots,i_n\} \equiv \{1,2,\ldots,n\}$. There are $(2m-1)!! \equiv (2m-1)(2m-3)\cdots 5\cdot 3\cdot 1$ terms in all for $n = 2m$.[16]

Thus, we see that $\hbar$ comes in the expansion with power $\frac{1}{2}(n_1 + \ldots + n_j) - j$, which is always a positive integer, since each n_i is at least 3. This proves that (21) is indeed an expansion in powers of $\hbar$, and we can write

$$K = K_{\mathrm{WKB}}(1 + \hbar K_1 + \hbar^2 K_2 + \cdots), \tag{23}$$

where the K_i's are ordinary finite-dimensional integrals over the time interval T. Polynomial potentials are best suited for this scheme, since the expansion of the action terminates at some finite n. However, it is important to note that regardless of the potential each term in the WKB expansion (coefficient of $\hbar^k$) is *always* a terminating series. For example, inspection of (21) shows that the first (post-WKB) term is, for

arbitrary potential,

$$\hbar K_1 = \left(\frac{-i}{\hbar}\right)\int_T \frac{dt}{4!} V^{(4)}(t)\int_{\mathscr{C}_0} x^4(t)dw_0(x) + \frac{1}{2}\left(\frac{-i}{\hbar}\right)^2 \int_{T^2} \frac{dt}{3!}\frac{ds}{3!} V^{(3)}(t)V^{(3)}(s) \times \int_{\mathscr{C}_0} x^3(t)x^3(s)dw_0(x), \tag{24}$$

and the moments formula gives

$$K_1 = \frac{-i}{8}\int_T V^{(4)}(t)G^2(t,t)dt + \frac{i}{24}\int_{T^2} V^{(3)}(t)V^{(3)}(s) \times [3G(t,t)G(t,s)G(s,s) + 2G^3(t,s)]dtds. \tag{25}$$

Let us study the structure of the coefficients K_j. In general, the $j = 1$ term in (19) is

$$\left(\frac{-i}{\hbar}\right)\sum_{n=3}^{\infty}\frac{dt}{n!}V^{(n)}(t)\int_{\mathscr{C}_0} x^n(t)dw_0(x) = \sum_{m=2}^{\infty}\frac{(i\hbar)^{m-1}}{m!2^m}\int_T V^{(2m)}(t)G^m(t,t)dt. \tag{26}$$

For arbitrary potentials, this is an infinite series in $\hbar$ with no constant term. Similarly, we find that:

(a) In the series for $j = 2$, the $n_1 = n_2 = 3$ term is the term proportional to $\hbar$, and the three terms $n_1 = n_2 = 4$; $n_1 = 3$, $n_2 = 5$; and $n_1 = 5$, $n_2 = 3$ are the ones proportional to $\hbar^2$. All the subsequent j series start out with $\hbar^k$ for $k \geqslant 2$.

(b) In the series for $j = 3$, the three terms $n_1 = n_2 = 3$, $n_3 = 4$; $n_1 = 4$, $n_2 = n_3 = 3$; and $n_1 = n_3 = 3$, $n_2 = 4$ are the only ones proportional to $\hbar^2$, and the $n_1 = n_2 = n_3 = 4$ term is the only one proportional to $\hbar^3$.

(c) In the series for $j = 4$, the term $n_1 = n_2 = n_3 = n_4 = 3$ is the only one proportional to $\hbar^2$.

(d) The series for $j = 5$ starts out with the $\hbar^3$ term.

Thus, we can write the term proportional to $\hbar^2$ in the expansion. It is:

$$\hbar^2 K_2 = \frac{-\hbar^2}{48}\int_T V^{(6)}(t)G^3(t,t)dt - \frac{1}{2\hbar^2}\int_{T^2}\frac{dt_1}{4!}\frac{dt_2}{4!} \times V^{(4)}(t_1)V^{(4)}(t_2)\int_{\mathscr{C}_0} x^4(t_1)x^4(t_2)dw_0(x) - \frac{1}{2\hbar^2} \times \int_{T^2}\frac{dt_1}{3!}\frac{dt_2}{5!} V^{(3)}(t_1)V^{(5)}(t_2)\int_{\mathscr{C}_0} x^3(t_1)x^5(t_2)dw_0(x) + \frac{i}{6\hbar^3}\int_{T^3}\frac{dt_1}{3!}\frac{dt_2}{3!}\frac{dt_3}{4!} V^{(3)}(t_1)V^{(3)}(t_2)V^{(4)}(t_3) \times \int_{\mathscr{C}_0} x^3(t_1)x^3(t_2)x^4(t_3)dw_0(x) + \frac{1}{4!}\frac{1}{\hbar^4}\int_{T^4}\frac{dt_1\cdots dt_4}{(3!)^4} \times V^{(3)}(t_1)V^{(3)}(t_2)V^{(3)}(t_3)V^{(3)}(t_4) \times \int_{\mathscr{C}_0} x^3(t_1)x^3(t_2)x^3(t_3)x^3(t_4)dw_0(x), \tag{27}$$

and the moments formula (22) gives the value of the path integrals in terms of Feynman's Green function and the classical path.

Application to the anharmonic potential

The anharmonic oscillator potential, given by (1), is $V(q) = m\omega^2 q^2/2 + \lambda q^4/4$. The first-order correction to the WKB approximation is then given by (25)

$$K_1 = \frac{-3\lambda i}{4}\int_T G^2(t,t)dt + \frac{3\lambda^2 i}{2}\int_{T^2} dtds q_c(t)q_c(s) \times [3G(t,t)G(t,s)G(s,s) + 2G^3(t,s)], \tag{28}$$

where $G(t,t')$ is given explicitly by (13) with J given by (11) and $q_c(t)$ by (4). The resulting integrals over the elliptic functions are all well-known and of the type tabulated in Ref. 9. Higher-order corrections can be generated at will, although they generally involve a large number of integrals. The WKB approximation is given by (17), with the classical action S_c given by (9) and the VVM function M given by (12), with J in (11). Therefore, every function entering the semiclassical expansion of the anharmonic oscillator propagator has been explicitly calculated, and the definite integrals giving the coefficients of the expansion have been explicitly exhibited. It is pointed out, again, that this treatment is nonperturbative, since the functions involved in the terms of the expansion [for example, q_c and G in (28)], depend implicitly on λ. This example illustrates the power of path integration without limiting procedure.

ACKNOWLEDGMENTS

I would like to thank Professor Cécile DeWitt-Morette for her guidance and the many interesting and informal discussions we have had at all stages of this work. I have profited from discussions with many physicists, including Richard P. Feynman, Peter A. Hogan, and Larry S. Schulman. Thanks also go to the theory group of the physics department of the University of California at Santa Barbara for their hospitality during the summer of 1974, when the first draft of this work was being written.

APPENDIX A: THE SMALL-DISTURBANCE EQUATION

This appendix will derive and generalize some results used in the text on the equation of small disturbances. The latter, resulting from the second variation of the action functional, is satisfied by the variation in a classical path resulting from a small change in the boundary conditions. For example, let $S[q] = S[\bar\beta(u)]$ be an action functional. Each path $q \equiv \bar\beta(u)$ is characterized by a parameter u: $q(t) = \bar\beta(u)(t) \equiv \beta(u,t)$. If the set $\{\bar\beta(u)\}$ is a set of classical paths $\{q_c(u)\}$ labeled by a parameter u (say a constant of integration), then $S'[\bar\beta(u)] = 0$ by definition of $\bar\beta(u)$. If we differentiate with respect to u, we get

$$S''[\bar\beta(u)]\,\frac{\partial\bar\beta(u)}{\partial u} = 0. \tag{A1}$$

This is the small-disturbance equation with its explicit solution in terms of the classical path: $S''[\bar\beta(u)]$ (second functional derivative of the action evaluated at the classical path) yields the small-disturbance operator; $\partial\bar\beta(u)/\partial u$ is its explicit solution, called a Jacobi field along the classical path $\bar\beta(u)$. Thus, the derivative of a classical solution with respect to a constant of integration is a solution of the small-disturbance equation. Note that if S is derived from a Lagrangian which does not contain the time explicitly, and we take the time derivative of the differential equations resulting from $S'[\bar\beta(u)] = 0$, we find that the classical velocity $\partial\bar\beta(u)(t)/\partial t$ is also a solution of (A1).

This method of "variation through geodesics" was studied extensively by Milnor.[17] The approach was generalized by DeWitt-Morette [5(c)] for arbitrary action functionals, and independently by the author[11] for Lagrangian actions. This method of generating solutions of (A1) was known to Jacobi.[12]

Lagrangian action

Let us consider the Lagrangian action in n dimensions as a specific example:

$$S[q] \equiv \int_{t_a}^{t_b} L(q(t),\dot q(t),t)\,dt.$$

One can show by straightforward differentiation with respect to u that the linear mapping $S'[\bar\beta(u)]$ maps x into

$$S'[\bar\beta(u)]x = \int_{t_a}^{t_b}\left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot q^i}\right)\right]_{q=\bar\beta(u)}(t)\,x^i(t)\,dt \tag{A2}$$

if $x(t_b) = x(t_a) = 0$ and there are no discontinuities in the momentum $\partial L/\partial\dot q^i$. (Sum over repeated indices is implied). Differentiating once more with respect to u yields

$$S''[\bar\beta(u)]\,\frac{\partial\bar\beta(u)}{\partial u}\,x = \int_T\left[\left(A_{ij}(t) + B_{ij}(t)\frac{\partial}{\partial t} + C_{ij}(t)\frac{\partial^2}{\partial t^2}\right)\times\frac{\partial\beta^j(u,t)}{\partial u}\right]x^i(t)\,dt, \tag{A3}$$

where

$$A_{ij}(t) \equiv \frac{\partial^2 L}{\partial q^i\partial q^j} - \frac{d}{dt}\left(\frac{\partial^2 L}{\partial\dot q^i\partial q^j}\right), \tag{A4a}$$

$$B_{ij}(t) \equiv \frac{\partial^2 L}{\partial q^i\partial\dot q^j} - \frac{\partial^2 L}{\partial\dot q^i\partial q^j} - \frac{d}{dt}\left(\frac{\partial^2 L}{\partial\dot q^i\partial\dot q^j}\right), \tag{A4b}$$

$$C_{ij}(t) \equiv -\frac{\partial^2 L}{\partial\dot q^i\partial\dot q^j}. \tag{A4c}$$

Note that the above matrices satisfy the relations

$$\tilde C = C, \qquad (B-\dot C)^{\sim} = -(B-\dot C),$$
$$B + \tilde B = 2\dot C, \quad A - \tilde A = \dot B - \ddot C = \tfrac{1}{2}(\dot B - \dot{\tilde B}).$$

We assume that $C(t)$, the Jacobian of the transformation from the $\dot q$'s to the p's, never vanishes, so that a canonical formalism exists.

If $\bar\beta(u)$ is a family of classical paths q_c, then both sides of (A2) and (A3) are zero for all $x(t)$:

$$\left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial\dot q^i}\right)\right]_{q=q_c} = 0, \tag{A5}$$

$$\left[A_{ij}(t) + B_{ij}(t)\frac{d}{dt} + C_{ij}(t)\frac{d^2}{dt^2}\right]_{q=q_c} f^j(t) = 0. \tag{A6}$$

The first equation is the familiar Euler–Lagrange equation, yielding the classical solutions $q_c(t,u)$, where u is any of the $2n$ constants of integration, or any other suitable parameter (e.g., t_a or t_b).

The second equation is the small-disturbance equation, and the bracketed second-order linear differential operator is the (Hermitian) small-disturbance operator. The equation is solved by $\partial q_c(t,u)/\partial u$.

Attempts at solving (A6) by "frontal assault" are sometimes found in the literature (see, e.g., Ref. 7), and usually yield only approximate solutions, if any at all.

A convenient set of solutions is obtained by using endpoint boundary conditions:

$$q_c{}^i(t_a) = q_a{}^i, \quad q_c{}^i(t_b) = q_b{}^i.$$

Thus, for any fixed $\{i,i'\} = \{1,2,\ldots,n\}$, the two sets

$$f^j{}_{(i)}(t) \equiv \frac{\partial q_c{}^j(t)}{\partial q_a{}^i}, \quad g^j{}_{(i')}(t) \equiv \frac{\partial q_c{}^j(t)}{\partial q_b{}^{i'}}$$

are sets of solutions of (A6) satisfying the obvious boundary conditions:

$$f^j{}_{(i)}(t_a) = \delta^j{}_i, \quad g^j{}_{(i)}(t_a) = 0,$$
$$f^j{}_{(i)}(t_b) = 0, \quad g^j{}_{(i)}(t_b) = \delta^j{}_i.$$

We can use these solutions as building blocks for other solutions, which can usually be written as linear combinations of them.

Two other sets of solutions can be obtained by differentiating $q_c(t)$ with respect to t_a or t_b:

$$h^j(t) \equiv \frac{\partial q_c{}^j(t)}{\partial t_a}, \quad k^j(t) \equiv \frac{\partial q_c{}^j(t)}{\partial t_b}.$$

They satisfy the boundary conditions:

$$h^j(t_a) = -\dot{q}_c^j(t_a), \quad k^j(t_a) = 0,$$
$$h^j(t_b) = 0, \quad k^j(t_b) = -\dot{q}_c^j(t_b).$$

Proof: The second and third are obvious since the operations, say, "$\partial/\partial t_a$" and "evaluate at t_b" commute. The first and fourth are more subtle. The first is derived as follows:

$$h^j(t_a) = -[h^j(t_b) - h^j(t_a)] = -\int_{t_a}^{t_b} \frac{\partial \dot{q}_c^j(t)}{\partial t_a}\, dt,$$

$$= -\frac{\partial}{\partial t_a}\int_{t_a}^{t_b} \dot{q}_c^j(t)dt - \dot{q}_c^j(t_a)$$

$$= -\frac{\partial}{\partial t_a}(q_b^j - q_a^j) - \dot{q}_c^j(t_a) = -\dot{q}_c^j(t_a). \quad \blacksquare$$

The fourth relation is derived in a similar manner.

Theorem: Let $x(t)$ and $y(t)$ be two solutions of the small disturbance equation (A6) in one dimension. Their Wronskian depends on t only through $C(t)$:

$$W(t) \equiv \dot{x}(t)y(t) - x(t)\dot{y}(t) = \alpha C(t_a)C^{-1}(t), \tag{A7}$$

where α is a constant, and it is assumed that $C(t)$ never vanishes. If $\alpha \neq 0$, x and y are linearly independent.

Proof:

$$\dot{W} = \ddot{x}y - \ddot{y}x = -C^{-1}(Ax + B\dot{x})y + C^{-1}(Ay + B\dot{y})x$$
$$= -BC^{-1}(\dot{x}y - \dot{y}x) = -BC^{-1}W$$
$$\Rightarrow W(t) = \alpha \exp\left[-\int_{t_a}^{t} B(s)C^{-1}(s)ds\right].$$

However, we can see from (A4) that $B = \dot{C}$ in one dimension, and the result follows.

Green functions

We now study the Green functions $G^{jk}(t,t')$ of the small-disturbance operator, which satisfy

$$\left[A_{ij}(t) + B_{ij}(t)\frac{\partial}{\partial t} + C_{ij}(t)\frac{\partial^2}{\partial t^2}\right]G^{jk}(t,t')$$
$$= \delta_i{}^k\delta(t - t'), \tag{A8}$$

where A, B, and C are given by (A4) for $q = q_c$. We restrict ourselves to one dimension.

Theorem: The advanced and retarded Green functions are unique and are given by

$$G^-(t,t') = G^+(t',t) = J(t,t')Y(t - t') \tag{A9}$$

where $J(t,t')$ is the Jacobi commutator:

$$J(t,t') = \left[\frac{\partial q_c(t)}{\partial \alpha_1}\frac{\partial q_c(t')}{\partial \alpha_2} - \frac{\partial q_c(t')}{\partial \alpha_1}\frac{\partial q_c(t)}{\partial \alpha_2}\right]$$
$$\times\left[\frac{\partial q_c(t_b)}{\partial \alpha_1}\frac{\partial p_c(t_b)}{\partial \alpha_2} - \frac{\partial p_c(t_b)}{\partial \alpha_1}\frac{\partial q_c(t_b)}{\partial \alpha_2}\right]^{-1}, \tag{A10}$$

α_1 and α_2 being the two constants of integration. t_b in the denominator can be replaced by t_a.

Proof: We look for the most general Green function of the form $G^-(t,t') = f(t,t')Y(t - t')$. Upon differentiation and use of the fact that $x\delta'(x) = -\delta(x)$, we have

$$\frac{\partial G^-}{\partial t} = f_{,1}(t,t')Y(t - t') + f(t,t')\delta(t - t'),$$
$$\frac{\partial^2 G^-}{\partial t^2} = f_{,11}(t,t')Y(t - t') + 2f_{,1}(t,t')\delta(t - t')$$
$$- f(t,t')\frac{\delta(t - t')}{t - t'},$$

$$D_t G^- \equiv \left[A(t) + B(t)\frac{\partial}{\partial t} + C(t)\frac{\partial^2}{\partial t^2}\right]G^-$$
$$= Y(t - t')D_t f(t,t') + \delta(t - t')$$
$$\times\left[B(t) + 2C(t)\frac{\partial}{\partial t} - \frac{C(t)}{t - t'}\right]f(t,t'),$$

where $f_{,1}(t,t')$ denotes the derivative with respect to the first argument, evaluated at (t,t').

Thus, G^- is a Green function if $D_t f(t,t') = 0$, i.e., if $f(t,t')$ is a homogeneous solution in t, and if the coefficient of the delta function at $t = t'$ is 1. If we expand about $t = t'$,

$$f(t,t') = f(t,t) + (t - t')f_{,1}(t,t) + \tfrac{1}{2}(t - t')^2 f_{,11}(t,t) + \cdots,$$

the second condition gives the boundary conditions on f:

(a) $f(t,t) = 0$, since $C(t) \neq 0$,

(b) $f_{,1}(t,t) = C^{-1}(t)$.

If $f(t,t')$ is a solution in t, then we can write

$$f(t,t') = \beta(t')x(t) + \gamma(t')y(t),$$

where x and y are two linearly independent solutions. If we insert the boundary conditions and remember (A7), which indicates that $\dot{x}y - x\dot{y} = \alpha C(t_a)C^{-1}(t)$, we have

$$f(t,t') = \frac{x(t)y(t') - x(t')y(t)}{C(t_a)[y(t_a)\dot{x}(t_a) - \dot{y}(t_a)x(t_a)]}. \tag{A11}$$

Let us choose $x(t) = \partial q_c(t)/\partial \alpha_1$ and $y(t) = \partial q_c(t)/\partial \alpha_2$. By definition of $C(t), C(t_a) \equiv -\partial^2 L/\partial \dot{q}_c^2(t_a) = -\partial p_c(t_a)/\partial \dot{q}_c(t_a)$, so $-C(t_a)\partial \dot{q}_c(t_a)/\partial \alpha_1 = \partial p_c(t_a)/\partial \alpha_1$. Inserting this in (A11) we see that $f(t,t')$ is given by $J(t,t')$ in (A10). Note that our $J(t,t')$ is what Bryce DeWitt[18] calls $\tilde{G}$ (he defines $\tilde{G}$ by $G^+ - G^-$, but since his Green functions are the negative of ours, $\tilde{G} = J$).

The greatest simplification in $J(t,t')$ as expressed in (A10) occurs when the constants of integration are initial (or final) position and momentum, for example, $\alpha_1 = q_c(t_a) = q_a$, $\alpha_2 = p_c(t_a) = p_a$. The denominator is then equal to 1, and

$$J(t,t') = \frac{\partial q_c(t)}{\partial q_a}\frac{\partial q_c(t')}{\partial p_a} - \frac{\partial q_c(t')}{\partial q_a}\frac{\partial q_c(t)}{\partial p_a}$$
$$\equiv \{q_c(t), q_c(t')\}_{(q_a,p_a)}.$$

This Poisson bracket becomes the commutator $[\mathbf{Q}(t),\mathbf{Q}(t')]/i\hbar$ when the system is quantized, whence the name of the function.

Feynman's Green function, which vanishes at t_a and t_b, can be built from G^- and G^+ as follows:

$$G(t,t') = G^-(t,t') + J(t,t_a)J(t_b,t')/J(t_a,t_b)$$

$$= G^+(t,t') + J(t,t_b)J(t_a,t')/J(t_a,t_b). \quad \text{(A12)}$$

Indeed, it is readily apparent that the additions to G^- and G^+ are homogeneous solutions, and that G vanishes when t or t' is t_a or t_b. Another form for G is shown in the main text (Equation 13).

Particle in a potential

Let us concentrate on the case of a particle in a potential in one dimension, with Lagrangian $L = mq^2/2 - V(q)$. The dynamical equation is

$$\ddot{q}_c(t) + m^{-1}V'[q_c(t)] = 0. \quad \text{(A13)}$$

The small-disturbance equation is

$$\left\{ -\frac{d^2}{dt^2} - m^{-1}V''[q_c(t)] \right\} f(t) = 0. \quad \text{(A14)}$$

Consider two linearly independent solutions of (A14), D and $\bar{D}$, satisfying

$$D(t_b) = 1, \quad \bar{D}(t_b) = 0,$$
$$\dot{\bar{D}}(t_b) = 0, \quad \dot{\bar{D}}(t_b) = -1. \quad \text{(A15)}$$

Their Wronskian $W = D\dot{\bar{D}} - \dot{D}\bar{D}$ is constant and equal to -1. D and $\bar{D}$ depend on t_b, t_a, q_b, and q_a through $q_c(t)$. The antisymmetric Jacobi commutator along the classical path $q_c(t)$ can be shown to be

$$J(t,t') = \bar{D}(t)D(t') - D(t)\bar{D}(t').$$

It is obviously a solution of (A14) in both t and t'.

Classical path in terms of Jacobi fields

$$q_c(t) = A\int_{t_a}^{t} D(s)ds + B\int_{t_a}^{t} \bar{D}(s)ds + q_a, \quad \text{(A16)}$$

where

$$A \equiv \frac{q_b - q_a - V'(q_b)\int_T \bar{D}(u)du}{\int_T D(s)ds}, \quad B \equiv V'(q_b).$$

Proof: $\dot{q}_c(t)$, being a derivative of the classical path, is a solution of the small-disturbance equation, and hence a linear combination of D and $\bar{D}$: $\dot{q}_c(t) = AD(t) + B\bar{D}(t)$. Integrating from t_a to t yields

$$q_c(t) = A\int_{t_a}^{t} D(s)ds + B\int_{t_a}^{t} \bar{D}(s)ds + q_c(t_a).$$

However, $q_c(t)$ is now the solution of a *third*-order differential equation. Therefore, we need a third boundary condition, other than $q_c(t_a) = q_a$ and $q_c(t_b) = q_b$. It is provided by the dynamical equation (A13) evaluated, say, at t_b. This gives A and B. Note that

$$\dot{q}_c(t_a) = AD(t_a) + B\bar{D}(t_a), \quad \text{(A17)}$$
$$\dot{q}_c(t_b) = A. \quad \text{(A18)}$$

Criterion for nonexistence of a classical path

What must the relationship between t_a, t_b, q_a, and q_b be in order for a classical path $q_c(t)$ such that $q_c(t_a) = q_a$ and $q_c(t_b) = q_b$ *not* to exist? The answer is given in terms of Jacobi fields. q_c will *not* exist if

$$\int_T D(s)ds = 0, \quad \textit{or} \quad \bar{D}(t_a) = M^{-1} = 0, \quad \text{(A19a)}$$

and

$$q_b - q_a - V'(q_b)\int_T \bar{D}(s)ds \neq 0. \quad \text{(A19b)}$$

This is easily proved by looking at (A16) which gives $q_c(t)$ in terms of the Jacobi fields. $q_c(t)$ is infinite if the denominator of A is zero (first condition) and the numerator of A is nonzero (second condition). That the two forms of the first condition are equivalent can be seen by differentiating (A17) with respect to q_b. On the right-hand side, we get $\partial\dot{q}_c(t_a)/\partial q_b = M = 1/\bar{D}(t_a)$, and on the left-hand side we get a fraction with denominator $[\int_T D(s)ds]^2$. Thus, whenever $\bar{D}(t_a)$ vanishes, $\int_T D(s)ds$ must also vanish.

For a general discussion of these conditions in the context of caustics and catastrophe theory, see Ref. 19.

Zero Jacobi field: The only Jacobi field vanishing at both t_a and t_b is $f(t) = 0$, unless

$$\bar{D}(t_a) = M^{-1} = 0 \quad \text{(A20a)}$$

and

$$q_b - q_a - V'(q_b)\int_T \bar{D}(s)ds = 0, \quad \text{(A20b)}$$

in which case $f(t) = a\bar{D}(t)$, where a is an arbitrary constant.

Proof: It is obtained by writing $f(t) = a\bar{D}(t) + bD(t)$ and inserting the boundary conditions. However, if $\bar{D}(t_a) = 0$, we may not have a classical path, in which case a Jacobi field is meaningless. Therefore the second condition is necessary to insure that one or more classical paths exist.

Example: the harmonic oscillator

We illustrate this with the harmonic oscillator ($V(q) = \frac{1}{2}\omega^2q^2$). The classical path, for arbitrary endpoints, is given by

$$q_c(t) = \frac{q_a \sin\omega(t_b - t) + q_b \sin\omega(t - t_a)}{\sin\omega T}$$

$$= A\cos(\omega t + \varphi), \quad \text{(A20c)}$$

where $T = t_b - t_a$ and

$$A = (\sin\omega T)^{-1}[q_a^2 + q_b^2 - 2q_aq_b\cos\omega T]^{1/2},$$

TABLE A1. Classical paths for the harmonic oscillator ($n = ..., -1,0,1,2,...$)

Harmonic Oscillator	$q_a \neq q_b$	$q_a = q_b = q_0 \neq 0$	$q_a = q_b = 0$
$\omega T \neq n\pi$	Unique $q_c(t)$ exists and is given by (A20c)	$q_c(t) = q_0 \dfrac{\cos\omega[t - (t_a + t_b)/2]}{\cos(\omega T/2)}$	$q_c(t) = 0$
$\omega T = 2n\pi$	$q_c(t)$ never exists	$q_c(t) = q_0 \dfrac{\cos\omega[t - (t_a + t_b)/2]}{\cos(n\pi)}$	Noncountably infinite number of classical paths given by $q_c(t) = A \sin\omega(t_b - t)$ (A arbitrary)
$\omega T = (2n+1)\pi$	$q_c(t)$ never exists	$q_c(t)$ never exists	

$$\varphi = \arccos\left(\frac{q_a \sin\omega t_b - q_b \sin\omega t_a}{[q_a^2 + q_b^2 - 2q_a q_b \cos\omega T]^{1/2}} \right).$$

It may fail to exist when $\sin(\omega T) = 0$ (the amplitude becomes infinite), except when $q_a = q_b = 0$, in which case there is an infinite number of q_c's.

The various cases are summarized in Table A1.

The Jacobi fields in this case are

$$D(t) = \cos\omega(t_b - t), \quad \bar{D}(t) = \frac{1}{\omega}\sin\omega(t_b - t).$$

We can quickly verify all our criteria. We have

(a) $$\int_T D(s)ds = \bar{D}(t_a) = \frac{1}{\omega}\sin\omega T,$$

(b) $$\int_T \bar{D}(s)ds = \frac{1}{\omega^2}(1 - \cos\omega T).$$

If $\omega T = n\pi$, we have no classical path, unless

$$q_b - q_a - V'(q_b)\omega^{-2}(1 - \cos\omega T) = 0,$$

i.e., if $q_a = q_b$ and $\omega T = 2n\pi$ (yielding one path), or if $q_a = q_b = 0$, which implies that $V'(q_b) = 0$ (yielding an infinite number of paths).

The commutator function

The dynamical equation (A13) can be solved by quadratures: If we substitute $\dot{q}_c(t) = u$, we obtain the energy $E \equiv \frac{1}{2}mu^2 + V(q_c) = \text{const.}$ as a first integral. A second integration gives

$$F(t,t_a,q_c,q_a,E) \equiv t - t_a - \left(\frac{m}{2}\right)^{1/2} \times \int_{q_a}^{q_c} \frac{dx}{[E - V(x)]^{1/2}} = 0,$$

which yields $t(q_c)$ rather than $q_c(t)$. In order to differentiate the classical path with respect to the constants of integration (here, the energy E and the initial position q_a), we will need the implicit function theorem. The latter states essentially that

$$F(x_1,...,x_n) = 0 \Rightarrow \frac{\partial x_i}{\partial x_j} = -\frac{\partial F/\partial x_j}{\partial F/\partial x_i} \quad (i \neq j).$$

This gives

$$\frac{\partial q_c}{\partial q_a} = \left[\frac{E - V(q_c)}{E - V(q_a)}\right]^{1/2}, \tag{A21a}$$

$$\frac{\partial q_c}{\partial E} = \tfrac{1}{2}[E - V(q_c)]^{1/2} \int_{q_a}^{q_c} [E - V(x)]^{-3/2} dx, \tag{A21b}$$

$$\frac{\partial q_c}{\partial t} = \dot{q}_c(t) = \left\{\frac{2}{m}[E - V(q_c)]\right\}^{1/2}, \tag{A21c}$$

$$\frac{\partial q_c}{\partial t_a} = -\left\{\frac{2}{m}[E - V(q_c)]\right\}^{1/2}, \tag{A21d}$$

$$\frac{\partial \dot{q}_c(t_b)}{\partial E} = \left(\frac{\partial}{\partial E}\frac{\partial q_c}{\partial t}\right)_{t=t_b} = \{2m[E - V(q_c(t_b))]\}^{-1/2}, \tag{A21e}$$

$$\frac{\partial \dot{q}_c(t_b)}{\partial q_a} = \left(\frac{\partial}{\partial q_a}\frac{\partial q_c}{\partial t}\right)_{t=t_b} = 0. \tag{A21f}$$

Substituting these in (A10), we obtain the commutator (here $p_c = m\dot{q}_c$, $\alpha_1 = q_a$, $\alpha_2 = E$):

$$J(t,t') = \left\{\frac{[E - V(q_c(t))][E - V(q_c(t'))]}{2m}\right\}^{1/2} \times \int_{q_c(t)}^{q_c(t')} \frac{dx}{[E - V(x)]^{3/2}}. \tag{A22}$$

If the constants of integration are initial position and momentum q_a and p_a, then $J(t,t')$ is still given by (A22) with E replaced by $p_a^2/2m + V(q_a)$. This is not a trivial statement [compare with (A26)], as we show below.

Proof: In terms of q_a and p_a, the solution is

$$F(t,t_a,q_c,q_a,p_a) \equiv t - t_a - \left(\frac{m}{2}\right)^{1/2}$$

$$\times \int_{q_a}^{q} \left[\frac{p_a^2}{2m} + V(q_a) - V(x) \right]^{-1/2} dx = 0. \tag{A23}$$

Then

$$\frac{\partial q_c}{\partial q_a} = - \frac{\partial F/\partial q_a}{\partial F/\partial q_c}$$

$$= \left[\frac{p_a^2}{2m} + V(q_a) - V(q_c) \right]^{1/2} \left\{ \frac{\sqrt{2m}}{p_a} + \frac{V'(q_a)}{2} \right.$$

$$\left. \times \int_{q_a}^{q_c} dx \left[\frac{p_a^2}{2m} + V(q_a) - V(x) \right]^{-3/2} \right\},$$

$$\frac{\partial q_c}{\partial p_a} = - \frac{\partial F/\partial p_a}{\partial F/\partial q_c}$$

$$= \frac{p_a}{2m} \left[\frac{p_a^2}{2m} + V(q_a) - V(q_c) \right]^{1/2}$$

$$\times \int_{q_a}^{q_c} \left[\frac{p_a^2}{2m} + V(q_a) - V(x) \right]^{-3/2} dx. \tag{A24}$$

Substituting the above in expression (15) for J, some terms cancel out and we get the result. The nontriviality of this result is illustrated by the fact that $\partial q_c/\partial q_a$ in (A24) is *not* obtained from $\partial q_c/\partial q_a$ in (A21a) by simply replacing E by $p_a^2/2m + V(q_a)$.

We can give the commutator in terms of the endpoints q_a and q_b. For this we have

$$F(t,t_a,q_c,q_a,E) \equiv t - t_a - \left(\frac{m}{2} \right)^{1/2} \int_{q_a}^{q_c} [E - V(x)]^{-1/2} dx, \tag{A25}$$

$$G(t_b,t_a,q_b,q_a,E)$$

$$\equiv t_b - t_a - \left(\frac{m}{2} \right)^{1/2} \int_{q_a}^{q_b} [E - V(x)]^{-1/2} dx,$$

that is, E in the first equation is really a function of q_a and q_b, given implicitly by the second equation. It is no longer an independent constant of integration, but q_b is. Thus, we have

$$\frac{\partial q_c}{\partial q_a} = - \frac{\partial F/\partial q_a}{\partial F/\partial q_c}$$

$$= \left\{ (m/2)^{1/2}[E - V(q_a)]^{-1/2} + \tfrac{1}{2}(m/2)^{1/2}(\partial E/\partial q_a) \right.$$

$$\left. \times \int_{q_a}^{q_c} [E - V(x)]^{-3/2} dx \right\} \{[(m/2)^{1/2}[E - V(q_c)]^{-1/2}]\}^{-1},$$

where $\partial E/\partial q_a$ is obtained by using the implicit function theorem on G:

$$\frac{\partial E}{\partial q_a} = - \frac{\partial G/\partial q_a}{\partial G/\partial E}$$

$$= \frac{-2}{[E - V(q_a)^{1/2}} \left\{ \int_{q_a}^{q_c} [E - V(x)]^{-3/2} dx \right\}^{-1}.$$

Finally,

$$\frac{\partial q_c}{\partial q_a} = \left[\frac{E - V(q_c)}{E - V(q_a)} \right]^{1/2} \cdot \frac{\int_{q_c}^{q_b} [E - V(x)]^{-3/2} dx}{\int_{q_a}^{q_b} [E - V(x)]^{-3/2} dx}.$$

Similarly, we find

$$\frac{\partial q_c}{\partial q_b} = \left[\frac{E - V(q_c)}{E - V(q_b)} \right]^{1/2}$$

$$\times \int_{q_a}^{q_c} [E - V(x)]^{-3/2} dx \cdot \left\{ \int_{q_a}^{q_b} [E - V(x)]^{-3/2} dx \right\}^{-1}.$$

As for the Van Vleck–Morette function $M = -m\partial \dot{q}_c(t_b)/\partial q_a$, we use

$$\dot{q}_c(t_b) = \left\{ \frac{2}{m} [E(q_a,q_b) - V(q_b)] \right\}^{1/2}.$$

This gives

$$M = - m \frac{\partial \dot{q}_c(t_b)}{\partial q_a}$$

$$= - \frac{m}{2} \left(\frac{2}{m} \right)^{1/2} [E - V(q_b)]^{-1/2} \frac{\partial E}{\partial q_a},$$

i.e.,

$$M = \left(\frac{m}{2} \right)^{1/2} \{ [E - V(q_a)][E - V(q_b)] \}^{-1/2}$$

$$\times \left\{ \int_{q_a}^{q_b} [E - V(x)]^{-3/2} dx \right\}^{-1}.$$

Finally, the commutator in terms of the endpoints is given by (A10) with $\alpha_1 = q_a$ and $\alpha_2 = q_b$:

$$J(t,t')$$

$$= \left(\frac{2}{m} \right)^{1/2} [E - V(q_c(t))]^{1/2} [E - V(q_c(t'))]^{1/2}$$

$$\times \left\{ \int_{q_a}^{q_b} [E - V(x)]^{-3/2} dx \right\}^{-1}$$

$$\times \left\{ \int_{q_c(t)}^{q_b} dx [E - V(x)]^{-3/2} \int_{q_a}^{q_c(t')} dy [E - V(y)]^{-3/2} \right.$$

$$- \int_{q_c(t')}^{q_b} dx [E - V(x)]^{-3/2}$$

$$\left. \times \int_{q_a}^{q_c(t)} dy [E - V(y)]^{-3/2} \right\}, \tag{A26}$$

where $q_c(t,q_a,q_b)$ and $E(q_a,q_b)$ are given implicitly by (A25).

[1]J.J. Loeffel, A. Martin, B. Simon, and A.S. Wightman, Phys. Lett. B **30**, 656–58 (1969), Barry Simon, Ann. Phys. (N.Y.) **58**, 76–136 (1970).
[2]Indeed, the dynamical equation for the $\lambda\varphi^4$ self-interaction, namely $(\Box - m^2)\varphi - \lambda\varphi^3 = 0$, can be readily reduced to the dynamical equation for the one-dimensional anharmonic oscillator, namely $\bar{\varphi}'' - m^2\bar{\varphi}/K^2 - \lambda\bar{\varphi}^3/K^2 = 0$, where $\varphi(x_1,x_2,x_3,x_4) \equiv \bar{\varphi}(K\cdot x)$, K being an arbitrary 4-vector (plane-wave solution). The elliptic functions which are solutions of this equation are periodic, and admit of a restricted superposition principle, rare for nonlinear equations: If an elliptic cosine (cn) with a certain modulus k_1 is a solution, and if an elliptic sine (sn) with another modulus k_2 is also a solution, then the linear combination (cn + $i$sn) is also a solution, but the common modulus k_3 is different from k_1 and k_2 [cf. Gérard Pétiau, Cah. Phys. **14**, 1–24 (1960), and D.F. Kurdgelaidzé, *ibid.*, No. 128, 149–57 (1961)].
[3]Carl M. Bender and Tai Tsun Wu, Phys. Rev. Lett. **21**, 406–09 (1968); Phys. Rev. **184**, 1231–60 (1969); Phys. Rev. Lett. **27**, 461–65 (1971); Phys. Rev. D **7**, 1620–36 (1973).
[4]P.M. Mathews and K. Eswaran, Lett. Nuovo Cimento **5**, 15–18 (1972).
[5]C. DeWitt-Morette, (a) Commun. Math. Phys. **28**, 47–67 (1972); (b) **37**, 63–81 (1974); (c) Ann. Phys. (N.Y.) **97**, 367–99 (1976); (d) "Path integration in non-relativistic quantum mechanics" (with A. Maheshwari and B. Nelson), to appear in Phys. Rep. 9 C.
[6]C.S. Lam, Nuovo Cimento, Ser. X, **47**, 451–69 (1967); **50**, 504–10 (1967).
[7]S. Sarkar, Phys. Rev. D **8**, 1060–67 (1973).
[8]P.M. Mathews and M.S. Seshadri, Int. J. Theor. Phys. **13**, 279–88 (1975).
[9]Paul F. Byrd and Morris D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer-Verlag, Berlin, 1954).
[10]We can better understand this question of physical boundary conditions from the simpler example of a particle in free fall with friction taken into account. The dynamical equation is $\ddot{x} = -(g + k\dot{x})$, with solution $x(t) = -gt/k + Ak^{-2}e^{-kt} + B$, where A and B are constants of integration. When $k \to 0$, we expect to retrieve free fall, namely $x(t) = -gt^2/2 + v_0t + x_0$. Instead, we find a "singularity" at $k = 0$ if A and B are numerically specified. However, numerical specification of A and B does not constitute *physical* boundary conditions. Physical boundary conditions, such as $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$, *always* make A and B depend on k in such a manner as to make the solution [namely, in this case $x(t) = x_0 - gt/k + (g + v_0k)(1 - e^{-kt})/k^2$] reduce properly when $k \to 0$.
[11]Maurice M. Mizrahi, "An Investigation of the Feynman Path Integral Formulation of Quantum Mechanics," Ph. D. thesis, University of Texas at Austin, 1975.
[12]C.G. Jacobi, "On the theory of the calculus of variations and of differential equations," Crelle's Math. J. **17**, 68–82 (1837), referred to in Oskar Bolza, *Lectures on the Calculus of Variations* (U. Chicago Press, Chicago, 1904; also Chelsea, New York, 1960, 1973).
[13]This relation and similar ones can be simply derived as follows. For any Lagrangian L in n dimensions and $u \equiv t_a, t_b, q_a$, or q_b, we have

$$\partial S_c/\partial u = (\partial/\partial u)\int_{t_a}^{t_b} L(q_c,\dot{q}_c,t)\,dt = (p_c)_i(t_b)[\partial q_c^i(t)/\partial u]_{t=t_b} - (p_c)_i(t_a)[\partial q_c^i(t)/\partial u]_{t=t_a} + \epsilon L(u),$$

where $\epsilon = 0$ *for* $u = q_a$ or $u = q_b$, 1 for $u = t_b$, and -1 for $u = t_a$, since $(p_c)_i(t) \equiv [\partial L/\partial \dot{q}^i]_{q=q_c}$. Given that the classical Hamiltonian is $H_c(p_c,q_c,t) \equiv (p_c)_i(\dot{q}_c)^i - L(q_c,\dot{q}_c,t)$, this gives the following four relations of Hamilton–Jacobi theory:

$$\partial S_c/\partial q_a^i = -(p_c)_i(t_a), \quad \partial S_c/\partial t_a = H_c(-\partial S_c/\partial q_a, q_a, t_a),$$

$$\partial S_c/\partial q_b^i = (p_c)_i(t_b), \quad \partial S_c/\partial t_b = -H_c(\partial S_c/\partial q_b, q_b, t_b).$$

[14]Maurice M. Mizrahi, J. Math. Phys. **19**, 298–307 (1978).
[15]Maurice M. Mizrahi, J. Math. Phys. **17**, 566–75 (1976).
[16]Among the specific moments needed are the following:

(a) $\int_{\text{all } w} x^{2n}(t)\,dw_0(x) = (2n)!\hbar^n i^n G^n(t,t)/2^n n!$,

(b) $\int_{\text{all } w} x^2(t)x^2(t')\,dw_0(x) = -\hbar^2[G(t,t)G(t',t') + 2G^2(t,t')]$,

(c) $\int_{\text{all } w} x^3(t)x^3(t')\,dw_0(x) = -i\hbar^3[9G(t,t)G(t,t')G(t',t') + 6G^3(t,t')]$,

(d) $\int_{\text{all } w} x^2(t)x^4(t')\,dw_0(x) = -i\hbar^3[12G^2(t,t')G(t',t') + 3G(t,t)G^2(t',t')]$,

(e) $\int_{\text{all } w} x^4(t)x^4(t')\,dw_0(x) = \hbar^4[9G^2(t,t)G^2(t',t') + 24G^4(t,t') + 72G(t,t)G(t',t')G^2(t,t')]$.

For higher moments, we need a more compact notation. We write $12\hbar^2G^2(t_1,t_2)G(t_2,t_2) \equiv 12(12)^2(22)$.
Then,

(f) $\int x^3(t_1)x^3(t_2)x^4(t_3)\,dw_0(x)$

$$= 27(11)(12)(22)(33)^2 + 18(12)^3(33)^2 + 72(31)^3(32)(22) + 72(31)(32)^3(11) + 108(33)(31)(32)(11)(22) + 108(33)(32)^2(11)(12) + 108(33)(31)^2(22)(21) + 216(31)^2(32)^2(12) + 216(33)(31)(32)(12)^2.$$

There are $(3 + 3 + 4 - 1)!! = 9\times7\times5\times3 = 945$ terms in all.
[17]J. Milnor, *Morse Theory*, based on lecture notes by M. Spivak and R. Wells, Annals of Mathematics Studies, No. 51(Princeton U. P., Princeton, N.J., 1969).
[18]Bryce S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
[19]C. DeWitt-Morette, in *Long-Term Predictions in Dynamics*, edited by V. Szebehely and B.D. Tapley (Reidel, Dordrecht, 1976), pp. 57–65, also pp. 67–70 (with Pete Tschumi).

A general setting for reduction of dynamical systems

G. Marmo

Istituto di Fisica Teorica, Università degli Studi
and Istituto Nazionale di Fisica Nucleare, Sezione di Napoli, Italy

E. J. Saletan[a)]

Istituto di Fisica Teorica, Università degli Studi
and Istituto Nazionale di Fisica Nucleare, Sezione di Napoli, Italy

A. Simoni

Istituto di Fisica Teorica, Università di Napoli, Italy
(Received 8 September 1978)

The reduction of dynamical systems is discussed in terms of projecting vector fields with respect to foliations of the manifold on which the dynamics take place. Examples of established reduction procedures are presented and are shown to be special cases of the general procedure described in this paper. Instances of other types of analysis of dynamical systems related to our projection procedure are briefly discussed.

1. Several papers[1-3] have recently been devoted to the problem of reducing a dynamical system to two or more systems of lower dimension. In these papers various procedures are used, usually based on constants of the motion (functions) or on vector fields which generate symmetries of the dynamics. In addition, requirements of one kind or another are placed on the objects (functions, vector fields) used to achieve the reduction. These conditions may be that the functions form a function group[4] or that the vector fields commute, close to form a Lie algebra, are Hamiltonian, or have some other property. When functions and vector fields are used together to achieve the reduction, the requirements may also be on the relations between them.

In this paper we present a general setting in which all of the reduction procedures known to us appear as special cases. It is based on the idea of a *foliation* of the manifold on which the dynamics is taking place, and involves invariant geometric objects other than functions and fields, namely *distributions* and p forms. As will be seen, the dynamical vector field is not required in general to be even locally Hamiltonian, although the usual kind of reduction of Hamiltonian dynamics[3] can also be understood in the terms we present here.

The plan of this paper is roughly the following. We first discuss projectability of a vector field (the dynamics) with respect to a foliation. The problem then is to find foliations with respect to which the dynamics is projectable, and essentially two methods for finding them are presented. Some examples of established procedures are discussed, including the use of an invariant application (from the original manifold to another) rather than invariant functions. Finally, some remarks are made about other procedures used in dynamics which may be viewed in terms of foliations and projections, followed by a brief discussion of the opposite procedure. The terms in which these matters are discussed are those of global differential geometry on manifolds.

2. A *dynamical systems*, a *dynamics*, is a vector field Δ on a manifold M of finite dimension n. The problem of mechanics is to *integrate* the dynamics, i.e., to obtain the integral curves of Δ. Such integral curves for a dynamical system can sometimes be obtained by integrating related dynamical systems on manifolds whose dimensions are lower than n. The object of this paper is to discuss ways of finding such dynamical systems of lower dimension, that is of *reducing* the original dynamics.

In the simplest possible case of such reduction, the integral curves of Δ can be found from the integral curves of two independent dynamical systems Δ_1 on M_1 and Δ_2 on M_2, but usually this is not possible. What is often possible is to obtain the integral curves of Δ from one dynamical system Δ_1 on M_1 and from a second system which is neither independent of Δ_1 nor exactly a dynamical system. We call the first case, of two independent systems, a *direct* reduction (or direct *splitting*), and the second case a *semidirect* reduction. A common example of a semidirect reduction is the usual one used in the Kepler problem in the plane (in two degrees of freedom), for which the "effective one-dimensional Hamiltonian" yields an independent dynamical system in r, p_r, and in which the remaining equation $d\theta/dt = l(mr^2)^{-1}$ is a first-order differential equation for the azimuth angle. This differential equation depends on the solution of the effective one-dimensional problem and is therefore not independent; moreover, it does not yield a vector field in the usual sense. In this paper we discuss semidirect splittings, direct ones appearing as a special case.

In all cases the reduction of the dynamical system will be achieved through a foliation.[5] That is, a foliation F of M will be found such that Δ can be projected onto the quotient space $M_1 = M/F$, and then the projected vector field is what we have called Δ_1 above. What we have called Δ_2 is then the rest of the motion, which we shall say is *on the leaves* of F. Since Δ is not projectable with respect to just every foliation of M, the problem of reducing the dynamics is essentially one of finding a suitable F.

[a)]Permanent address: Physics Department, Northeastern University, Boston, Mass. 02115.

 0022-2488/79/050856-05$01.00

Remark: In most cases that are of interest to us, these foliations turn out to be fibrations. That is, the leaves are all diffeomorphic to each other and the foliation can be written as a direct product of the leaf and U for a neighborhood U of each point in the quotient space.

Let $\pi : M \to M/F = M_1$ be the projection with respect to the foliation F. We shall assume in all cases that we deal with that this foliation is C^∞, though in some applications it turns out not to be. A vector field $X \in \mathfrak{X}(M)$ is said to be *projectable* iff for each $f \in \mathcal{F}(M_1)$ there exists an $f' \in \mathcal{F}(M_1)$ such that

$$L_X(f \circ \pi) = f' \circ \pi. \tag{1}$$

What this means is that the Lie derivative with respect to X of any function $f \circ \pi$ which is constant on leaves is itself a function $f' \circ \pi$ which is constant on leaves. Roughly speaking, those components of X which point out of the leaves are the same everywhere on each leaf. When X is projectable with respect to the foliation F, one can define a new vector field $\tilde{X} \in \mathfrak{X}(M_1)$ by using the fact that the assignment of $f \to f'$ is a *derivation*: it is linear, and for $f, g \in \mathcal{F}(M_1)$ one obtains

$$L_X(fg \circ \pi) = L_X(f \circ \pi \cdot g \circ \pi) = L_X(f \circ \pi) g \circ \pi + f \circ \pi L_X(g \circ \pi)$$

$$= f' \circ \pi \cdot g \circ \pi + f \circ \pi \cdot g' \circ \pi = (f'g + fg') \circ \pi.$$

This means that there exists a unique vector field $\tilde{X} \in \mathcal{F}(M_1)$ such that $f' = L_{\tilde{X}} f$. We write $\tilde{X} = \tilde{\pi} X = T\pi \cdot X$.

Suppose that Δ is projectable with respect to a foliation F. Then the integral curves of Δ are projected onto the integral curves of $\tilde{\Delta} = \tilde{\pi} \Delta$. Formally this may be seen as follows. Let $c : \mathbb{R} \to M$ be an integral curve of Δ, and consider its projection $\pi c : \mathbb{R} \to M_1 \equiv \pi(M)$. Then

$$\tilde{\Delta} = (\pi c(t)) = (T\pi \cdot \Delta)(\pi c(t)) = T\pi(\Delta c(t))$$

$$= T\pi(Tc(t,1)) = T(\pi c)(t,1).$$

[Recall that c is an integral curve of X iff $X(c(t)) = (Tc)(t,1)$.] Thus by integrating the reduced dynamics one obtains information about the initial dynamics. What remains to be determined is the motion along the leaves, but we shall not discuss that here in any detail.

The problem now is to find a foliation with respect to which Δ is projectable. Foliations are usually obtained in one of two ways: from functions and from involutive distributions. Consider first functions.

3. Let $f = \{f_1, \ldots, f_k\} : M \to \mathbb{R}^k$ be of class C^∞, and consider the sets $V_a = f^{-1}(a) \equiv \{m \in M \mid f(m) = a\}$, where $a = \{a_1, \ldots, a_k\} \in \mathbb{R}^k$. of M, and $M = \cup_a V_a$. This occurs, for instance, if the values of f are all regular, which is not always true (there are often even topological obstructions to such regularity[6]). A theorem by Sard states that the set R_f of regular values of f is dense in $f(M)$, and in most applications one can proceed by removing isolated points or submanifolds from M and considering them separately. At any rate, let us assume that f yields a foliation as described above. The projection $\pi : M \mapsto M_1 : m \mapsto f(m)$ with respect to this foliation maps each V_a onto a.

Let $g \in \mathcal{F}(M_1)$, and form the function $g \circ f \equiv g \circ \pi \in \mathcal{F}(M)$; we want to apply L_X to this function in order to study the projectability of a field $X \in \mathfrak{X}(M)$ with respect to the foliation given by f. We have

$$L_X(g \circ \pi) = L_X(g \circ f) = d(g \circ f)(X) = dg(Tf \cdot X).$$

What is important now is that $Tf \cdot X$ is linear in the $L_X f_j$. Indeed, in calculating $Tf \cdot X$ one forms the vector field in $\mathfrak{X}(M_1)$ whose components are the $df_j(X) = L_X f_j$. In fact in local form, when we may write $d(g \circ f) = \Sigma [\partial (g \circ f)/\partial f_j] df_j$, the calculation becomes

$$L_X(g \circ f) = d(g \circ f)(X) = \sum \frac{\partial (g \circ f)}{\partial f_j} df_j(X)$$

$$= \sum \frac{\partial (g \circ f)}{\partial f_j} L_X f_j. \tag{2}$$

Now assume that f is a constant of the motion for Δ, so that $L_\Delta f_j = 0 \forall j$. Then according to (2), $L_\Delta (g \circ \pi) = 0$, which satisfies the condition of Eq. (1), and Δ is therefore projectable with respect to the foliation. In fact Δ is projected onto the null vector field in $\mathfrak{X}(M_1)$, for the function g' that it assigns to g is just the null function. This is because f is a constant of the motion, and Δ is therefore tangent to the leaves $V_a = f^{-1}(a)$ and can be thought of as a collection of vector fields, each on its own V_a. This is a direct, rather than a semidirect splitting, moreover one in which Δ_1 or $\tilde{\Delta}$ is the null field. Only the motion on the leaves remains.

A common example of such splitting through functions is the use of the energy function. Consider, for instance, the Kepler problem after it has already been reduced to the equivalent problem in one degree of freedom. For simplicity, take the angular momentum $l > 0$ fixed, and consider only the radial part of the problem in phase space, which in this case is $T^*\mathbb{R}^+$. If the energy function $f = \frac{1}{2}(p_r^2 + l^2/r^2)/m - k/r$ is used to foliate the two-dimensional phase space, the V_a are compact closed curves for $a < 0$, and are noncompact (infinite) curves for $a \geqslant 0$.[7] The foliation is not a fibration. The quotient manifold is the semiline from the minimum energy $a = -\frac{1}{2} mk^2/l^2$ to $a = \infty$, and on $f(M)$ the dynamics is stationary: $\tilde{\Delta}$ is the null vector field and its integral curves are fixed points.

4. We now turn to foliations associated with distributions. A foliation on M defines a set of submanifolds, one of which passes through each point $m \in M$ and consequently defines at each such point a subspace of $T_m M$, namely the tangent space of the submanifold. To go from a distribution to a submanifold is to attempt the converse of this: given a set of vectors at each $m \in M$, is it possible to integrate these to obtain a foliation? More specifically, let a subspace E_m of $T_m M$ be given at every $m \in M$, each subspace of the same dimension $k < n$. The set $\mathcal{D}$ of these E_m is called a *distribution* if the transition from point to point in M is smooth: about each $m \in M$ there is a neighborhood U_m in which there exists a set of $X_i \in \mathfrak{X}(U_m)$, $i \in \{1, \ldots, k\}$, such that the $X_i(m')$ span $E_{m'}$ for $m' \in U_m$. The necessary and sufficient condition that $\mathcal{D}$ can be integrated to yield a foliation is that it be *involutive*, that is that $X, Y \in \mathcal{D} \Rightarrow [X, Y] \in \mathcal{D}$ at each point $m \in M$. For the

proof of this consequence of Frobenius's theorem and for further details about distributions, see Ref. 8. In our applications distribution will usually be given by a set of vector fields in $\mathfrak{X}(M)$, and then $X, Y \in \mathscr{D} \Rightarrow (fX + gY) \in \mathscr{D}$, $f,g \in \mathscr{F}(M)$, and involutivity becomes a global condition. What we shall use is that an involutive distribution in this sense defines a foliation.

Remark: A function $f: M \to \mathbb{R}^k$ can yield a foliation F, as we have seen, and this foliation defines a distribution $\mathscr{D}$, necessarily involutive. But it is not always true that the foliation $F_{\mathscr{D}}$ one then obtains from $\mathscr{D}$ is the same as F. For example F may have disconnected leaves, while $F_{\mathscr{D}}$ only connected ones; each connected component of a leaf of F can itself be a leaf of $F_{\mathscr{D}}$. Actually this cautionary remark depends on the definition of a foliation. If leaves of a foliation must always be connected by definition, $V_a = f^{-1}(a)$ will not always be a single leaf.

For foliations defined in terms of distributions, the test contained in Eq. (1) for projectability of a vector field is simplified. If $Y \in \mathscr{D}$, then Y is tangent to the leaves of the foliation. If $g \in \mathscr{F}(M)$ is constant on leaves, then $L_Y g = 0$ $\forall y \in \mathscr{D}$ and conversely. Moreover, every such function can be written as a function over the quotient space $M_1 = \pi(M)$ in the form $g = f \circ \pi$, $f \in \mathscr{F}(M_1)$, and every function of this form is constant on the leaves; that is $L_Y(f \circ \pi) = 0 \forall Y \in \mathscr{D}$. Thus the test for projectability becomes: X is projectable with respect to the foliation induced by $\mathscr{D}$ iff for $f \in \mathscr{F}(M)$

$$L_Y f = 0 \forall Y \in \mathscr{D} \Rightarrow L_Y L_X f = 0 \forall Y \in \mathscr{D}. \tag{3}$$

A distribution $\mathscr{D}$ will be called *invariant* under the dynamics Δ iff $[\Delta, Y] \in \mathscr{D}$ $\forall Y \in \mathscr{D}$, that is, if $L_\Delta Y \in \mathscr{D}$ $\forall$ $Y \in \mathscr{D}$. If $\mathscr{D}$ is invariant under Δ, then Δ is projectable with respect to the foliation induced by $\mathscr{D}$. Indeed, let $f \in \mathscr{F}(M)$ be such that $L_Y f = 0 \forall Y \in \mathscr{D}$. Then (3) is satisfied, for

$$L_Y L_\Delta f = [L_Y L_\Delta - L_\Delta L_Y] f = -L_{[\Delta, Y]} f = 0,$$

since $[Y, \Delta] \in \mathscr{D}$. It is thus seen that an invariant distribution will provide a splitting of Δ.

Remark: If Δ itself is in an involutive distribution $\mathscr{D}$, then $\mathscr{D}$ is automatically invariant. Moreover, $L_Y f = 0$ $\forall$ $Y \in \mathscr{D}$ implies that $L_\Delta f = 0$, and Δ is trivially projectable according to (3) with $\tilde{\pi}\Delta = 0$. As in the case of foliations defined by functions, Δ is projected onto the null vector field and all the motion is on the leaves.

5. The next question, then, is how to find invariant distributions. Finding them is hardly ever as simple or intuitive as finding constants of the motion, but it can be made somewhat easier by the use of differential forms. Let $\alpha \in \Omega^p(M)$ be a p form; we will be interested now in p forms with nonnull kernels. The kernel of α, $\ker\alpha \equiv \{X \in \mathfrak{X}(M) \mid i_X \alpha = 0\}$, is a distribution $\mathscr{D}_\alpha$ and it can be shown that $\mathscr{D}_\alpha$ is involutive iff $\alpha \wedge d\alpha = 0$. In particular, if $d\alpha = 0$, i.e., if α is closed, $\mathscr{D}_\alpha$ is involutive. If α is, moreover, *conformally invariant under* Δ, that is, if

$$L_\Delta \alpha = f\alpha, \quad f \in \mathscr{F}(M) \tag{4}$$

(when $f = 0$, α is said to be *invariant under* Δ), then $\mathscr{D}_\alpha$ is invariant under Δ. Indeed, $Y \in \mathscr{D}_\alpha$ implies $[\Delta, Y] \in \mathscr{D}_\alpha$, for

$$i_{[\Delta, Y]}\alpha = [L_\Delta i_Y - i_Y L_\Delta]\alpha = -i_Y(f\alpha) = -f i_Y \alpha = 0.$$

Thus a conformally invariant closed p form provides an involutive distribution and hence a foliation with respect to which Δ is projectable, and thereby reduces the dynamics.

This procedure may seem very contrived, but it is a general one and most, if not all, of the standard procedures used to reduce dynamical systems are in fact special cases. We discuss some of these.

Let Δ be a Hamiltonian vector field on a symplectic manifold (M,ω). Assume that it has already been reduced with the aid of a constant of the motion $f: M \to \mathbb{R}^k$ as described above. Then each $V_a = f^{-1}(a)$ can be said to contain a part of the dynamics Δ, as has been mentioned already. Let us call each such part Δ_a; this is a vector field on V_a. A closed invariant 2-form can also be defined on each V_a as follows. Since every $X_a \in \mathfrak{X}(V_a)$ can be obtained by restriction from at least one $X \in \mathfrak{X}(M)$ which is tangent to V_a, the equation

$$\omega_a(X_a, Y_a)(m) = \omega(X, Y)(m), \quad m \in V_a$$

defines the 2-form ω_a on V_a which is closed because ω is closed and invariant under Δ_a because ω is invariant under Δ. It is generally not regular, however (V_a may even be of odd dimension), so that $\ker\omega_a$ need not be empty. Thus V_a can be foliated and the dynamics Δ_a reduced by the procedure we have just described, but with M replaced by V_a and α by ω_a. It can be shown that the quotient space inherits a symplectic structure from ω_a (from ω) and that the projected vector field $\tilde{\Delta}_a$ is Hamiltonian with respect to this structure.[1]

This is the usual procedure used, for example, in reducing the three-dimensional Kepler problem (the phase space has six dimensions) in a two-step process. First the angular momentum vector is used to foliate the phase space with submanifolds V_a of dimension three, and then the 2-form ω_a obtained from ω is used to foliate each V_a. In the resulting semidirect splitting, $\tilde{\Delta}_a$ is the vector field of the equivalent radial problem which was discussed above (and which we split further by using the energy function). Note that it is, as is well known, Hamiltonian. The rest of the motion, as was also mentioned before, is on the leaves, the motion of the azimuth angle.

As another example, consider the dynamical vector field on $\mathbb{R}^4$ given by

$$\Delta = f_1 \frac{\partial}{\partial q_1} + g_1 \frac{\partial}{\partial p_1} + f_2 \frac{\partial}{\partial q_2} + g_2 \frac{\partial}{\partial q_2},$$

where f_1 and g_1 are functions only of the variables q_1, p_1, while f_2 and g_2 are functions of all four variables. This system is presented already in a semidirect split form, and the procedure for integrating it is apparent: solve first for the variables q_1, p_1 and insert the solution into the equations one obtains for the other variables q_2, p_2. A clear understanding of what is happening is obtained in our terms when one notes that $\alpha = dq_1 \wedge dp_1$ is conformally invariant under Δ:

$$L_\Delta \alpha = \left(\frac{\partial f_1}{\partial q_1} + \frac{\partial g_1}{\partial p_1} \right) \alpha. \tag{4'}$$

The kernel of α is the distribution

$$\mathscr{D}_\alpha = \left\{ \varphi \frac{\partial}{\partial q_2} + \psi \frac{\partial}{\partial p_2} ; \varphi, \psi \in \mathscr{F}(\mathbb{R}^4) \right\}.$$

All vector fields in $\mathscr{D}_\alpha$ are parallel to the planes whose equations are $q_1 = \text{const}, p_1 = \text{const}$: these planes provide the foliation. The quotient of $\mathbb{R}^4$ with respect to this foliation is the plane of all values of q_1 and p_1, and on this quotient manifold (isomorphic to what one might call the q_1, p_1 plane, namely $q_2 = 0, p_2 = 0$) the dynamics $\tilde{\Delta}$ is independent. Note that the foliation in this case is associated with no constants of the motion. The functions q_1 and p_1, for example, are not constants of the motion.

Remark: According to Eq. (4) $\tilde{\Delta}$ is Hamiltonian with respect to the symplectic form $dq_1 \wedge dp_1 = \alpha$ iff α is invariant, for then $L_\Delta \alpha = di_\Delta \alpha = 0$, and a function $h \in \mathscr{F}(\mathbb{R}^2)$ exists such that $i_\Delta \alpha = dh$, or such that $f_1 = \partial h / \partial p_1$, $g_1 = -\partial h / \partial q_1$.

A recent paper[2] uses a technique which mixes these general procedures. Let $f : M \to \mathbb{R}^k$ be a constant of the motion, and let $\mathbf{S}$ be a set of globally linearly independent vector fields $\{S_1, \ldots, S_r\}$ which are symmetries for Δ, that is such that $[S_j, \Delta] = 0$. With certain assumptions, among which are that $\mathbf{S}$ define an involutive distribution $\mathscr{D}_S$ and that the $L_{S_j} f_k = T_{kj} \in \mathscr{F}(M)$ can be written as functions of the f_k, it is shown that a reduction of the dynamics is obtained. In our terms this can be seen as follows. Let $\mathscr{D}_f$ be the distribution given by the foliation induced by the constant of the motion f. That is, $\mathscr{D}_f = \{X \in \mathfrak{X}(M) \mid L_X f = 0\}$. Then it can be shown that $\mathscr{D}_S \cap \mathscr{D}_f \equiv \mathscr{D}$ is an invariant distribution, necessarily involutive, and it can be used to reduce the dynamics.

6. Having generalized from foliations associated with functions (constants of the motion) to more general foliations associated with distributions and through them with p forms, one may try proceeding also in another way to foliations associated with C^∞ applications of the form $\phi : M \to N$, where N is a differential manifold of dimension $k < n$. In almost exact analogy with the case of functions, we say that a vector field $X \in \mathfrak{X}(M)$ is projectable with respect to the foliation associated with ϕ iff for every $f \in \mathscr{F}(N)$ there exists an $f' \in \mathscr{F}(N)$ such that

$$L_X \phi_* f = \phi_* f'. \tag{5}$$

If, moreover, ϕ is onto (or, one could say, if $\phi(M)$ is a differential manifold, which could then be defined as N), there exists a unique vector field $\tilde{X} \in \mathfrak{X}(N)$ such that $L_{\tilde{X}} f = f'$. Now, $L_X \phi_* f$ can be calculated as in the case of constants of the motion, and again in exact analogy one obtains

$$L_X \phi_* f = df(T\phi \cdot X). \tag{6}$$

As in the case of functions, a foliation with respect to which Δ is projectable can be found if ϕ is a *constant of the motion*, that is, if $\phi(c(t))$ is a fixed point in N, where $c(t)$ is an integral curve of Δ. In that case

$$T\phi \cdot \Delta(c(t)) = T\phi(Tc(t,1)) = T(\phi c)(t,1) = 0,$$

since $\phi(c(t))$ is fixed in N and hence has zero derivative. Since (we assume) an integral curve passes through each point $m \in M$, it follows that $T\phi \cdot \Delta(m) = 0$. Then according to (6) $L_\Delta \phi_* f = 0 \forall f \in \mathscr{F}(N)$, and Δ is trivially projectable by Eq. (5); $\tilde{\Delta}$ is the null vector field, as in the analogous case in which N is replaced by $\mathbb{R}^k$, and the motion is entirely on the leaves.

A very simple example of this situation is the case of the vector field

$$\Delta = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}$$

On $M = \mathbb{R}^2 - \{0\}$, with the application $\phi : M \to S^1$. To describe ϕ we use the angle θ around the circle S^1:

$$\phi : (x,y) \mapsto \theta = \arctan(y/x).$$

(Of course, this requires more than one chart.) Then if $X \in \mathfrak{X}(M)$ has components ξ, η, it is easily shown that

$$T\phi : (x,y;\xi,\eta) \to \left(\theta, \frac{-\xi y + x\eta}{x^2 + y^2} \right).$$

The components of Δ are $\xi = x$, $\eta = y$, so that at each point (x,y) this vector field is mapped into the null vector field by $T\phi$, and thus ϕ is a constant of the motion. The leaves of the foliation, in this case orbits of the dynamics, are obtained by fixing θ, or from $y/x = \text{const}$. They are the straight lines passing through the origin.

7. In conclusion we mention several procedures used in analyzing dynamical systems, procedures which, from the geometrical point of view, are closely related to projecting vector fields with respect to foliations.

In action-angle variables, the Hamiltonian is written as a function only of the action variables J_i, constants of the motion. The function $J = \{J_1, \ldots, J_k\} : M \to \mathbb{R}^k$ defines a foliation with respect to which the dynamics is projected onto the null vector field, as is always the case with constants of the motion. The leaves of this foliation, at least locally, are tori, and on them the motion is extremely simple $\dot{w}_i = \nu_i(J)$. As this procedure is generally applied, moreover, $k = n/2$, so that the leaves and the quotient manifold have the same dimension. For the simple harmonic oscillator the ν_i are independent of the J_i: the splitting is then direct.

Certain types of perturbation methods can also be understood in the light of this reduction procedure. Suppose, for instance, that our manifold M is foliated by a certain distribution $\mathscr{D}$ which is invariant with respect to the vector field Δ. Suppose, moreover, that Δ can be written in the form $\Delta = \Delta_0 + \mu\Delta'$, where $\Delta_0 \in \mathscr{D}$, so that Δ_0 projects down to the null vector field on the quotient manifold, and that $\mathscr{D}$ is therefore invariant under Δ'. It then follows that $\tilde{\pi}\Delta = \tilde{\pi}\Delta'$, and the global dynamics can be thought of as a motion that takes place on the leaves, while the leaves are carried into each other by Δ'. If μ is a very small parameter, this motion *of* the leaves may be very slow as compared to the motion *on* them, and thus in a sense one can study each motion separately, almost as though they were decoupled. In fact this is true also if μ is very large, so that the motion takes place mostly on the quotient manifold, remaining relatively stationary on the leaves.

Somewhat more specifically, let $\mathscr{D}$ be invariant under Δ, and let us suppose that Δ can be decomposed into a field

$\Delta_0 \in \mathscr{D}$ and $Z = \Delta - \Delta_0$, where Z is "small" with respect to Δ_0. We can then consider the average of Z on a leaf [9]

$$\langle Z \rangle = \int_{\pi^{-1}(m_\alpha)} Z \, d\mu.$$

We shall call $\langle Z \rangle$ the *mean field* induced by Δ on the quotient manifold, for in a sense what we have done is to replace Z by a constant field which reproduces its action in an average way. When we replace Δ by $\Delta - Z + \langle Z \rangle = \Delta_0 + \langle Z \rangle$, we obtain a field whose components pointing out of the leaves are constant, in agreement with the remark after Eq. (1). Of course the projection of this field along the leaves is not the same as the projection of Δ along the leaves but it is possible to obtain estimates of the error made.[10] When M is a symplectic manifold, this is in fact the procedure used in studying adiabatic invariance.[11]

Something like the reverse of this reduction procedure can sometimes be used to investigate difficult dynamical systems. Suppose that a reasonably complicated dynamical system Δ is given on a manifold M, and suppose that a larger manifold M' can be constructed so that M appears as a quotient with respect to some foliation in M'. To simplify the considerations, let us assume that foliation to be a fibration, so that (M', π, M) is a fiber bundle in which π is the projection $\pi : M' \to M$. It may be possible to choose motion on the fibers in such a way that it becomes easy to integrate the total dynamics on M', consisting of the motion on the fibers plus the dynamics on the base. Then the initial dynamics on the base can be integrated by projecting with π. A simple example of this procedure, cooked up for this demonstration, is the following. Let M be the contangent bundle of the semiaxis, $T^*\mathbb{R}^+$, with the natural symplectic structure $\omega = dq \wedge dp$, $q \in \mathbb{R}^+$. Consider the dynamical field whose Hamiltonian is

$$H = \tfrac{1}{2}(p^2 + q^2) + \tfrac{1}{2} l^2/q^2, \tag{7}$$

where l is a constant. Now extend M in two steps. First at each point of $T^*\mathbb{R}^+$ we define the fiber S^1 (the circle), and on the fiber consider the motion whose equation is

$$\frac{d\theta}{dt} = \frac{l}{q^2},$$

where θ is the usual (local) angle coordinate on the circle. We have now extended M to $S^1 \times T^*\mathbb{R}^+ = S^1 \times \mathbb{R}^+ \times \mathbb{R} = (\mathbb{R}^2 - \{0\}) \times \mathbb{R}$, in which (q, θ) give the local polar chart on $\mathbb{R}^2 - \{0\}$, and the remaining variable p gives a chart on $\mathbb{R}$. What we have so far is related in an obvious way to $M' = T^*(\mathbb{R}^2 - \{0\})$: we extend our manifold to M' by forming the contangent bundle over S^1, and then the natural symplectic form $\omega' = \omega + \omega_S$, where ω_S is the natural symplectic form of T^*S^1, can be written in our local chart as

$$\omega = dq \wedge dp + d\theta \wedge dp_\theta.$$

Now consider the isotropic harmonic oscillator on $\mathbb{R}^2 - \{0\}$, whose Hamiltonian can be written in a Cartesian chart in the form

$$H' = \tfrac{1}{2}(p_x^2 + x^2) + \tfrac{1}{2}(p_y^2 + y^2). \tag{8}$$

The dynamical system Δ' obtained from H' can be reduced by the procedure which uses the angular momentum in the way we have described for the Kepler problem. The resulting Hamiltonian dynamics projected onto $T^*\mathbb{R}^+$ is obtained from the Hamiltonian of Eq. (7). But Δ' is an extremely easy problem to deal with. In fact, as in the example of Eq. (4) it is already split, in this case directly.

What we are describing here is a sort of unfolding of the initial dynamics by imbedding it in a larger one which is easier to integrate, and then projecting the solution back to the initial manifold. In the example we have given, the larger dynamical system is easier to integrate because after being constructed it can be projected in a different way, onto different submanifolds than the original one. It is conceivable that the larger system is easy to integrate for other reasons. We go no further into this procedure. A systematic description and a more detailed treatment will be found in Ref. 12.

ACKNOWLEDGMENT

We thank Professor Bruno Vitale for many helpful discussions during the course of this work.

[1] S. Smale, "Topology and Mechanics I, II," Inv. Math. **10**, 305 (1970); and **11**, 45 (1970); J.M. Sourieau, *Structure des Systêmes Dynamiques* (Dunod, Paris, 1970); J. Marsden and A. Weinstein, "Reduction of Symplectic Manifolds with Symmetry," Rep. Math. Phys. **5**, 122 (1974); G.M. Marle, "Symplectic Manifolds, Dynamical Groups and Hamiltonian Mechanics," edited by M. Cahen and M. Flato (Reidel, Boston, 1976); N.N. Nehorosov, Trans. Moscow Math. Soc. **26**, 121 (1972); K.R. Meyer, *Symmetries and Integrals in Mechanics Dynamical Systems*, edited by M. Peixoto (Academic, New York, 1973), p. 259.

[2] F. Gonzales-Gascon and F. Moreno-Insertis, "Symmetries, First Integrals and Splittability of Dynamical Systems," Lett. Nuovo Cimento **21**, 253 (1978).

[3] G. Marmo, E.J. Saletan, and A. Simoni, "Reduction of Symplectic Manifolds through Constants of the Motion," to appear in Nuovo Cimento.

[4] L.P. Eisenhardt, "Continuous Groups of Transformations," Dover reprint, 1963; L.P. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Vol. 5.

[5] C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, 1957); F. Brickell, and R.S. Clark, *Differentiable Manifolds* (Van Nostrand, Reinhold, London, 1970); R. Palais, *A Global Formulation of Lie Theory of Transformation Groups* Mem. 22 (Amer. Math. Soc., Providence, R.I., 1975).

[6] R. Thom, "Quelques Proprietés Globals des Variétés Differentiables," Comm. Math. Helv. **28**, 17–85 (1954).

[7] The minimum energy $a = -\frac{1}{2} mk^2/l^2$ is not a regular value, for $df = 0$ at this value of a. The corresponding V_a is the point $r = l^2/(mk)$, $p_r = 0$ and is of dimension zero, rather than one, which is the dimension of the V_a for all the other values of a. Thus this point has to be treated separately; it corresponds to the circular orbit in configuration space.

[8] F.W. Warner *Foundations of Differentiable Manifolds and Lie Groups* (Scott, Foresman, Glenview, Ill., 1971); see also Brickell and Clark, Ref. 5.

[9] The integral in this equation is to be understood symbolically. Its accurate definition requires some notion of transport of vectors on the leaf, so that vectors at two different points can be added. Such a notion may be obtained from a group action on the leaf or something like that, such as a connection.

[10] V. Arnol'd, *Les Méthodes Mathematiques de la Mechanique Classique* (Mir, Moscow, 1976).

[11] T. Levi-Civita, *Opere Complete* (Zanichelli, Bologna), Vol. IV. pp. 466–545; Vol. V, pp. 417–52.

[12] D. Kazhdan, B. Kostant, and S. Sternberg, "Hamiltonian Group Actions and Dynamical Systems of Calogero Type," Commun. Pure Appl. Math. (to appear, 1978).

On an inhomogeneous Schrödinger equation and its solutions in scattering theory

H. van Haeringen

Natuurkundig Laboratorium der Vrije Universiteit, Amsterdam, The Netherlands
and Institute for Theoretical Physics, P.O. Box 800, University of Groningen, The Netherlands
(Received 19 July 1978)

We prove that $\psi_{s,l}$, the partial-wave projection of the irregular Coulomb wavefunction ψ_s, is a solution of an inhomogeneous Schrödinger equation. New expressions for $\psi_{s,l}$ and ψ_s are obtained in terms of the Coulomb Green functions $G_{C,l}$ and G_C, respectively. We discuss irregular solutions, the analogs of ψ_s, for Coulomb-like and short-range potentials. We find that in general these functions do not approach asymptotically the scattering amplitude times an outgoing spherical wave, in contrast to the pure Coulomb function ψ_s.

1. INTRODUCTION

The physical three-dimensional Coulomb scattering wavefunction $\psi^{(+)}$ is customarily split up into an "incoming part" ψ_i and a "scattered part" ψ_s. Each one of these three functions is a solution of the Schrödinger equation, $\psi^{(+)}$ is regular, ψ_i and ψ_s are irregular. In Ref. 1 we have derived closed expressions for $\psi_l^{(+)}$,$\psi_{i,l}$, and $\psi_{s,l}$, the partial wave (p.w.) projections of $\psi^{(+)}$, ψ_i, and ψ_s, respectively. We proved that $\psi_{i,l}$ and $\psi_{s,l}$ are *no* solutions of the p.w. projected Schrödinger equation.

The function ψ_s asymptotically approaches the Coulomb scattering amplitude times a Coulomb-modified outgoing spherical wave [cf. Eq. (5.1)]. The question arises whether there also exists for other potentials a function which

(i) is an irregular solution of the three-dimensional Schrödinger equation, and

(ii) asymptotically approaches the scattering amplitude times an outgoing spherical wave (possibly modified).

In this paper we shall discuss a large class of irregular solutions of the three-dimensional Schrödinger equation with a local potential. Their asymptotic behavior is easily obtained when the potential is spherically symmetric. In this case we are able to show that the condition (ii) is not satisfied in general. It seems that the *pure* Coulomb potential is a remarkable exception in this respect.

In Sec. 2 we shall prove that $\psi_{s,l}$ is a solution of an "inhomogeneous Schrödinger equation," see Eq. (2.3). With the help of this result we deduce in Sec. 3 a new expression for $\psi_{s,l}$, in terms of the Coulomb Green function $G_{c,l}$, Eq. (3.1). In the second part of Sec. 3 we investigate the behavior of $\psi_{s,l}(r)$ for $r\to 0$, starting from different equivalent expressions. When $l=0$ this function diverges like $\ln r$, but for $l>0$ it has a finite limit for $r\to 0$ [Eq. (3.17)].

Summation of the p.w. series with $\psi_{s,l}$ leads in a natural way to an expression for $\psi_s(\mathbf{k},\mathbf{r})$ in terms of the three-dimensional Coulomb–Green function G_C, Eq. (4.1). We define in Eq. (4.2) a class of irregular solutions $\psi_w(\mathbf{k},\mathbf{r})$ of the three-dimensional Schrödinger equation for a not necessarily spherically symmetric potential in analogy to ψ_s, and study these functions in Sec. 4. We also discuss here the connection with a line charge distribution on the positive z axis.

The most interesting feature of the Coulomb irregular solution ψ_s is, as we said before, that it asymptotically approaches a Coulomb-modified outgoing spherical wave times the Coulomb scattering amplitude. In Sec. 5 we discuss the question whether such an irregular solution with a similar asymptotic behavior can be found for other potentials. We successively consider the Coulomb, Coulomb-like, and short-range potentials, first with the "Coulomb-choice" for w, i.e., $w(r)$ proportional to e^{ikr}, and afterwards for other functions w. We have not been able to find an irregular solution ψ_w with the desired property of giving the scattering amplitude, so it seems to be fortuitous that ψ_s yields asymptotically the scattering amplitude. Therefore, although the regular physical wavefunction $\psi^{(+)}(\mathbf{k},\mathbf{r})$ for any local potential can be expressed as the sum of two irregular solutions, $\psi^{(+)}=\psi_i+\psi_s$, this splitting seems to be useful only in the pure Coulomb case.

We shall work throughout in the coordinate representation and restrict ourselves to local potentials. As usual we take $\hbar=2m=1$, $E=(k+i\epsilon)^2$ with $\epsilon\downarrow 0$, and we suppress the energy dependence of G, G_0, and T. We will often use the subscript C to denote Coulomb quantities.

The p.w. "outgoing" physical scattering state is denoted by $|kl+\rangle$, cf. Eq. (11.13) of Taylor.[2] Its connection with Newton's $\psi_l^{(+)}$ and φ_l follows from

$$\langle r|kl+\rangle=(2/\pi)^{1/2}(kr)^{-1}i^l\psi_l^{(+)}(k,r) \tag{1.1}$$

and [Eq. (12.145) of Ref. 3]

$$\psi_l^{(+)}(k,r)=k^{l+1}\varphi_l(k,r)f_l^{-1}(k)/(2l+1)!!, \tag{1.2}$$

where $f_l(k)$ is the Jost function. Furthermore we will use the symbols $|kl\uparrow\rangle$ and $|kl\downarrow\rangle$ to denote the Jost solutions of the p.w. Schrödinger equation, see Ref. 1. We have

$$\langle r|kl\uparrow\rangle=(2/\pi)^{1/2}(kr)^{-1}f_l(k,r), \tag{1.3}$$

$$2i|kl+\rangle=e^{2i\delta_l}|kl\uparrow\rangle-|kl\downarrow\rangle, \tag{1.4}$$

and

$$\langle r|kl+\rangle=(-)^l\langle r|kl-\rangle^*=(-)^l\langle kl-|r\rangle. \tag{1.5}$$

The Coulomb Jost solution is denoted by $\langle r|kl\uparrow\rangle_C$, and for

 0022-2488/79/050861-07$01.00

$V \equiv 0$ we have[1]

$$\langle r|kl\uparrow\rangle_0 = (2/\pi)^{1/2} i^l h_l^{(+)}(kr). \tag{1.6}$$

We shall suppress l when $l = 0$. In particular,

$$\langle r|k\,0\uparrow\rangle_0 = \langle r|k\uparrow\rangle_0 = (2/\pi)^{1/2} e^{ikr}/(kr).$$

The subscript 0 to a bra or ket signifies $V = 0$, whereas for a function, e.g., in $f^*_{C,0}$, it means $l = 0$. The behavior of $\langle r|kl+\rangle$ and $\langle r|kl\uparrow\rangle$ at $r = 0$ follows from

$$\lim_{r\to 0}(\pi/2)^{1/2}(2ikr)^{-l}\langle r|kl+\rangle(2l+1)!/l!$$
$$= f_l^{-1}(k) \tag{1.7}$$

and

$$\lim_{r\to 0}(\pi/2)^{1/2}(-2ikr)^l kr\langle r|kl\uparrow\rangle l!/(2l)!$$
$$= f_l(k), \tag{1.8}$$

respectively. These equations are valid for Coulomb-like as well as for (nonsingular) short-range potentials.

For a local central potential $V_l(r)$ is independent of l. Therefore, we shall occasionally suppress the subscript l here.

2. COULOMB FUNCTIONS SATISFYING AN INHOMOGENEOUS SCHRÖDINGER EQUATION

In this section we shall prove that χ_l (see Ref. 1) is a solution of the following inhomogeneous differential equation of the Schrödinger type,

$$(k^2 - H_{C,l})\chi_l(kr) = -\langle r|V_C|k\uparrow\rangle_0 f^*_{C,0}, \tag{2.1a}$$

that is, written in a more explicit form,

$$\left(k^2 + \frac{1}{r}\frac{d^2}{dr^2}r - \frac{l(l+1)}{r^2} - \frac{2k\gamma}{r}\right)\chi_l(kr)$$
$$= -\frac{2k\gamma}{r}\left(\frac{2}{\pi}\right)^{1/2}\frac{e^{ikr}}{kr}\frac{e^{\pi\gamma/2}}{\Gamma(1-i\gamma)}. \tag{2.1b}$$

Here V_C is the Coulomb potential and $f^*_{C,0}$ is the complex conjugate of the Coulomb Jost function for $l = 0$ (e.g., Ref. 3),

$$f_{C,l} = f_{C,l}(k) = e^{\pi\gamma/2}\Gamma(l+1)/\Gamma(l+1+i\gamma).$$

The function χ_l has been defined in Ref. 1 by

$$\chi_l(kr) = e^{2i\sigma_l}\langle r|kl\uparrow\rangle_C - 2i\psi_{s,l}(r), \tag{2.2}$$

where $\langle r|kl\uparrow\rangle_C$ is the Jost solution for the p.w. Schrödinger equation with the Coulomb potential. It follows that if Eq. (2.1) is valid, we also have

$$(k^2 - H_{C,l})\psi_{s,l}(r) = \langle r|V_C|k\uparrow\rangle_0 f^*_{C,0}/(2i). \tag{2.3}$$

As we said in the Introduction, $\psi_{s,l}$ is the p.w. projection of ψ_s [see Eq. (4) of Ref. 1],

$$\psi_s(\mathbf{k},\mathbf{r}) = -(2\pi)^{-(3/2)}e^{\pi\gamma/2}(\Gamma(1+i\gamma)/\Gamma(-i\gamma))$$
$$\times e^{ikr}U(1+i\gamma,1,i\mathbf{k}\cdot\mathbf{r} - ikr). \tag{2.4}$$

In order to prove Eq. (2.1), we substitute the following closed expression for χ_l [see Eq. (A.17) of Ref. 1],

$$\chi_l(kr) = (2/\pi)^{1/2}\frac{\exp(\pi\gamma/2)}{\Gamma(1-i\gamma)}\frac{\exp(ikr)}{kr}$$
$$\times {}_1F_1\left(-l,l+1,1;1-i\gamma;\frac{1}{2ikr}\right), \tag{2.5}$$

and introduce the new variables $z = (2ikr)^{-1}$ and $\mu = -i\gamma$. After some manipulations the equation to be proved reduces to

$$-z^3F''(z) + z(1-2z)F'(z) + [\mu + l(l+1)z]F(z) = \mu. \tag{2.6}$$

Here

$$F(z) \equiv {}_1F_1(-l,l+1,1;\mu+1;z)$$
$$= \sum_{n=0}^{l} z^n(-l)_n(l+1)_n/(\mu+1)_n$$

is a polynomial, so the proof of Eq. (2.6) is obtained in a straightforward way.

3. A NEW EXPRESSION FOR $\psi_{s,l}$

In this section we shall prove the equation

$$G_{C,l}V_{C,l}|k\uparrow\rangle_0 f^*_{C,0} = 2i\psi_{s,l}, \tag{3.1}$$

where

$$2i\psi_{s,l} = e^{2i\sigma_l}|kl\uparrow\rangle_C - \chi_l. \tag{3.2}$$

The left-hand side of (3.1) gives a new expression for $\psi_{s,l}$. Further we shall investigate the behavior of $\psi_{s,l}(k,r)$ for $r\to 0$, see Eq. (3.17).

Note, however, that

$$G_{C,l}V_{C,l}|k\downarrow\rangle_0$$

is even not defined. This can be easily deduced from our discussion below [the integral $\int_r^\infty$ in Eq. (3.8) would be divergent in this case], but it also follows from the equality $G_{C,l}V_{C,l} = G_{0,l}T_{C,l}$ and the well-known fact that the half-shell Coulomb T matrix, that is $T_{C,l}|kl\rangle$, is not defined.

For the proof of Eq. (3.1) we use

$$\langle r|G_{C,l}|r'\rangle = (-)^{l+1}\tfrac{1}{2}\pi k\langle r_<|kl+\rangle\langle r_>|kl\uparrow\rangle, \tag{3.3}$$

where $r_<$ is the smaller one and $r_>$ the larger one of the pair r,r'. Such a representation of the Green function holds for any local central potential, as is well known.

A natural and direct way to prove Eq. (3.1) would consist of inserting (3.3) and using the known explicit expressions for the regular and irregular Coulomb wavefunctions, i.e.,

$$\langle r|kl+\rangle_C = (2/\pi)^{1/2}e^{-\pi\gamma/2}[\Gamma(l+1+i\gamma)/\Gamma(2l+2)]$$
$$\times(2ikr)^l e^{-ikr}\,{}_1F_1(l+1-i\gamma;2l+2;2ikr), \tag{3.4a}$$

and

$$\langle r|kl\uparrow\rangle_C = (2/\pi)^{1/2}e^{\pi\gamma/2+ikr}(kr)^{-1}(-2ikr)^{l+1}$$

$$\times U(l+1+i\gamma,2l+2,-2ikr). \tag{3.4b}$$

However, it turns out that this approach is somewhat complicated. We have been able to prove Eq. (3.1) in this way only for $l=0$ and for $l=1$. In order to show the complications arising here, we now briefly discuss the $l=0$ case. By using

$$\frac{d}{dz}\,{}_1F_1(-i\gamma;1;z) = -i\gamma\ {}_1F_1(1-i\gamma;2;z),$$

$$\frac{d}{dz}\,e^{z}U(1+i\gamma,1,z) = -e^{z}U(1+i\gamma,2,z),$$

and

$${}_1F_1(-i\gamma;1;z)$$
$$= -i\gamma\ {}_1F_1(1-i\gamma;2;z) + (1+i\gamma)\ {}_1F_1(-i\gamma;2;z),$$
$$U(1+i\gamma,1,z)$$
$$= U(1+i\gamma,2,z) - (1+i\gamma)U(2+i\gamma,2,z),$$

we obtain

$$\langle r|G_C V_C|k\uparrow\rangle_0$$
$$= e^{\pi\gamma/2}\Gamma(1+i\gamma)\langle r|k\uparrow\rangle_C + (2/\pi)^{1/2}2ie^{ikr}\Gamma(1+i\gamma)$$
$$\times[{}_1F_1(-i\gamma;1;2ikr)U(1+i\gamma,2,-2ikr)$$
$$+ i\gamma\ {}_1F_1(1-i\gamma;2;2ikr)U(1+i\gamma,1,-2ikr)]. \tag{3.5}$$

The expression between the square brackets can be reduced by noting that the Wronskian W for the functions

$$f(z)\equiv{}_1F_1(-i\gamma;1;z)$$

and

$$g(z)\equiv e^{z}U(1+i\gamma,1,-z)$$

is equal to

$$W(f,g)\equiv fg' - f'g = z^{-1}\exp[z + i\pi\,\mathrm{sgn}(\mathrm{Im}z)]/\Gamma(1+i\gamma).$$

In this way we get from Eq. (3.5),

$$\langle r|G_C V_C|k\uparrow\rangle_0 f^*_{C,0} = e^{2i\sigma_0}\langle r|k\uparrow\rangle_C$$
$$-(2/\pi)^{1/2}e^{\pi\gamma/2}(kr)^{-1}e^{ikr}/\Gamma(1-i\gamma),$$

which is just Eq. (3.1) for $l=0$.

For $l>1$ the above procedure is rather complicated. Therefore, we resort to a different approach.

In the preceding section we have proved

$$(k^2 - H_{C,l})(2i\psi_{s,l}) = V_C|k\uparrow\rangle_0 f^*_{C,0}. \tag{3.6}$$

This equation follows from Eq. (3.1), but not vice versa. We shall nevertheless prove Eq. (3.1) with the help of Eq. (3.6). To this end we first observe that the quantity $G_{C,l}V_C|k\uparrow\rangle_0 f^*_{C,0}$ is a solution of the same inhomogeneous differential equation,

$$(k^2 - H_{C,l})G_{C,l}V_C|k\uparrow\rangle_0 f^*_{C,0} = V_C|k\uparrow\rangle_0 f^*_{C,0}.$$

Therefore, this quantity equals the sum of a particular solution of this equation and some solution of the corresponding homogeneous differential equation. According to Eq. (2.1), $-\chi_l$ is a particular solution. Further, we know that any solution of the homogeneous differential equation is a linear combination of $|kl\uparrow\rangle_C$ and $|kl\downarrow\rangle_C$. Therefore,

$$G_{C,l}V_C|k\uparrow\rangle_0 f^*_{C,0} = -\chi_l + C_1|kl\uparrow\rangle_C + C_2|kl\downarrow\rangle_C. \tag{3.7}$$

We shall prove that $C_2=0$ and $C_1 = e^{2i\sigma_l}$ by establishing the behavior of the left-hand side for $r\to\infty$ and for $r\to 0$, respectively.

Substitution of (3.3) in the left-hand side of Eq. (3.7) yields

$$\langle r|G_{C,l}V_C|k\uparrow\rangle_0 f^*_{C,0}$$
$$= (-)^{l+1}\tfrac{1}{2}\pi k f^*_{C,0}\Big[\langle r|kl\uparrow\rangle_C\int_0^r\langle r'|kl+\rangle_C\ V_C(r')$$
$$\times\langle r'|k\uparrow\rangle_0 r'^2dr' + \langle r|kl+\rangle_C\int_r^\infty\langle r'|kl\uparrow\rangle_C$$
$$\times V_C(r')\langle r'|k\uparrow\rangle_0 r'^2dr'\Big]. \tag{3.8}$$

We further use Eq. (1.4) for the Coulomb case,

$$2i|kl+\rangle_C = e^{2i\sigma_l}|kl\uparrow\rangle_C - |kl\downarrow\rangle_C$$

and

$$\langle r|kl\uparrow\rangle_C \sim (2/\pi)^{1/2}(kr)^{-1}\exp[ikr - i\gamma\ln(2kr)],\quad r\to\infty. \tag{3.9}$$

It follows that for $r\to\infty$ the second term on the right-hand side of Eq. (3.8) is negligible. For the first term we find, for $r\to\infty$,

$$-(2/\pi)^{1/2}f^*_{C,0}(kr)^{-1}e^{ikr} + \text{const}\ r^{-1-i\gamma}e^{ikr}.$$

Clearly this implies that we have $C_2=0$ in Eq. (3.7).

In order to prove $C_1 = e^{2i\sigma_l}$, we consider the expressions in Eq. (3.8) for $r\to 0$. With the help of Eqs. (1.7) and (1.8) one easily verifies that

$$\langle r|G_{C,l}V_C|k\uparrow\rangle_0 = O(\ln r),\qquad r\to 0, \text{when} l=0,$$
$$= O(1),\quad r\to 0, \text{when} l = 1,2,3,\cdots. \tag{3.10}$$

Finally we use Eq. (3.4b), where (Ref. 4, p. 288, corrected)

$$U(a,c,z) = z^{1-c}\Gamma(c-1)/\Gamma(a) + O(|z|^{2-\mathrm{Re}c}),$$
$$z\to 0,\ \mathrm{Re}c\geqslant 2,\ c\neq 2, \tag{3.11}$$

and deduce from Eq. (26) of Ref. 1 that

$$\chi_l(kr)\simeq(2/\pi)^{1/2}e^{\pi\gamma/2}(kr)^{-1}$$
$$\times(-2ikr)^{-l}\Gamma(2l+1)/\Gamma(l+1-i\gamma),\quad r\to 0. \tag{3.12}$$

With the help of these expressions we obtain $C_1 = e^{2i\sigma_l}$. This completes the proof of Eq. (3.1).

The behavior of $\langle r|G_{C,l}V_C|k\uparrow\rangle_0$ at $r\to 0$, as given by Eq. (3.10) is somewhat peculiar. The function $\psi_{s,l}$ has the same

behavior, according to Eq. (3.1), that has just been proved. It may be interesting to deduce this behavior of $\psi_{s,l}$ at $r=0$ in an independent manner. We shall do this in two ways: (i) by starting from $\psi_s(\mathbf{k},\mathbf{r})$, and (ii) by using an integral representation for $\psi_{s,l}$ which we have obtained previously.[1] These considerations give at the same time a more precise expression for $\psi_{s,l}$ at $r=0$.

First we note that Eqs. (3.2) and (3.4b) may be used for our purpose, but this approach is not simple for $l>0$. So let us start with $\psi_s(\mathbf{k},\mathbf{r})$, a closed form for which has already been given in Eq. (2.4). By using

$$U(1+i\gamma,1,z)\simeq -(2C+\psi(1+i\gamma)+\ln z)/\Gamma(1+i\gamma), \quad z\to 0, \tag{3.13}$$

where C is Euler's constant and ψ the digamma function, we get

$$\psi_s(\mathbf{k},\mathbf{r})\simeq(2\pi)^{-(3/2)}\exp(ikr+\pi\gamma/2)\times\ln(kr-\mathbf{k}\cdot\mathbf{r})/\Gamma(-i\gamma), \quad kr-\mathbf{k}\cdot\mathbf{r}\to 0. \tag{3.14}$$

The p.w. projection of ψ_s is given by

$$\psi_{s,l}(r)\equiv 2\pi\int_{-1}^{1}P_l(x)\psi_s(\mathbf{k},\mathbf{r})dx,$$

with $x=\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}$. We now use the equalities

$$\int_{1}^{1}P_l(x)\ln(1-x)dx = 2\ln 2-2, \quad l=0$$

$$= -2/[l(l+1)], \quad l=1,2,3,\cdots, \tag{3.15}$$

that follow easily with the help of (e.g., Ref. 4, p. 239)

$$\sum_{n=1}^{\infty}(n^{-1}+(n+1)^{-1}P_n(x) = -1+\ln 2-\ln(1-x), \quad -1\leqslant x<1. \tag{3.16}$$

In this way we obtain, for $r\to 0$,

$$(\pi/2)^{1/2}e^{-\pi\gamma/2}\Gamma(-i\gamma)\psi_{s,l}(r)\simeq\ln r, \quad l=0$$

$$\simeq -1/[l(l+1)], \quad l=1,2,3\cdots. \tag{3.17}$$

This expression not only agrees with Eq. (3.10) but also gives more information.

Finally we will deduce the expression (3.17) from the following integral representation for $\psi_{s,l}$ [Eq. (18) of Ref. 1],

$$\psi_{s,l}(r) = -(2/\pi)^{1/2}i^{-l}e^{\pi\gamma/2}[\Gamma(-i\gamma)]^{-1}\times\int_0^{\infty}j_l(krt)e^{ikr(1+t)}t^{i\gamma}(1+t)^{-1-i\gamma}dt. \tag{3.18}$$

We use the new variable $z=krt$ and see that we have to investigate the following integral for small r,

$$I_l(r)\equiv\int_0^{\infty}j_l(z)e^{iz}(z+kr)^{-1}(1+kr/z)^{-i\gamma}dz. \tag{3.19}$$

When $l>0$ we may put $r=0$ in the integrand because of $j_l(z)=O(z^l)$, $z\to 0$. In this case we obtain

$$\lim_{r\to 0}I_l(r)=\int_0^{\infty}j_l(z)e^{iz}z^{-1}dz=i^l/[l(l+1)], \quad l=1,2,3,\cdots, \tag{3.20}$$

which follows by using formula 6.621.1 of Ref. 5. For $l=0$ we have

$$I_0(r)+\ln kr = \int_0^{\infty}\sin z e^{iz}z^{-1}(z+kr)^{-1}(1+kr/z)^{-i\gamma}dz - \int_0^{1-kr}(z+kr)^{-1}dz,$$

which clearly has a finite limit for $r\to 0$, so

$$I_0(r) = -\ln kr+O(1), \quad r\to 0. \tag{3.21}$$

By substituting the above results in Eq. (3.18) we obtain the second proof of Eq. (3.17).

4. IRREGULAR SOLUTIONS IN THE GENERAL CASE

In the preceding section we have expressed $\psi_{s,l}$ in terms of the Coulomb Green function $G_{C,l}$, see Eq. (3.1). By summing the p.w. series for both sides of this equation we obtain

$$\psi_s(\mathbf{k},\mathbf{r}) = \int_0^{\infty}\langle\mathbf{r}|G_C|\hat{\mathbf{k}}r'\rangle V_C(r')\langle r'|k\uparrow\rangle_0 r'^2dr'f^*_{C,0}/(2i). \tag{4.1}$$

In this section we shall discuss irregular solutions ψ_w for a general potential V, not necessarily spherically symmetric. To this end we define, in close analogy to Eq. (4.1),

$$\psi_w(\mathbf{k},\mathbf{r})\equiv\int_0^{\infty}\langle\mathbf{r}|G|\hat{\mathbf{k}}r'\rangle w(r')dr', \tag{4.2}$$

where $G=(k^2+\Delta-V)^{-1}$ and the function w is arbitrary to the extent that the above integral is well defined. For convenience we assume w to be continuously differentiable. By a formal application of G^{-1} it is easily seen that ψ_w satisfies

$$(k^2+\Delta-V)\psi_w(\mathbf{k},\mathbf{r})=r^{-2}w(r)\delta(\hat{\mathbf{r}},\hat{\mathbf{k}}). \tag{4.3}$$

The Dirac delta function is defined by

$$\int f(\hat{\mathbf{r}})\delta(\hat{\mathbf{r}},\hat{\mathbf{k}})d\hat{\mathbf{r}}=f(\hat{\mathbf{k}}),$$

where the domain of integration is the surface of the unit sphere.

We will show that ψ_w in general has a logarithmic singularity in the forward direction ($\hat{\mathbf{k}}=\hat{\mathbf{r}}$). By inserting $G=G_0+G_0VG$ in (4.2) one can show that this singularity in general comes from G_0. So we replace G by G_0 in Eq. (4.2) and use

$$\langle\mathbf{r}|G_0|\mathbf{r}'\rangle = -(4\pi)^{-1}|\mathbf{r}-\mathbf{r}'|^{-1}\exp(ik|\mathbf{r}-\mathbf{r}'|).$$

It follows that the singular part of ψ_w is given by

$$\psi_w = -(4\pi)^{-1}\int_0^{\infty}w(r')\exp(iky)/ydr'+O(1), \quad x\to 1,$$

with $y=(r^2+r'^2-2rr'x)^{1/2}$ and $x=\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}$ as before. The singularity comes from the integrand at the point $r'=r$. In or-

der to investigate its behavior in this region we introduce the new variable $z = r'/r$. Then one can show that for any positive a,

$$\int_a^1 (1 - 2xz + z^2)^{-1/2} f(z)dz = \mp \tfrac{1}{2} f(1)\ln(1 - x) + O(1), \quad x \uparrow 1, \quad a \lesssim 1, \tag{4.4}$$

for a continuously differentiable function $f(z)$. With the help of Eq. (4.4) we obtain

$$\psi_w = (4\pi)^{-1} w(r)\ln(1 - x) + O(1), \quad x \uparrow 1. \tag{4.5}$$

This expression gives the logarithmic singularity of the irregular solution ψ_w for a general local potential V.

Now we will briefly discuss the singular behavior of ψ_w at $r = 0$. In this case we assume $x \neq 1$. Since $(z^2 - 2xz + 1)^{1/2} \sim z$ for $z \to \infty$ we have

$$\psi_w = -(4\pi)^{-1} \int_1^\infty e^{ikrz} w(rz) \frac{dz}{z} + O(1), \quad r \to 0. \tag{4.6}$$

When w is constant we use

$$\Gamma(0, -ikr) = \int_1^\infty e^{ikrz} \frac{dz}{z} = -\ln kr + O(1), \quad r \to 0,$$

where Γ is the incomplete gamma function, and obtain from Eq. (4.6),

$$\psi_w = (4\pi)^{-1} w(0)\ln kr + O(1), \quad r \to 0. \tag{4.7}$$

When w is proportional to e^{ikr} [cf. Eq. (4.13)] we get exactly the same expression, (4.7).

We note that Eqs. (4.5) and (4.7) can be combined,

$$\psi_w(\mathbf{k},\mathbf{r}) = (4\pi)^{-1} w(r)\ln(kr - \mathbf{k}\cdot\mathbf{r}) + O(1), \tag{4.8}$$

for $\hat{\mathbf{k}}\cdot\hat{\mathbf{r}} \to 1$ as well as for $r \to 0$. This expression may be compared with Eq. (3.14).

If we now restrict ourselves to spherically symmetric potentials, $\psi_w(\mathbf{k},\mathbf{r})$ is a function of k, r and $\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}$ only. In this case it is possible to consider the p.w. projection of (4.2),

$$\psi_{w,l}(k,r) = \int_0^\infty \langle r|G_l|r'\rangle w(r')dr'. \tag{4.9}$$

In order to deduce the behavior of $\psi_{w,l}$ at $r = 0$, we use Eq. (3.3) which is valid for any local central potential. Then Eq. (4.9) may be rewritten as

$$\psi_{w,l}(k,r) = (-)^{l+1} \tfrac{1}{2}\pi k \left[\langle r|kl\uparrow\rangle \int_0^r \langle r'|kl+\rangle w(r')dr' + \langle r|kl+\rangle \int_r^\infty \langle r'|kl\uparrow\rangle w(r')dr' \right]. \tag{4.10}$$

By using Eqs. (1.7) and (1.8) we obtain

$$\psi_{w,l}(k,r) \simeq -(2l+1)^{-1} w(0) \left[r^{-l-1} \int_0^r r'^l dr' + r^l \int_r^{r_0} r'^{-l-1} dr' \right], \quad r \to 0, \tag{4.11}$$

where r_0 is an unimportant constant. Therefore,

$$\psi_{w,l}(k,r) = w(0)\ln r + O(1), \quad r \to 0, \quad l = 0,$$
$$= -w(0)/[l(l+1)] + o(1), \quad r \to 0, \quad l = 1,2,3,\cdots. \tag{4.12}$$

One easily verifies that the p.w. projection of both sides of Eq. (4.8) yields expressions for $\psi_{w,l}$ that are in agreement with Eq. (4.12).

We note that for the Coulomb case, $G = G_C$, ψ_w is just equal to the irregular Coulomb wave ψ_s given by Eq. (2.4) if we choose the function w as

$$w(r) = -i\gamma(2/\pi)^{1/2} f^*_{C,0} e^{ikr}. \tag{4.13}$$

We conclude this section with a remark on the logarithmic singularity of ψ_w, given by Eq. (4.8). We see from Eq. (4.3) that the delta function singularity must be generated by the Laplace operator acting on $\ln(kr - \mathbf{k}\cdot\mathbf{r})$, so

$$\Delta \ln(kr - \mathbf{k}\cdot\mathbf{r}) \simeq 4\pi r^{-2}\delta(\hat{\mathbf{r}},\hat{\mathbf{k}}). \tag{4.14}$$

It is interesting to note that one can verify that Eq. (4.14) holds with an equality sign.

In order to show this, let us take $\hat{\mathbf{k}}$ along the positive z-axis as before. Then the right-hand side of (4.14) describes a uniform line charge density along the positive z-axis. In view of the symmetry in the problem it is natural to use cylindrical coordinates R, z, φ, where $R^2 = r^2 - z^2$. Then we have

$$2\pi r^{-2}\delta(\hat{\mathbf{r}},\hat{\mathbf{z}}) = r^{-2}\delta(1 - \cos\zeta) = R^{-1}\delta(R)\theta(z),$$

where θ is the unit step function. Further,

$$kr - \mathbf{k}\cdot\mathbf{r} = k(r - z) = k((R^2 + z^2)^{1/2} - z).$$

The electrostatic potential for a uniform charge distribution on the positive z axis is just proportional to the logarithmic term discussed above. Poisson's equation reads in this case

$$\Delta \ln((R^2 + z^2)^{1/2} - z) = 2R^{-1}\delta(R)\theta(z). \tag{4.15}$$

This equation shows that (4.14) holds with an equality sign. So we see that the inhomogeneous term in Eq. (4.3) may be compared with a line charge distribution along the positive z axis with density $w(r)$ or $w(z)$.

5. ON THE CONNECTION WITH THE SCATTERING AMPLITUDE

The function $\psi_s(\mathbf{k},\mathbf{r})$ [see Eq. (2.4)] is called the scattered part of the complete physical scattering wavefunction $\psi^{(+)}(\mathbf{k},\mathbf{r})$ for the Coulomb potential because of its asymptotic behavior, which is given by [cf. Eq. (40) of Ref. 1]

$$\psi_s(\mathbf{k},\mathbf{r}) \sim f^C(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}})(2\pi)^{-(3/2)} r^{-1}\exp(ikr - i\gamma\ln 2kr), \quad r \to \infty. \tag{5.1}$$

Here f^C is the Coulomb scattering amplitude,

$$f^C(x) = -\frac{\gamma}{2k} e^{2i\sigma_0} (\tfrac{1}{2} - \tfrac{1}{2}x)^{-1 - i\gamma}.$$

One may compare (5.1) with the well-known short-range potential formula,

$$\psi^{(+)}(\mathbf{k},\mathbf{r}) \sim (2\pi)^{-3/2}(e^{i\mathbf{k}\cdot\mathbf{r}} + f(\theta)e^{ikr}/r), \quad r \to \infty.$$

From Eq. (4.1).

$$\psi_s(\mathbf{k},\mathbf{r}) = \int_0^\infty \langle \mathbf{r}|G_C|\hat{\mathbf{k}}r'\rangle V_C(r')\langle r'|k\uparrow\rangle_0 r'^2 dr' f^*_{C,0}/(2i),$$

we see that this "scattered part" equals an integral involving the Green operator G_C.

It is interesting to investigate whether Eq. (4.1) can be generalized to other potentials. The problem is, how to find an irregular solution, such that its asymptotic behavior equals the scattering amplitude times a (possibly modified) outgoing spherical wave, just as in Eq. (5.1).

Let us first consider again the p.w. function $\psi_{s,l}$ for the pure Coulomb potential. Below we shall consider the generalization to Coulomb-like and other potentials. From Eqs. (3.1) and (3.8) we have, for $r\to\infty$,

$$\psi_{s,l}(k,r)\sim\tfrac{1}{4}i\pi k f^*_{C,0}\langle r|kl\uparrow\rangle_C\int_0^r {}_C\langle kl-|r'\rangle \times V_C(r')\langle r'|k\uparrow\rangle_0 r'^2dr'. \tag{5.2}$$

We split the integral in two parts, $\int_0^R+\int_R^r$, where R is so large that the asymptotic behavior of ${}_C\langle kl-|r'\rangle$ can be used. With the help of Eqs. (1.4) and (3.9) we obtain a term with the asymptotic behavior $-(2/\pi)^{1/2}f^*_{C,0}e^{ikr}/(2ikr)$. According to Eq. (3.2),

$$\psi_{s,l}(k,r)=(-\chi_l+e^{2i\sigma_l}\langle r|kl\uparrow\rangle_C)/(2i),$$

this term is $-\chi_l/(2i)$. The rest of $\psi_{s,l}$ is proportional to $\langle r|kl\uparrow\rangle_C$. By using Eq. (3.1) we deduce

$$e^{2i\sigma_l}=f^*_{C,0}\lim_{R\to\infty}\left[(2kR)^{i\gamma}-\tfrac{1}{2}\pi k\int_0^R {}_C\langle kl-|r\rangle V_C(r)\times\langle r|k\uparrow\rangle_0 r^2dr\right]. \tag{5.3}$$

It is interesting to replace $\langle r|k\uparrow\rangle_0$ by $\langle r|q\uparrow\rangle_0$ here, where as before

$$\langle r|q\uparrow\rangle_0=(2/\pi)^{1/2}e^{iqr}/(qr),$$

with $\mathrm{Im}q>0$ and consider the limit for $q\to k$. When $q\neq k$ the integral $\int_0^R\cdots$ is convergent for $R\to\infty$ and may be denoted in this case by ${}_C\langle kl-|V_{C,l}|q\uparrow\rangle_0$. We have been able to obtain the following closed expressions,

$$\begin{aligned}{}_C\langle kl-|V_{C,l}|q\uparrow\rangle_0&=\frac{4i\gamma}{\pi q}e^{\pi\gamma/2}\left(\frac{q+k}{q-k}\right)^{i\gamma/2}Q_l^{i\gamma}(q/k)\\&=-\frac{2}{\pi q}e^{-\pi\gamma/2}\frac{\Gamma(1+i\gamma)\Gamma(1-i\gamma)\Gamma(l+1)}{\Gamma(l+1-i\gamma)}\\&\quad\times\left[P_l^{(i\gamma,-i\gamma)}(q/k)-\left(\frac{q+k}{q-k}\right)^{i\gamma}P_l^{(-i\gamma,i\gamma)}(q/k)\right].\end{aligned} \tag{5.4}$$

Here $Q_l^{i\gamma}$ is Legendre's function of the second kind, and $P_l^{(\cdots)}$ is Jacobi's polynomial. In the particular case $l=0$ this expression agrees with Eq. (7) of Ref. 6 that we used for the derivation of the Coulomb off-shell Jost function in closed form.

When $q\sim k$, Eq. (5.4) can be simplified. By inserting

$$P_l^{(i\gamma,-i\gamma)}(1)=\Gamma(l+1+i\gamma)/[\Gamma(1+i\gamma)\Gamma(l+1)], \tag{5.5}$$

we obtain

$${}_C\langle kl-|V_{C,l}|q\uparrow\rangle_0\underset{q\to k}{\sim}-\frac{2}{\pi k}e^{-\pi\gamma/2}\left[e^{2i\sigma_l}\Gamma(1-i\gamma)-\Gamma(1+i\gamma)\left(\frac{q+k}{q-k}\right)^{i\gamma}\right]. \tag{5.6}$$

The second term on the right-hand side contains the factor $(q-k)^{-i\gamma}$ and is therefore singular for $q\to k$. It may be compared with the "correction factor" ω of Ref. 6, Eq. (2). Note also the similarity with the so-called Coulombian asymptotic state of Ref. 7, Eq. (16), where the typical factor $f^*_{C,0}(p+k)^{i\gamma}(p-k)^{-i\gamma}$ occurs.

This singular term corresponds to that part of the integral on the right-hand side of Eq. (5.3) which contains the (for $R\to\infty$) divergent factor $(2kR)^{i\gamma}$. The other term is continuous for $q\to k$ and this one corresponds just to the "convergent part" of the integral in (5.3).

A natural generalization of the expression $\langle r|G_{C,l}V_{C,l}|k\uparrow\rangle_0$ to other central potentials is

$$\psi_l(k,r)\equiv\langle r|G_lV_l|k\uparrow\rangle_0, \tag{5.7}$$

where G_l is the Green function for V_l. So ψ_l corresponds to the Coulomb function $\psi_{s,l}$ of Eq. (3.1) [we have omitted the constant factor $f^*_{C,0}/(2i)$ which is irrelevant here]. We first assume that V_l is a Coulomb plus short-range potential, $V_{C,l}+V_{s,l}$. In order to investigate the asymptotic behavior of ψ_l, we use the expression [cf. Eq. (3.3)]

$$\langle r|G_l|r'\rangle=-\tfrac{1}{2}\pi k\langle kl-|r_<\rangle\langle r_>|kl\uparrow\rangle.$$

It may be noted that $\langle r|kl\uparrow\rangle$ has exactly the same asymptotic behavior as $\langle r|kl\uparrow\rangle_C$, which is given by Eq. (3.9). Furthermore we have [cf. Eq. (1.4)]

$$2i\langle kl-|=\exp[2i(\sigma_l+\delta_l^C)]\langle kl\downarrow|-\langle kl\uparrow|, \tag{5.8}$$

where δ_l^C is the Coulomb-modified phase shift. We proceed in the same way as in the pure Coulomb case, and find that ψ_l can again be split up in two parts, $\psi_l=\psi_l^{(1)}+\psi_l^{(2)}$, which have different asymptotic behavior. For the first term we get

$$\psi_l^{(1)}(r)\sim-(2/\pi)^{1/2}e^{ikr}/(kr),\quad r\to\infty,\quad l=0,1,2,.... \tag{5.9}$$

Obviously this is the analog of the function χ_l. Since the right-hand side of (5.9) is independent of l, it follows that the sum of the p.w. series,

$$\sum_{l=0}^{\infty}(4\pi)^{-1}(2l+1)P_l(x)\psi_l^{(1)}(r),$$

is proportional to $\delta(1-x)$ for $r\to\infty$.

For the second term we obtain

$$\psi_l^{(2)}(r)\sim\langle r|kl\uparrow\rangle\lim_{R\to\infty}\left[(2kR)^{i\gamma}-\tfrac{1}{2}\pi k\int_0^R\langle kl-|r'\rangle V(r')\langle r'|k\uparrow\rangle_0 r'^2dr'\right],\quad r\to\infty. \tag{5.10}$$

The integral $\int_0^R$ is divergent for $R\to\infty$. In this limit it has exactly the same singular behavior as for the pure Coulomb case, which can be verified with the help of Eq. (5.8). It is therefore natural to split off the pure Coulomb part. We do this by using the two-potential formalism; in the notation of Ref. 8 we have

$$V_l|kl+\rangle = V_{C,l}|kl+\rangle_C + (1+T_{C,l}G_{0,l})t_{Cs,l}|kl+\rangle_C, \tag{5.11a}$$

or

$$\langle kl-|V_l = {}_C\langle kl-|V_{C,l} + {}_C\langle kl-|t_{Cs,l}(1+G_{0,l}T_{C,l}). \tag{5.11b}$$

Here $t_{Cs,l}$ satisfies the equation

$$t_{Cs,l} = V_{s,l} + V_{s,l}G_{C,l}t_{Cs,l},$$

so it is a "short-range operator." Substitution of (5.11) in (5.10) yields

$$\psi_l^{(2)}(r)\sim\langle r|kl\uparrow\rangle[f_{C,0}^{*-1}e^{2i\sigma_l} - \tfrac{1}{2}\pi k_C\langle kl-| \times t_{Cs,l}(1+G_{0,l}T_{C,l})|k\uparrow\rangle_0], \tag{5.12}$$

where we have used Eq. (5.3). The phase shift for $V_{C,l}+V_{s,l}$ is related to $t_{Cs,l}$ in the following well-known way,

$${}_C\langle kl-|t_{Cs,l}|kl+\rangle_C = i(\pi k)^{-1}e^{2i\sigma_l}[\exp(2i\delta_l^C)-1]. \tag{5.13}$$

Comparison with Eq. (5.12) shows that the p.w. series $\Sigma_l(4\pi)^{-1}(2l+1)P_l(x)\psi_l^{(2)}(r)$ is not proportional to the scattering amplitude in general. Therefore, also

$$\psi(\mathbf{k},\mathbf{r}) = \sum_{l=0}^{\infty}(4\pi)^{-1}(2l+1)P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}})\psi_l(r)$$

does not in general have the desired asymptotic behavior [recall that the p.w. series with $\psi_l^{(1)}$ is proportional to $\delta(1-x)=\delta(1-\hat{\mathbf{k}}\cdot\hat{\mathbf{r}})$, for $r\to\infty$].

For a short-range potential we obtain for $r\to\infty$, following the same procedure,

$$\psi_l(r)\sim -\tfrac{1}{2}\pi k\langle r|kl\uparrow\rangle\langle kl-|V_l|k\uparrow\rangle_0 = -\tfrac{1}{2}\pi k\langle r|kl\uparrow\rangle\langle kl|T_l|k\uparrow\rangle_0. \tag{5.14}$$

In this case the phase shift δ_l is given by

$$\langle kl|T_l|kl\rangle = i(\pi k)^{-1}(e^{2i\delta_l}-1).$$

Apparently the p.w. series with the ψ_l of (5.14) will in general not be proportional to the scattering amplitude, for $r\to\infty$.

The procedure described above can be repeated for the function $\psi_{w,l}$ of Sec. 4. That is, we replace $\langle r|k\uparrow\rangle_0$ by a rather arbitrary function $w(r)$ and consider the asymptotic behavior of $\psi_{w,l}(r)$, see Eq. (4.9). Again we are not able to find a function w for any potential (except for V_C), such that $\psi_w(\mathbf{k},\mathbf{r})$ for $r\to\infty$ approaches the scattering amplitude times an outgoing spherical wave.

So it seems that the pure Coulomb function ψ_s in unique in having the property (5.1). This would mean that the useful property (5.1) of the irregular solution ψ_s is merely a coincidence. Therefore, although the regular physical wavefuntion $\psi^{(+)}(\mathbf{k},\mathbf{r})$ for any potential can be expressed as the sum of two irregular solutions $\psi^{(+)}=\psi_i+\psi_s$, this splitting seems to be useful only in the pure Coulomb case.

ACKNOWLEDGMENT

This investigation forms a part of the research program of the Foundation for Fundamental Research of Matter (FOM), which is financially supported by the Netherlands Organization for Pure Scientific Research (ZWO).

[1] H. van Haeringen, Nuovo Cimento B **34**, 53 (1976).
[2] J.R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[3] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[4] W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966).
[5] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[6] H. van Haeringen, J. Math. Phys. **19**, 1379 (1978).
[7] H. van Haeringen, J. Math. Phys. **17**, 995 (1976).
[8] H. van Haeringen and R. van Wageningen, J. Math. Phys. **16**, 1441 (1975).

Electromagnetic quadripotential for the pure-radiation field generated by a classical charged point-particle

Francesco Miglietta

Istituto di Fisica dell'Universitá, Pavia, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, Italy
(Received 22 May 1978)

We obtain a simple expression for the electromagnetic quadripotential corresponding to the pure-radiation field generated by a classical charged point-particle. The solution does not satisfy the Lorentz condition, and has interesting properties. It propagates inside the light cone of the particle and has a discontinuity across the sheet of the light cone itself. This discontinuity is responsible for the correct propagation of the electromagnetic effects with the velocity of light. A similar result is also obtained for the velocity-dependent field.

I. INTRODUCTION

The electromagnetic quadripotential generated by a classical charged point particle is given, up to guage transformations, by the well known Liénard–Wiechert solution, which in covariant form reads[1,2]

$$A^{i}_{\binom{R}{A}}(x) = 2e\int_{-\infty}^{+\infty} ds'\,\vartheta(\pm R^{0})\delta(R_m R^{m})u_i(s')$$

$$= \pm\left|\frac{eu_i}{R}\right|_{s'=s_{\binom{R}{A}}}, \tag{1.1}$$

where the subscripts R,A refer to retarded, advanced solutions. In Eq. (1.1) we have indicated with e the charge, with u^i the quadrivelocity, and with R^i the quadrivector,

$$R^{i} = x^{i}\text{-}z^{i}(s'), \tag{1.2}$$

$x^i = z^i(s')$ being the equation of the world line of the particle. We have introduced also the scalar R given by

$$R = R_i u^{i} \tag{1.3}$$

and we have indicated with $s_{\binom{R}{A}}$ the value of the proper time s' which is solution to the following set of equations:

$$R_i R^{i} = 0, \quad R^{0} \lesseqgtr 0. \tag{1.4}$$

The corresponding electromagnetic field tensor F_{ik}, can be split, in a covariant way, into a velocity-dependent field

$$F^{(\mathrm{vel})}_{\binom{R}{A}^{ik}} = \pm\frac{e}{R^{3}}(u_k R_i - u_i R_k)\,|_{s'=s_{\binom{R}{A}}} \tag{1.5}$$

and a pure-radiation field

$$F^{(\mathrm{rad})}_{\binom{R}{A}^{ik}} = \pm\frac{e}{R^{2}}\Big[R_i(a_k - a_R u_k)$$

$$- R_k(a_i - a_R u_i)\Big]\,|_{s'=s_{\binom{R}{A}}}. \tag{1.6}$$

In Eq. (1.6) a_R is given by

$$a_R = (1/R)a^{i}R_i, \tag{1.7}$$

a^i being the quadriacceleration of the particle.

It is commonly believed that it is impossible to split the quadripotential in a similar way. In this paper we obtain, on the contrary, separate expressions for the quadripotential $A^{(\mathrm{rad})}$, which corresponds to the pure-radiation field of Eq. (1.6), and for $A^{(\mathrm{vel})}$, which corresponds to the velocity-dependent field of Eq. (1.5).

II. QUADRIPOTENTIAL FOR THE PURE-RADIATION FIELD

The starting point is the following identity:

$$\frac{1}{R}(a_i - a_R u_i) = \frac{\partial}{\partial R^{i}}a_R = \frac{\partial}{\partial x^{i}}a_R, \tag{2.1}$$

which allows us to write Eq. (1.6) in the following way:

$$F^{(\mathrm{rad})}_{\binom{R}{A}^{ik}}(x) = 2e\int_{-\infty}^{+\infty} ds'\,\vartheta(\pm R^{0})\delta(R_m R^{m})$$

$$\times\left(R_i\frac{\partial}{\partial R^{k}} - R_k\frac{\partial}{\partial R^{i}}\right)a_R. \tag{2.2}$$

Equation (2.2) suggests immediately for $A^{(\mathrm{rad})}$ the following expression:

$$A^{(\mathrm{rad})}_{\binom{R}{A}^{ik}}(x) = e\int_{-\infty}^{+\infty} ds'\,\vartheta(\pm R^{0})\vartheta(R_m R^{m})\frac{\partial}{\partial R^{i}}a_R. \tag{2.3}$$

By a simple calculation one can verify that the field tensor F_{ik}, generated by the quadripotential (2.3), is indeed the pure-radiation field given in Eq. (2.2). One can verify that $A^{(\mathrm{rad})}$, given in Eq. (2.3), does not satisfy the Lorentz condition.

The mathematical and physical properties of the solution given in Eq. (2.3) are quite interesting and peculiar. For simplicity we refer to the retarded solution only. From the mathematical point of view we observe that the extremely singular δ function, which enters the Liénard–Wiechert quadripotential of Eq. (1.1), is absent in Eq. (2.3). From the physical point of view we observe that the quadripotential, given in Eq. (2.3), propagates inside the future light cone of the particle. The discontinuous behavior of the ϑ function across the sheet of the future light cone is responsible for the

 0022-2488/79/050868-03$01.00

physically correct propagation of the electromagnetic effects with the velocity of light.

III. QUADRIPOTENTIAL FOR THE VELOCITY-DEPENDENT FIELD

The velocity-dependent field can be treated in a similar way. Instead of Eq. (2.1) we have

$$-\frac{u_i}{R^2}=\frac{\partial}{\partial R^i}\frac{1}{R}=\frac{\partial}{\partial x^i}\frac{1}{R}. \tag{3.1}$$

In perfect analogy with Eq. (2.3) we obtain, in this case,

$$A^{(\mathrm{vel})}_{\binom{R}{A}}(x)=e\int ds'\,\vartheta(\pm R^0)\vartheta(R_m R^m)\frac{\partial}{\partial R^i}\frac{1}{R}. \tag{3.2}$$

The physical and mathematical properties of the solution given in Eq. (3.2) are similar to those discussed in the previous section.

IV. TWO EXAMPLES

As a first example we calculate the retarded potential of Eq. (3.2) for a particle at rest. We assume

$$z^0=s',\quad z^\alpha=0\quad(\alpha=1,...,3), \tag{4.1}$$

which implicates $u^i=(1,0,0,0)$. Equation (3.2) gives, for the retarded solution,

$$\begin{aligned}A^{(\mathrm{vel})}_{(R)0}&=-e\int_{-\infty}^{+\infty}dz^0\,\vartheta(x^0-z^0-r)(x^0-z^0)^{-2}\\&=-e\int_{-\infty}^{+\infty}dz^0\,\vartheta(R^0-r)(R^0)^{-2}\\&=-e\int_{-r}^{+\infty}dR^0(R^0)^{-2}=-e/r,\end{aligned} \tag{4.2}$$

$$A^{(\mathrm{vel})}_{(R)1}=A^{(\mathrm{vel})}_{(R)2}=A^{(\mathrm{vel})}_{(R)3}=0, \tag{4.3}$$

which is just the Coulomb potential.

As a second example we calculate, using Eq. (2.3), the retarded potential for the pure-radiation field generated by a harmonically oscillating particle. We assume

$$z^0=-s',\quad z^1=\sin\omega z^0,\quad z^2=z^3=0,$$

and, for simplicity, we assume also

$$|b\omega|\ll 1 \tag{4.4}$$

(nonrelativistic velocity). Equation (4.3) together with the condition (4.4) gives

$$u^0\cong 1,\quad u^1\cong b\omega\cos\omega z^0\quad u^2=u^3=0, \tag{4.5}$$

and

$$a^0\cong 0,\quad a^1\cong -b\omega\sin\omega z^0,\quad a^2=a^3=0. \tag{4.6}$$

Assuming also

$$r\gg b \tag{4.7}$$

we obtain

$$R^0=x^0-z^0,\quad R^\alpha\cong x^\alpha(\alpha=1,...,3). \tag{4.8}$$

A simple calculation gives

$$R\cong x^0-z^0 \tag{4.9}$$

and

$$a_R=b\omega^2[x/(x^0-z^0)]\sin\omega z^0. \tag{4.10}$$

With these assumptions, Eq. (2.3) gives

$$A^{(\mathrm{rad})}_{(R)2}=A^{(\mathrm{rad})}_{(R)3}=0, \tag{4.11}$$

$$\begin{aligned}A^{(\mathrm{rad})}_{(R)0}&=-eb\omega^2x\int_{-\infty}^{+\infty}dx^0\,\vartheta(x^0-z^0-r)(x^0-z^0)^{-2}\sin\omega z^0\\&=-eb\omega^2(x/r)\sin\omega(x^0-r)\\&\quad-eb\omega^3x\int_{+\infty}^{r}dR^0\frac{1}{R^0}\cos\omega(x^0-R^0),\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}A^{(\mathrm{rad})}_{(R)1}&=eb\omega^2\int_{-\infty}^{+\infty}dz^0\,\vartheta(x^0-z^0-r)(x^0-z^0)^{-1}\sin\omega z^0\\&=-eb\omega^2\int_{+\infty}^{r}dR^0\frac{1}{R^0}\sin\omega(x^0-R^0).\end{aligned} \tag{4.13}$$

Explicit calculations for A_0 and A_1 give[3]

$$\begin{aligned}A^{(\mathrm{rad})}_{(R)0}&=-eb\omega^2\frac{x}{r}\sin\omega(x^0-r)\\&\quad-eb\omega^3x\Big\{\cos(\omega x^0)\mathrm{Ci}(\omega r)\\&\quad+\sin(\omega x^0)\Big[\mathrm{Si}(\omega r)-\frac{\pi}{2}\Big]\Big\},\end{aligned} \tag{4.14}$$

$$\begin{aligned}A^{(\mathrm{rad})}_{(R)1}&=-eb\omega^2\Big\{\sin(\omega x^0)\mathrm{Ci}(\omega r)\\&\quad-\cos(\omega x^0)\Big[\mathrm{Si}(\omega r)-\frac{\pi}{2}\Big]\Big\}.\end{aligned} \tag{4.15}$$

The electromagnetic field tensor corresponding to this solution, can be calculated easily from Eqs. (4.11) to (4.13). One obtains

$$\begin{aligned}F_{01}&=E_x=eb\omega^2r^{-1}(1-x^2/r^2)\sin\omega(x^0-r),\\F_{02}&=E_y=-eb\omega^2(xy/r^3)\sin\omega(x^0-r),\\F_{03}&=E_z=-eb\omega^2(xz/r^3)\sin\omega(x^0-r),\\F_{12}&=-B_z=eb\omega^2(y/r^2)\sin\omega(x^0-r),\\F_{13}&=B_y=eb\omega^2(z/r^2)\sin\omega(x^0-r),\\F_{23}&=-B_x=0,\end{aligned} \tag{4.16}$$

which is the correct answer, as one can verify directly using Eq. (1.6).

V. CONCLUSION

We have obtained very simple expressions for the quadripotentials corresponding both to the pure-radiation field and to the velocity-dependent field generated by a classical point particle. Both of the potentials vanish outside the light cone of the particle, but curiously they do not vanish inside it.

Collecting Eq. (3.2) and Eq. (2.3) together, we obtain the following expression for the complete quadripotential generated by a classical point particle:

$$A_{\binom{R}{A}}(x) = e\int_{-\infty}^{+\infty} ds'\,\vartheta(\pm R^0) \times \vartheta(R_m R^m)\frac{\partial}{\partial R^i}\left(\frac{1}{R} + a_R\right). \tag{5.1}$$

Equation (5.1) gives the same field strengths as the Liénard–Wiechert potential of Eq. (1.1).[4]

A solution which is physically equivalent to Eq. (5.1), and then to Eq. (1.1), is given by the following curious expression:

$$\bar{A}_{\binom{R}{A}}(x) = -e\int_{-\infty}^{+\infty} ds'[1 - \vartheta(\pm R^0)\vartheta(R_m R^m)] \times \frac{\partial}{\partial R^i}\left(\frac{1}{R} + a_R\right). \tag{5.2}$$

which apparently violates causality, since it gives propagation, for the quadripotential, in spacelike directions and backwards in time also. It can be immediately verified that A and $\bar{A}$ are connected by the following gauge transformation:

$$A_{\binom{R}{A}}(x) = \bar{A}_{\binom{R}{A}}(x) + \frac{\partial}{\partial x^i} e\int_{-\infty}^{+\infty} ds'\left(\frac{1}{R} + a_R\right). \tag{5.3}$$

ACKNOWLEDGMENT

I thank Bruno Bertotti for his valued comments on the manuscript.

[1] F. Rohrilich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965).

[2] A. OBarut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964).

[3] For the definitions of sine- and cosine-integral see, e.g., M. Abramowitz and I. A. Stegun; *Handbook of Mathematical Functions* (Applied Mathematics Series 55, National Bureau of Standards, Washington, D.C., 1970).

[4] We recall that the quadripotential has no direct physical meaning in classical electrodynamics. Only the field strengths are physically measurable quantities.

On the complete integrability of the stationary, axially symmetric Einstein equations

Dieter Maison

Max-Planck Institut für Physik und Astrophysik, Munich, Fed. Rep. Germany
(Received 25 July 1978)

A linear eigenvalue problem in the spirit of Lax is constructed for the nonlinear differential equations describing stationary, axially symmetric Einstein spaces. In suitable variables these equations yield a generalization of the well-known sine-Gordon equation. The similarity of the system to the nonlinear σ

Most of the striking results in black hole physics[1] are based on our knowledge of certain exact solutions of Einstein's equations. Therefore, it is desirable to enlarge the known families of exact solutions. On the other hand, a systematic, nonperturbative solution method for the highly nonlinear Einstein equations is badly missing. That even stays true for certain restricted classes of solutions, like the stationary, axially symmetric ones. Recent investigations of the latter class by Geroch[2] and Kinnersley[3] revealed an amazing algebraic structure, which raised the hope that a more complete treatment of this important family of solutions might be possible. These authors find an infinite-dimensional Lie algebra of infinitesimal symmetry transformations, the origin of which remains, however, rather mysterious. It is my intention to "explain" their results in relating them to similar properties of so-called "completely integrable Hamiltonian systems."[4]

In fact I conjecture that the Einstein spaces admitting a two-parameter Abelian group of isometries constitute themselves such a completely integrable system. This conjecture is based on the Lax type linear system (cf. Ref. 4) constructed in Sec. 4 of this paper. This also raises the hope that linear methods like the inverse scattering method[4] may be employed to solve the nonlinear equations.

Although from a physical point of view the stationary, axially symmetric Einstein spaces are the more interesting ones, I shall keep the discussion general and treat arbitrary two-parameter Abelian isometry groups. The case of two spacelike Killing vector fields is mathematically interesting, as it leads—in a special case—to the equations of the nonlinear σ-model, known from elementary particle physics. Expressed in suitable variables, the latter is equivalent to the well-known sine–Gordon theory.[5]

The aim of the present paper is to construct a system of linear equations in the spirit of Lax (cf. Ref. 4). This is achieved in several steps. For the convenience of the reader I repeat in Sec. 1 the standard reduction of the problem to a two-dimensional one.[2]

In Sec. 2 the infinite set of conservation laws found by Geroch[2] and developed in a more systematic way by Kinnersley[3] is used to derive an equation similar to one obtained by Lüscher and Pohlmeyer[6] for the nonlinear σ-model. In Sec. 3 a one-parameter family of solutions is generated from any given one, to be used in Sec. 4 for the construction of the linear eigenvalue problem along the ideas of Lund.[7]

1. REDUCTION TO A TWO-DIMENSIONAL PROBLEM

Geroch[2] has shown how an Einstein space E admitting a two-parameter Abelian group G_2 of isometries may be described covariantly on a two-dimensional manifold S. The space S is obtained from E as the space of orbits of the group G_2 in E. At least locally the projection Π from E onto S induces the structure of a smooth manifold on S. Let $^*\Pi$ denote the corresponding linear projection of the tangent spaces. Introducing two independent Killing vector fields ξ_i^μ $(i = 1,2)$ that generate the group G_2 the linear map $^*\Pi$ can be represented by the matrix

$$^*\Pi^\mu{}_\nu = \delta^\mu{}_\nu - \lambda^{ik}\xi_i^\mu\xi_{k\nu}, \tag{1.1}$$

where λ^{ik} is the inverse of the 2×2 matrix

$$\lambda_{ik} \equiv \xi_i^\mu \xi_k^\nu g_{\mu\nu}. \tag{1.2}$$

The projection of the metric $g_{\mu\nu}$ of E defines a metric tensor h_{ab} $(a,b = 1,2)$ on S. There are two cases to be distinguished:

(A) Both Killing vector fields ξ_i^μ are spacelike; hence $\operatorname{sgn}(\lambda_{ik}) = (+,+)$ and $\operatorname{sgn}(h_{ab}) = (-,+)$:

(B) one of the Killing vector fileds is timelike (stationarity) and hence $\operatorname{sgn}(\lambda_{ik}) = (-,+)$ and $\operatorname{sgn}(h_{ab}) = (+,+)$.

Whenever necessary I shall discuss the two cases separately referring to them as (A) and (B).

As a consequence of $R_{\mu\nu} = 0$ the vector fields

$$\Omega_{ik,\mu} \equiv \epsilon_{\mu\nu\kappa\lambda}\xi_i^\nu \nabla^\kappa \xi_k^\lambda \quad (i,k = 1,2) \tag{1.3}$$

are curl-free,

$$\Omega_{ik\,[\mu,\nu]} = 0. \tag{1.4}$$

Therefore, they can be derived from potentials Ω_{ik}:

$$\Omega_{ik\mu} = \partial_\mu \Omega_{ik}. \tag{1.5}$$

In order to be well defined on S the Ω_{ik} have to be constant along the orbits of G_2, i.e.,

$$\mathcal{L}_{\xi_j}\Omega_{ik} = \xi_j^\mu \Omega_{ik\mu} = 0, \tag{1.6}$$

which will be assumed from now on.

As discussed in Ref. 2, the equations $R_{\mu\nu} = 0$ on E are equivalent to the following equations on S:

$$R^{(h)}_{ab} = \tfrac{1}{2}\operatorname{Tr}(\lambda^{-1}D_aD_b\lambda) - \tfrac{1}{4}\operatorname{Tr}(\lambda^{-1}D_a\lambda\lambda^{-1}D_b\lambda), \tag{1.7a}$$

 0022-2488/79/050871-07$01.00

$$D^a(\tau\lambda^{-1}D_a\lambda) = 0, \tag{1.7b}$$

where $R^{(h)}_{ab}$, resp. D_a, are the Ricci tensor, resp. covariant derivative, corresponding to the metric h_{ab}. λ is the matrix (λ_{ik}) of Eq. (1.2) and

$$\tau^2 \equiv |\det\lambda|. \tag{1.8}$$

In two-dimensional spaces it is always (at least locally) possible to choose coordinates in which $h_{ab} = h\eta_{ab}$ with

$$(\eta_{ab}) = \begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} \quad \text{case (A)},$$
$$(\eta_{ab}) = \begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix} \quad \text{case (B)}. \tag{1.9}$$

In these special coordinates Eq. (1.7b) decouples from Eq. (1.7a). Therefore, one can first solve Eq. (1.7b) and then use the solution for λ in Eq. (1.7a), which can easily be integrated.[8] From now on the choice Eq. (1.9) for h_{ab} will be made, and I will restrict myself to the study of Eq. (1.7b).

The relation

$$-(\det\lambda)\lambda^{-1} = \epsilon\lambda\epsilon, \tag{1.10}$$

where ϵ is the matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, allows us to write Eq. (1.7b) in the form ($\mu \equiv \epsilon\lambda$):

$$\partial^a(\tau^{-1}\mu\partial_a\mu) = 0; \tag{1.11}$$

μ has the properties

$$\mathrm{Tr}\mu = 0 \quad \text{and} \quad \mu^2 = \begin{cases} -\tau^2\mathbb{1} & \text{case(A)}, \\ \tau^2\mathbb{1} & \text{case(B)}. \end{cases} \tag{1.12}$$

Taking the trace of Eq. (1.11) yields

$$\partial^a\partial_a\tau = 0. \tag{1.13}$$

Because of $\mathrm{Tr}\mu = 0$ the matrix μ is an element of the Lie algebra sl(2,R). Using the basis

$$Q_1 = i\sigma_2, \quad Q_2 = \sigma_1, \quad Q_3 = \sigma_3 \tag{1.14}$$

for sl(2,R), μ can be expanded as

$$\mu = \sum_{i=1}^{3} q^i Q_i. \tag{1.15}$$

The Killing form of sl(2,R) yields a pseudonorm for the 3-vector $q = (q_i)$:

$$q\cdot q \equiv -q_1^2 + q_2^2 + q_3^2 = \begin{cases} -\tau^2 & \text{case(A)}, \\ \tau^2 & \text{case(B)}. \end{cases} \tag{1.16}$$

In terms of q Eq. (1.11) can be written as

$$\partial^a(\tau^{-1}q\times\partial_a q) = 0, \tag{1.17}$$

where the cross product is the usual one in a three-dimensional space with the metric given in Eq. (1.16).

2. THE INFINITELY MANY CONSERVATION LAWS OF GEROCH

Since there occur some relevant differences in signs I prefer to treat the cases (A) and (B) separately.

Case (A)

Equation (1.11) can be read as the integrability condition for ($\tilde\partial_a \equiv \epsilon_{ab}\partial^b$)

$$\partial_a\omega = \tau^{-1}\mu\tilde\partial_a\mu \tag{2.1}$$

with some 2×2 matrix ω [which incidentally turns out to be closely related to the matrix of potentials Ω of Eq. (1.5), namely $\omega = -\epsilon\Omega$].

Taking the trace of Eq. (1) yields ($\sigma \equiv -\frac{1}{2}\mathrm{Tr}_\omega$)

$$\partial_a\sigma = \tilde\partial_a\tau \tag{2.2}$$

and hence $\partial^a\partial_a\sigma = 0$, i.e., $\tau = (\det\lambda)^{1/2}$ and σ ($\sigma \equiv -\frac{1}{2}\mathrm{Tr}\omega$) are a pair of conjugate solutions of the wave equation in two dimensions.

Geroch[2] has demonstrated that the "potentials" μ and ω give rise to the recursive construction of an infinite sequence of new potentials μ_n and ω_n via an infinite sequence of conservation laws involving these potentials. I prefer to follow the somewhat more appropriate presentation of Kinnersley[3] and define

$$\partial\mu_{n+1} = -\partial(\mu\omega_n) - \partial\omega\mu_n - \mu\partial\omega_n + 2\sigma\partial\mu_n, \tag{2.3a}$$
$$\partial\omega_{n+1} = \partial(\mu\mu_n) - \partial\omega\omega_n + \mu\partial\mu_n + 2\sigma\partial\omega_n, \quad n = 0,1,2..., \tag{2.3b}$$

with the initial data

$$\mu_0 = 0, \quad \mu_1 = \mu, \quad \omega_0 = -\mathbb{1}, \quad \omega_1 = \omega. \tag{2.3c}$$

The integrability conditions for the existence of the potentials μ_n and ω_n are easily proved by induction, starting from Eq. (2.1).

A more concise form of Eqs. (2.3) is obtained using generating functions, defined by

$$V(s) \equiv \sum_{n=0}^{\infty} s^n\mu_n, \quad U(s) \equiv \sum_{n=0}^{\infty} s^n\omega_n. \tag{2.4}$$

In terms of $V(s)$ and $U(s)$ Eqs. (2.3) read

$$s^{-1}\partial V(s) = -\partial(\mu U(s)) - \partial\omega V(s) - \mu\partial U(s) + 2\sigma\partial V(s), \tag{2.5a}$$
$$s^{-1}\partial U(s) = \partial(\mu V(s)) - \partial\omega U(s) + \mu\partial V(s) + 2\sigma\partial U(s), \tag{2.5b}$$

In turns out that $V(s)$ can be eliminated from these equations by the following ansatz:

$$V(s) = f(s,\tau,\sigma)\mu U(s), \tag{2.6}$$

where f is a real function to be determined. Putting the ansatz for $V(s)$ into Eq. (2.5a) yields the compatible system of equations

$$\tau^2 f^2 - 2\sigma f + s^{-1}f + 1 = 0, \tag{2.7a}$$
$$\tau^2 f\frac{\partial f}{\partial\tau} - 2\sigma\frac{\partial f}{\partial\tau} - \tau\frac{\partial f}{\partial\sigma} + s^{-1}\frac{\partial f}{\partial\tau} = 0, \tag{2.7b}$$
$$\tau^2 f\frac{\partial f}{\partial\sigma} - 2\frac{\partial(\sigma f)}{\partial\sigma} - \tau\frac{\partial f}{\partial\tau} + s^{-1}\frac{\partial f}{\partial\sigma} = 0. \tag{2.7c}$$

The solution with $f(0,\tau,\sigma) = 0$ is given by

$$f(s,\tau,\sigma) = \frac{1}{2s\tau^2}\left[2s\sigma - 1 + \sqrt{(2s\sigma-1)^2 - 4s^2\tau^2}\,\right]. \tag{2.8}$$

Using this function in the ansatz equation (2.6) for $V(s)$, Eq. (2.5a) is identically fulfilled, irrespective what $U(s)$ may be.

Equation (2.5b) becomes now

$$\partial U(s) = \frac{\tau^{-1}f}{1-\tau^2 f^2}\Big(\mu\tilde{\partial}\mu - \tau f\mu\partial\mu + 2\tau^2\frac{\partial(\tau f)}{\partial\tau}\partial\tau + 2\tau^3\frac{\partial f}{\partial\sigma}\partial\sigma\Big)U(s). \tag{2.9}$$

The last two terms in Eq. (2.9) can be eliminated putting

$$U'(s)\equiv(1-\tau^2 f^2)U(s), \tag{2.10}$$

for which

$$\partial U' = \frac{\tau^{-1}f}{1-\tau^2 f^2}(\mu\tilde{\partial}\mu - \tau f\mu\partial\mu)U' \tag{2.11}$$

is obtained.

This equation takes a particularly simple form in light-cone coordinates on S defined by

$$\xi = \frac{x^1+x^2}{2}, \quad \eta = \frac{x^1-x^2}{2} \tag{2.12}$$

and with the function

$$\gamma^{-1}(s_1\cdots)\equiv\frac{1-\tau f}{1+\tau f} = \left(\frac{1+2s(\tau-\sigma)}{1-2s(\tau+\sigma)}\right)^{1/2} \tag{2.13}$$

instead of f. The result of these changes is

$$\partial_\xi U' = \tfrac{1}{2}(1-1/\gamma)\tau^{-2}\mu\partial_\xi\mu U', \tag{2.14a}$$

$$\partial_\eta U' = \tfrac{1}{2}(1-\gamma)\tau^{-2}\mu\partial_\eta\mu U'. \tag{2.14b}$$

These equations show a striking similarity with Eqs. (9) of Lüscher and Pohlmeyer[6] in their work on the O(3) nonlinear σ-model and in fact goes over into their equation in the special case $\tau=1, \sigma=0$, if one identifies μ with their ψ. A more detailed discussion of this similarity will follow in Sec. 4.

Case (B)

Due to $\mu^2=\tau^2$ and the different signature of h_{ab}, there are some changes in sign compared to case (A).

τ and σ are now a pair of conjugate harmonic functions Eqs. (2.7) become

$$\tau^2 f^2 + 2\sigma f - s^{-1}f - 1 = 0, \tag{2.15a}$$

$$\tau^2 f\frac{\partial f}{\partial\tau} + 2\sigma\frac{\partial f}{\partial\tau} - \tau\frac{\partial f}{\partial\sigma} - s^{-1}\frac{\partial f}{\partial\tau} = 0, \tag{2.15b}$$

$$\tau^2 f\frac{\partial f}{\partial\sigma} + 2\frac{\partial(\sigma f)}{\partial\sigma} + \tau\frac{\partial f}{\partial\tau} - s^{-1}\frac{\partial f}{\partial\sigma} = 0, \tag{2.15c}$$

with the solution

$$f(s,\tau,\sigma) = \frac{1}{2s\tau^2}\left[1-2s\sigma - \sqrt{(1-2s\sigma)^2+4s^2\tau^2}\,\right]\}. \tag{2.16}$$

The resulting equation for U reads

$$\partial U = -\frac{\tau^{-1}f}{1+\tau^2 f^2}\Big(\mu\tilde{\partial}\mu + \tau^{-1}f\mu\partial\mu + 2\tau^2\frac{\partial(\tau f)}{\partial\tau}\partial\tau + 2\tau^3\frac{\partial f}{\partial\sigma}\partial\sigma\Big)U. \tag{2.17}$$

respectively $[U'\equiv(1+\tau^2 f^2)U]$

$$\partial U' = -\frac{\tau^{-1}f}{1+\tau^2 f^2}(\mu\tilde{\partial}\mu + \tau f\mu\partial\mu)U'. \tag{2.18}$$

Convenient coordinates for this case are

$$\xi = x^1 + ix^2, \quad \bar{\xi} = x^1 - ix^2,$$

in which Eq. (2.18) becomes

$$\partial_\xi U' = -\tfrac{1}{2}(1-1/\gamma)\tau^{-2}\mu\partial_\xi\mu U', \tag{2.19a}$$

$$\partial_{\bar{\xi}} U' = -\tfrac{1}{2}(1-\gamma)\tau^{-2}\mu\partial_{\bar{\xi}}\mu U', \tag{2.19b}$$

where the function

$$\gamma^{-1}(s_1\cdots)(s,\tau,\sigma)\equiv\frac{1+i\tau f}{1-i\tau f} = \left(\frac{1-2s(\sigma+i\tau)}{1-2s(\sigma-i\tau)}\right)^{1/2} \tag{2.20}$$

has been introduced instead of f.

3. THE GENERATION OF ONE-PARAMETER FAMILIES OF SOLUTIONS

In view of the special role played by τ (which together with σ could be introduced as coordinate in S) it is convenient to normalize the matrix μ, putting

$$\bar{\mu}\equiv\tau^{-1}\mu$$

with

$$\bar{\mu}^2 = \begin{cases} -\mathbb{1} & \text{in case (A),} \\ \mathbb{1} & \text{in case (B).} \end{cases} \tag{3.1}$$

Equation (1.11) yields

$$\partial_\eta(\tau\bar{\mu}\partial_\xi\bar{\mu}) + \partial_\xi(\tau\bar{\mu}\partial_\eta\bar{\mu}) = 0 \quad \text{in case (A),} \tag{3.2a}$$

$$\partial_\xi(\tau\bar{\mu}\partial_{\bar{\xi}}\bar{\mu}) + \partial_{\bar{\xi}}(\tau\bar{\mu}\partial_\xi\bar{\mu}) = 0 \quad \text{in case (B).} \tag{3.2b}$$

Equations (2.14), resp. (2.19), become

$$[\bar{U}\equiv(-s/f)^{1/2}U']$$

$$\partial_\xi\bar{U} = \tfrac{1}{2}(1-1/\gamma)\bar{\mu}\partial_\xi\bar{\mu}\bar{U}, \tag{3.3a}$$

$$\partial_\eta\bar{U} = \tfrac{1}{2}(1-\gamma)\bar{\mu}\partial_\eta\bar{\mu}\bar{U} \tag{3.3b}$$

to be supplemented by

$$\partial_\xi\partial_\eta\tau = 0, \tag{3.3c}$$

resp.,

$$\partial_\xi\bar{U} = -\tfrac{1}{2}(1-1/\gamma)\bar{\mu}\partial_\xi\bar{\mu}\bar{U}, \tag{3.4a}$$

$$\partial_{\bar{\xi}}\bar{U} = -\tfrac{1}{2}(1-\gamma)\bar{\mu}\partial_{\bar{\xi}}\bar{\mu}\bar{U}, \tag{3.4b}$$

to be supplemented by

$$\partial_\xi\partial_{\bar{\xi}}\tau = 0. \tag{3.4c}$$

Supposing one is able to solve Eqs. (3.3), resp. (3.4), for $\bar{U}$ given some solution μ of Eq. (3.2a), resp. (3.2b), then a lengthy but elementary computation shows that

$$\bar{\mu}^{(s)} = \bar{U}(s)^{-1}\bar{\mu}\bar{U}(s), \quad s\in\mathbb{R}, \tag{3.5a}$$

$$\tau^{(s)} = \left(\frac{\gamma(s)-\gamma(s)^{-1}}{4s\tau}\right)^2\tau \tag{3.5b}$$

yields a one-parameter family of solutions of Eq. (3.2a), resp. (3.2b). $\bar{\mu}^{(s)}$ has the further property

$$\partial_\xi\bar{\mu}^{(s)} = (1/\gamma)\bar{U}^{-1}\partial_\xi\bar{\mu}\bar{U} \tag{3.6a}$$

and

$$\partial_\eta\bar{\mu}^{(s)} = \gamma\bar{U}^{-1}\partial_\eta\bar{\mu}\bar{U} \tag{3.6b}$$

in case (A) and similarly for (B), which also entails

$$(\partial_\xi \bar{\mu}^{(s)})^2 = \frac{1}{\gamma^2}(\partial_\xi \bar{\mu})^2 \quad \text{and} \quad (\partial_\eta \bar{\mu}^{(s)})^2 = \gamma^2(\partial_\eta \bar{\mu})^2. \tag{3.7}$$

Similarly $\tau^{(s)}$ obeys

$$\tau^{(s)-1}\partial_\xi \tau^{(s)} = \frac{1}{\gamma^2}\tau^{-1}\partial_\xi \tau \tag{3.8a}$$

and

$$\tau^{(s)-1}\partial_\eta \tau^{(s)} = \gamma^2 \tau^{-1}\partial_\eta \tau \tag{3.8b}$$

in case (A), and analogous relations in case (B)

The family $\bar{\mu}^{(s)}$ is to be compared to the family $q^{(\gamma)}$ of Ref. 6. The unitarity of $U^{(\gamma)}$ in Ref. 6 is replaced here by the condition $\det \bar{U} = 1$, which plays an analogous role for the group SL(2,R) as does the unitarity for the group SU(2). In order to see this, one defines for an arbitrary, nonsingular 2×2 matrix a

$$a^x \equiv -\epsilon a^T \epsilon = (\det a) a^{-1}. \tag{3.9}$$

The conjugation $\times$ has the following properties:

$$\begin{aligned} &a^{xx} = a, \quad (\lambda \cdot \mathbb{1})^x = \lambda \cdot \mathbb{1}, \quad \lambda \in \mathbb{C}, \\ &(ab)^x = b^x a^x, \quad (a+b)^x = a^x + b^x, \\ &a = -a^x \Longleftrightarrow \mathrm{Tr}\, a = 0, \\ &a^x a = \mathbb{1} \Longleftrightarrow \det a = 1. \end{aligned} \tag{3.10}$$

In particular $\bar{\mu}^x = -\bar{\mu}$, which together with Eqs. (3.3), resp. (3.4), leads to

$$\partial(\bar{U}^x \bar{U}) = \partial(\bar{U}\bar{U}^x) = 0 \tag{3.11}$$

showing that Eqs. (3.3), resp. (3.4), are compatible with the normalization $\bar{U}^x \bar{U} = \bar{U}\bar{U}^x = 1$, i.e., $\bar{U} \in \mathrm{SL}(2,R)$.

The transformation $\mu \to \mu^{(s)}$ is very similar, but not equal, to the one found by Geroch.[2] The infinitesimal form of Eq. (3.5) is

$$\delta\bar{\mu} = [\omega, \bar{\mu}], \quad \delta\tau = 4\sigma\tau, \tag{3.12}$$

whereas Geroch's transformation is

$$\delta\bar{\mu} = [[\omega, k], \bar{\mu}] + 2\sigma[\bar{\mu}, k], \quad \delta\tau = 0 \tag{3.13}$$

with some constant matrix $k \in \mathrm{sl}(2,R)$.

Clearly one can also supplement the transformation $\mu \to \mu^{(s)}$ by a constant SL(2,R) "rotation" $U_0(k)$ in the form $\bar{\mu}^{(s)} = [\bar{U}(s)U_0(k)]^{-1}\bar{\mu}\bar{U}(s)U_0(k)$.

The use of the transformation $\mu \to \mu^{(s)}$ is twofold. On the one hand, it provides for a method to generate new solutions of Eq. (1.11) and hence of Eqs. (1.7) from old ones, if it is possible to compute $\bar{U}(s)$. This has already been undertaken successfully[9] with Geroch's transformation Eq. (3.13). On the other hand, the family $\bar{\mu}^{(s)}$ will be essential for the construction of the linear eigenvalue problem in the next section.

4. CONSTRUCTION OF THE LINEAR EIGENVALUE PROBLEM

This section is devoted to the derivation of a linear eigenvalue problem in the spirit of Lax (cf. Ref. 4), using the family $\mu^{(s)}$ originating from each solution μ of Eq. (1.11). For that purpose the representation of $\bar{\mu}^{(s)}$ as a 3-vector [cf. Eq. (1.15)]

$$\bar{\mu} = \sum_{i=1}^{3} q^i Q_i \tag{4.1}$$

with the normalization

$$q^2 = \begin{cases} -1 & \text{in case (A)} \\ +1 & \text{in case (B)} \end{cases}$$

[recall $q^2 = -(q^1)^2 + (q^2)^2 + (q^3)^2$] is used.

Case (A)

From $q^2 = -1$ it is clear that

$$(\partial_\xi q)^2 > 0 \quad \text{and} \quad (\partial_\eta q)^2 > 0.$$

Therefore, the functions

$$A \equiv \sqrt{(\partial_\xi q)^2}, \quad B \equiv \sqrt{(\partial_\eta q)^2} \tag{4.2}$$

are real. Let α be defined through

$$\cos\alpha = \frac{\partial_\xi q \cdot \partial_\xi q}{AB}. \tag{4.3}$$

As a consequence of Eq. (3.2a) q obeys

$$\partial_\xi \partial_\eta q + \tfrac{1}{2}\tau^{-1}(\partial_\xi \tau \partial_\eta q + \partial_\eta \tau \partial_\xi q) - (\partial_\xi q \cdot \partial_\eta q)q = 0, \tag{4.4}$$

which in turn implies for the invariants A, B and α the equations

$$\partial_\eta A + \tfrac{1}{2}\tau^{-1}\partial_\xi \tau B \cos\alpha + \tfrac{1}{2}\tau^{-1}\partial_\eta \tau A = 0, \tag{4.5a}$$

$$\partial_\xi B + \tfrac{1}{2}\tau^{-1}\partial_\xi \tau B + \tfrac{1}{2}\tau^{-1}\partial_\eta \tau A \cos\alpha = 0, \tag{4.5b}$$

$$\partial_\xi \partial_\eta \alpha + AB \sin\alpha - \tfrac{1}{2}\partial_\eta\left(\tau^{-1}\partial_\eta \tau \frac{A}{B}\sin\alpha\right) - \tfrac{1}{2}\partial_\xi\left(\tau^{-1}\partial_\xi \tau \frac{B}{A}\sin\alpha\right) = 0. \tag{4.5c}$$

Equation (4.5c) is a generalization of the well-known sine–Gordon equation to which it reduces for $A = B = \tau = 1$.

According to Eq. (3.7) the vectors $q^{(s)}$ corresponding to $\bar{\mu}^{(s)}$ have the invariants

$$A^{(s)} = \frac{1}{\gamma(s)}A, \quad B^{(s)} = \gamma(s)B, \quad \alpha^{(s)} = \alpha. \tag{4.6}$$

The vectors $q^{(s)}$, $\partial_\xi q^{(s)}$ and $\partial_\eta q^{(s)}$ can be orthonormalized with respect to the metric

$$\begin{pmatrix} - & & \\ & + & \\ & & + \end{pmatrix}$$

yielding the basis

$$\begin{aligned} Z_1^{(s)} &= q^{(s)}, \quad Z_2^{(s)} = \frac{\gamma B \partial_\xi q^{(s)} + \gamma^{-1} A \partial_\eta q^{(s)}}{2AB\cos(\alpha/2)}, \\ Z_3^{(s)} &= \frac{\gamma B \partial_\xi q^{(s)} - \gamma^{-1} A \partial_\eta q^{(s)}}{2AB\sin(\alpha/2)}. \end{aligned} \tag{4.7}$$

Let $Z^{(s)}$ be the matrix built from the rows $Z_i^{(s)}$

$$Z^{(s)}=\begin{pmatrix} Z_1^{(s)} \\ Z_2^{(s)} \\ Z_3^{(s)} \end{pmatrix}. \tag{4.8}$$

By a somewhat tiring computation one derives the following equations for $Z^{(s)}$

$$\partial_\eta Z^{(s)} = C_1 Z^{(s)}, \tag{4.9a}$$

$$C_1=\begin{pmatrix} 0, \gamma B\cos\frac{a}{2}, -\gamma B\sin\frac{\alpha}{2} \\ \gamma B\cos\frac{\alpha}{2}, 0, -\frac{1}{2}\partial_\eta\alpha+\frac{1}{2}\tau^{-1}\partial_\xi\tau\frac{B}{A}\sin\alpha \\ -\gamma B\sin\frac{\alpha}{2}, \frac{1}{2}\partial_\eta\alpha-\frac{1}{2}\tau^{-1}\partial_\xi\tau\frac{B}{A}\sin\alpha, 0 \end{pmatrix},$$

$$\partial_\xi Z^{(s)} = C_2 C^{(s)}, \tag{4.9b}$$

$$C_2=\begin{pmatrix} 0, \frac{1}{\gamma}A\cos\frac{\alpha}{2}, \frac{1}{\gamma}A\sin\frac{\alpha}{2} \\ \frac{1}{\gamma}A\cos\frac{\alpha}{2}, 0, \frac{1}{2}\partial_\xi\alpha-\frac{1}{2}\tau^{-1}\partial_\eta\tau\frac{A}{B}\sin\alpha \\ \frac{1}{\gamma}A\sin\frac{\alpha}{2}, -\frac{1}{2}\partial_\xi\alpha+\frac{1}{2}\tau^{-1}\partial_\eta\tau\frac{A}{B}\sin\alpha, 0 \end{pmatrix},$$

The matrices C_i are elements of the vector representation of the Lie algebra so(2,1). The compatibility condition of Eqs. (4.9a), (4.9b)

$$\partial_\eta C_2 - \partial_\xi C_1 = [C_1, C_2] \tag{4.10}$$

is equivalent to Eqs. (4.5).

Introducing a basis for the vector representation of so(2,1),

$$I_1=\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix},\quad I_2=\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},\quad I_3=\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.11}$$

one can expand the C_i as

$$C_i = \sum_{k=1}^{3} \omega_i{}^k I_k \tag{4.12}$$

with

$$(\omega_1{}^k) = \left(-\frac{1}{2}\partial_\eta\alpha+\frac{1}{2}\tau^{-1}\partial_\xi\tau\frac{B}{A}\sin\alpha, -\gamma B\sin\frac{\alpha}{2}, \gamma B\cos\frac{\alpha}{2}\right), \tag{4.13a}$$

$$(\omega_2{}^k) = \left(\frac{1}{2}\partial_\xi\alpha-\frac{1}{2}\tau^{-1}\partial_\eta\tau\frac{A}{B}\sin\alpha, \frac{1}{\gamma}A\sin\frac{\alpha}{2}, \frac{1}{\gamma}A\cos\frac{\alpha}{2}\right).$$

The final step (cf. Ref. 10) is now to go over to the spinor representation of the matrices C_i.

The matrices

$$\tilde{Q}_1 = -\frac{i}{2}\sigma_3,\quad \tilde{Q}_2=\frac{1}{2}\sigma_1,\quad \tilde{Q}_3=\frac{1}{2}\sigma_2 \tag{4.14}$$

may be taken as the representatives of the basis (I_k) in the spinor representation of so(2,1). Hence the spinor representatives of the matrices C_i are

$$c_1 = \sum_{k=1}^{3} \omega_1{}^k \tilde{Q}_k = \frac{1}{2}\begin{pmatrix} (i/2)\partial_\eta\alpha - (i/2)\tau^{-1}\partial_\xi\tau(B/A)\sin\alpha, & -i\gamma B e^{-i\alpha/2} \\ i\gamma B e^{i\alpha/2}, & -(i/2)\partial_\eta\alpha + (i/2)\tau^{-1}\partial_\xi\tau(B/A)\sin\alpha \end{pmatrix} \tag{4.15a}$$

$$c_2 = \sum_{k=1}^{3} \omega_1{}^k \tilde{Q}_k = \frac{1}{2}\begin{pmatrix} -(i/2)\partial_\xi\alpha - (i/2)\tau^{-1}\partial_\eta\tau(A/B)\sin\alpha, & -(i/\gamma)Ae^{i\alpha/2} \\ (i/\gamma)Ae^{-i\alpha/2}, & (i/2)\partial_\xi\alpha - (i/2)\tau^{-1}\partial_\eta\tau(A/B)\sin\alpha \end{pmatrix}. \tag{4.15b}$$

Introducing the normalized two-component spinor ψ ($\psi^+\sigma_3\psi = 1$) yields finally the desired linear eigenvalue problem

$$\partial_\eta\psi = c_1\psi, \tag{4.16a}$$

$$\partial_\xi\psi = c_2\psi \tag{4.16b}$$

with c_i from Eq. (4.15)

The matrices c_i depend parametrically on the real parameter s which plays the role of an "eigenvalue" in Eqs. (4.16). Unfortunately, the dependence on the "eigenvalue" s is rather more complicated than in the known examples of the Lax equations (cf. Ref. 4 for a survey of examples). The comparison with the nonlinear σ-model becomes simpler if one changes from the variable s to a variable u defined by

$$u \equiv \left(\frac{1+2s}{1-2s}\right)^{1/2}, \tag{4.17}$$

which gives for the function

$$\gamma(s(u)) = u\left(\frac{1+(1-1/u^2)(\tau-\sigma-1)}{1+(1-u^2)(\tau+\sigma-1)}\right)^{1/2}, \tag{4.18}$$

reducing to $\gamma = u$ for $\tau = 1$, $\sigma = 0$.

Case (B)

In contrast to case (A) the vector q has now positive square $q^2 = 1$, because of $\mu^2 = \tau^2 > 0$. Furthermore, the vectors $\partial_\xi q$ and $\partial_{\bar\xi} q$ are now complex. This does however not prevent one from making a similar construction as in case (A).

Choosing the square root appropriately, the functions

$$A \equiv \sqrt{(\partial_\xi q)^2} \quad \text{and} \quad B \equiv \sqrt{(\partial_{\bar\xi} q)^2} \tag{4.19}$$

are complex conjugate to each other. The angle α, defined through

$$\cos\alpha = \frac{\partial_\xi q\cdot\partial_{\bar\xi} q}{AB}, \tag{4.20}$$

turns out to be real. The analog of Eq. (4.4) is

$$\partial_\xi\partial_{\bar\xi} q + \frac{1}{2}\tau^{-1}(\partial_\xi\tau\partial_{\bar\xi} q + \partial_{\bar\xi}\tau\partial_\xi q) + (\partial_\xi q\cdot\partial_{\bar\xi} q)q = 0 \tag{4.21}$$

leading to the equations

$$\partial_{\bar\xi} A + \frac{1}{2}\tau^{-1}\partial_\xi\tau B\cos\alpha + \frac{1}{2}\tau^{-1}\partial_{\bar\xi}\tau A = 0, \tag{4.22a}$$

$$\partial_\xi B + \frac{1}{2}\tau^{-1}\partial_\xi\tau B + \frac{1}{2}\tau^{-1}\partial_{\bar\xi}\tau A\cos\alpha = 0, \tag{4.22b}$$

$$\partial_\xi\partial_{\bar\xi}\alpha + AB\sin\alpha - \partial_{\bar\xi}\left(\frac{1}{2}\tau^{-1}\partial_{\bar\xi}\tau\frac{A}{B}\sin\alpha\right) - \partial_\xi\left(\frac{1}{2}\tau^{-1}\partial_\xi\tau\frac{B}{A}\sin\alpha\right) = 0. \tag{4.22c}$$

A suitable real basis in the space of q's is now

$$Z_1^{(s)} = q^{(s)}, \quad Z_2^{(s)} = \frac{\gamma B \partial_\xi q^{(s)} + \gamma^{-1} A \partial_{\bar\xi} q^{(s)}}{2AB\cos(\alpha/2)}, \quad Z_3^{(s)} = \frac{-\gamma B \partial_\xi q^{(s)} - \gamma^{-1} A \partial_{\bar\xi} q^{(s)}}{2iAB\sin(\alpha/2)}. \tag{4.23}$$

The equations for $Z^{(s)}$ become

$$\partial_{\bar\xi} Z^{(s)} = \begin{pmatrix} 0, & \gamma B\cos(\alpha/2), & i\gamma B\sin(\alpha/2) \\ -\gamma B\cos(\alpha/2), & 0, & (i/2)\partial_{\bar\xi}\alpha - (i/2)\tau^{-1}\partial_\xi\tau(B/A)\sin\alpha \\ i\gamma B\sin(\alpha/2), & (i/2)\partial_{\bar\xi}\alpha - (i/2)\tau^{-1}\partial_\xi\tau(B/A)\sin\alpha, & 0 \end{pmatrix} Z^{(s)} \equiv C_1 Z^{(s)}, \tag{4.24a}$$

$$\partial_\xi Z^{(s)} = \begin{pmatrix} 0, & (1/\gamma)A\cos(\alpha/2), & -(i/\gamma)A\sin(\alpha/2) \\ -(A/\gamma)\cos(\alpha/2), & 0, & -(i/2)\partial_\xi\alpha + (i/2)\tau^{-1}\partial_{\bar\xi}\tau(A/B)\sin\alpha \\ -(i/\gamma)A\sin(\alpha/2), & -(i/2)\partial_\xi\alpha + (i/2)\tau^{-1}\partial_{\bar\xi}\tau(A/B)\sin\alpha, & 0 \end{pmatrix} Z^{(s)} \equiv C_2 Z^{(s)}, \tag{4.24b}$$

Taking

$$I_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad I_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4.25}$$

as a basis for the vector representation of the Lie algebra so(2,1) the C_i can be expanded as

$$C_i = \sum_{\epsilon=1}^{3} \omega_i{}^k I_k$$

with

$$(\omega_1{}^k) = \left(\frac{i}{2}\partial_{\bar\xi}\alpha - \frac{i}{2}\tau^{-1}\partial_{\bar\xi}\tau\frac{B}{A}\sin\alpha, i\gamma B\sin\frac{\alpha}{2}, B\gamma\cos\frac{\alpha}{2}\right), \tag{4.26a}$$

$$(\omega_2{}^k) = \left(-\frac{i}{2}\partial_\xi\alpha + \frac{i}{2}\tau^{-1}\partial_\xi\tau\frac{A}{B}\sin\alpha, -\frac{i}{\gamma}A\sin\frac{\alpha}{2}, \frac{1}{\gamma}A\cos\frac{\alpha}{2}\right). \tag{4.26b}$$

The spinor representation of the basis (I_k) is given by

$$\tilde Q_1 = \tfrac{1}{2}\sigma_3, \quad \tilde Q_2 = \tfrac{1}{2}\sigma_1, \quad \tilde Q_3 = -\frac{i}{2}\sigma_2 \tag{4.27}$$

yielding the spinor representation of the matrices C_i

$$c_1 = \tfrac{1}{2}\begin{pmatrix} (i/2)\partial_{\bar\xi}\alpha - (i/2)\tau^{-1}\partial_{\bar\xi}\tau(B/A)\sin\alpha, & -\gamma B e^{-i\alpha/2} \\ \gamma B e^{i\alpha/2}, & -(i/2)\partial_{\bar\xi}\alpha + (i/2)\tau^{-1}\partial_\xi\tau(B/A)\sin\alpha \end{pmatrix}, \tag{4.28a}$$

$$c_2 = \tfrac{1}{2}\begin{pmatrix} -(i/2)\partial_\xi\alpha + (i/2)\tau^{-1}\partial_{\bar\xi}\tau(A/B)\sin\alpha, & -(1/\gamma)Ae^{(i\alpha/2)} \\ (1/\gamma)Ae^{-i\alpha/2}, & (i/2)\partial_\xi\alpha - (i/2)\tau^{-1}\partial_{\bar\xi}\tau(A/B)\sin\alpha \end{pmatrix}. \tag{4.28b}$$

The linear "eigenvalue problem" (with "eigenvalue" s hidden in γ) having eqs. (4.22) as compatibility conditions is given by

$$\partial_{\bar\xi}\psi = c_1\psi, \tag{4.29a}$$

$$\partial_\xi\psi = c_2\psi, \tag{4.29b}$$

where ψ is a complex two-component spinor, normalized to $\psi^*\sigma_2\psi = 1$.

Equations (4.16), resp. (4.29), differ in two respects from the Lax equations of the examples of completely integrable systems like the sine–Gordon equation or the Korteweg–de Vries equation.

First the dependence on the "eigenvalue" s is rather involved [cf. Eqs. (2.13), resp. (2.20)]. In particular the expansion of the "wavefunction" ψ around $\gamma = 0$ or $\gamma = \infty$ used in the sine–Gordon theory in order to derive an infinite sequence of local conservation laws and to prove the complete integrability is in general not possible here. Second, the interesting solutions are the asymptotically Minkowskian ones, for which the linear problem has a different asymptotic behavior than the one assumed usually, leading to asymptotic free wave solutions for ψ. Nevertheless, the asymptotic behavior in the present case is simple enough that one may hope a method similar to the "inverse scattering method"[4] may be invented to reduce the nonlinear problem to a sequence of linear ones.

At any rate it is clear that the linear Eqs. (4.16), resp. (4.29), are a very special feature of the system and contain a lot of information, which may be sufficient to prove the complete integrability of the system, a point to be clarified by further investigations.

ACKNOWLEDGMENTS

I am indebted to Dr. P. Breitenlohner, Dr. K. Pohlmeyer, and Dr. S. Schlieder for many clarifying discussions.

[1]*Black Holes* (*Les Houches* 1972), edited by C. Dewitt and B.S. DeWitt (Gordon and Breach, New York, 1973).
[2]R. Geroch, J. Math. Phys. **13**, 394 (1972).
[3]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977).
[4]A.C. Scott, E.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[5]K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[6]M. Lüscher and K. Pohlmeyer, Nucl. Phys. B **137**, 46 (1978).
[7]F. Lund, Phys. Rev. D **15**, 1540 (1977).
[8]T. Lewis, Proc. Roy. Soc. A **136**, 176 (1932).
[9]W. Kinnersley and D.M. Chitre, J. Math. Phys. **19**, 1926 (1978).
[10]A. Neveu and N. Papanicolaou, Commun. Math. Phys. **58**, 31 (1978).

An attempt to separate the long and short range forces by Gaussian method

Otieno Malo[a)]

Department of Physics, University of Nairobi, Nairobi, Kenya
(Received 24 February 1978)

In the study of phase transition problems, short range forces (SRF) play a dominant role. A constructive and rigorous study of the effects of short range forces has yet to be given. It is suggested in the present paper that by separating long range forces (LRF) from the short range forces, it would be possible to estimate contributions to the virial expansion of the collective oscillations due to short range forces. The method of Stratonovich *or* the functional integration technique is employed in the treatment of the interaction term of the partition function.

I. INTRODUCTION

In the present paper an explicit form of $V(r)$ is irrelevant since no particular physical system is under investigation. Our problem is concerned with the mathematics of this technique leading to an explicit expression of the SR and LR terms of the $V(r)$. From there on, it is hoped we shall be in a position to study specific physical problems with definite forms of $V(r)$ for solids, liquids, and gases leading to a new study of first and second order phase transition or order–disorder transitions.[1]

The problem of obtaining thermodynamic properties of a system with long range interaction (LRF) is confronted in many branches of physics, e.g., in plasma or in electrolyte theory.[2,3,4] Such forces cannot be treated in the same way as short range forces (SRF), since a straightforward calculation of the virial coefficients leads to a divergent answer so that either the divergent virial series has to be manipulated into a finite answer or a different approach has to be applied as in the Debye–Huckel electrolyte theory.[5,6] These methods have a number of setbacks especially when we need accuracy greater than 1st approximation and it is therefore necessary to produce or invent a method which gives the higher order corrections in a simple way and also present the possibility of investigating these systems at high densities, solids. A system with purely LRF has both mathematical and physical drawbacks since mathematically, the grand partition function will have an essential singularity when treated as a function of interaction terms or of temperature and will only exist in the absence of attractive forces. But physically, for example in an electrolyte, forces change their nature drastically at short distances.

The Gaussian method or the functional integration technique has application to a fairly wide class of statistical mechanics problems[7]; the great sucess of statistical mechanics is invariably associated with systems in which interaction between the particles is either neglected or can be transformed away.[8] Systems under conditions in which interaction plays an essential role are vastly more difficult to analyze rigorously and meaningful approximations are hard to obtain. Unfortunately almost all phase transition problems—melting, condensation, ferromagnetic transition, order–disorder transition, and the like fall into this category. At present we do not have a complete understanding of any of these phase transitions as they occur in nature and it is interesting to note that we understand even less about boiling water than about liquid helium.

The immediate effect of interparticle interaction is that it will hinder the drift motion of the individual particles but contrary to this enhances their collective motion. The most direct method of studying the effects of intermolecular potentials between molecular systems that do not form stable compounds under ordinary conditions to date is the molecular beam scattering experiment (MBS).[9] However for studying effects of intermolecular potentials of stable systems, the MBS method cannot compete with optical spectroscopy (OS) which gives to a large extent only information on the attractive portion of the potential. It should however, be noted that the MBS is more universal and covers the entire energy range. Intermolecular potentials are of basic importance for the understanding of many macroscopic properties of matter. It is at the basis of all theories on the equation of state of gases, liquids and solids. In fact the two-body potential is the starting point of the theoretical description of gas kinetic processes. Thus once the potential curves (hypersurfaces) are available, nonequilibrium statistical mechanics–Boltzmann equation–or the simple equilibrium statistical mechanics can be used to compute all the transport properties or, when molecules are involved, relaxation times for rotational and vibrational degrees of freedom.[10,11] It is important to point out that previously information on the intermolecular potential has been obtained from measurements of these and other macroscopic properties. These experimental methods have the main disadvantage that since the observed values are averaged over the behavior of many molecular interactions, the data are frequently not sensitive to important details of the potential.

We believe that the method we are developing in this paper can prove to be very powerful in theoretical investigations of the problem of order–disorder transitions in alloys and lattice gas of Yang and Lee. The Ising model[12] which was initially developed as a model of ferromagnetism can be applied. It assumes that the energy of a lattice of spins is given,

a)Present address: Quantum Chemistry Group, Box 518, S-75120, Uppsala 1, Sweden.

 0022-2488/79/050878-09$01.00

in the absence of an external magnetic field, by

$$H = \tfrac{1}{2} \sum_{ij}{}' V_{ij} \mu_i \mu_j,$$

where μ_i is the spin variable which the model assumes to be the only significant variable of the spin at the lattice site indexed i. μi assumes the values ± 1 (spin up and spin down); V_{ij} denotes the interaction energy between the particles i and j when their spins are parallel. The prime on the double sum indicates the constraint $i \neq j$. The partition function for an Ising model of N sites is therefore

$$Z_N = \sum_{\{\mu\}} \exp\left(-\beta \tfrac{1}{2} \sum_{ij}{}' V_{ij} \mu_i \mu_j \right),$$

where $\Sigma_{\{\mu\}}$ denotes summation over all spin configurations and $\beta = 1/kT$. The free energy per spin is given by

$$F = -kT \lim_{n \to \infty} \frac{1}{N} \log Z_N.$$

From the knowledge of this function we could compute all thermodynamic properties of the system.

The shape of a typical intermolecular potential in the radial region most effective at thermal energies is well known in literature. Scattering experiments indicate that this potential shape is typical for collision partners of which at least one has a closed shell. And for such systems, the potential minima are located between 3.0 and 6.0 Å while the depths lie between 1.10^{-3} and 60×10^{-3} eV. Thus the attraction at long distances can be attributed to the long range Coulomb coupling between the electrons in the two systems—London dispersion forces. Therefore, the electron motion are correlated in such a way as to reduce the potential energy. Theoretical calculations predict for ground state atom–atom interactions an R^{-6} behavior. The strong repulsion at short distances can be largely attributed to a repulsion of the electron clouds, due to the Pauli principle mutual exclusion of electrons, and to the electrostatic repulsion of the nuclei. In view of the difficulties encountered in calculation of potential surfaces and especially of the well location and depth, theoretical understanding is still based on aproximate methods. At long ranges, beyond the minimum $R \geqslant 2Rm$ reliable semiempirical methods[13,14] are available for obtaining constant C in the expression $V(R) = -C/R^6$. Unfortunately, there are no reliable methods available for estimating well location Rm and depth ϵ. However, at very short ranges (~ 2 Å) the Thomas–Fermi–Dirac method has been shown to give reliable results.[15]

II. CLASSICAL PARTITION FUNCTION

A. Thermodynamic properties

Let us consider a system of interacting particles with the following Hamiltonian,

$$H = \sum_i \frac{P_i^2}{2m_i} + \sum_{ij} V(r_i - r_j), \tag{II.1}$$

where P_i and m_i are the momentum and mass of the ith particle respectively. $V(r_i - r_j)$ is the interaction energy between particles at positions r_i and r_j.

The classical partition function for such a system is given by the following configuration integral,[16]

$$Z = \prod_i \int dP_i \, dr_i \, e^{-\theta H}$$

$$= \left(\frac{m}{2\pi KT} \right)^{3N/2} \prod_i \int dr_i \, e^{-\theta V(r_i)} \tag{II.2}$$

after carrying out integration on the kinetic energy term.

It is evident from Eq. (II.2) above that our problem reduces to the evaluation of the following integral,

$$Q = \prod_i \int dr_i \, e^{-\theta V(r_i)}. \tag{II.3}$$

The above integral involves both short and long range terms of the interaction potential. Our primary aim is to try and separate these two terms. In this way, we believe that we shall have introduced a new method for concrete and accurate investigation of the points of phase transitions.

Let us define a Fourier transform in a box Ω (periodic boundary condition),

$$f(x) = \frac{1}{\Omega} \sum_K e^{-iKx} f(K), \tag{II.4}$$

with the inverse

$$f(K) = \int_\Omega e^{iKx} f(x) \, dx \tag{II.5}$$

such that $f(K)$ is of order unity.

We note that

$$\frac{1}{\Omega} \sum_K e^{iKx} = \delta^3 x$$

and

$$\frac{1}{\Omega} \int_\Omega e^{iKx} \, d^3x = \delta_{K,0}$$

the number density is $\rho(x)$. Its Fourier transform is

$$\rho_x = \frac{1}{\Omega} \sum_K \rho_K e^{-iKx} \tag{II.6}$$

with the inverse

$$\rho_K = \int_\Omega e^{iKx} \rho(x) \, d^3x = \sum_n e^{iKx_n}. \tag{II.7}$$

Let us now consider the following potential,

$$\sum_{m,n} V(x_m - x_n) = \sum_{m,n} \sum_K \exp[iK(x_m - x_n)] \tilde{V}(K), \tag{II.8}$$

and from the definition of the number density Eq. (II.6), we come to

$$\sum_{m,n} V(x_m - x_n) = \frac{1}{\Omega} \sum_K \tilde{V}(K) \rho(K) \rho(-K)$$

$$= \left(\frac{1}{2\pi} \right)^3 \int d^3K \tilde{V}_K \rho_K \rho_{\quad K}. \tag{II.9}$$

Putting the constraint $m \neq n$ we obtain

$$\sum_{m \neq n} V(x_m - x_n) = \left(\frac{1}{2\pi}\right)^3 \int d^3K \tilde{V}_K \rho_K \rho_{-K} + NV_0. \tag{II.10}$$

Here N is the number of particles in the system.

In this formalism, the expression for Q takes the form

$$Q = Q_0 \prod_n \int d^3x \exp\left(- \frac{\theta}{\Omega} \sum_{K \neq 0} \tilde{V}_K \rho_K \rho_{-K} \right), \tag{II.11}$$

where

$$Q_0 = \exp\left(\theta \frac{N}{2} V_0 + \theta \frac{N^2}{2\Omega} \tilde{V}_0 \right)$$

which corresponds to the self energy part of the system.

We shall now introduce new variables by representing the number density in the form

$$\rho_K = C_K + iS_K, \quad \rho_{-K} = C_K - iS_K, \tag{II.12}$$

where

$$S_K = \frac{1}{2i} [\rho_K - \rho_{-K}] = \sum_n \sin K x_n$$

$$C_K = \tfrac{1}{2}[\rho_K + \rho_{-K}] = \sum_n \cos K x_n \tag{II.13}$$

and Eq. (II.11) takes the form

$$Q = Q_0 \prod_{n=1}^{N} \int d^3x_n \exp\left(- \frac{\theta}{2} \sum_{K \neq 0} \tilde{V}_K (C_K^2 + S_K^2) \right) \tag{II.14}$$

as evidenced by Eq. (II.12).

If we now call

$$\frac{\theta}{2} \tilde{V}_K = \lambda_K^2,$$

then Eq. (II.14) becomes

$$Q = Q_0 \prod_{n=1}^{N} \int d^3x_n \exp\left[- \tfrac{1}{2} \sum_{k \neq 0} \lambda_K^2 (C_K^2 + S_K^2) \right]. \tag{II.15}$$

Equation (II.15) is a standard form of integral and we shall now apply the following identity,

$$\exp(-\tfrac{1}{2} C_K^2 \lambda_K^2)$$

$$= \exp(\log \sqrt{\pi}) \int_{-\infty}^{\infty} \exp(-\alpha_K^2) \exp(i\alpha_K \lambda_K C_K \sqrt{2}) d\alpha_K, \tag{II.16}$$

and come directly to

$$Q = Q_0 \int \prod_K d\alpha_K d\beta_K \exp\left[- \sum_K (\alpha_K^2 + \beta_K^2 + \log\pi) \right]$$

$$\times \prod_n \int d^3x_n \exp\left[i \sum_K (C_K \alpha_K + S_K \beta_K) \lambda_K \sqrt{2} \right]. \tag{II.17}$$

Using Eq. (II.13) above, we can transform the last integral in Eq. (II.17) to be of the form

$$\prod_n^N \int \exp\left[i \sum_K (C_K \alpha_K + S_K \beta_K) \lambda_K \sqrt{2} \right] d^3x_n$$

$$= \left[\int d^3x \exp[iF(x)] \right]^N, \tag{II.18}$$

where

$$F(x) = \sum_{K \neq 0} \lambda_K \sqrt{2} (\alpha_K \cos Kx + \beta_K \sin Kx).$$

Now assuming that $F(x)$ is large only for a finite region about the origin and write

$$\left[\int_\Omega d^3x e^{iF(x)} \right]^N \underset{n \to \infty}{\Rightarrow} \Omega^N \exp\left(\frac{N}{\Omega} \int_\Omega d^3x \{ e^{iF(x)} - 1 \} \right). \tag{II.19}$$

Thus we have

$$Q = Q_0 \Omega^N \int \prod_K d\alpha_K d\beta_K \exp\left(- \sum_{k \neq 0} (\alpha_K^2 + \beta_K^2 + \log\pi) \right)$$

$$\times \exp\left(\frac{N}{\Omega} \int_\Omega \{ e^{iF(x)} - 1 \} d^3x \right). \tag{II.20}$$

The factor Ω^N occurs already for the free particles. It may be advantageous to remove the peculiar Ω dependence of λ_K assuming that

$$\tilde{\Phi}(K) = \sqrt{\theta \tilde{V}(K)}\, \sqrt{2}$$

and introducing new variables

$$\alpha_K = \alpha'_K / \sqrt{\Omega}, \quad \beta_K^+ = \beta'_K / \sqrt{\Omega},$$

Eq. (II.20) will thus be transformed to

$$Q = Q_0 \int \prod_K d\alpha_K \, d\beta_K$$

$$\times \exp\left(- \frac{1}{\Omega} \sum_{K \neq 0} (\alpha_K^2 + \beta_K^2 + \Omega \log\pi) \right)$$

$$\times \exp\left(\frac{N}{\Omega} \int_\Omega d^3x \{ e^{iF(x)} - 1 \} \right), \tag{II.21}$$

where we have dropped the primes and now

$$F(x) = \frac{1}{\Omega} \sum_{K \neq 0} (\alpha_K \cos Kx + \beta_K \sin Kx) \tilde{\Phi}_K.$$

To illustrate results obtainable in this formalism, we first expand,

$$e^{iF(x)} = 1 + iF(x) - \frac{F^2(x)}{2!} + \cdots. \tag{II.22}$$

Then

$$\int_\Omega d^3x F(x) = 0.$$

In order to compute $\int d^3x F^2(x)$ it is imperative to note that

$$F(x) = \frac{1}{\Omega} \sum_{K \neq 0} F(K) e^{-iKx}$$

with

$$F(K) = (\alpha_K + i\beta_K + \alpha_{-K} - i\beta_{-K}) \tfrac{1}{2} \Phi(K),$$

we have

$$\int d^3x F^2(x) = \left(\frac{1}{\Omega} \sum_{K \neq 0} F(K) F(-K) \right).$$

Thus the part of the partition function that is integrated over is the exponential of minus

$$\frac{1}{\Omega} \sum_{K \neq 0} (\alpha_K^2 + \beta_K^2) + \frac{N}{2\Omega^2} \sum_{K \neq 0} F(K) F(-K)$$
$$= \frac{1}{\Omega} \sum_{K \neq 0} \left((\alpha_K^2 + \beta_K^2) + \frac{N}{2\Omega} \Phi \tfrac{1}{4} (\alpha_K + \alpha_{-K})^2 \right.$$
$$\left. + \frac{N}{8\Omega} (\beta_K - \beta_{-K})^2 \Phi \right). \tag{II.23}$$

This is a Bogolyibov quadratic form. However we only need the Gaussian integral. Thus considering a pair K and $-K$ for α_K and β_K variables and performing all the necessary rearrangement and certain basic manipulations, we obtain the following result:

$$Q = Q_0 \exp\left(\sum_K \log \Omega \right)$$
$$\times \exp\left[-\tfrac{1}{2} \sum_K \log\left(1 + \frac{N}{\Omega} \Phi(K) \right) \right], \tag{II.24}$$

or finally

$$Q = \Omega^N \exp\left[-\left(\frac{N}{2} \theta V_0 + \frac{\theta}{2\Omega} \tilde{V}_0 N^2 \right) \right]$$
$$\times \exp\left[-\tfrac{1}{2} \sum_K \log\left(1 + \frac{N}{\Omega} \sqrt{2\theta V_K} \right) \right]. \tag{II.25}$$

From Eq. (II.25), we can now compute Helmholtz free energy for the system,

$$A(\Omega, \theta) = -KT \log Q + \text{free kinetic energy term.} \tag{II.26}$$

We can now write down expressions for certain thermodynamic properties of such a system. For example, pressure P will be given by

$$P = -\left(\frac{\partial A}{\partial \Omega} \right)_T, \tag{II.27}$$

$$P = \frac{N^2}{2\Omega^2} \tilde{V}_0 + \frac{N}{2\theta\Omega}$$
$$+ \frac{N}{2\theta\Omega^2} \sum_{K \neq 0} \frac{\sqrt{2\theta \tilde{V}_K}}{1 + (N/\Omega)(2\theta/\sqrt{\tilde{V}_K})},$$

and entropy S of the system will be of the form

$$S = -\left(\frac{\partial A}{\partial \theta} \right)_\Omega$$
$$= \frac{1}{2\theta^2} \left\{ \sum_{K \neq 0} \left[\log\left(1 + \frac{N}{\Omega} \sqrt{2\theta \tilde{V}_K} \right) \right.\right.$$
$$\left.\left. - \frac{2\theta (N/\Omega) \tilde{V}_K}{\sqrt{2\theta \tilde{V}_K} (1 + (N/\Omega)\sqrt{2\theta \tilde{V}_K})} \right] - \log \Omega^N \right\}. \tag{II.28}$$

From the knowledge of the free energy of the system as given by Eq. (II.26), one is, in principle, able to calculate all the thermodynamic properties of the system and compare the values with experimental results.

B. Change of variables

Here we shall use the relevant Gaussian quadrature formulas.[17] Let us introduce new variables:

$$\xi_K^{\pm} = \frac{\alpha_K \pm \alpha_{-K}}{\sqrt{2}}, \quad \eta^{\pm} = \frac{\beta_K \pm \beta_{-K}}{\sqrt{2}}.$$

The Jacobian of transformation will evidently take the form

$$\begin{vmatrix} \frac{\partial \xi_K^+}{\partial \alpha_K} = \frac{1}{\sqrt{2}} & \frac{\partial \xi_K^+}{\partial \alpha_{-K}} = \frac{1}{\sqrt{2}} \\ \frac{\partial \xi_K^-}{\partial \alpha_K} = \frac{1}{\sqrt{2}} & \frac{\partial \xi_K^-}{\partial \alpha_{-K}} = -\frac{1}{\sqrt{2}} \end{vmatrix} = -1. \tag{II.29}$$

In terms of the new variables, the Gaussian averaging becomes

$$\epsilon \langle \cdots \rangle$$
$$= \int \prod_{K \neq 0} d\alpha_K d\beta_K \exp\left[-\sum_{K \neq 0} (\alpha_K^2 + \beta_K^2 + \log \pi) \right] \langle \cdots \rangle$$
$$= \int \prod_{K \neq 0} d\xi_K^+ d\xi_K^- d\eta_K^+ d\eta_K^- \langle \cdots \rangle$$
$$\times \exp\left[-\sum_{K \neq 0} (\xi_K^{2+} + \xi_K^{2-} + \eta_K^{2+} + \eta_K^{2-} + 2\log \pi) \right], \tag{II.30}$$

where

$$\langle \cdots \rangle = \exp\left(\frac{N}{\Omega} \int_\Omega d^3x (e^{iF(x)} - 1) \right) \tag{II.31}$$

and

$$F(x) = 2 \sum_{K \neq 0} \lambda_K (\xi_K^+ \cos Kx + \eta_K^- \sin Kx).$$

We can thus get rid of ξ_K^- and η_K^+ by simply integrating

Eq. (II.30) and obtaining the following,

$$\epsilon\langle\cdots\rangle = \int \sum_{K\neq 0} d\xi_K^+ d\eta_K^- \langle\cdots\rangle$$

$$\times \exp\Big[- \sum_{K\neq 0} (\xi_K^2 + \eta_K^2 + 2\log\pi)\Big]. \qquad \text{(II.32)}$$

In an external field $U(x)$, the partition function Eq. (II.21) will take the form below,

$$Q = Q_0 \Omega^N \exp\Big(\frac{N}{\Omega}\theta V_0\Big)$$

$$\times \epsilon\Big\langle \exp\Big[\frac{N}{\Omega}\int d^3x (e^{iF(x)} e^{-\theta U(x)} - 1)\Big]\Big\rangle. \qquad \text{(II.33)}$$

C. Formal relations for external field response

By retracing the steps outlined earlier, we could write an expression for the partition function in the presence of an external potential $U(x)$ in the form

$$\frac{Q}{Q_0\Omega^N} = \exp\Big(-\frac{N}{\Omega}U_0\Big)$$

$$\times \epsilon\Big\langle \exp\Big[\frac{N}{\Omega}\int d^3x (e^{iF(x)} e^{-\theta U(x)} - 1)\Big]\Big\rangle. \qquad \text{(II.34)}$$

On the other hand,

$$Q = \int\cdots\int dx_1\cdots dx_n \exp\Big(-\frac{\theta}{2}\sum_{i\neq j} V_{ij}\Big)$$

$$\times \exp\Big[-\theta\int U(x)\sum_{i=1}^{N}\delta(x_i - x)\, d^3x\Big]. \qquad \text{(II.35)}$$

Now considering Q as a functional of $U(x)$ and noting that the local density is

$$n(x) = \sum_{i=1}^{N}\delta(x_i - x), \qquad \text{(II.36)}$$

then the variation Q with respect to $U(x)$ is

$$\frac{1}{Q}\frac{\delta Q}{\delta U(s)} = -\theta\,\overline{n(s)}, \qquad \text{(II.37)}$$

$$\frac{1}{Q}\frac{\delta^2 Q}{\delta U(s)\delta U(s')} = \theta^2\overline{n(s)n(s')}. \qquad \text{(II.38)}$$

But since

$$\overline{n(s)n(s')} = \overline{(\bar{n}(s) + \delta n)(\bar{n}(s') + \delta n)}$$

$$= \overline{n(s)}\;\overline{n(s')} + \overline{\delta n(s)\delta n(s')}, \qquad \text{(II.39)}$$

it follows then from Eq. (II.37), Eq. (II.38), and Eq. (II.39) that

$$\theta^2\overline{\delta n(s)\delta n(s')} = -\theta\frac{\delta\overline{n(s')}}{\delta U(s)}. \qquad \text{(II.40)}$$

We recognize in Eq. (II.40) that $\overline{n(s')}$ is a functional of $U(s)$.

Next we consider Q as a functional of $\overline{n(s)}$ expecting this to be more correlated with it. Thus

$$\int \frac{\delta\bar{n}(s)}{\delta U(s')}\frac{\delta U(s')}{\delta\bar{n}(s')}\,ds' = \delta(s - s'') \qquad \text{(II.41)}$$

and

$$\frac{1}{Q}\frac{\delta Q}{\delta\bar{n}(s)} = -\theta\int \bar{n}(s'')\frac{\delta U(s'')}{\delta\bar{n}(s)}\,ds''. \qquad \text{(II.42)}$$

Expressions for higher order and mixed derivatives can be found if desired.

D. Separation of short and long range interactions

This paper sets out on an attempt to separate long and short range parts of interaction potential $V(xi - xj)$. In doing so we believe that it is necessary to define the regime and limits of the short range part of the potential. The region of long range forces includes certain aspects of short range interactions. So by determining the short range region, we aim at renormalizing the long range part of the interaction potential. Our Gaussian average denoted by ϵ was performed in reciprocal K space corresponding to the Fourier transform of function $F(x)$, see Eq. (II.23).

Now assuming that short range forces act in K space for values of $K \geqslant K_0$, then we could draw a sharp distiction between $K > K_0$ and $K < K_0$ the latter corresponding to purely long range interactions.

With the above in mind, let us denote by ϵ_L the average carried out over α_K and β_K for $K < K_0$ corresponding to long range forces. Then we write

$$\epsilon^W = \epsilon\Big\langle \exp\Big(\frac{N}{\Omega}\int d^3x (e^{iF(x)} - 1)\Big)\Big\rangle$$

$$= \epsilon\Big\langle \exp\Big(\frac{N}{\Omega}\int d^3x (e^{iF_l(x)} - 1)\Big)$$

$$\times \exp\Big(\varepsilon\frac{N}{\Omega}\int e^{iF_l(x)}(e^{iF_s(x)} - 1) d^3x\Big)\Big\rangle. \qquad \text{(II.43)}$$

We have inserted ε which is later set equal to 1. Here

$$F(x) = F_L(x) + F_S(x),$$

representing long $F_L(x)$ and short $F_S(x)$ range forces, respectively.

Let us now concentrate on the short range part of the potential by taking the average over α_K and β_K denoted by ϵ_S for $K > K_0$. Equation (II.43) can be written in the form

$$e^W = \epsilon_L\Big\langle \exp\Big(\frac{N}{\Omega}\int d^3x [e^{iF_l(x)} - 1] e^A\Big)\Big\rangle \qquad \text{(II.44)}$$

with the short range average

$$e^{A} = \epsilon_S \left\langle \exp\left(\varepsilon \frac{N}{\Omega} \int e^{iF_L(x)} (e^{iF_S(x)} - 1)\, d^3x \right) \right\rangle. \quad \text{(II.45)}$$

Further we consider e^A and treat it by a virial development in ε in order to determine the contributions from different terms, linear etc. The $F_L(x)$ will remain and e^A is a function of α_K and β_K for $K < K_0$.

Develop

$$A = A_0 + \varepsilon \left.\frac{dA}{d\varepsilon}\right|_{\varepsilon=0} + \frac{\varepsilon^2}{2!} \left.\frac{d^2A}{d\varepsilon^2}\right|_{\varepsilon=0} + \cdots. \quad \text{(II.46)}$$

Here the constant term $A_0 = 0$ corresponds to a system with purely long range forces.

Considering the linear term in the expansion given by Eq. (II.46),

$$\left.\frac{dA}{d\varepsilon}\right|_{\varepsilon_0} = \frac{N}{\Omega} \mathscr{H}(K_0) \int e^{iF_L(x)}\, d^3x. \quad \text{(II.47)}$$

$$\mathscr{H}(K_0) = \epsilon_S \langle e^{iF_S(x)} - 1 \rangle$$

$$= \left\langle \exp\left(- \frac{1}{\Omega} \sum_{K \geq K_0} \frac{|\Phi(K)|^2}{4} - 1 \right) \right\rangle. \quad \text{(II.48)}$$

Equation (II.48) describes a system with purely short range forces. Since we will develop $\exp[iF_L(x)]$ up to $F_L^2(x)$ we see that the short range forces induce a renormalization of the quadratic form of the long range part.

In the quadratic approximation to the short range force, we need the following term,

$$\left.\frac{d^2A}{d\varepsilon^2}\right|_{\varepsilon_0} = \frac{N^2}{\Omega^2} \int d^3x\, d^3y \exp\{i[F_L(x) + F_L(y)]\} Z(y - x), \quad \text{(II.49)}$$

where

$$Z(y - x) = \exp\left[- \frac{1}{\Omega} \sum_{K \geq K_0} \frac{|\Phi_K|^2}{2} \right] \times \left[\exp\left(- \frac{1}{\Omega} \sum_{K \geq K_0} \frac{|\Phi_K|^2}{2} \cos K(y - x) \right) - 1 \right]. \quad \text{(II.50)}$$

Thus to this order

$$A = \varepsilon \frac{N}{\Omega} \mathscr{H}(K_0) \int e^{iF_L(x)} d^3x + \frac{\varepsilon}{2!} \frac{N^2}{\Omega^2} \int d^3x\, d^3y \times \exp\{i[F_L(x) + F_L(y)]\} Z(y - x). \quad \text{(II.51)}$$

Let us now go back and reexamine Eq. (II.44) by expanding it in powers of $F_L(x)$ up to quadratic term. Thus

$$W = W_0 + W_1 + W_2.$$

W_0 is the constant part of the exponential and corresponds to a system with purely short range interactions,

$$W_0 = \varepsilon \frac{N}{\Omega} \mathscr{H}(K_0) + \frac{\varepsilon^2}{2!} \frac{N^2}{\Omega} \int Z(y)\, dy. \quad \text{(II.52)}$$

The linear term $W_1 = 0$, simply because

$$\int F_L(x)\, d^3x = 0,$$

the absence of long range forces in the region $K > K_0$, i.e.,

$$W_1 = \frac{\varepsilon^2}{2!} \frac{N^2}{\Omega^2} 2i \times \int F_L(x) Z(|y - x|)\, d^3x\, d^3y = 0. \quad \text{(II.53)}$$

W_2 is the quadratic part and has the form

$$W_2 = - \frac{N}{2\Omega} \int F_L^2(x)\, d^3x \{1 + \varepsilon \mathscr{H}(K_0)\} + \frac{\varepsilon^2}{2!} \frac{N^2}{\Omega^2} \int d^3x\, d^3y [F_L(x) + F_L(y)]^2 Z(y - x). \quad \text{(II.54)}$$

The obvious physical intepretation of the quadratic term Eq. (II.54) is that a system with long range forces has a certain amount of short range interactions.

We could go on with this scheme and obtain higher order terms in W. However from what we have done above, the trend looks pretty obvious and in fact we already know what terms to choose from our expansions in order to deal with a system with purely short range or long range forces. Our next task would be to reexpress the partition function Q in terms of the short range and long range Gaussian averages.

Thus thermodynamic properties of a system with purely short range forces can be computed from the following partition function,

$$Q = Q_0 e^{W_0}, \quad \text{(II.55)}$$

where W_0 is given by Eq. (II.52). And similarly from

$$Q = Q_0 e^{W_2}, \quad \text{(II.56)}$$

for systems exhibiting long range interactions. W_2 is given by Eq. (II.54).

III. THE GRAND PARTITION FUNCTION

A. General remarks

It seems easier to make contact with the standard treatment (e.g., virial series) via the grand partition function. The grand partition function $\Xi(\theta,\Omega,z)$ is defined by

$$\Xi(\theta,\Omega,z) = \sum_{N>0} Q_N(\Omega,\theta) \frac{z^N}{N!}. \quad \text{(III.1)}$$

Here

$$Q_N(\Omega,\theta) = \int_{\Omega} \cdots \int dx_1 \cdots dx_N e^{-\theta V} \quad \text{(III.2)}$$

and

$$z = \frac{e^{-\theta\mu}}{\Lambda^3}, \quad \Lambda = \left(\frac{h^2\theta}{2\pi m}\right)^{1/2}.$$

We go back to the beginning and express Q_N in terms of Gaussian quadratures. However, to include the $\tilde{V}(0)$ and $V(0)$ terms, we use variables α_0 and β_0. Extend $F(x)$ to include $K = 0$ terms and let $\epsilon\langle\cdots\rangle$ involve Gaussian averages over α_0 and β_0 as well. Then

$$Q_N = \epsilon\left\langle\left[\int d^3x e^{iF(x)}\right]^N\right\rangle\{e^{\theta V(0)/2}e^{\theta\tilde{V}(0)/2\Omega}\}^N \quad \text{(III.3)}$$

such that

$$\Xi(\theta,\Omega,z) = \epsilon\left\langle\exp\left(z_1\int d^3x e^{iF(x)}\right)\right\rangle \quad \text{(III.4)}$$

with

$$z_1 = z\exp\left(\frac{\theta}{2}V_{x=0}\right)\exp\left(\frac{\theta}{2\Omega}\tilde{V}_{K=0}\right)$$

and we expect the term $(\theta/2\Omega)V_K = 0$ to vanish since it comes from $N(N-1)/2$.

In what follows, we drop that term and Eq. (III.3) will take the form

$$Q_N = \epsilon\left\langle\exp\left(\frac{\theta}{2}V_{x=0}\right)\int d^3x e^{iF(x)}\right\rangle.$$

The ordinary virial development gives pressure in the form

$$p\theta = \frac{1}{\Omega}\log\Xi = b_1 z + b_2 z^2 + b^3 z^3 + \cdots, \quad \text{(III.5)}$$

where $b_1, b_2, \cdots$ are the virial coefficients. It follows from Eq. (III.1) above and Eq. (II.43) that

$$\frac{\omega}{\Omega} = \frac{1}{\Omega}\log\Xi$$

$$= z + \frac{z^2}{2!}\int f(x)\,d^3x + \cdots. \quad \text{(III.6)}$$

Here

$$f(x) = \{e^{-\theta V(x)} - 1\}$$

and z is determined from $\bar{N}/\Omega$.

It is well known from statistical thermodynamics that

$$\bar{N} = \theta\left(\frac{\partial\log\Xi}{\partial\mu}\right)_{\theta,\Omega}$$

$$= z\left(\frac{\partial\log\Xi}{\partial z}\right)_{\theta,\Omega}. \quad \text{(III.7)}$$

B. Separation of long range forces in the grand partition function

In this subsection we shall try to compare results obtainable in Sec. II by separating long and short range interactions in the grand partition function. From Sec. III A we have

$$\Xi = \epsilon\left\langle\exp\left\{z\int d^3x\exp[iF(x)]\exp\left[\frac{\theta}{2}V(0)\right]\right\}\right\rangle \quad \text{(III.8)}$$

and putting

$$F = F_S + F_L,$$
$$R_S = \exp[\theta V_S(0)/2],$$
$$R_L = \exp[\theta V_L(0)/2],$$

we can show that

$$\Xi = \epsilon_L\left\langle\exp\left(z\int e^{iF_L(x)}R_L d^3x\right)\right\rangle$$

$$\times\epsilon_S\left\langle\exp\left[z\int R_L e^{iF_l}(e^{iF_S}R_S - 1)d^3x\right]\right\rangle. \quad \text{(III.9)}$$

Next we perform the cumulant expansion which is simple since

$$\epsilon_S\langle e^{iF_S(x)}R_S - 1\rangle = 0,$$

for all x. Thus

$$\epsilon_S\left\langle\exp\left[z\int e^{iF_l}R_L(e^{iF_S}R_S - 1)d^3x\right]\right\rangle$$

$$\equiv \frac{z^2}{2!}X_2 + \frac{z^3}{3!}X_3$$

$$= \exp\left(\frac{z^2}{2!}X_2 + \frac{z^3}{3!}X_3\right) \quad \text{(III.10)}$$

with

$$X_2 = \int\int dx\,dy\,R_L^2\exp\{i[F_L(x) + F_L(y)]\}f(x-y), \quad \text{(III.11)}$$

$$X_3 = \int\int\int dx\,dy\,dz\,R_L^3\exp\{i[F_L(x) + F_L(y) + F_L(z)]\}$$
$$\times[f(x-y)f(y-z)f(z-x) + 3f(x-y)f(y-z)].$$

Here

$$f(r) = e^{-\theta V(r)} - 1$$

refers to the short range part of the interaction potential.

Our next task is to carry the development of the long range part

$$e^{iF_l(x)}R_L,$$

and briefly mention the physics they represent.

As in Sec. II, the constant part corresponds to a system with purely short range forces

$$W_S = z\Omega R_L + \frac{z^2}{2!}\int f(x)R_L^2\,d^3x$$
$$\times\frac{z^3}{3!}R_L^3\int\int\int dx\,dy\,dz\,(f_{12}f_{23}f_{31} + 3f_{12}f_{23}). \quad \text{(III.12)}$$

For the time being we have kept R_L unexpanded.

We therefore have

$$\frac{z^2}{2!}X_2 = \frac{z^2}{2!}R_L\left\{\Omega\tilde{f}(0) + 2i\tilde{f}(0)\int F_L(x)\,d^3x - \int\int F_L(x)[\tilde{f}(0)\delta(x-y) + f(x-y)]F_L(y)\,dx\,dy\right\} \tag{III.13}$$

for the short range forces, as can be seen in Eq. (III.10), and we could also write an expression for the next term $(z^3/3!)X_3$, which is rather long.

The linear part W_1 all refers to the α_0 and β_0 modes and can be treated separately.

The quadratic part of the expansion admittedly contains a lot of algebraic manipulations and we are not going to reproduce that here. We just mention that with

$$F_L(x) = \frac{1}{\Omega}\sum_K \tilde{F}_L(K)e^{-iKx}$$

we would split off the $K = 0$ term and the result would be essentially the same result we obtained in Sec. II. Namely, the long range forces involve some amount of short range interactions. So the contribution to the grand partition function from terms $K \neq 0$ are

$$\exp\left(-\tfrac{1}{2}\sum_{\substack{K\neq 0\\ K<K_0}} \log[1 + \theta\tilde{V}_K\{A + 2\tilde{B}(K)\}]\right). \tag{III.14}$$

Here

$$\tilde{B}(K) = \int_\Omega B(x)e^{iKx}\,d^3x$$

and

$$A = \frac{z^2}{2}R_L^2 + \frac{z^2}{2}R_L^2\tilde{f}(0) + \frac{z^3}{4}R_L^3\left\{3\tilde{f}^2(0) + \frac{1}{\Omega}\sum_m \tilde{f}(m)\tilde{f}(-m)\tilde{f}(m)\right\}. \tag{III.15}$$

Now turning to the zero mode, the quadratic part is proportional to $1/\Omega$ and does not contribute. The linear part requires completing the square and gives

$$\exp\left(-\Omega\frac{\theta}{2}\tilde{V}(0)\{1 + zR_L\tilde{f}(0) + \frac{z^2}{2}R_L^2[f_3 + 3\tilde{f}^2(0)]\right\}^2 z^2R_L^2\Big). \tag{III.16}$$

Lastly, putting all these approximations together, we come to the following virial formula,

$$\frac{1}{\Omega}\log\Xi = zR_L\left(1 + \frac{zR_L}{2}\int f(x)\,d^3x + zR_L^2 f_3\right) - \frac{1}{2}\frac{1}{\Omega}\sum_{\substack{K\neq 0\\ K<K_0}} \log\{1 + 2\theta\tilde{V}(K)[A + \tilde{B}(K)]\} - \frac{\theta}{2}\tilde{V}(0)\left(1 + zR_L\tilde{f}(0) + \frac{z^2}{2}R_L^2[f_3 + 3\tilde{f}(0)]\right)^2 \times z^2R_L^2, \tag{III.17}$$

which in effect gives the pressure of the system.

We have thus shown that for both the classical partition and the grand partition function, the constant term of expansion W_0 of Gaussian averages describes a system with purely short range forces while the quadratic term W_2 corresponds to a system with a mixture of short and long range interactions which is to be expected.

ACKNOWLEDGMENTS

The author wishes to sincerely thank the following: Professor E.P. Gross for having drawn his attention to the problem, certain members of the Quantum Chemistry Group, Uppsala University for stimulating discussions, the referee whose comments were very helpful in revising this paper, and finally the Swedish Institute whose generous grant made my visit to Sweden possible during my sabbatical leave.

APPENDIX: THE USE OF THE GRAND PARTITION FUNCTION

Let us consider a system with the following coordinate independent Hamiltonian,

$$\mathscr{H} = \sum_{i=1}^N \frac{P_i^2}{2m} + \frac{N(N-1)^2}{2\Omega}\tilde{V}(0)\varepsilon^2 - \frac{N}{2}V(0)\eta^2 \tag{A1}$$

The coefficient ε and η will be set equal to unity later. The classical partition function for such a system will take the form

$$Z_N = \left(\frac{m}{2\pi KT}\right)^{3N/2}\int dx_1\cdots dx_N e^{-\theta V} = \left(\frac{m}{2\pi KT}\right)^{3N/2}\Omega^N \times \exp\left(-\frac{\theta N(N-1)}{2\Omega}\tilde{V}(0)\varepsilon^2 + \frac{\theta N}{2}V(0)\eta^2\right). \tag{A2}$$

Thus the free energy for the canonical ensemble is given by

$$A = \frac{N}{\theta}\left(\log\frac{N}{\Omega}\Lambda^3 - 1\right) - \frac{1}{\theta}\log\left(\frac{Z_N}{\Omega^N}\right), \tag{A3}$$

and therefore its variation is given by

$$\delta A = \frac{N(N-1)}{2\Omega}\tilde{V}(0) + \frac{N}{2}V(0). \tag{A4}$$

Pressure for such a system will be

$$P = \frac{N}{\Omega}\frac{1}{\theta} + \frac{1}{\theta}\left[\frac{\partial}{\partial\Omega}\log\left(\frac{Z_N}{\Omega^N}\right)\right]_{\theta,N} \tag{A5}$$

and its variation is

$$\delta_P = \frac{N(N-1)}{2\Omega^2}\tilde{V}(0). \tag{A6}$$

The chemical potential for the system is

$$\mu = \frac{1}{\theta}\log\left(\frac{N\Lambda^3}{\Omega}\right) - \frac{1}{\theta}\left[\frac{\partial}{\partial N}\log\left(\frac{Z_N}{\Omega^N}\right)\right] \tag{A7}$$

with a variation of

$$\delta\mu = \frac{N-\frac{1}{2}}{\Omega}\tilde{V}(0) - \tfrac{1}{2}V(0). \tag{A8}$$

The grand partition function for such a system is of the form

$$\Xi(\Omega,\theta,z) = \sum_{N\geq 0} Z_N(\Omega,\theta)z^N/N! \tag{A9}$$

and it is clear from the above that

$$Z_N z^N = \exp\left(-\frac{\theta}{2\Omega}\tilde{V}(0)\varepsilon^2N^2\right)[\Omega,z_1]^N \tag{A10}$$

with

$$z_1 = z\exp\left(\frac{\theta}{2\Omega}\tilde{V}(0)\varepsilon\right)\exp\left(\frac{\theta}{2}V(0)\eta^2\right). \tag{A11}$$

Using Gaussian quadratures we can write the following,

$$\exp\left(-\frac{\theta}{2\Omega}\tilde{V}(0)\varepsilon^2N^2\right) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp(-\alpha_0^2)$$

$$\times\exp\left[i\alpha_0\sqrt{2}\left(\theta\frac{\tilde{V}(0)}{\Omega}\varepsilon^2\right)^{1/2}N\right]d\alpha_0 \tag{A12}$$

such that

$$Z_N z^N = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\alpha_0^2}$$

$$\times\left\{\Omega z_1\exp\left[i\alpha_0\left(2\theta\frac{\tilde{V}(0)}{\Omega}\varepsilon^2\right)^{1/2}\right]\right\}^N d\alpha_0. \tag{A13}$$

Thus the grand partition function will be of the form

$$\Xi = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\alpha_0^2}d\alpha_0\exp\left[\Omega z_1\exp\left(i\alpha_0\frac{\tilde{\Phi}}{\sqrt{\pi}}\right)\right], \tag{A14}$$

where

$$\tilde{\Phi}(0) = [2\theta\tilde{V}(0)\varepsilon^2]^{1/2}.$$

It follows readily from Eq. (A14) that the grand partition function satisfies the following functional equation,

$$\frac{d\Xi(z_1)}{dz_1} = \Omega\exp\left(\frac{-\tilde{\Phi}^2}{4\Omega}\right)\Xi\left[z_1\exp\left(-\frac{\tilde{\Phi}^2}{2\Omega}\right)\right] \tag{A15}$$

with

$$\Xi(0) = 1.$$

Now to show how the grand partition works, we use the maximum term,

$$\frac{Z_{\bar{N}}z^{\bar{N}}}{\bar{N}!}.$$

On applying Stirling's approximation, we find that

$$\log\left(\frac{\Omega z_1}{\bar{N}}\right) = \frac{\bar{N}}{\Omega}\frac{\tilde{\Phi}^2}{2} \tag{A16}$$

and this yields

$$z_1(\bar{N}) = \frac{\bar{N}}{\Omega}\exp\left(\frac{\bar{N}}{\Omega}\frac{\tilde{\Phi}^2}{2}\right)$$

$$\Rightarrow\exp\left(-\frac{\theta}{2}\tilde{V}(0)\right)\frac{\bar{N}}{\Omega}\exp\left(\frac{\bar{N}}{\Omega}\frac{\tilde{\Phi}^2}{2}\right).$$

Thus the value of Ξ is

$$\Xi = \exp\left(\frac{\bar{N}_2}{\Omega}\frac{\tilde{\Phi}^2}{4}\right)e^{\bar{N}}. \tag{A17}$$

The equation

$$\bar{N} = z\left(\frac{\partial\log\Xi}{\partial z}\right)$$

is an identity and

$$\frac{\log\Xi}{\Omega} = \frac{\bar{N}^2}{\Omega^2}\frac{\tilde{\Phi}^2}{4} + \frac{\bar{N}}{\Omega}. \tag{A18}$$

[1] D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953); S.F. Edwards, Phil. Mag. **3**, 1119 (1958).
[2] E. Montroll and J. Ward, Phys. Fluids **1**, 55 (1958).
[3] L. Landau and E.M. Lifshitz, *Statistical Physics* (Pergamon, New York, 1958).
[4] A.A. Abrikosov, L.P. Gorkov, and I.E. Dyaloshinski, J. Exp. Theor. Phys. **36**, 900 (1959).
[5] P. Debye, *Polar Molecules* (Dover, New York, 1929).
[6] J.M. Ziman, *The Principles of the Theory of Solids* (Cambridge U.P., Cambridge, 1964).
[7] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[8] R. Brout, *Phase Transitions* (Benjamin, New York, 1965).
[9] H. Pauly and J.P. Toennies, Adv. Atom. Mol. Phys. **1**, 195 (1965); Method Exp. Phys. A **7**, 227 (1968).
[10] H.G. Benewitz, R. Haerten, O. Klais, and G. Muller, Z. Phys. **249**, 168 (1971).
[11] S.M. Freund, G.A. Fisk, D.R. Herschbach, and W. Klemperer, J. Chem. Phys. **54**, 2510 (1971).
[12] E. Ising, Z. Phys. **31**, 253 (1925).
[13] P.W. Langhoff and M. Karplus, *The Padé Aproximations in Theoretical Physics* (Academic, New York, 1970).
[14] A. Dalgarno, Adv. Chem. Phys. **12**, 143 (1967).
[15] A.A. Abrahamson, Phys. Rev. **178**, 76 (1969).
[16] J.R. Shrieffer, Summer School, University of Pennyslvania (1969).
[17] E. Montroll, Commun. Pure Appl. Math. **5**, 437 (1952).

S-wave off-shell T and K matrices for the Yukawa potential by Ecker–Weizel approximations[a)]

U. Das, S. Chakravarty,[b)] and B. Talukdar

Department of Physics, Visva-Bharati University, Santiniketan 731235 West Bengal, India
(Received 12 April 1978)

Expressions for s-state off-shell wavefunctions associated with outgoing and standing wave boundary conditions are derived for the Yukawa potential by using Ecker–Weizel approximations. The results are used to relate the T and K matrix elements to tabulated transcendental functions.

I. INTRODUCTION

In a recent paper[1] (hereafter cited as paper I) two of us have obtained s-state eigenfunctions of the Yukawa Hamiltonian associated with Jost and regular boundary conditions. We have accomplished this by solving the relevant Schrödinger equation using Ecker–Weizel approximations.[2] The purpose of the present paper is to show how, using the approximations of paper I, it is possible to derive the off-energy-shell results for scattering by a Yukawa potential. We formulated the problem in terms of van Leeuwan–Reiner approach[3] to off-shell scattering as used by Fuda and Whiting[4] and by our group.[5] In this approach expressions for off-shell T and K matrices are derived by using off-shell wavefunctions associated with outgoing wave and standing wave boundary conditions. Both wavefunctions are expressed in terms of Jost solutions and Jost functions. In addition to the ordinary Jost function there also appears an off-shell Jost function. In close analogy with the ordinary Jost functions the off-shell Jost function is determined by the behavior of the off-shell wavefunction irregular at the origin. We consider the s-wave scattering of a particle by a central potential $V(r)$. Let $\mathbf{k}$ denote the on-shell momentum related to the energy by $E = k^2 + i\epsilon$, $\epsilon \ll 1$, and $\mathbf{q}$, an off-shell momentum. The radial van Leeuwen–Reiner equation for the Jost solution is given by[4]

$$\left(\frac{d^2}{dr^2} + k^2 - V(r)\right) f(k,q,r) = (k^2 - q^2)\exp(iqr). \tag{1}$$

In writing Eq. (1) we omit, for brevity, the subscript $l = 0$ and use units in which $\hbar^2/2m$ is unity. For this case the off-shell Jost functions can be obtained as

$$f(k \pm q) = f(k, \pm q, 0). \tag{2}$$

The ordinary Jost solutions and Jost functions are related to off-shell ones by

$$f(\pm k, r) = \lim_{q \to k} f(k, \pm q, r), \tag{3a}$$

and

$$f(\pm k) = \lim_{q \to k} f(k, \pm q). \tag{3b}$$

The off-shell wavefunctions $\Psi^+(k,q,r)$ and $\Psi^{(-)}(k,q,r)$ satisfying outgoing and standing wave boundary conditions are given by the following expressions:

$$\Psi^+(k,q,r) = -\tfrac{1}{2}\pi q T(k,q,k^2) f(k,r) + \frac{1}{2i}[f(k,q,r) - f(k,-q,r)], \tag{4a}$$

and

$$\Psi^{(-)}(k,q,r) = -\tfrac{1}{4}\pi q K(k,q,k^2)[f(k,r) + f(-k,r)] + \frac{1}{2i}[f(k,q,r) - f(k,-q,r)]. \tag{4b}$$

In Eqs. (4a) and (4b), $T(k,q,k^2)$ and $K(k,q,k^2)$ represent the half-off-shell T and K matrices. We have

$$T(k,q,k^2) = \frac{f(k,q) - f(k,-q)}{i\pi q f(k)}, \tag{5a}$$

and

$$K(k,q,k^2) = \frac{2}{i\pi q} \cdot \frac{f(k,q) - f(k,-q)}{f(k) + f(-k)} = \frac{2}{\pi q} \cdot \frac{\mathrm{Im} f(kq)}{|f(k)|\cos\delta(k)}. \tag{5b}$$

Here $\delta(k)$ stands for the negative of the phase of the Jost function. Obviously, it is the phase shift induced by the potential.

The Yukawa potential behaves like $1/r$ near the origin. Off-shell T and K matrices for such a potential can be calculated by using the relations[6,7]

$$T(p,q,k^2) = T(k,q,k^2) + \frac{2}{\pi}\frac{k^2 - p^2}{pq}\int_0^\infty dr \sin pr \times [\Psi^+(k,q,r) - \sin qr + \tfrac{1}{2}\pi q T(k,q,k^2)\exp(ikr)], \tag{6}$$

and

$$K(p,q,k^2) = K(k,q,k^2) + \frac{2}{\pi}\frac{k^2 - p^2}{pq}\int_0^\infty dr \sin pr \times [\Psi^{(-)}(k,q,r) - \sin qr + \tfrac{1}{2}\pi q K(k,q,k^2)\cos kr]. \tag{6b}$$

[a)]Partially supported by the Department of Atomic Energy, Government of India.

[b)]One of the authors (SC) acknowledges the receipt of a Junior Research Fellowship of the University Grants Commission, Goverment of India.

 0022-2488/79/050887-04$01.00

Equations (6a) and (6b) represent the basic formulas for computing off-shell T and K matrices for potentials singular at the origin. These equations do not involve the potential explicitly.

In Sec. II we solve Eq. (1) for the Yukawa potential using Ecker–Weizel approximations and write down the expression for $f(k,q)$. In Sec. III we relate the Yukawa T and K matrices to tabulated transcendental functions. We conclude by noting that the approximate results presented in this paper satisfy the usual relations between T and K matrices as derived by Kouri and Levin.[8]

II. OFF-SHELL JOST FUNCTIONS

For the attractive Yukawa potential $-V_0 \exp(-\mu r)/r$, Eq. (1) reads

$$\left[\frac{d^2}{dr^2} + k^2 + \frac{V_0 \exp(-\mu r)}{r}\right] f(k,q,r) = (k^2 - q^2)\exp(iqr). \tag{7}$$

The standard substitutions

$$f(k,q,r) = \exp(ikr)v(r) \tag{8a}$$

and

$$\mu r = -\ln(1-x) \tag{8b}$$

transform Eq. (7) in the form

$$\left[x(1-x)\frac{d^2}{dx^2} - x\left(1 - \frac{2ik}{\mu}\right)\frac{d}{dx} - \frac{V_0 x}{\mu \ln(1-x)}\right] v(x) = \frac{(k^2-q^2)}{\mu^2} x(1-x)^{i(k-q)/\mu - 1}. \tag{9}$$

In the Ecker–Weizel approximation one proceeds by assuming that $V_0 x/\mu \ln(1-x)$ is a slowly varying function of x and to a first approximation it is a constant $-\gamma$. Recently Lam and Varshni[9] has discussed the rationale of this approximation for a considered quantum state of a fermion. Working within the framework of Ecker–Weizel model, we transform the independent variable in Eq. (9) by substituting $x = 1 - y$. We thus obtain

$$\left\{y(1-y)\frac{d^2}{dy^2} + \left[1 - \frac{2ik}{\mu} - \left(1 - \frac{2ik}{\mu}\right)y\right]\frac{d}{dy} + \gamma\right\} v(y) = \frac{(k^2-q^2)}{\mu^2}\left[y^{i(k-q)/\mu - 1} - y^{i(k-q)/\mu}\right]. \tag{10}$$

A particular solution of Eq. (10) is given by[10]

$$v(y) = \frac{(k^2-q^2)}{\mu^2}\left[f_\sigma(a,b;c;y) - f_{\sigma+1}(a,b;c;y)\right], \tag{11}$$

where

$$\sigma = \frac{i(k-q)}{\mu}, \tag{12}$$

$$a = -\frac{i}{\mu}\left[k - (k^2 - \gamma\mu^2)^{1/2}\right], \tag{13a}$$

$$b = -\frac{i}{\mu}\left[k + (k^2 - \gamma\mu^2)^{1/2}\right], \tag{13b}$$

and

$$c = 1 - \frac{2ik}{\mu}. \tag{13c}$$

The function $f_\sigma(a,b;c;y)$ is related to a generalized hypergeometric function by

$$f_\sigma(a,b;c;y) = \frac{y^\sigma}{\sigma(\sigma + c - 1)} \times {}_3F_2(1, \sigma + a, \sigma + b; \sigma + 1, \sigma + c; y). \tag{14}$$

This series converges when $|y| < 1$; it converges when $|y| = 1$ provided that $\mathrm{Re}(c - a - b) > 0$ which is true in our case.

From Eqs. (8a) and (11), the off-shell Jost solution is

$$f(k,q,r) = \frac{k^2 - q^2}{\mu^2}\exp(ikr)\left[f_\sigma(a,b;c;\exp(-\mu r)) - f_{\sigma+1}(a,b;c;\exp(-\mu r))\right]. \tag{15}$$

Making use of the recurrence relation[10]

$$(\sigma + a)(\sigma + b) f_{\sigma+1}(a,b;c;Z) = \sigma(\sigma + c - 1) f_\sigma(a,b;c;Z) - Z^\sigma, \tag{16}$$

Eq. (15) can be rewritten in a convenient form

$$f(k,q,r) = \exp(iqr) + ab\exp(ikr) f_{\sigma+1}(a,b;c;\exp(-\mu r)). \tag{17}$$

From Eqs. (14) and (17) it is easy to see that asymptotically $f(k,q,r) \sim \exp(iqr)$. In the on-energy-shell limit, Eq. (17) yields

$$f(k,r) = \exp(ikr)[1 + ab f_1(a,b;c;\exp(-\mu r))]. \tag{18}$$

This result agrees with our previous result in Paper I obtained by solving the relevant Schrödinger equation. The off-shell Jost function obtained from Eq. (17) is given by

$$f(k,q) = \frac{\Gamma(1+\sigma)\Gamma(c+\sigma)}{\Gamma(1+a+\sigma)\Gamma(1+b+\sigma)}. \tag{19a}$$

For the on-shell case one has $\sigma = 0$. Thus

$$f(k) = \frac{\Gamma(c)}{\Gamma(1+a)\Gamma(1+b)}. \tag{19b}$$

III. T AND K MATRICES

The off-shell Jost functions in Eqs. (19a) and (19b) can be used to write the half off-shell T and K matrices given by Eq. (5) in closed form in terms of Γ functions. In order to obtain the fully off-shell T and K matrices it is necessary to combine Eqs. (4a), (4b), (17) and (18). It is easy to show that

$$\Psi^+(k,q,r) - \sin qr + \tfrac{1}{2}\pi q T(k,q,k^2)\exp(ikr) = \frac{ab}{2i}\exp(ikr)\left[f_{\sigma+1}(a,b;c;\exp(-\mu r)) - f_{\sigma'+1}(a,b;c;\exp(-\mu r))\right] - \tfrac{1}{2}ab\pi q T(k,q,k^2) \times \exp(ikr) f_1(a,b;c;\exp(-\mu r)), \tag{20}$$

and that

$$\Psi^{(-)}(k,q,r) - \sin qr + \tfrac{1}{2}\pi q K(k,q,k^2)\cos kr$$
$$= \frac{ab}{2i}\exp(ikr)[f_{\sigma+1}(a,b;c;\exp(-\mu r))$$
$$- f_{\sigma'+1}(a,b;c;\exp(-\mu r))]$$
$$- \tfrac{1}{2}\pi q ab K(k,q,k^2) f_1(a,b;c;\exp(-\mu r))\cos kr. \quad (21)$$

Making use of Eqs. (20) and (21) in (6a) and (6b), we change the variable by substituting $y = \exp(-\mu r)$. We thus obtain off-shell T and K matrices as

$$T(p,q,k^2)$$
$$= T(k,q,k^2) - \frac{(k^2-p^2)ab}{2\pi\mu pq}\left\{\int_0^1 dy\, y^{\rho-1}[f_{\sigma+1}(a,b;c;y)\right.$$
$$- f_{\sigma'+1}(a,b;c;y) - i\pi q T(k,q,k^2) f_1(a,b;c;y)]$$
$$- \int_0^1 dy\, y^{\rho'-1}[f_{\sigma+1}(a,b;c;y) - f_{\sigma'+1}(a,b;c;y)$$
$$\left. - i\pi q T(k,q,k^2) f_1(a,b;c;y)]\right\}, \quad (22)$$

and

$$K(p,q,k^2)$$
$$= K(k,q,k^2) + \frac{(k^2-p^2)ab}{2\pi i\mu pq}\left\{\frac{\pi q}{2} K(k,q,k^2)\right.$$
$$\times\left[\int_0^1 y^{\rho''-1} f_1(a,b;c;y)\,dy + \int_0^1 y^{\rho'-1} f_1(a,b;c;y)\,dy\right.$$
$$\left. - \int_0^1 y^{\rho'''-1} f_1(a,b;c;y)\,dy - \int_0^1 y^{\rho-1} f_1(a,b;c;y)\,dy\right]$$
$$- i\left[\int_0^1 y^{\rho-1} f_{\sigma+1}(a,b;c;y)\,dy - \int_0^1 y^{\rho-1} f_{\sigma'+1}(a,b;c;y)\,dy\right.$$
$$\left. - \int_0^1 y^{\rho'-1} f_{\sigma+1}(a,b;c;y)\,dy + \int_0^1 y^{\rho'-1} f_{\sigma'+1}(a,b;c;y)\,dy\right]\Bigg\}, \quad (23)$$

with

$$\rho = -i(p+k)/\mu, \quad \rho' = i(p-k)/\mu,$$
$$\rho'' = -\rho, \quad \text{and} \quad \rho''' = -\rho'. \quad (24)$$

All the integrations in Eqs. (22) and (23) can be carried out by using the result

$$\int_0^1 dy\, y^{\alpha-1} f_\beta(a,b,c;y)$$
$$= [\beta(\beta+c-1)(\alpha+\beta)]^{-1}{}_4F_3(1,\beta+a,\beta+b,\alpha+\beta;$$
$$\beta+1,\beta+c,\alpha+\beta+1;1). \quad (25)$$

The final results for the Yukawa T and K matrices are given by

$$T(p,q,k^2)$$
$$= \frac{\gamma(k^2-p^2)}{2\pi\mu pq}\{i\pi q T(k,q,k^2)[X(-p,k) - X(p,k)]$$
$$+ [Y(p,q,k) - Y(p,-q,k) - Y(-p,q,k)$$
$$+ Y(-p,-q,k)]\}, \quad (26)$$

and

$$K(p,q,k^2) = \frac{\gamma(k^2-p^2)}{2\pi\mu pq}\left\{\frac{i\pi q}{2} K(k,q,k^2)[X(-p,-k)\right.$$
$$+ X(-p,k) - X(p,-k) - X(p,k)]$$
$$+ [Y(p,q,k) - Y(p,-q,k) - Y(-p,q,k)$$
$$\left. + Y(-p,-q,k)]\right\}, \quad (27)$$

where

$$X(p,k) = \frac{1}{ab\rho}\times{}_3F_2(a,b,\rho;c;1+\rho;1), \quad (28a)$$

and

$$Y(p,q,k) = [(\sigma+1)(\sigma+c)(\sigma+\rho+1)]^{-1}$$
$$\times{}_4F_3(1,\sigma+a+1,\sigma+b+1,\sigma+\rho+1;$$
$$\sigma+2,\sigma+c+1,\rho+\sigma+2;1). \quad (28b)$$

In writing Eqs. (26) and (27), we have also used

$${}_4F_3(1,1+a,1+b,1+c;2,1+e,1+f;Z)$$
$$= \left(\frac{ef}{abcZ}\right)[{}_3F_2(a,b,c;e,f;Z) - 1], \quad (29)$$

which follows directly from the infinite series representation of the generalized hypergeometric function.

In conclusion we note that the results presented in Eqs. (26) and (27) for the Yukawa T and K matrices are formally similar to those for the Hulthén potential treated in Refs. 6 and 7. In an interesting work Kouri and Levin[8] have obtained a relation connecting the K operator and the real part of T. This can be used as a check on complicated expressions one usually obtains for the off-shell T and K matrices for local potentials. Our approximate expressions in Eqs. (26) and (27) are seen to satisfy this relation. Thus there is no physical uncertainity in pursuing Ecker–Weizel approximations to the off-energy-shell region. The results presented in this paper will be useful as starting points of any perturbation-theoretic calculation.

[1]B. Talukdar, R.N. Chaudhuri, U. Das, and P. Banerjee, J. Math. Phys. **19**, 1654 (1978).
[2]G. Ecker and W. Weizel, Ann. Physik (Leipzig) **17**, 126 (1956).
[3]J.M.J. van Leeuwen and A.S. Reiner, Physica **27**, 99 (1961).
[4]M.G. Fuda and J.S. Whiting, Phys. Rev. C **8**, 1255 (1973).
[5]B. Talukdar, N. Mallick, and S. Mukhopadhyaya, Phys. Rev. C **15**, 1252 (1977); B. Talukdar and M.N. Sinha Ray, J. Math. Phys. **18**, 2097 (1977).
[6]O.P. Bahethi and M.G. Fuda, J. Math. Phys. **12**, 2076 (1971).
[7]B. Talukdar, N. Mallick, S.B. Sarkar, and L. Mondal, Phys. Rev. C **15**, 1257 (1977).
[8]D.J. Kouri and F.S. Levin, Phys. Lett. B **48**, 203 (1974).
[9]C.S. Lam and Y.P. Varshni, Phys. Lett. A **59**, 363 (1975).
[10]A.W. Babister, *Transcendental Functions Satisfying Non-homogeneous Linear Differential Equations* (Macmillan, New York, 1967), Chap. 6.

The classical limit of quantum nonspin systems

R. Gilmore

Physics Department, University of South Florida, Tampa, Florida 33620
(Received 28 August 1978)

The classical limit of operators X belonging to any compact Lie algebra $\mathfrak{g}$ is computed. If $X\in\mathfrak{g}$, the classical limit in the representation Γ^{Λ}, whose highest weight is Λ, is $\lim\Gamma^{\Lambda}(X/N)=\Sigma s_i g(\mathbf{f}_i,X,\Omega)$, where the limit is taken as $N\to\infty$, the sum runs from $i=1$ to $r=$ rank $\mathfrak{g}$, $\Lambda=\Sigma\mu_i\mathbf{f}_i,\mathbf{f}_i$ are the highest weights of the r fundamental representations of $\mathfrak{g}$, $s_i=\lim\mu_i/N$, and $g(\mathbf{f}_i,X,\Omega)$ is the expectation value of X with respect to the coherent states $|\mathbf{f}_i,\Omega\rangle$ in the representation $\Gamma^{\mathbf{f}_i}$. Examples and applications are given.

1. INTRODUCTION

The classical limit of quantum spin [SU(2)] systems has recently become a powerful tool for rigorously studying the ground state and thermodynamic critical properties of some physical systems. A rigorous justification for the use of the classical limit involves the use of atomic coherent states for SU(2)[1] to put lower and upper[2] bounds on the quantum partition function. These bounds are obtained by replacing the spin Hamiltonian by its Q and P representatives[1] (functions defined over the sphere surface) and the trace over operator-valued functions by an integral over the sphere. The Q and P representatives of all irreducible spherical tensor operators are also known.[3] In the limit of large N ("thermodynamic limit") the Q and P representatives per particle become equal, so all spin operators can be replaced by their classical limits

$$\lim_{N\to\infty}\Gamma^{J}[(X/N)^{K}]=\left[s\left\langle\begin{matrix}\frac{1}{2}\\ \theta\phi\end{matrix}\middle|X\middle|\begin{matrix}\frac{1}{2}\\ \theta\phi\end{matrix}\right\rangle\right]^{K},\quad 0\leqslant s=2J/N\leqslant 1, \tag{1}$$

where K is finite, $|\frac{1}{2};\theta\phi\rangle$ are the coherent states of SU(2) in the representation $j=\frac{1}{2}$, and $X=J_3,J_+,J_-$. The classical limit has reduced to a simple algorithm the problem of studying the critical properties of systems depending on N identical particles whose internal dynamical group is SU(2):

(1) Write down the Hamiltonian per particle in terms of intensive[4,5] angular momentum operators,

$$H/N=h_Q(\mathbf{J}/N).$$

(2) Replace the angular momentum operators by their classical limits (1). This converts the operator h_Q into a c-number "potential" h_C. The appropriate finite temperature potential is $\Phi(\beta)=h_C-(\beta N)^{-1}\ln Y(N,J)$, where $Y(N,J)$ is an SU(2) multiplicity factor.

(3) The minima of h_C and $\Phi(\beta)$ $(\beta=1/kT)$ rigorously give the ground state energy per particle and the free energy per particle in the thermodynamic limit, respectively. The critical properties of the system are determined by studying how the minima of h_C and $\Phi(\beta)$ change as a function of changing interaction parameters and temperature, respectively.

Step 2 involves Lie group theory, and in particular the use of coherent states. Step 3 involves catastrophe theory, and in particular local and nonlocal bifurcation theory. This algorithm[2] has been successfully applied to study the ground state and thermodynamic critical properties of a large class of quantum optics Hamiltonians of Dicke type[6] and nuclear Hamiltonians of pseudospin type.[7,8]

2. CLASSICAL LIMITS FOR COMPACT ALGEBRAS

Powerful though this algorithm is, its use is restricted to model systems constructed from operators belonging to the Lie algebra su(2). The Bogoliubov and Lieb inequalities, on which Step 2 of this algorithm is based, are independent of specific group-theoretic details, except that the Lieb inequaltiy requires a compact domain. Therefore, the only obstruction to extending this algorithm to other groups is the lack of a classical limit for operators belonging to Lie algebras more complicated than su(2) ("nonspin").

Coherent states for general Lie groups have been introduced[9,10] and extensively studied.[11] If G is a compact semisimple Lie group with Lie algebra $\mathfrak{g}$, Γ^{Λ} is an irreducible representation characterized by highest weight (or any extremal weight) Λ, $|\Lambda,\Lambda\rangle$ is the state of highest weight, and H is the stability group of $|\Lambda,\Lambda\rangle$, then the coherent states $|\Lambda,\Omega\rangle$ are defined by[11]

$$\left|\begin{matrix}\Lambda\\ \Omega\end{matrix}\right\rangle=\Omega\left|\begin{matrix}\Lambda\\ \Lambda\end{matrix}\right\rangle,\quad \Omega\in G/H. \tag{2}$$

Here Ω is a group element and also a coset representative. The coset generally depends on the representation, or equivalently on Λ. The Q-representative of an operator $X\in\mathfrak{g}$ can be determined from the generating function,

$$\left\langle\begin{matrix}\Lambda\\ \Omega\end{matrix}\middle|e^{\gamma X}\middle|\begin{matrix}\Lambda\\ \Omega\end{matrix}\right\rangle=\left\langle\begin{matrix}\Lambda\\ \Lambda\end{matrix}\middle|\Omega^{-1}e^{\gamma X}\Omega\middle|\begin{matrix}\Lambda\\ \Lambda\end{matrix}\right\rangle. \tag{3}$$

The product of group elements can be written in a more convenient form using Baker–Campbell–Hausdorff formulas,[11,12,13]

$$\Omega^{-1}e^{\gamma X}\Omega=e^{S}e^{\mathbf{d}\cdot\mathbf{H}}e^{S_+}. \tag{4}$$

Here S_+ is a sum over the "shiftup" operators in $\mathfrak{g}$, corre-

 0022-2488/79/050891-03$01.00

sponding to positive roots in the algebra,[13] so that $S_+|\Lambda,\Lambda\rangle = 0$. S_- is a "shiftdown" operator, and $\mathbf{d}\cdot\mathbf{H}$ is diagonal. Using (4) in (3), we find

$$(3) = \left\langle \begin{matrix}\Lambda\\ \Lambda\end{matrix}\right| e^{S_-} e^{\mathbf{d}\cdot\mathbf{H}} e^{S_-} \left| \begin{matrix}\Lambda\\ \Lambda\end{matrix}\right\rangle = e^{\mathbf{d}\cdot\Lambda}. \tag{5}$$

The coefficients d depend on $\Omega\in G/H$, $X\in\mathfrak{g}$, and the parameter γ.

The highest weight can be written[14]

$$\Lambda = \sum_{i=1}^{r} \mu_i \mathbf{f}_i, \tag{6}$$

where r is the rank of $\mathfrak{g}$, $\mathbf{f}_i$ are the highest weights of the r fundamental irreducible representations, and μ_i are nonnegative integers. The right-hand side of (5) factors accordingly,

$$(5) = e^{\mathbf{d}\cdot(\Sigma\mu_i\mathbf{f}_i)} = \prod_{i=1}^{r} (e^{\mathbf{d}\cdot\mathbf{f}_i})^{\mu_i}. \tag{7}$$

The functions $\exp(\mathbf{d}\cdot\mathbf{f}_i)$ are easily computed, for

$$\left\langle \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right| e^{\gamma X} \left| \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right\rangle = e^{\mathbf{d}\cdot\mathbf{f}_i}. \tag{8}$$

By combining (8) with (3) we have a concise form for the generating function,

$$\left\langle \begin{matrix}\Lambda\\ \Omega\end{matrix}\right| e^{\gamma X} \left| \begin{matrix}\Lambda\\ \Omega\end{matrix}\right\rangle = \prod_{i=1}^{r} \left[\left\langle \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right| e^{\gamma X} \left| \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right\rangle\right]^{\mu_i}. \tag{9}$$

In order to compute the classical limit of X, it is useful first to write the matrix elements in (9) as follows:

$$\begin{aligned}\left\langle \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right| e^{\gamma X} \left| \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right\rangle &= 1 + \gamma\left\langle \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right| X \left| \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right\rangle + \frac{\gamma^2}{2!}\left\langle \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right| X^2 \left| \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right\rangle + \cdots \\ &= 1 + \gamma g(\mathbf{f}_i, X, \Omega, \gamma).\end{aligned} \tag{10}$$

Then (9), considered as a generating function, can be written

$$\begin{aligned}\sum_{k=0}^{\infty} \left\langle \begin{matrix}\Lambda\\ \Omega\end{matrix}\right| \frac{\gamma^k X^k}{k!} \left| \begin{matrix}\Lambda\\ \Omega\end{matrix}\right\rangle &= \prod_{i=1}^{r} (1 + \gamma g(\mathbf{f}_i, X, \Omega, \gamma))^{\mu_i} \\ &= \prod_{i=1}^{r} \sum_{k_i} \frac{\mu_i!}{(\mu_i - k_i)! k_i!} [\gamma g(\mathbf{f}_i, X, \Omega, \gamma)]^{k_i}.\end{aligned} \tag{11}$$

The expectation value of the Kth power of X/N is

$$\begin{aligned}\left\langle \begin{matrix}\Lambda\\ \Omega\end{matrix}\right| (X/N)^K \left| \begin{matrix}\Lambda\\ \Omega\end{matrix}\right\rangle &= \prod_{i=1}^{r} \sum_{k_i} \left(\frac{\mu_i!}{(\mu_i - k_i)! N^{k_i}}\right) \frac{1}{k_i!} \frac{1}{N^{K - \Sigma k_i}} \\ &\times \left(\frac{d}{d\gamma}\right)^K [\gamma g(\mathbf{f}_i, X, \Omega, \gamma)]^{k_i}|_{\gamma=0}.\end{aligned} \tag{12}$$

This simplifies considerably in the limit of large N, for $\Sigma k_i \leqslant K$ and

$$\lim_{N\to\infty} \frac{1}{N^{K - \Sigma k_i}} = \delta_{K,\Sigma k_i},$$

$$\lim_{N\to\infty} \frac{\mu_i!}{(\mu_i - k_i)! N^{k_i}} = \lim_{N\to\infty} \left(\frac{\mu_i}{N}\right)^{k_i} = s_i^{k_i}, \tag{13}$$

so that (12) becomes simply

$$\begin{aligned}(12) &\xrightarrow{N\to\infty} \prod_{i=1}^{r} \sum_{\Sigma k_i = K} \left(\frac{\mu_i}{N}\right)^{k_i} \frac{K!}{k_i!} [g(\mathbf{f}_i, X, \Omega, \gamma = 0)]^{k_i}, \\ \lim_{N\to\infty} \left\langle \begin{matrix}\Lambda\\ \Omega\end{matrix}\right| (X/N)^K \left| \begin{matrix}\Lambda\\ \Omega\end{matrix}\right\rangle &= \left(\sum_{i=1}^{r} s_i \left\langle \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right| X \left| \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right\rangle\right)^K.\end{aligned} \tag{14}$$

In short, the classical limit of $(X/N)^K$ is determined from the Q representative of X in each of the fundamental representations of $\mathfrak{g}$.

The classical limit of $(X/N)^K (Y/N)^L$ can be determined by constructing the appropriate generating function and following the procedure described in Eqs. (3)–(14). The result is simply that the classical limit of the operator product is the product of the classical limits of the operators. Nor does it matter in what order the operators occur, for the commutator $[X/N, Y/N] = N^{-1}([X,Y]/N)$ vanishes in this limit.

The result (14) is valid for all compact semisimple Lie groups. The proof can be extended to noncompact semisimple Lie groups, provided we deal with their square-integrable representations.[11]

3. EXAMPLES AND APPLICATIONS

To illustrate how the classical limit (14) is used and why it is useful, we consider two examples and an application. These all deal with the groups SU(r), so we consider the coherent states for this group first. The stability group H depends on the class of representations used. For the fully symmetric representations, $H = U(r-1)$ and the coherent states have already been explicitly constructed.[9] If the generators for SU(r) obey commutation relations[13] $[E_{ij}, E_{rs}] = E_{is}\delta_{jr} - E_{rj}\delta_{si}$ and the extremal state is $|1\rangle$, then

$$\Omega = \exp\left\{\frac{1}{2}\sum_{j=2}^{r} \theta_j (e^{-i\phi_j} E_{j1} - e^{i\phi_j} E_{1j})\right\}, \tag{15}$$

and the expectation values of the operators E_{ij} in the first fundamental representation are

$$\begin{aligned}s_i \left\langle \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right| E_{ji} \left| \begin{matrix}\mathbf{f}_i\\ \Omega\end{matrix}\right\rangle &= \bar{h}_j h_i, \\ h_i &= \cos\theta/2, \quad i = 1, \\ &= \frac{\theta_i}{2} e^{-i\phi_i} \sin\theta/2, \quad i > 1, \\ \left(\frac{\theta}{2}\right)^2 &= \sum_{i=2}^{r} \left(\frac{\theta_i}{2}\right)^2.\end{aligned} \tag{16}$$

These results are valid only in the fully symmetric representations of SU(r). We remark that Hamiltonians constructed from operators belonging to su(r) have been studied in nuclear physics for $r = 2$,[15] $r = 3$,[16] and $r = 6$.[17,18]

Example 1: For SU(2), $J_z = (E_{11} - E_{22})/2$, $J_+ = E_{12}$, $J_- = E_{21}$, so that

$$\left\langle \begin{matrix} J \\ \theta\phi \end{matrix} \right| (J_+/N)^K \left| \begin{matrix} J \\ \theta\phi \end{matrix} \right\rangle = (s\tfrac{1}{2}\sin\theta\, e^{i\phi})^K, \quad 0 \leqslant s = K/N \leqslant 1. \tag{17}$$

The classical limit of J_- is $(s/2)(e^{-i\phi}\sin\theta)$ and for J_z it is $(s/2)\cos\theta$. This is the basis for the algorithm described above.

Example 2: For SU(3), in the representation characterized by Young partition (λ_1,λ_2), $\Lambda = \mu_1 f_1 + \mu_2 f_2$, where $(\mu_1,\mu_2) = (\lambda_1 - \lambda_2, \lambda_2)$ and $\mathbf{f}_1$ and $\mathbf{f}_2$ are the highest weights $(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3})$ and $(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})$ of the 3 and $\bar{3}$ representations,

$$\left\langle \begin{matrix} \Lambda \\ \Omega \end{matrix} \right| (X/N)^K \left| \begin{matrix} \Lambda \\ \Omega \end{matrix} \right\rangle = \left(s_1 \left\langle \begin{matrix} 3 \\ \Omega \end{matrix} \right| X \left| \begin{matrix} 3 \\ \Omega \end{matrix} \right\rangle + s_2 \left\langle \begin{matrix} \bar{3} \\ \Omega \end{matrix} \right| X \left| \begin{matrix} \bar{3} \\ \Omega \end{matrix} \right\rangle \right)^K, \tag{18}$$

where $s_i = \lim_{N\to\infty} \mu_i/N$ and $X \in su(3)$.

Application: We assume the Hamiltonian describing a system of N nucleons, each with r available states, has the form

$$\hat{h}_Q = \hat{H}/N = (\text{diagonal}) + (\text{interaction}). \tag{19}$$

For the diagonal contribution we take

$$(\text{diagonal}) = \sum_{i=1}^{r} \epsilon_i (H_i/N), \quad \epsilon_1 < \epsilon_2 < \cdots < \epsilon_r, \tag{20}$$

where H_i and E_{ii} are elements in the Cartan subalgebra of su(r). We assume a quadrupole form for the interaction, as follows,

$$(\text{interaction}) = \frac{Q}{2}\left[\left(\frac{E_{1j}}{N}\right)^2 + \left(\frac{E_{j1}}{N}\right)^2\right], \quad 1 < j \leqslant r, \tag{21a}$$

$$\text{or} \quad = \frac{Q}{2}\left[\left(\frac{E_{ij}}{N}\right)^2 + \left(\frac{E_{ji}}{N}\right)^2\right], \quad 1 < i < j \leqslant r, \tag{21b}$$

where $Q \geqslant 0$. Expression (21a) describes an interaction between the ground and an excited state, (21b) between two excited states. The Hamiltonian (19) can be regarded as the multilevel extension of the Lipkin–Meshkov–Glick pseudospin Hamiltonian.[15]

The ground state of (19) belongs to the fully symmetric representation of SU(r), with $\Lambda = N\mathbf{f}_1$, so that $s_1 = 1$, $s_2 = \cdots = s_{r-1} = 0$ in (14). Therefore the classical limits of (20) and (21) can be read directly from (16):

$$(20) \to \epsilon_1 \cos^2\frac{\theta}{2} + \sum_{i=2}^{r} \epsilon_i \left(\frac{\theta_i}{\theta}\sin\frac{\theta}{2}\right)^2, \tag{22}$$

$$(21a) \to \frac{Q}{2}\left(\frac{\theta_j}{\theta}\sin\frac{\theta}{2}\cos\frac{\theta}{2}\right)^2 (e^{-2i\phi_j} + e^{2i\phi_j}), \tag{23a}$$

$$(21b) \to \frac{Q}{2}\left(\frac{\theta_i\theta_j}{\theta^2}\sin^2\frac{\theta}{2}\right)^2 (e^{-2i(\phi_i - \phi_j)} + e^{2i(\phi_i - \phi_j)}). \tag{23b}$$

Equations (20) and (21) represent Step 1 in the extension to "nonspin" systems of the algorithm described in Sec. I. Equations (22) and (23) represent Step 2.

For the Hamiltonian, (20) + (21a), the ground state energy per nucleon is obtained by minimizing (22) + (23a). A simple local bifurcation analysis[6] shows that the system is disordered (all $\theta_i = 0$) for $Q < \Delta_j = \epsilon_j - \epsilon_1$, but a second-order phase transition occurs as Q becomes larger than Δ_j and the system becomes ordered, with $\cos\theta_j = \Delta_j/Q$, all other $\theta_i = 0$ ($i \neq j$). This result is exact in the thermodynamic limit.

For the Hamiltonian (20) + (21b), the ground state energy per nucleon is obtained by minimizing (22) + (23b). A nonlocal bifurcation occurs that may be treated by a catastrophe theory analysis.[6,8] The ground state is disordered for $Q < \Delta_j - \Delta_i$. There is a locally stable ordered state for $Q > \Delta_j - \Delta_i$ which is metastable for $\sqrt{Q} < \sqrt{\Delta_j} + \sqrt{\Delta_i}$ and globally stable for $\sqrt{Q} > \sqrt{\Delta_j} + \sqrt{\Delta_i}$. The corresponding phase transition is first order. The order parameters of the locally stable ordered state are defined by $e^{2i(\phi_i - \phi_j)} = -1$, $\phi_i + \phi_j$ arbitrary, $\theta_i^2 + \theta_j^2 = \pi^2$, $\theta_j^2/\theta_i^2 = [Q - (\Delta_j - \Delta_i)/Q + (\Delta_j - \Delta_i)]$, $\theta_k = 0$ and ϕ_k arbitrary ($k \neq i$, $k \neq j$). The saddle barrier separating the ordered state from the disordered state has height $\Delta_j\Delta_i/Q$.

4. SUMMARY AND CONCLUSIONS

A simple formula (14) is derived in Sec. 2 for the classical limits of operators belonging to compact Lie algebras. This allows the extension of a simple algorithm (Sec. 1) for determining the ground state and thermodynamic critical properties of systems whose Hamiltonians are constructed from operators belonging to such algebras. In its first application to groups other than SU(2), the algorithm reveals that the multilevel extension of the Lipkin–Meshkov–Glick pseudospin Hamiltonian supports second or first order ground state energy phase transitions, depending on whether the quadrupole interaction between two levels does or does not involve the ground state.

[1] F.T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[2] E.H. Lieb, Commun. Math. Phys. **31**, 327 (1973); K. Hepp and E.H. Lieb, Phys. Rev. A **8**, 2517 (1973).
[3] R. Gilmore, J. Phys. A **9**, L65 (1976).
[4] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[5] R. Gilmore, Physica A **86**, 137 (1977).
[6] R. Gilmore, J. Math. Phys. **18**, 17 (1977).
[7] R. Gilmore and D.H. Feng, Phys. Lett. B **76**, 26 (1978).
[8] R. Gilmore and D.H. Feng, Nucl. Phys. A **301**, 189 (1978).
[9] R. Gilmore, Ann. Phys. (NY) **74**, 391 (1972).
[10] A.M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[11] R. Gilmore, Rev. Mex. de Fisica **23**, 143 (1974).
[12] R. Gilmore, J. Math. Phys. **15**, 2090 (1974).
[13] R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).
[14] H. Weyl, *The Classical Groups* (Princeton, N.J., 1939).
[15] H.J. Lipkin, N. Meshkov, and A.J. Glick, Nucl. Phys. **62**, 188 (1965).
[16] J.P. Elliott, Proc. Roy. Soc. (London) Ser. A **245**, 128 (1958).
[17] A. Arima and F. Iachello, Phys. Rev. Lett. **35**, 1069 (1975).
[18] A. Arima, T. Ohtsuka, F. Iachello, and I. Talmi, Phys. Lett. B **66**, 205 (1977).

Path integrals for waves in random media[a)]

Roger Dashen

Institute for Advanced Study, Princeton, New Jersey 08540
(Received 9 September 1977)

The problem of wave propagation in a random medium is formulated in terms of Feynman's path integral. It turns out to be a powerful calculational tool. The emphasis is on propagation conditions where the rms (multiple) scattering angle is small but the log-intensity fluctuations are of order unity—the so-called saturated regime. It is shown that the intensity distribution is then approximately Rayleigh with calculable corrections. In an isotropic medium, the local or Markov approximation which is commonly used to compute first and second (at arbitrary space–time separation) moments of the wave field is explicitly shown to be valid whenever the rms multiple scattering angle is small. It is then shown that in the saturated regime the third and higher moments can be obtained from the first two by the rules of Gaussian statistics. There are small calculable corrections to the Gaussian law leading to "coherence tails." Correlations between waves of different frequencies and the physics of pulse propagation are studied in detail. Finally it is shown that the phenomenon of saturation is physically due to the appearance of many Fermat paths satisfying a perturbed ray equation. For clarity of presentation much of the paper deals with an idealized medium which is statistically homogeneous and isotropic and is characterized by fluctuations of a single typical scale size. However, the extension to inhomogeneous, anisotropic, and multiple scale media is given. The main results are summarized at the beginning of the paper.

1. INTRODUCTION AND SUMMARY OF RESULTS

The problem of propagation of waves in a random medium appears in a number of areas of research and applied science. Some examples are atmospheric optics, radio astronomy, and underwater sound. The problem is furthermore an old one which has been studied extensively. The earlier work (summarized in the monographs of Tatarskii[1] and Chernov[2]) employed the Rytov approximation. In this approximation the logarithm of the amplitude is computed using first order perturbation theory. The Rytov method is applicable whenever the intensity fluctuations are small. When the wavelength is small it reduces to first order geometric optics or WKB. More recently, a different approximation which reduces the problem to a Markov process has lead to considerable progress in cases where the intensity fluctuations are not small. This method is explained in Tatarskii's second book[3] and in two excellent reviews of the recent literature.[4,5] Nevertheless, important problems remain. In particular, there does not exist a global view of what is going on in the so-called saturated regime where the intensity fluctuations are important.

In this paper Feynman's path integral[6] is applied to the problem of wave propagation in a random medium. It provides a natural and systematic method for attacking the problem, especially when the intensity fluctuations are large and the Rytov approximation fails. The path integral is widely used in quantum mechanics and statistical mechanics but it is expected that many readers will not be familiar with it, thus the paper is meant to be self-contained. The reader who desires further background information on path integrals will do well to consult the book of Feynman and Hibbs.[6]

Because some readers will not be familiar with path integrals there are some peculiarities in the organization of this paper. In real situations, random media are often statistically inhomogeneous or anisotropic and frequently have a power law spectrum in the scale size of fluctuations. Path integrals are capable of handling all these complications. (In fact the author first developed the method for propagation of sound in the ocean,[7] a problem which has these complications and more.) However, it is vastly easier to explain the path integral method for an idealized medium which is statistically homogeneous and isotropic and whose fluctuations are characterized by a single[8] typical scale size L (small compared to the distance R of propagation). The bulk of the paper is therefore devoted to a study of this idealized situation. Once this has been done the transition to realistic media is relatively simple. However, this manner of presentation has a defect for which only an apology can be offered. Because of the temporary restriction to a single scale size , results which are directly applicable to atmospheric optics do not appear until late in the paper (specifically, Secs. 7 and 8). Finally, to illustrate the power of the path integral method (and, hopefully, motivate the reader), a number of results for the idealized problem will be summarized below. The translation of these results to more complicated cases is generally straightforward: The details are given in the text. Listing the results will require the definition of some symbols. This will (temporarily) be done in terms of the idealized problem and the reader who has worked on propagation in a random medium will find that they are familiar objects, e.g., Tatarskii's phase structure function D. For other readers, the motivation for these definitions will become apparent in Secs. 2 and 3.

Actually, there are two distinct kinds of problems of propagation in a random medium, corresponding to whether the scattering angles, single and/or multiple, are large or small. If the fluctuations are weak so that a single scattering approximation (Born approximation) applies, there is little

[a)]This work was done under the sponsorship of the Advanced Research Projects Agency Contract DAHC15-73-C-0370.

 0022-2488/79/050894-27$01.00

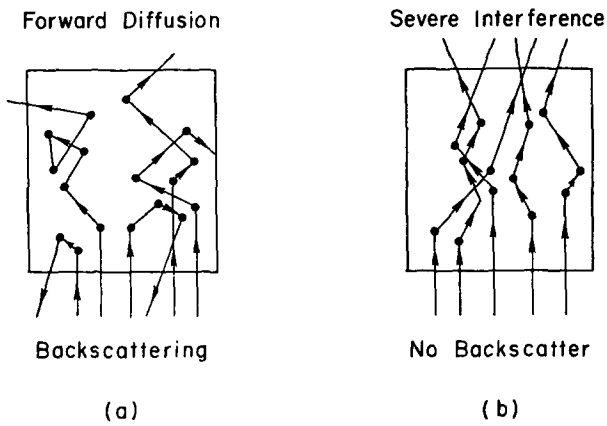

FIG. 1. The difference between multiple large angle scattering (a) and multiple small angle scattering (b).

distinction between the two cases. However in a multiple scattering regime, which is the case of interest here, the two kinds of problems are very different. This is illustrated in Fig. 1. The considerations of this paper will be restricted to situations where the single *and* multiple scattering angles are small. This is sufficient to cover the applications mentioned above. The large angle multiple scattering situation is like a problem in radiative transport and is most efficiently treated by other methods.

It will be assumed that the problem can be reduced to a scalar wave equation with an index of refraction $n(\mathbf{x},t)$ which may depend on the frequency $\omega = ck$. In a homogeneous medium $\langle n\rangle$ is a constant and for waves of a fixed frequency can be set equal to unity. Defining

$$\mu(\mathbf{x},t) = 1 - n(\mathbf{x},t), \tag{1.1}$$

μ will be taken to have a zero mean and a covariance

$$\langle \mu(\mathbf{x},t)\mu(\mathbf{x}',t')\rangle = \rho(|\mathbf{x}-\mathbf{x}'|,t-t'). \tag{1.2}$$

It will be further assumed that either μ is a Gaussian[9] random field *or* that $kL\langle\mu^2\rangle^{1/2}$ is small, in which case the distribution need not be specified.

Let the two-dimensional vector $\mathbf{r}_0 = (x_0,y_0)$ label the location of a point source[10] in the plane $z=0$. Then in a plane of constant $z>0$, the signal will be $E(z,\mathbf{r},\mathbf{r}_0,t)$ where $\mathbf{r}=(x,y)$ specifies the transverse coordinates of the observation point. The total range of propagation will be denoted by R and for $|\mathbf{r}|, |\mathbf{r}_0| \ll R$, and a CW source it is useful to define a complex envelope $\mathscr{E}$ by

$$E(z,\mathbf{r},\mathbf{r}_0,t) = \mathrm{Re}[\mathscr{E}(z,\mathbf{r},\mathbf{r}_0,t)e^{i(kz-\omega t)}]. \tag{1.3}$$

The time dependence of $\mathscr{E}$ is due to fluctuations in the medium. It will be assumed that the full wave equation for $\mathscr{E}$ can be approximated by the parabolic wave equation[3-7,11]

$$\left(i\frac{\partial}{\partial z} + \frac{1}{2k}\nabla^2 - k\mu(\mathbf{r},z,t)\right)\mathscr{E}(\mathbf{r},\mathbf{r}_0,z,t) = 0, \tag{1.4}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

plus a boundary condition at $z=0$

$$\mathscr{E} \to (4\pi z)^{-1}\exp\left(\frac{ik(\mathbf{r}-\mathbf{r}_0)^2}{2z}\right). \tag{1.5}$$

If L and T are the characteristic lengths and times over which μ changes, the validity condition for Eq. (1.4) are (i) $kL \gg 1$, (ii) $kL \ll \omega T$, and (iii) that the rms multiple scattering angle $(\langle\mu^2\rangle R/L)^{\frac{1}{2}}$ should be small.

Feynman's path integral gives the solution to the parabolic wave equation in terms of a (strictly speaking) infinite dimensional integral. It turns out that this integral can be studied in almost exactly the same way as Mercier[12] originally attacked the phase screen integral. The result is that propagation in a statistically homogeneous medium is very similar to the phase screen problem. This will continue to be true in rather general inhomogeneous media, of which the phase screen is a special case.

In order to indicate what can be learned from the path integral it is necessary to review some known features of propagation in a random medium. The qualitative character of $\mathscr{E}$ is determined by two parameters Φ and Ω defined by

$$\Phi^2 = k^2\left\langle\left(\int_0^R \mu(\mathbf{e}_z z,t)dz\right)^2\right\rangle$$

$$= k^2 R\int_{-\infty}^{\infty} \rho(|z|,0)dz + O(L/R), \tag{1.6}$$

where $\mathbf{e}_z$ is a unit vector in the z direction and in the second line it has been assumed that $R \gg L$ and

$$\Omega = 6kL^2/R. \tag{1.7}$$

The parameter Φ is just the rms phase fluctuation as computed in first order geometric optics[1] and serves as a measure of the strength of the fluctuations. The other parameter Ω is essentially the square of the ratio of the scale size L to the extent of a Fresnel zone. As shown in Fig. 2 if Φ is less than one *or* less than Ω, then the Rytov[1-5] approximation is valid. In the region where the Rytov approximation is valid, the problem can be considered to have been solved years ago. The intensity fluctuations (scintillations) are small and the relation between $\mathscr{E}$ and μ is simple and direct. Also, as

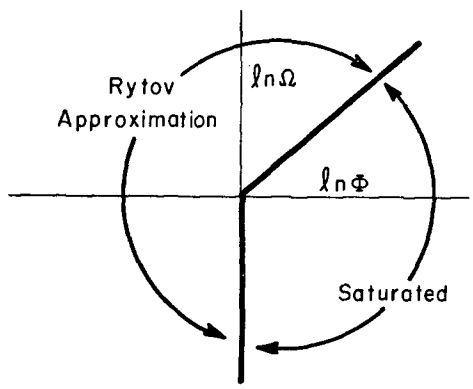

FIG. 2. Parameter regimes in $\Phi - \Omega$ space.

shown in Fig. 2, when both Φ and Φ/Ω are greater than unity, the fluctuations in $\mathscr{E}$ saturate.[13] In particular, the variance of $\ln|\mathscr{E}|^2 = \ln I$ approaches a constant of order unity and the properties of $\mathscr{E}$ are determined more by statistical considerations than by the detailed properties of μ.

Path integral methods have nothing new to add when the Rytov approximation is valid. The considerations of this paper will therefore be restricted to the saturated regions. There is then a small parameter $\alpha = \Omega/\Phi$ whose order of magnitude[14] is

$$\alpha \sim \frac{6L^{3/2}}{R^{3/2}\langle\mu^2\rangle^{\frac{1}{2}}}. \tag{1.8}$$

The path integral allows the calculation of any moment of $\mathscr{E}$ as an asymtotic series in α. The result is that $\mathscr{E}$ is uniformly distributed in phase and that the moments of intensity $I \equiv |\mathscr{E}|^2$ are given by

$$\langle I^n\rangle = n!\langle I\rangle^n[1 + \tfrac{1}{2}n(n-1)C\alpha + O(\alpha^2)], \tag{1.9}$$

where C is a calculable constant of order unity whose precise value depends on the spectrum of μ. In the limit $\alpha = 0$ the distribution is therefore Rayleigh[15] with

$$P(I) = \frac{1}{\langle I\rangle}\exp\left[-\frac{I}{\langle I\rangle}\right]. \tag{1.10}$$

However, the correction grows with n and cannot be neglected for $n \gtrsim (2/\alpha)^{\frac{1}{2}}$. It follows that there must be significant deviations from a Rayleigh distribution when $I/\langle I\rangle$ is greater than $\sim(2/\alpha)^{\frac{1}{2}}$.

In addition to the distribution of $\mathscr{E}$, one also wants to know the coherences in space and time. Recent work on coherences has been greatly facilitated by the observation[3-5] that under certain conditions the problem can be replaced by a simpler local or Markov one where, in effect, one makes the replacement

$$\rho(\mathbf{x},t) \to \delta(z)\hat{\rho}(|\mathbf{r}|,t), \tag{1.11}$$

with $\mathbf{r} = (x,y)$ and

$$\hat{\rho}(|\mathbf{r}|,t) = \int_{-\infty}^{\infty} \rho((\mathbf{r}^2 + z^2)^{\frac{1}{2}},t)\,dz. \tag{1.12}$$

Note that Φ^2 is equal to $k^2R\hat{\rho}(0,0)$ and it will be convenient to use the function $\hat{\rho}$ to *define* T and L by the expansion

$$k^2R\hat{\rho}(|\mathbf{r}|,t) = \Phi^2\left(1 - \frac{\mathbf{r}^2}{2L^2} - \frac{t^2}{2T^2} + \cdots\right). \tag{1.13}$$

Within factors of order unity, the L and T so defined will be equal to the length and time over which the original covariance ρ is nonvanishing.

It has been pointed out by several authors[3-5] that in the Markov approximation the coherence of $\mathscr{E}^*$ and $\mathscr{E}$ can be computed exactly. It is

$$\frac{\langle\mathscr{E}^*(\mathbf{r}',\mathbf{r}_0',t')\mathscr{E}(\mathbf{r},\mathbf{r}_0,t)\rangle}{\mathscr{E}_0^*(\mathbf{r}',\mathbf{r}_0')\mathscr{E}_0(\mathbf{r},\mathbf{r}_0)} = \exp[-\tfrac{1}{2}D(\mathbf{r}-\mathbf{r}',\mathbf{r}_0-\mathbf{r}_0',t-t')], \tag{1.14}$$

where

$$\mathscr{E}_0(\mathbf{r},\mathbf{r}_0) = \langle I\rangle^{\frac{1}{2}}\exp\left[ik\frac{(\mathbf{r}-\mathbf{r}_0)^2}{2R}\right] \tag{1.15}$$

and D is the phase structure function of first order geometric optics[1]

$$D(\mathbf{r},\mathbf{r}_0,t) = 2k^2R\int_0^1 [\hat{\rho}(0,0) - \hat{\rho}(|u\mathbf{r} + (1-u)\mathbf{r}_0|,t)]\,du. \tag{1.16}$$

The phase structure function always appears in an exponential and in the saturated region where Φ^2 is large, D can be approximated by an expansion in $\mathbf{r},\mathbf{r}_0$, and t

$$D(\mathbf{r},\mathbf{r}_0,t) = \Phi^2\left[\frac{\mathbf{r}^2 + \mathbf{r}_0^2 + \mathbf{r}\cdot\mathbf{r}_0}{3L^2} + \frac{t^2}{T^2} + \cdots\right]. \tag{1.17}$$

Coherences are then characterized by two parameters Φ/T and Φ/L.

The literature is somewhat confusing as to the validity conditions for Eq. (1.14). It turns out that the approximation leading to Eq. (1.14) has a very simple interpretation in the path integral formalism. In the next section it will become evident that for the isotropic medium under consideration, Eq. (1.14) is valid as long as the parabolic wave equation is valid. From the path integral one can actually compute the first correction to Eq. (1.14). It is of order of the rms multiple scattering angle $(R/L)^{1/2}\langle\mu^2\rangle^{1/2}$ which must be small if the parabolic wave equation is valid.

For small α the path integral also allows the calculation of $\langle\mathscr{E}^*(\mathbf{r},\mathbf{r}_0,t)\mathscr{E}^*(\mathbf{r}',\mathbf{r}_0',t')\mathscr{E}(\mathbf{r}'',\mathbf{r}_0'',t'')\mathscr{E}(\mathbf{r}''',\mathbf{r}_0''',t''')\rangle$ and more generally an arbitrary $2n$th order moment. In the limit $\alpha = 0$, the real and imaginary parts of $\mathscr{E}$ are jointly Gaussian. To see the use of this result, let us consider a typical question of practical interest. Take a fixed source and receiver so that $\mathscr{E}$ is a function only of time and suppose that at $t = 0$ $\mathscr{E}/\mathscr{E}_0$ is known to have a value η. An interesting practical question is then what is the probability $P(\eta')$ that $\mathscr{E}(t)/\mathscr{E}_0$ will take on the value η'. Since $\mathscr{E}$ has a Gaussian distribution, $P(\eta')$ is simply

$$P(\eta') = \exp\left(-\frac{|\eta' - e^{-D(t)/2}\eta|^2}{1 - e^{-D(t)}}\right)\Big/\pi(1 - e^{-D(t)}), \tag{1.18}$$

where $D(t) = D(\mathbf{0},\mathbf{0},t) \approx \Phi^2(t/T)^2$. The qualitative behavior of $P(\eta')$ is indicated in Fig. 3. It is evident that the signal stays in one quadrant of the complex plane and is therefore coherent over a time of order T/Φ. A further property of Gaussian statistics and a covariance of the form $\exp[-\frac{1}{2}(\Phi t/T)^2]$ is that the signal will move in a straight line for times less than $\sim T/\Phi$. One can ask the more general question of given that $\mathscr{E}(\mathbf{r},\mathbf{r}_0,t)/\mathscr{E}_0(\mathbf{r},\mathbf{r}_0)$ is equal to η, what is the probability that $\mathscr{E}(\mathbf{r}',\mathbf{r}_0',t')/\mathscr{E}_0(\mathbf{r}',\mathbf{r}_0')$ will be equal to η'. The result is just Eq. (1.18) with $D(t)$ replaced by $D(\mathbf{r}-\mathbf{r}',\mathbf{r}_0-\mathbf{r}_0',t-t')$.

As stated above the Gaussian statistics leading to $P(\eta')$ are obtained by computing moments. Again the approximation scheme breaks down for moments of order $(2/\alpha)^{1/2}$ and Eq. (1.18) is valid only for $|\eta|$ and $|\eta'|$ less than $(2/\alpha)^{1/4}$. Actually the order α corrections to any moment are calculable. They are most important for intensity correlations where they lead to coherence tails[4] of order α which are small but fall much less rapidly than e^{-D}.

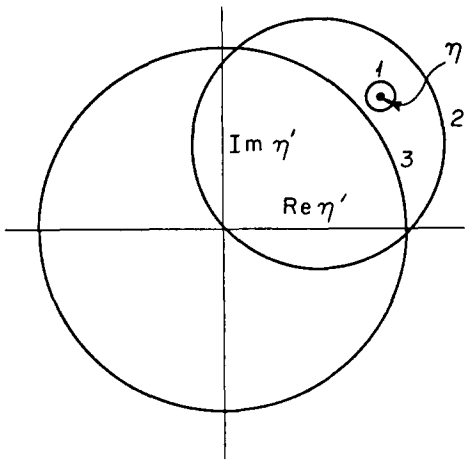

FIG. 3. Illustrating Eq. (1.18). The signal η' will lie, with 90% probability, within the circles: (1) for $\Phi t/T$ small, (2) for $\Phi t/T \sim 1$, and (3) for $\Phi t/T$ large. The location of the signal η at $t=0$ was an unlikely one lying outside the 90% probability circle for a Rayleigh distribution.

The path integral also provides a simple method for calculating the correlation between waves of different frequencies.[4,5] In the saturated region, where $\Phi > 1$ the result is, for $|\omega-\omega'|$ small compared to $\bar{\omega}\equiv\frac{1}{2}(\omega+\omega')$

$$\frac{\langle \mathscr{E}^*(\omega')\mathscr{E}(\omega)\rangle}{\mathscr{E}_0^*(\omega')\mathscr{E}_0(\omega)} = \exp\left[-\frac{1}{2}\left(\frac{\omega-\omega'}{\omega_g}\right)^2\right]\Lambda(\omega-\omega'), \tag{1.19}$$

where ω_g is

$$\omega_g^{-2} = \left\langle\left(\int_0^R \frac{d}{d\omega}(k\mu(\mathbf{e}_z z,t))dz\right)^2\right\rangle \tag{1.20}$$

and for a single scale medium

$$\Lambda(\omega) = \left(\frac{6i\omega}{\omega_0\alpha}\right)^{1/2}/\sin\left(\frac{6i\omega}{\omega_0\alpha}\right)^{1/2} \tag{1.21}$$

with $\omega_0^{-2} = c_g^{-2}R\hat{\rho}(0,0)$, where c_g is the unperturbed group velocity. For a nondispersive medium $\omega_0=\omega_g$ When α is very small the second factor on the right-hand side of Eq. (1.19) falls much more rapidly than the first one. The first factor

$$\exp\left[-\frac{1}{2}\left(\frac{\omega-\omega'}{\omega_g}\right)^2\right]$$

can than be replaced by unity. In the limit $\alpha=0$ the higher order correlations in frequency are Gaussian. One can then obtain probability distributions in frequency from Eq. (1.18) with $\exp[-D/2]$ replaced by the right-hand side of Eq. (1.19) and e^{-D} replaced by its absolute value squared. It is worth noting that a first order geometric optics calculation misses the second and dominant factor on the right-hand side of Eq. (1.19) and therefore vastly overestimates the range of coherence in frequency.

It can be seen from the path integral that saturation corresponds to the appearance of multiple Fermat paths which satisfy a perturbed ray equation. The signal tends to propagate along these Fermat paths and because there are many of them, they interfere and produce Gaussian statistics. They will become manifest in an experiment with a pulsed source where the received signal will tend to show several arrivals. These multiple Fermat paths are responsible for the factor Λ in Eq. (1.19).

With one exception these results can easily be extended to statistically inhomogeneous or anisotropic media and to media with multiple scales. The exception is that $\Lambda(\omega)$ defined in Eq. (1.19) cannot be computed for certain multiple scale media. Actually, the path integral yields further information in the case of multiple scale media. It appears to be only partially understood[3-5] that in this case there are two distinct saturated regimes. An examination of the path integral shows that there are indeed two, one of which (the fully saturated regime) is analogous to the saturated regime in single scale media and another one (the partially saturated regime) is new. Many experiments in atmospheric optics lie in the partially saturated regime and this case is treated in some detail (Sec. 8). The fundamental distinction between the fully and partially saturated regimes shows up in correlations between waves of different frequency. In the fully saturated regime the real and imaginary parts of $\mathscr{E}(\omega)$ are jointly Gaussian random variables. For partial saturation $\mathscr{E}(\omega)$ acts like a random phase times a Gaussian object. A consequence is that propagation of narrow pulses is qualitatively different in the two regimes. Depending on the medium there may be further qualitative differences between full and partial saturation.

The detailed organization of the paper is as follows. Sections 2–6 and Appendices A and B are devoted to the idealized homogeneous, isotropic medium with a single scale size. In Sec. 2 the path integral is introduced and applied to the calculation of the first and second moments. Appendix A contains the calculation of the error in Eq. (1.14). Section 3 is devoted to the calculation of higher moments when α is small and Sec. 4 summarizes the statistics of $\mathscr{E}$ in the limit $\alpha=0$. Special attention is given to statistics in frequency and pulse propagation. The corrections to the limiting statistics are derived in Appendix B and discussed in Sec. 5. The appearance of multiple Fermat paths is demonstrated in Sec. 6. Media with multiple scales are introduced in Sec. 7 and the distinction between full and partial saturation is made. In the fully saturated case there is a simple modification of the results for a single scale medium (Table II). The partially saturated regime is more difficult. Section 8 is devoted to partial saturation in a medium like that encountered in atmospheric optics. Appendix C contains some calculations relevant to Sec. 8 and Appendix D discusses some other kinds of multiple scale media. Methods for handling inhomogeneous and anisotropic media are given in Sec. 9 and Appendices E and F.

2. FIRST AND SECOND MOMENTS FROM THE PATH INTEGRAL

Feynman[6] pointed out that the solution to Eq. (1.4) with the boundary condition in Eq. (1.5) is given by an infinite dimenisonal integral. It is defined as the limit of a finite dimensional integral with $2n-2$ integration variables corre-

sponding to the Cartesian components of $n-1$ two-dimensional vectors $\mathbf{r}_j j=1,2,\ldots,n-1$. With the convention that $\mathbf{r}_j|_{j=0}$ and $\mathbf{r}_j|_{j=n}$ are the source $\mathbf{r}_0$ and receiver $\mathbf{r}$ coordinates, Feynman's integral is

$$\mathscr{E}(\mathbf{r},\mathbf{r}_0,t)=\lim_{n\to\infty}\frac{i}{2k}\int\left(\prod_{j=1}^{n-1}d^2r_j\right)\left(\frac{kn}{2\pi iR}\right)^n \times\exp\left\{\frac{ikR}{n}\sum_{j=1}^{n}\left[\frac{n^2}{2}\left(\frac{\mathbf{r}_j-\mathbf{r}_{j-1}}{R}\right)^2-\mu(\mathbf{r}_j+\mathbf{e}_z z_j,t)\right]\right\}, \tag{2.1}$$

where each component of $\mathbf{r}_j j=1,\ldots,n-1$, is integrated over the range $-\infty$ to $+\infty$ and $z_j=jR/n$. In μ, $\mathbf{r}_j$ is understood to be a vector in the xy plane and $\mathbf{e}_z$ is a unit vector in the z direction. At each point $\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_{n-1}$ in the integration volume, the $n-1$ points in space $(\mathbf{r}_1,z_1),(\mathbf{r}_2,z_2),\ldots(\mathbf{r}_{n-1},z_{n-1})$ can be thought of as discrete points along a path $\mathbf{r}(z)$ connecting $(\mathbf{r}_0,0)$ to $(\mathbf{r},R)$ with $\mathbf{r}_j=\mathbf{r}(z_j)$, see Fig. 4. In this sense Feynman's integral is an integral over paths. Associating R/n with a differential increment dz in range the argument of the exponential has a continuum limit

$$\frac{ikR}{n}\sum_{j=1}^{n}\left[\frac{n^2}{2}\left(\frac{\mathbf{r}_j-\mathbf{r}_{j-1}}{R}\right)^2-\mu(\mathbf{r}_j+\mathbf{e}_z z_j,t)\right] \to ik\int_0^R\left[\tfrac{1}{2}(\mathbf{r}'(z))^2-\mu(\mathbf{r}(z)+\mathbf{e}_z z,t)\right]dz, \tag{2.2}$$

where $\mathbf{r}'\equiv d\mathbf{r}/dz$. The path integral for $\mathscr{E}$ can then be schematically written as

$$\mathscr{E}(\mathbf{r},\mathbf{r}_0,t)=\frac{i}{2k}\int d\,(\text{paths}) \times\exp\left\{ik\int_0^R\left[\tfrac{1}{2}(\mathbf{r}'(z))^2-\mu(\mathbf{r}(z)+\mathbf{e}_z z,t)\right]dz\right\}, \tag{2.3}$$

where the integration is over all paths connecting $(\mathbf{r}_0,0)$ to $(\mathbf{r},R)$ and the volume element in path space d (paths) is the coefficient of the exponential in Eq. (2.1).

We will be computing averages of products of path integrals and the following formula will be needed. Let $\mathbf{r}_n(z)$, $n=1,2,\cdots$ be some set of paths and $\xi_n=\pm1$ corresponding phases. Then if *either* μ is a Gaussian random field *or* $kL\langle\mu^2\rangle^{1/2}\ll1$ and its statistics are arbitrary, it is well known[1-6] that

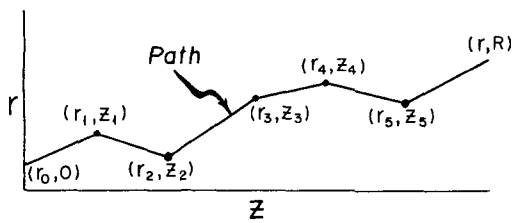

FIG. 4. A path in the path integral for $n=6$.

$$\left\langle\exp\left[-ik\sum_n\xi_n\int_0^R\mu(\mathbf{r}_n(z)+\mathbf{e}_z z,t_n)\,dz\right]\right\rangle = \exp\left(-\frac{k^2}{2}\sum_{m,n}\xi_m\xi_n\int_0^R\int_0^R\rho(\{[\mathbf{r}_n(z)-\mathbf{r}_m(z')]^2 +((z'-z)^2\}^{1/2},t_n-t_m)\,dzdz'\right). \tag{2.4}$$

As a first application of the path integral we can compute $\langle\mathscr{E}\rangle$. This will not turn out to be a particularly interesting quantity but the calculation is simple and it will show how path integrals work and where the Markov approximation comes in. Bringing the average inside the path integral and using Eq. (2.4) yields

$$\langle\mathscr{E}\rangle=\frac{i}{2k}\int d\,(\text{paths})\exp\left(\frac{ik}{2}\int_0^R(\mathbf{r}')^2dz-\frac{k^2}{2} \times\int_0^R\int_0^R\rho([(\mathbf{r}(z)-\mathbf{r}(z'))^2+(z-z')^2]^{1/2},0)\,dzdz'\right). \tag{2.5}$$

The Markov approximation now appears as follows. The parabolic wave equation assumes that the normals to the wave fronts point in directions that are close to the z axis. In terms of the path integral this means that for the important paths $|\mathbf{r}'|=|d\mathbf{r}/dz|$ must be small. It then follows that for important paths $(\mathbf{r}(z)-\mathbf{r}(z'))^2+(z-z')^2\approx|z-z'|^2$ and Eq. (2.5) becomes

$$\langle\mathscr{E}\rangle=\frac{i}{2k}\exp\left(-\frac{k^2}{2}\int_0^R\int_0^R\rho(|z-z'|,0)\,dzdz'\right) \times\int d\,(\text{paths})\exp\left(\frac{ik}{2}\int_0^R(\mathbf{r}')^2dz\right). \tag{2.6}$$

The remaining path integral is just the path integral for $\mathscr{E}_0$ and for $R\gg L$ the double integral over ρ can be replaced by $R\hat{\rho}(0,0)$. The final result is then

$$\langle\mathscr{E}\rangle=\mathscr{E}_0\exp[-\tfrac{1}{2}\Phi^2]. \tag{2.7}$$

This is the usual formula obtained in the Markov approximation.[3-5] What we have seen here is that this approximation has a very natural interpretation in terms of the path integral and that it is valid as long as the parabolic wave equation is valid.

Since Φ^2 is large in the saturated region $\langle\mathscr{E}\rangle$ is exponentially small and therefore not particularly interesting. The same is true for $\langle\mathscr{E}\mathscr{E}\rangle$ and its complex conjugate $\langle\mathscr{E}^*\mathscr{E}^*\rangle$. The path integral for $\langle\mathscr{E}\mathscr{E}\rangle$ will be a double path integral over two paths $\mathbf{r}_1(z)$ and $\mathbf{r}_2(z)$ and will contain a factor

$$\exp\left(-\frac{k^2}{2}\int_0^R\int_0^R\{\rho([(\mathbf{r}_1(z)-\mathbf{r}_1(z'))^2+(z-z')^2]^{1/2},0) +\rho([(\mathbf{r}_2(z)-\mathbf{r}_2(z'))^2+(z-z')^2]^{1/2},0)$$

$$+2\rho[(\mathbf{r}_1(z)-\mathbf{r}_2(z'))^2+(z'-z)^2,t\,]^{1/2}\}\,dzdz'), \tag{2.8}$$

where t is the time difference between the two $\mathscr{E}$'s in the average. This factor is of order $\exp[-\Phi^2]$ in all important regions of path space and $\langle\mathscr{E}\mathscr{E}\rangle$ is exponentially small.

A more interesting quantity is $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$ where $\mathscr{E}(1)$ is a shorthand notation for $\mathscr{E}(\mathbf{r}_1,\mathbf{r}_{01},t_1)$ and $\mathscr{E}^*(2)$ for $\mathscr{E}'^*(\mathbf{r}_2,\mathbf{r}_{02},t_2)$. The formula for this object is

$$\frac{1}{4k^2}\int d^2(\text{paths})\exp\left(\frac{ik}{2}\int_0^R[(\dot{\mathbf{r}}_1(z))^2-(\dot{\mathbf{r}}_2(z))^2]\,dz-V\right), \tag{2.9}$$

where the path integral is a "double path integral" over two paths $\mathbf{r}_1(z)$ and $\mathbf{r}_2(z)$ connecting $(\mathbf{r}_{01},0)$ to $(\mathbf{r}_1,R)$ and $(\mathbf{r}_{02},0)$ to $(\mathbf{r}_2,R)$, respectively and

$$\begin{aligned}V=\frac{k^2}{2}\int_0^R\int_0^R\{&\rho([(\mathbf{r}_1(z)-\mathbf{r}_1(z'))^2+(z-z')^2]^{1/2},0)\\&+\rho([(\mathbf{r}_2(z)-\mathbf{r}_2(z'))^2+(z-z')^2]^{1/2},0)\\&-2\rho([(\mathbf{r}_1(z)-\mathbf{r}_2(z'))^2+(z-z')^2]^{1/2},t_1-t_2)\}\,dzdz'.\end{aligned} \tag{2.10}$$

There is now a region in path space where the integrand is not exponentially small. It is $\mathbf{r}_1(z)\approx\mathbf{r}_2(z)$ and almost all of the path integral will come from this region. As before, $(\mathbf{r}_1(z)-\mathbf{r}_1(z'))^2$ and $(\mathbf{r}_2(z)-\mathbf{r}_2(z'))^2$ can be neglected relative to $(z-z')^2$ and in the same spirit $[(\mathbf{r}_1(z)-\mathbf{r}_2(z'))^2+(z'-z)^2]^{1/2}$ can be approximated by $[(\mathbf{r}_1(\bar{z})-\mathbf{r}_2(\bar{z}))^2+(z'-z)^2]^{1/2}$, where $\bar{z}=\frac{1}{2}(z+z')$. Then for $R\gg L$ the integral over $z-z'$ can be done and

$$V=\int_0^R d(|\mathbf{r}_1(z)-\mathbf{r}_2(z)|,t_1-t_2)dz, \tag{2.11}$$

where

$$d(|\mathbf{r}|,t)=k^2[\hat{\rho}(0,0)-\hat{\rho}(|\mathbf{r}|,t)]. \tag{2.12}$$

At this point it is convenient to change variables to paths $\mathbf{u}(z)$ and $\mathbf{v}(z)$,

$$\begin{aligned}\mathbf{u}(z)&=\tfrac{1}{2}(\mathbf{r}_1(z)+\mathbf{r}_2(z))-\tfrac{1}{2}(\mathbf{r}_1+\mathbf{r}_2)\frac{z}{R}-\tfrac{1}{2}(\mathbf{r}_{01}+\mathbf{r}_{02})\left(1-\frac{z}{R}\right),\\ \mathbf{v}(z)&=\mathbf{r}_1(z)-\mathbf{r}_2(z),\end{aligned} \tag{2.13}$$

which satisfy the end point conditions $\mathbf{u}(0)=\mathbf{u}(R)=\mathbf{0}$ and $\mathbf{v}(0)=(\mathbf{r}_{01}-\mathbf{r}_{02})$, $\mathbf{v}(R)=(\mathbf{r}_1-\mathbf{r}_2)$. From the finite form of the path integral in Eq. (2.1) it is clear that this change of variables is allowed and that the associated Jacobian is equal to unity. After integrating the first term in the exponential by parts and using the end point conditions, the path integral for $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$ becomes

$$\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle=\left(\frac{2\pi R}{k}\right)^2\mathscr{E}_0^*(2)\mathscr{E}_0(1)\int d^2(\text{paths})$$

$$\times\exp\left[-ik\int_0^R\mathbf{u}(z)\cdot\mathbf{v}''(z)dz\right.$$

$$\left.-\int_0^R d(|\mathbf{v}(z)|,t)dz\right], \tag{2.14}$$

where $t=t_1-t_2$. In analogy with the formula

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{ixy}dy=\delta(x), \tag{2.15}$$

the integral over the path $\mathbf{u}$ in Eq. (2.14) will produce a "δ functional" which forces $\mathbf{v}''$ to vanish identically.[16] With the end point conditions given above $\mathbf{v}(z)$ must then be

$$\mathbf{v}(z)=(\mathbf{r}_1-\mathbf{r}_2)\frac{z}{R}+(\mathbf{r}_{01}-\mathbf{r}_{02})\left(1-\frac{z}{R}\right). \tag{2.16}$$

In d the path $\mathbf{v}$ can then be replaced by the right-hand side of Eq. (2.16) and the factor containing d then becomes just $\exp[-\frac{1}{2}D]$. The remaining path integral is equal to $(2\pi R/k)^{-2}$ and the result[17] reproduces Eq. (.14),

$$\frac{\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle}{\mathscr{E}_0^*(2)\mathscr{E}_0(1)}=\exp[-\tfrac{1}{2}D(1,2)], \tag{1.14'}$$

where $D(1,2)$ is a shorthand notation for $D(\mathbf{r}_1-\mathbf{r}_2,\mathbf{r}_{01}-\mathbf{r}_{02},t_1-t_2)$.

Appendix A contains an explicit calculation of the first correction to the Markov approximation for $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$. It is shown to be proportional to the rms multiple scattering angle $(\langle\mu^2\rangle R/L)^{\frac{1}{2}}$ which must be small if the parabolic wave equation is valid. Henceforth, all calculations will be done in this Markov approximation. The general prescription is that whenever $\rho([(\mathbf{r}_i(z)-\mathbf{r}_j(z'))^2+(z-z')^2],t_i-t_j)$ appears, it is to be replaced by

$$\delta(z-z')\hat{\rho}\left(\left|\mathbf{r}_i\left(\frac{z+z'}{2}\right)-\mathbf{r}_j\left(\frac{z+z'}{2}\right)\right|,t_i-t_j\right).$$

Turning now to the calculation of $\langle\mathscr{E}^*(\omega')\mathscr{E}(\omega)\rangle$, the path integral for this quantity will contain [with $k=k(\omega)$ and $k'=k(\omega')$]

$$\left\langle\exp\left[-ik\int_0^R\mu_\omega(\mathbf{r}_1(z)+\mathbf{e}_z z)dz+ik'\right.\right.$$
$$\left.\left.\times\int_0^R\mu_{\omega'}(\mathbf{r}_2(z)+\mathbf{e}_z z)dz\right]\right\rangle, \tag{2.17}$$

where the time dependence of μ has been suppressed and the subscript indicates that for a dispersive medium μ can depend on ω. Let us first compute this average in the absence of dispersion. When μ_ω is independent of ω it is, in the Markov approximation

$$\exp(-\tfrac{1}{2}(k-k')^2R\hat{\rho}(0,0)-kk'$$
$$\times\int_0^R[\hat{\rho}(0,0)-\hat{\rho}(|\mathbf{r}_1(z)-\mathbf{r}_2(z)|,0)]dz). \tag{2.18}$$

For paths which make a significant contribution to the path integral, the second term in the argument of the exponential

must be of order unity or less. In this term one can therefore approximate kk' by $\bar{k}^2$ where $\bar{k}=\frac{1}{2}(k+k')$. Generalizing to dispersive media, one finds that in the same approximation the result is just Eq. (2.18) with $R\hat{\rho}(0,0)$ replaced by $(c_g/\omega_g)^2$, where ω_g was defined in Eq. (1.20). The path integral will also contain a factor

$$\exp\left[\frac{ik}{2}\int_0^R(\mathbf{r}_1'(z))^2dz-\frac{ik'}{2}\int_0^R(\mathbf{r}_2'(z))^2dz\right], \tag{2.19}$$

which can be simplified by making an orthogonal transformation to paths $\mathbf{u}$ and $\mathbf{v}$ defined by

$$\mathbf{r}_1(z)=\mathbf{u}(z)-\frac{k'\mathbf{v}(z)}{k-k'},\quad \mathbf{r}_2(z)=\mathbf{u}(z)-\frac{k\mathbf{v}(z)}{k-k'}. \tag{2.20}$$

After making this transformation the path integral factors into a product of integrals over $\mathbf{u}$ and $\mathbf{v}$. Upon dividing by $\mathscr{E}_0^*(\omega')\mathscr{E}_0(\omega)$ the integral over $\mathbf{u}$ cancels and the final result is, for $|\omega-\omega'|$ small compared to $\bar{\omega}=\frac{1}{2}(\omega+\omega')$,

$$\frac{\langle\mathscr{E}^*(\omega')\mathscr{E}(\omega)\rangle}{\mathscr{E}_0^*(\omega')\mathscr{E}_0(\omega)}=\exp\left[-\frac{1}{2}\left(\frac{\omega-\omega'}{\omega_g}\right)^2\right]\Lambda(\omega-\omega'), \tag{2.21}$$

where

$$\Lambda(\omega-\omega')=\int d\,(\text{paths})\,\exp\left[-\frac{i\bar{k}^2}{2(k-k')}\int_0^R(\mathbf{v}'(z))^2dz-\bar{k}^2\int_0^R[\hat{\rho}(0,0)-\hat{\rho}(\mathbf{v}(z),0)]dz\right]\Big/d\int(\text{paths})\times\exp\left(-\frac{\bar{k}^2}{2(k-k')}\int_0^R(\mathbf{v}'(z))^2dz\right). \tag{2.22}$$

In the saturated region where Φ is large, $|\mathbf{v}(z)|$ will be very small for the important paths and the expansion

$$\hat{\rho}(0,0)-\hat{\rho}(|\mathbf{v}(x)|,0)\approx\tfrac{1}{2}\hat{\rho}(0,0)\left(\frac{\mathbf{v}(x)}{L}\right)^2 \tag{2.23}$$

can be used. The path integral for Λ is then

$$\Lambda(\omega-\omega')=\int d\,(\text{paths})\,\exp\left(-\frac{i\bar{k}^2}{2(k-k')}\times\int_0^R(\mathbf{v}'(z))^2dz-\frac{\bar{k}^2\hat{\rho}(0,0)}{2L^2}\int_0^R(\mathbf{v}(z))^2dz\right)\Big/\int d\,(\text{paths})\,\exp\left[-\frac{i\bar{k}^2}{2(k-k')}\int_0^R(\mathbf{v}'(z))^2dz\right]. \tag{2.24}$$

This type of path integral was evaluated by Feynman and setting $k-k'=(\omega-\omega')/c_g$, it is equal to

$$\Lambda(\omega-\omega')\left(6i\frac{\omega-\omega'}{\omega_0\alpha}\right)^{1/2}/\sin\left(6i\frac{\omega-\omega'}{\omega_0\alpha}\right)^{1/2}, \tag{2.25}$$

where $\omega_0^{-2}=R\hat{\rho}(0,0)/c_g^2$ and $\alpha=6(L^4/R^3\hat{\rho}(0,0))^{1/2}=\Omega/\Phi$.

Combining Eqs. (2.21) and (2.25) yields Eq. (1.19). Some features of these correlations in ω were mentioned in the Introduction. We will return to their interpretation in Sec. 6.

Except for the explicit verification of the validity of the Markov approximation, the above results could be obtained by more familiar techniques which do not employ the path integral. The power of the path integral will become apparent in the next section when higher order moments are computed. They are extremely difficult to treat by the usual techniques.

3. HIGHER MOMENTS FOR SMALL α

When Φ is large, the average of any path integral will be exponentially small unless there is a region of path space where each path associated with an $\mathscr{E}$ is close to a path associated with an $\mathscr{E}^*$. Such a region does not exist for $\langle\mathscr{E}\rangle$ or $\langle\mathscr{E}^2\rangle$ and we have already seen that they are exponentially small. More generally, any moment with an unequal number of $\mathscr{E}$'s and $\mathscr{E}^*$'s will be vanishingly small.

Beyond $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$ the first nontrivial object is $\langle\mathscr{E}^*(4)\mathscr{E}(3)\mathscr{E}^*(2)\mathscr{E}(1)\rangle$. It is given by the quadruple path integral over four paths $\mathbf{r}_1(z),\cdots\mathbf{r}_4(z)$,

$$\langle\mathscr{E}^*(4)\mathscr{E}(3)\mathscr{E}^*(2)\mathscr{E}(1)\rangle=(2k)^{-4}\int d^4(\text{paths})\exp\left(-\frac{ik}{2}\sum_{j=1}^4(-1)^j\times\int_0^R(\mathbf{r}_j'(z))^2dz-M\right). \tag{3.1}$$

where

$$M=-\tfrac{1}{2}\sum_{i,j=1}^4(-1)^{i+j}\int_0^R d\,(|\mathbf{r}_i(z)-\mathbf{r}_j(z)|,t_i-t_j)dz. \tag{3.2}$$

There are two regions of path space where M is of order unity or smaller. They are: (a) $|\mathbf{r}_1(z)-\mathbf{r}_2(z)|<L/\Phi$, $|\mathbf{r}_3(z)-\mathbf{r}_4(z)|<L/\Phi$, with the distance between pairs of paths arbitrary and (b) $|\mathbf{r}_1(z)-\mathbf{r}_4(z)|<L/\Phi$, $|\mathbf{r}_3(z)-\mathbf{r}_2(z)|<L/\Phi$ again with the distance between pairs of paths arbitrary. In region (a) where $|\mathbf{r}_1(z)-\mathbf{r}_2(z)|$ is of order L/Φ, the oscillating factor

$$\exp\left(\frac{ik}{2}\int_0^R[(\mathbf{r}_1'(z))^2-(\mathbf{r}_2'(z))^2]dz\right)\sim\exp\left[-\frac{ik}{2}\int_0^R(\mathbf{r}_1(z)-\mathbf{r}_2(z))\cdot(\mathbf{r}_1''(z)+\mathbf{r}_2''(z))dz\right] \tag{3.3}$$

in the path integral will restrict $|\mathbf{r}_1''(z)+\mathbf{r}_2''(z)|$ to be of order $2\Phi/(kLR)$. For a typical path $|\mathbf{r}_1(z)+\mathbf{r}_2(z)|$ will then be roughly

$$\frac{1}{3}\left(\frac{R}{2}\right)^2|\mathbf{r}_1''(z)+\mathbf{r}_2''(z)|\sim\frac{\Phi R}{6kL}.$$

The centroid of the other path $\mathbf{r}_3(z)+\mathbf{r}_4(z)$ will be restrained in a similar way. It follows that most paths will be such that the ratio of the distance between the pairs to the scale length L is roughly $\Phi R/(6kL^2)=\alpha^{-1}$, where α is the parameter de-

fined in the Introduction. For small α the pairs are separated by many times L and therefore are uncorrelated. In region (a) M then reduces to

$$M \approx \int_0^R d\,(|\mathbf{r}_1(z)-\mathbf{r}_2(z)|, t_1-t_2)dz + \int_0^R d\,(|\mathbf{r}_3(z)-\mathbf{r}_4(z)|, t_3-t_4)dz \tag{3.4}$$

and in region (b) it becomes

$$M \approx \int_0^R d\,(|\mathbf{r}_3(z)-\mathbf{r}_2(z)|, t_3-t_2)dz + \int_0^R d\,(|\mathbf{r}_1(z)-\mathbf{r}_4(z)|, t_1-t_4)dz. \tag{3.5}$$

Thus in each of the two important regions of path space, the quadruple path integral factors into the product of two double path integrals, each of which is precisely the integral encountered in the calculation of $\langle \mathscr{E}^*\mathscr{E}\rangle$. The result is that

$$\langle \mathscr{E}^*(4)\mathscr{E}(3)\mathscr{E}^*(2)\mathscr{E}(1)\rangle \approx \langle \mathscr{E}^*(4)\mathscr{E}(3)\rangle\langle \mathscr{E}^*(2)\mathscr{E}(1)\rangle + \langle \mathscr{E}^*(2)\mathscr{E}(3)\rangle\langle \mathscr{E}^*(4)\mathscr{E}(1)\rangle, \tag{3.6}$$

where the two terms come from the two regions (a) and (b).

In Appendix B the error in Eq. (2.6) is obtained by computing the first correction. It is of order α and will be discussed in detail in Sec. 6.

Generalizing to an arbitrary moment is easy. The general nonvanishing moment is

$$\left\langle \prod_{j=1}^{n} \mathscr{E}^*(j) \prod_{i=1}^{n} \mathscr{E}(i) \right\rangle$$

and can be written as an integral over $2n$ paths $\mathbf{r}_j(z)$ and $\mathbf{r}_i(z)$. There will now be $n!$ important regions of path space corresponding to the number of ways paths $\mathbf{r}_j(z)$ can be paired with the paths $\mathbf{r}_i(z)$. In each of these regions the $2n$-tuple path integral can be approximated by a product of n double path integrals. Some simple combinatorics shows that the result will be as follows. Let $\bar{i}$ be a permutation of the indices i. For example, if $n=3$ and the permutation is $(1,2,3)\to(3,1,2)$, then $\bar{1}=3$, $\bar{2}=1$, and $\bar{3}=2$ or if the permutation is $(1,2,3)\to(2,1,3)$, then $\bar{1}=2$, $\bar{2}=1$ and $\bar{3}=3$. With this notation

$$\left\langle \prod_{j=1}^{n} \mathscr{E}^*(j) \prod_{i=1}^{n} \mathscr{E}(i) \right\rangle \approx \sum_{\text{perms}} \prod_{i,j=1}^{n} \langle \mathscr{E}^*(j)\mathscr{E}(\bar{i})\rangle, \tag{3.7}$$

where the sum is over all $n!$ possible permutations of the indices i.

The same result holds for correlations in frequency. Extending the notation $\mathscr{E}(j)$ to include a frequency label ω_j we have

$$\frac{\langle \mathscr{E}^*(j)\mathscr{E}(i)\rangle}{\mathscr{E}_0^*(j)\mathscr{E}_0(i)} = \exp\left[-\tfrac{1}{2}D(i,j) - \frac{1}{2}\left(\frac{\omega_i-\omega_j}{\omega_g}\right)^2\right]\Lambda(\omega_i-\omega_j), \tag{3.8}$$

which holds in the saturated region where only small values of $\omega_i-\omega_j$, $|\mathbf{r}_i-\mathbf{r}_j|$, etc., are interesting. The same construction that led to Eq. (3.7) for equal frequencies then shows that it holds for unequal frequencies as well.

The interpretation of Eq. (3.7) will be given in the next section. A final remark here is that the arguments leading to Eq. (3.7) do not depend on the validity of the Markov approximation. The latter is needed only when $\langle \mathscr{E}^*(1)\mathscr{E}(2)\rangle$ is explicitly evaluated.

4. THE STATISTICS OF $\mathscr{E}$ IN THE LIMIT $\alpha=0$

The moments of Eq. (3.7) correspond to a complex Gaussian distribution. The probability that $\mathscr{E}(j)/\mathscr{E}_0(j)$ will be equal to η_j for $j=1,\ldots,n$ is then

$$P_n(\eta_1,\ldots,\eta_n) = (\det[\pi M])^{-1} \exp\left[-\sum_{i,j=1}^{n} \eta_i^*(M)_{ij}^{-1}\eta_j\right], \tag{4.1}$$

where the n by n matrix M is

$$M_{ji} = \frac{\langle \mathscr{E}^*(i)\mathscr{E}(j)\rangle}{\mathscr{E}_0^*(i)\mathscr{E}_0(j)} = \exp\left[-\tfrac{1}{2}D(i,j) - \frac{1}{2}\left(\frac{\omega_j-\omega_i}{\omega_g}\right)^2\right]\Lambda(\omega_j-\omega_i). \tag{4.2}$$

Equation (1.10) corresponds to the special case $n=1$ and Eq. (1.18) is obtained by dividing $P_2(\eta,\eta')$ by $P_1(\eta)$. The measure is $d^2\eta = d(\mathrm{Im}\eta)d(\mathrm{Re}\eta)$.

In principle, Eqs. (4.1) and (4.2) determine all the statistical properties of $\mathscr{E}$. For example, it follows from Gaussian statistics that for $\mathscr{E}(j)=A(j)e^{i\phi(j)}$ the correlations of amplitude and rate of phase $\dot{\phi}=d\phi/dt$ are[18]

$$\frac{\langle A(1)A(2)\rangle}{\langle I\rangle} = E(|M_{12}|) - \tfrac{1}{2}(1-|M_{12}|^2)K(|M_{12}|), \tag{4.3}$$

where E and K are the complete elliptic integrals of the first and second kinds and

$$\langle \dot{\varphi}(1)\dot{\varphi}(2)\rangle = -\frac{\ddot{D}(1,2)}{4}\ln(1-|M_{12}|) = -\frac{1}{2}\frac{\Phi^2}{T^2}\ln(1-|M_{12}|), \tag{4.4}$$

where in the second line in the expansion of D in Eq. (1.17) has been used. Equation (4.4) can be extended to the derivatives of φ with respect to $\mathbf{r}$ and $\mathbf{r}_0$ in the obvious way. Intensity correlations $I(j)=|A(j)|^2$ are simpler with

$$\langle I(1)I(2)\rangle = \langle I\rangle^2 + |\langle \mathscr{E}^*(1)\mathscr{E}(2)\rangle|^2. \tag{4.5}$$

The appearance of Gaussian statistics in frequency is

somewhat unfamiliar. To see what it implies, let us compare the saturated regime to a simpler unsaturated one. When simple first order geometric optics applies and the medium is nondispersive, $\mathscr{E}(\omega)$ is equal to $\mathscr{E}_0(\omega)e^{i\omega\hat{\tau}}$, where $\hat{\tau}$ is a fluctuating time shift independent of ω. Under such propagation conditions, the statistics in ω are essentially trivial. The envelope $\mathscr{E}(\omega)$ fluctuates but does so in such a way that at a fixed time when $\hat{\tau}$ has a definite value a knowledge of $\mathscr{E}$ at one value of ω determines $\mathscr{E}$ for all ω. Another way to say the same thing is that a pulse will be subjected to a random time shift but will not be distorted in shape. For propagation in the saturated regime the statistics of $\mathscr{E}(\omega)$ are nontrivial and things are completely different. At one fixed time a knowledge of $\mathscr{E}(\omega)$ at one ω yields only statistical information about $\mathscr{E}$ at nearby frequencies. Correspondingly, the medium will distort a pulse in a way that is predictable only statistically. A peculiarity is that $\langle\mathscr{E}^*(\omega')\mathscr{E}(\omega)\rangle$ and has a phase corresponding to an average retardation.

The above remarks about $\mathscr{E}(\omega)$ are most easily made quantitative in terms of pulse propagation. It is worth going into this in some detail both because the physics is intetesting and because it will connect with the Fermat paths of Sec. 6. For simplicity the unperturbed medium will be assumed to be nondispersive with $\omega=ck$. Let the transmitted signal be $f_0(\tau)=\int e^{-i\omega\tau}\tilde{f}_0(\omega)d\omega$ where $\tilde{f}_0(-\omega)=\tilde{f}_0^*(\omega)$. Taking the unperturbed arrival time as the origin, the received signal will then be $f_r(\tau)=\int e^{-i\omega\tau}\mathscr{E}(\omega)\tilde{f}_0(\omega)d\omega$. The signal $f_r(\tau)$ is a Gaussian random variable whose complete statistics are determined by the covariance of $\tilde{f}_0\mathscr{E}$. Assuming[19] that variations in $\mathscr{E}_0(\omega)\tilde{f}_0(\omega)$ over a frequency corresponding to the width of $\langle\mathscr{E}^*(-\frac{1}{2}\omega)\mathscr{E}(\frac{1}{2}\omega)\rangle$ can be neglected, this covariance is

$$\langle\tilde{f}_0^*(\omega')\mathscr{E}^*(\omega')\tilde{f}_0(\omega)\mathscr{E}(\omega)\rangle=|\tilde{f}_0(\tfrac{1}{2}(\omega+\omega'))|^2\Lambda(\omega-\omega'), \tag{4.6}$$

with

$$\Lambda(\omega)=\left(i\frac{\omega}{\omega_1}\right)^{\frac{1}{2}}/\sin\left(i\frac{\omega}{\omega_1}\right)^{\frac{1}{2}},$$

where the small α limit of Eq. (1.19) has been used and $\omega_1=\omega_0\alpha/6=c_\chi L^2/R^2\hat{\rho}(0,0)$.

Denoting the received intensity $f_r^2(\tau)$ by $\mathscr{I}(\tau)$, the average $\langle\mathscr{I}(\tau)\rangle$ is a measure of the distribution of energy over arrival times. According to Eq. (4.6) it is

$$\frac{\langle\mathscr{I}(\tau)\rangle}{\left\langle\int_{-\infty}^{\infty}\mathscr{I}(\tau)d\tau\right\rangle}=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega\tau}\Lambda(\omega)d\omega$$
$$=\begin{bmatrix}0 & \tau<0\\ -2\omega_1\sum_{n=1}^{\infty}(-1)^n(n\pi)^2\exp(-n^2\pi^2\omega_1\tau) & \\ & \tau>0\end{bmatrix} \tag{4.7}$$

and vanishes for $\tau<0$ because the integrand is analytic in the upper-half plane. Evidently, all the energy comes in after the unperturbed arrival time and is confined to a region $0\lesssim\tau\lesssim(\omega_1\pi^2)^{-1}$. The net retardation is consistent with what was said above about the phase of $\langle\mathscr{E}^*(\omega')\mathscr{E}(\omega)\rangle$. The complete absence of energy for $\tau<0$ is peculiar to the limit of small α and will later be seen to have a simple physical interpretation (see Sec. 6).

For a sharp transmitted pulse the distribution of energy over arrival times can be thought of as being due to two effects. One is the wander in arrival time of the center of the pulse and the other is spreading of the pulse around its center. The two effects are in principle distinct. For the simple case of propagation in an unsaturated regime where $\mathscr{E}(\omega)=e^{i\omega\hat{\tau}}$, the wander is of order $\langle\hat{\tau}^2\rangle^{1/2}$ while the spread is just the width of the transmitted pulse. As we will now see, in the saturated regime the spread and wander are roughly equal. The width of $\langle\mathscr{I}(\tau)\rangle$ measures the sum of spread and wander. A quantity which measures the spreading, independent of wander, is

$$P(\tau)=\frac{\left\langle\int_{-\infty}^{\infty}\mathscr{I}(\tau+\tau')\mathscr{I}(\tau')d\tau'\right\rangle}{\left(\int_{-\infty}^{\infty}\langle\mathscr{I}(\tau')\rangle d\tau'\right)^2}. \tag{4.8}$$

When $\tilde{f}_0\mathscr{E}$ has a Gaussian distribution $P(\tau)$ is

$$P(\tau)=\frac{P_0(\tau)}{2\pi}\int_{-\infty}^{\infty}|\Lambda(\omega)|^2d\omega+\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega\tau}|\Lambda(\omega)|^2d\omega, \tag{4.9}$$

where

$$P_0(\tau)=\frac{\left(\int_{-\infty}^{\infty}f_0(\tau+\tau')f_0(\tau')d\tau'\right)^2}{\left(\int_{-\infty}^{\infty}f_0^2(\tau')d\tau'\right)^2}. \tag{4.10}$$

The two terms in $P(\tau)$ have the same height at $\tau=0$. The spike proportional to $P_0(\tau)$ falls rapidly leaving the second term whose width is a measure of the spread. Comparing with Eq. (4.7) one sees that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega\tau}|\Lambda(\omega)|^2d\omega=\frac{\int_{-\infty}^{\infty}\langle\mathscr{I}(\tau+\tau')\rangle\langle\mathscr{I}(\tau')\rangle d\tau'}{\left(\int_{-\infty}^{\infty}\langle\mathscr{I}(\tau')\rangle d\tau'\right)^2} \tag{4.11}$$

and it is clear that the spread and wander are essentially the same. A physical interpretation of the two pieces of $P(\tau)$ will be given in Sec. 5. Finally, a useful formula is

$$|\Lambda(\omega)|^2=\left|\frac{\omega}{\omega_1}\right|/\left(\sin^2\left(\frac{\omega}{2\omega_1}\right)^{1/2}+\sinh^2\left(\frac{\omega}{2\omega_1}\right)^{1/2}\right). \tag{4.12}$$

It is interesting to ask why it is that the square of the autocorrelation of f_0 rather than the autocorrelation of f_0^2 appears in P_0. The answer is that when $\tilde{f}_0(\omega)\mathscr{E}(\omega)$ has a Gaussian distribution, the medium cannot transmit any information that is not contained in the coherence $\langle\tilde{f}_0^*(\omega')\mathscr{E}^*(\omega')\tilde{f}_0(\omega)\mathscr{E}(\omega)\rangle$. As given by Eq. (4.6) this coherence depends only on $|\tilde{f}(\omega)|^2$ and the medium can only transmit information about the autocorrelation of f.

The statistics of the signal as a function of spatial wave numbers can be analyzed in a similar way. Multiplying

$\mathscr{E}(\mathbf{r},\mathbf{r}_0,t)$ by a suitable function of $\mathbf{r}_0$ and integrating over $\mathbf{r}_0$ one can represent a boundary condition at $z=0$ corresponding to, say, a plane wave emerging from a finite aperture. For such a signal, the Fourier transform

$$\tilde{\mathscr{E}}_0(\mathbf{l}) = \frac{1}{(2\pi)^2}\int e^{-i\mathbf{r}\cdot\mathbf{l}}\mathscr{E}_0(\mathbf{r})d^2r \tag{4.13}$$

will be sharply peaked around some $\mathbf{l}=\mathbf{l}_0$. With the correspondences $t\to\mathbf{l},\omega\to\mathbf{r}$ one can proceed as above and discuss spread and wander in $\mathbf{l}$. Again the medium can only transmit information contained in $\langle\mathscr{E}^*(\mathbf{r})\mathscr{E}(\mathbf{r}')\rangle$ which will typically depend only on $|\mathscr{E}_0(\frac{1}{2}(\mathbf{r}+\mathbf{r}'))|^2$.

5. CORRECTION TO THE $\alpha=0$ LIMIT

The leading corrections to the $\alpha=0$ limit are computed in Appendix B. The main results are as follows.

The order α correction to $\langle I^n\rangle$ is dominated by fluctuations near the transmitter and receiver. This is not unexpected since near their end points the paths cannot be separated into uncorrelated pairs. Explicitly $\langle I^n\rangle$ is

$$\langle I^n\rangle = n!\langle I\rangle^n[1+\tfrac{1}{2}n(n-1)\alpha C+O(\alpha^2)], \tag{5.1}$$

where

$$C = \frac{(3\pi)^{\frac{1}{2}}}{4}\frac{L\int_0^\infty q^2\tilde{\rho}(q,0)dq}{\int_0^\infty q\tilde{\rho}(q,0)dq} \tag{5.2}$$

and $\tilde{\rho}$ is the three-dimensional Fourier transform of ρ

$$\rho(|\mathbf{x}|,t) = \frac{4\pi}{|\mathbf{x}|}\int_0^\infty q\sin(q|\mathbf{x}|)\tilde{\rho}(q,t)dq. \tag{5.3}$$

The consequences of the fact that the error grows with n were noted in the Introduction. Note that the correction to $\langle I^n\rangle$ is positive. This means that the intensity fluctuations overshoot (i.e., become larger than Rayleigh) near the boundaries of the saturated regime.

The correction to a general correlation can also be computed. They are always fractionally small. For example, $\langle\mathscr{E}^*(2)\mathscr{E}^*(2)\mathscr{E}(1)\mathscr{E}(1)\rangle$ is proportional to $e^{-D(1,2)}$ in the $\alpha=0$ limit and the correction to it is of order $\alpha e^{-D(1,2)}$. The corrections to intensity correlations are the most interesting. In the $\alpha=0$ limit $\langle I(t_1)I(t_2)\rangle$ is equal to $\langle I\rangle^2\{1+\exp[-D(t_1-t_2)]\}$. At $t_1=t_2$ the order α correction is given by Eq. (5.1). However, at large $|t_1-t_2|$, $\langle I(t_1)I(t_2)\rangle$ must approach $\langle I\rangle^2$ and the corrections must go to zero. It turns out that half the correction dies like $\exp[-D(t_1-t_2)]$ but the other half falls much more slowly, leading to a coherence tail. (Note that this is consistent with what was said above about the corrections always being fractionally small.) For the general intensity correlation the coherence tail[4] is

$$\langle I(\mathbf{r}_1,\mathbf{r}_{01},t_1)I(\mathbf{r}_2,\mathbf{r}_{02},t_2)\rangle$$
$$= \langle I\rangle^2\Big[1+e^{-D(1,2)}+(\alpha\sqrt{3\pi}/8)L$$
$$\int_0^\infty q^2\tilde{\rho}(q,t_1-t_2)[J_0(|\mathbf{r}_1-\mathbf{r}_2|q)$$
$$+J_0(|\mathbf{r}_{01}-\mathbf{r}_{02}|q)]dq/\int_0^\infty q\tilde{\rho}(q,0)dq\Big], \tag{5.4}$$

where J_0 is a Bessel function and specializing to $\mathbf{r}_1=\mathbf{r}_2$ and $\mathbf{r}_{01}=\mathbf{r}_{02}$ produces

$$\langle I(t_1)I(t_2)\rangle$$
$$= \langle I\rangle^2\Big[1+\exp[-D(t_1-t_2)]$$
$$+\frac{\alpha\sqrt{3\pi}}{4}\frac{L\int_0^\infty q^2\tilde{\rho}(q,t_1-t_2)dq}{\int_0^\infty q\tilde{\rho}(q,0)dq}\Big]. \tag{5.5}$$

Equation (5.1) is not reproduced at $t_1=t_2$ because a term of order $\alpha e^{-D(1,2)}$ has been dropped from both Eqs. (5.4) and (5.5).

6. FERMAT PATHS

There is an interesting connection between averages of the path integral and averages over Fermat paths which satisfy the perturbed ray equation

$$\mathbf{r}''(z)+\nabla\mu(\mathbf{r}(z)+\mathbf{e}_z z)=0, \tag{6.1}$$

where $\nabla=(\partial/\partial x,\partial/\partial y)$. This will be illustrated for the special case of sources and receivers located at $\mathbf{r}=\mathbf{r}_0=0$ so that $\mathscr{E}$ is a function only of time. The path integral for $\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle$ is

$$\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle = \Big\langle\frac{1}{4k^2}\int d^2(\text{paths})$$
$$\times\exp\Big[ik\int_0^R[\tfrac{1}{2}(\mathbf{r}_1'(z))^2-\tfrac{1}{2}(\mathbf{r}_2'(z))^2$$
$$-\mu(\mathbf{r}_1(z)+\mathbf{e}_z z,0)+\mu(\mathbf{r}_2 z+\mathbf{e}_z(z),t)]dz\Big]\Big\rangle. \tag{6.2}$$

In the saturated region we know that for $\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle$ to be nonvanishing, t must be small and that only paths for which $|\mathbf{r}_1(z)-\mathbf{r}_2(z)|$ is small ($\sim L/\Phi$) contribute. Changing variables to $\mathbf{w}(z)=\frac{1}{2}(\mathbf{r}_1(z)+\mathbf{r}_2(z))$ and $\mathbf{v}(z)=\mathbf{r}_1(z)-\mathbf{r}_2(z)$, the path integral can then be approximated by

$$\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle$$
$$= \Big\langle\frac{1}{4k^2}\int d^2(\text{paths})\exp\Big(-ik\int_0^R\Big\{\mathbf{v}(z)\cdot\Big[\mathbf{w}''(z)$$
$$+\nabla\mu[\mathbf{w}(z)+\mathbf{e}_z z,0]\Big]-t\dot{\mu}(\mathbf{w}(z)+\mathbf{e}_z z,0)\Big\}dz\Big)\Big\rangle, \tag{6.3}$$

where the first term in the argument of the exponential has been integrated by parts and a dot indicates differentiation with respect to time. The integration over the path $\mathbf{v}(z)$ produces a "δ functional" which forces $\mathbf{w}''+\nabla\mu$ to vanish for all z. Thus the integral over $\mathbf{w}(z)$ is restricted to paths which satisfy the ray equation (6.1). This is a general feature of the saturated region. Higher order correlations are dominated by configurations where paths $\mathbf{r}_i(z)$ and $\mathbf{r}_{i'}(z)$ are pairwise close. A similar analysis shows that for each such pair, the path $\mathbf{w}_i(z)=\frac{1}{2}[\mathbf{r}_i(z)+\mathbf{r}_{i'}(z)]$ satisfies Eq. (6.1).

Equation (6.3) can be further analyzed. For most media $\dot{\mu}$ and μ are satistically independent. The average, $\langle\ \rangle$, can then be thought of as two independent averages $\langle\ \rangle_\mu$ and $\langle\ \rangle_{\dot{\mu}}$ over μ and $\dot{\mu}$. The δ function of $\mathbf{w}''+\nabla\mu$ that is produced by the integration over $\mathbf{v}$ is effected only by the average $\langle\ \rangle_\mu$, while the phase $\exp[ikt\int\dot{\mu}]$ is effected only by the other average $\langle\ \rangle_{\dot{\mu}}$. It is therefore possible to write $\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle$ as the integral over paths $\mathbf{w}$ of a μ-averaged δ functional which can be interpreted as the probability that a given path $\mathbf{w}$ will satisfy Eq. (6.1) times a phase which is to be averaged over $\dot{\mu}$. To do this correctly, it is necessary to go back to the definition of the path integral in Eq. (2.1). The integration variables $\mathbf{v}_k$ and $\mathbf{w}_k$, $k=1,2,\ldots,n$ are then discrete and the mathematics is straightforward. The integration over the $\mathbf{v}$'s can be done trivially and after some manipulation, one finds

$$\langle\mathscr{E}^*(t)\mathscr{E}(o)\rangle=\int d\,(\text{paths})'P(\text{path})\times\left\langle\exp\left[ikt\int_0^R\dot{\mu}(\mathbf{w}(z)+\mathbf{e}_z z,0)dz\right]\right\rangle_{\dot{\mu}},\tag{6.4}$$

where the integration $d\,(\text{paths})'$ is over paths $\mathbf{w}(z)$ with a modified volume element

$$d\,(\text{paths})'=\frac{1}{(4\pi)^2}\left(\prod_{j=1}^{n}d^2w_j\right)\left(\frac{n}{R}\right)^{2n},\tag{6.5}$$

which does not contain k and

$$P(\text{path})=\left(\frac{n}{R}\right)^{2n-2}\left\langle\prod_{j=1}^{n-1}\delta^2\left(\frac{n^2}{R^2}(\mathbf{w}_{j+1}+\mathbf{w}_{j-1}-2\mathbf{w}_j)+\nabla\mu(\mathbf{w}_j+\mathbf{e}_z z_j,0)\right)\right\rangle_\mu\tag{6.6}$$

is the probability that $\mathbf{w}$ will satisfy the finite difference approximation

$$\frac{n^2}{R^2}(\mathbf{w}_{j+1}+\mathbf{w}_{j-1}-2\mathbf{w}_j)+\nabla\mu(\mathbf{w}_j+\mathbf{e}_z z_j,0)=0\tag{6.7}$$

to the ray equation. In the limit $n\to\infty$, $P(\text{path})$ is the probability [with a measure $d\,(\text{paths})'$] that $\mathbf{w}$ will satisfy Eq. (6.1). Equation (6.4) shows explicitly that $\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle$ is a sum over Fermat paths with fluctuating phases $kt\int\dot{\mu}$. Finally, bringing the average of $\dot{\mu}$ inside the exponential yields

$$\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle=\int d\,(\text{paths})'P(\text{path})\times\exp\left[\frac{-k^2t^2}{2}\left\langle\left(\int_0^R\dot{\mu}(\mathbf{w}(z)+\mathbf{e}_z z,0)dz\right)^2\right\rangle_{\dot{\mu}}\right].\tag{6.8}$$

In the Markov approximation where $\mathbf{w}(z)$ is neglected relative to $\mathbf{e}_z z$ in the average of $\dot{\mu}$, Eq. (6.8) becomes

$$\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle=\exp\left[-\tfrac{1}{2}\Phi^2\left(\frac{t}{T}\right)^2\right]\int d\,(\text{paths})'P(\text{path})=\exp\left[-\tfrac{1}{2}\Phi^2\left(\frac{t}{T}\right)^2\right]\langle I\rangle\tag{6.9}$$

which is the standard result.

This provides a new way to look at the Markov approximation. It requires that an average like $\langle(\int_0^R\dot{\mu}(\mathbf{w}(z)+\mathbf{e}_z z,0)dz)^2\rangle$ along a path which satisfies the perturbed ray equation (6.1) should be well approximated by the corresponding average $\langle(\int_0^R\dot{\mu}(\mathbf{e}_z z,0)dz)^2\rangle$ along the unperturbed ray. For a homogeneous and isotropic medium, this will be the case as long as the rms multiple scattering angle $(\langle\mu^2\rangle R/L)^{\frac{1}{2}}$ is small.

According to Eq. (6.8), $\langle\mathscr{E}^*(t)\mathscr{E}(0)\rangle$ can in principle be computed by a geometric optics method which searches out the rays which satisfy the perturbed ray equation. Geometric optics corresponds to an approximate evaluation of the path integral by the method of stationary phase.[6] In the saturated region the stationary phase approximation will in fact be valid since for $\Phi>1$ the phase $k\int_0^R\mu dz$ is necessarily large. To get Gaussian statistics for $\mathscr{E}$, it is necessary that there be several rays connecting a given source and receiver. In path integral language this means that there will be multiple stationary phase points and $\mathscr{E}$ will be a discrete sum $\Sigma_k A_k e^{i\varphi_k}$ over contributions, one from each stationary phase point or ray. The phases φ_k and amplitudes A_k as well as the number of rays will fluctuate with μ yielding Gaussian statistics for $\mathscr{E}$.

It is difficult to prove rigorously that there are always multiple rays in the saturated regime. However there is a simple construction which shows the essential physics. At one fixed time the rays are stationary points of the path length S defined by

$$S=k\int_0^R[\tfrac{1}{2}(\mathbf{r}'(z))^2-\mu(\mathbf{r}(z)+\mathbf{e}_z z)]dz.\tag{6.10}$$

Let $S(\mathbf{r})$ be S evaluated for the special paths that go in a straight line from the source at $(\mathbf{0},0)$ to an arbitrary point $(\mathbf{r},z_0)$ with $0<z_0<R$ and then follow another straight line from (r,z_0) to the receiver at $(\mathbf{0},R)$. Multiple stationary points of $S(\mathbf{r})$ as a function of $\mathbf{r}$ will be indicative of multiple stationary points of the complete functional in Eq. (6.10) and the spacing of such points in $\mathbf{r}$ will be similar to the spacing between multiple rays. Now doing a simple integral shows that $S(\mathbf{r})$ can be written as

$$S(\mathbf{r})=\tfrac{1}{2}(\mathbf{r})^2B-S_1(\mathbf{r}),\tag{6.11}$$

where $B=kR/z_0(R-z_0)$ and S_1 is $k\mu$ integrated along the above mentioned path. To simplify $S(\mathbf{r})$, B^{-1} can be replaced by its average value $R/(6K)$. Then defining $\mathbf{r}=L\mathbf{u}$ and $S_1(\mathbf{r})=\Phi f(\mathbf{u})$ the quantity to be studied is

$$\tfrac{1}{2}\Omega\mathbf{u}^2-\Phi f(\mathbf{u}) \tag{6.12}$$

and we are interested in its stationary points which satisfy

$$\Omega\mathbf{u}-\Phi\nabla f(\mathbf{u})=0. \tag{6.13}$$

By construction f is a random function of order unity which changes by order one when its argument changes by order one; i.e., $|\nabla f|\sim 1$ and ∇f changes sign roughly each unit in $\mathbf{u}$. For $\Omega\gg\Phi$ the first term in Eq. (6.13) dominates and there will be a single solution near $\mathbf{u}=0$. This is the unsaturated regime. In the saturated regime, $\Phi\gg\Omega$, the random character of f guarantees that there will generally be many solutions, spaced by about one unit in $\mathbf{u}$ (a distance l in $\mathbf{r}$) and filling up the interval $0<|\mathbf{u}|<\Phi/\Omega$ $(0<|\mathbf{r}|<\Phi L/\Omega)$. To find the other boundary of the saturated regime, $\Phi>1$, we have to ask when the multiple rays are physically meaningful. From their interpretation as stationary phase points of the path integral it can be verified that two rays will be physically distinct if S varies by a quarter cycle, i.e., order unity between the two. The variation in S between two solutions of Eq. (6.13) will be roughly Φ and if they are to represent physically distinct rays Φ must be greater than unity.

An experiment with a pulsed source will tend to see several arrivals corresponding to the multiple Fermat paths. This random multipathing is the origin of the rapid falloff of frequency coherence which takes place in the saturated regime. To see how the orders of magnitude work, the difference in travel time between the ray nearest $\mathbf{u}=0$ and the furthest one out at $|\mathbf{u}|\sim\Phi/\Omega$ is $t_0=\omega^{-1}(\Phi^2/2\Omega\pm\Phi)\approx\Phi^2/(2\Omega\omega)$ where the two terms come from the two terms in Eq. (6.12) and it has been assumed that $\Phi\gg\Omega$. Frequencies which differ by more than t_0^{-1} will then be incoherent in agreement with Eqs. (1.19) and (1.21). Note that t_0 is positive. This is why in the limit $\alpha=0$ all the energy arrives after the unperturbed arrival time and $\langle\mathscr{I}(\tau)\rangle$ vanishes for $\tau<0$. Also the two terms in $P(\tau)$ (Sec. 4) can easily be interpreted in terms of fluctuating multipath. The spike $P_0(\tau)$ is the autocorrelation of each arrival with itself and the broad second term is the autocorrelation of different arrivals. Finally a word of caution. The above construction vastly underestimates the number of rays. In reality the number of rays is probably an exponential of Φ/Ω rather than Φ/Ω as the construction would imply. It may be extremely difficult to actually resolve the arrivals.

It is interesting to consider the transition into the saturated regime in terms of propagation of a pulse. Consider first crossing the line $\Phi=\Omega$ from the region where both Φ and Ω are large but $\Omega<\Phi$. With Φ and Ω large but well outside the saturated region, one knows from the Rytov approximation that the receiver will see a single arrival with a considerable wander in time of arrival. At the boundary of the saturated region the pulse will begin to split into several arrivals and well inside the saturated region there will be many arrivals that are spread out over a time long compared to the original wander in the single pulse. Crossing the boundary $\Phi=1$ from the region where both Φ and Ω are small is rather different. In this case one knows that well outside the saturated region, there will be a single arrival with no discernible wander in time of arrival accompanied by a small scattered wave spread over a continuum of arrival times. As the boundary of the saturated region is approached, the single peak will shrink and the scattered wave will grow in amplitude. Well inside the saturated region, the original peak will have disappeared completely and the now large scattered wave will have broken up into a number of discrete arrivals.

7. MEDIA WITH MULTIPLE SCALES

So far it has been assumed that the fluctuations in μ can be characterized by a single scale size L. Technically, this requires that the expansion of $\hat{\rho}$,

$$k^2R\hat{\rho}(|\mathbf{r}|,0)=\Phi^2\left(1-\frac{\mathbf{r}^2}{2L^2}+a\frac{\mathbf{r}^4}{4L^4}+\cdots\right) \tag{7.1}$$

through order $\mathbf{r}^4$ exists and that the coefficient a is of order unity. There are cases of practical importance where this is not true. For example, optical index of refraction fluctuations induced by Kolomogorov turbulence have the property that the (three-dimensional) Fourier transform $\tilde{\rho}(q)$ of ρ behaves like $|q|^{-11/3}$ over a long interval in q and the expansion in Eq. (7.1) makes sense only when the cutoff (inner scale) is taken into account and then a is very large. This and the following section are devoted to these media with multiple scales. It will be assumed that $\tilde{\rho}(q)$ goes like $|q|^{-2-p}$ for large q where $4>p>1$. (If p is greater than four, the medium acts like one with a single scale size and for $p<1$ it is so singular that $\langle\mu^2\rangle$ does not exist.) In practice there is always some physical cutoff at large q (inner scale). However, the effects of such a cutoff will be ignored in what follows.

For $p>2$, the length parameter L will be defined by Eq. (1.13) as before and in the case $p<2$, L will be defined by

$$\hat{\rho}(|\mathbf{r}|,0)=\hat{\rho}(0,0)\left[1-\frac{1}{2}\left|\frac{\mathbf{r}}{L}\right|^p\right] \tag{7.2}$$

for small $|\mathbf{r}|$. For Kolomogorov turbulence, p is equal to and $\hat{\rho}(0,0)$ and L are related to Tatarskii's C_n by $2.91C_n^2=\hat{\rho}(0,0)L^{-5/3}$. The parameters Φ and Ω continue to be defined by Eqs. (1.6) and (1.7).

The main qualitative difference between propagation in single and multiple scale media is that in the latter case there is more than one saturated regime. In terms of the Fermat paths of the last section, it turns out that in a multiple scale medium the smaller scale inhomogeneities can make multiple Fermat paths before the large ones do. This leads to a new kind of saturated regime. Even in a single scale medium with $p>4$ the line $\Phi=\Omega$ is not a sharp boundary. In reality there is a transition zone where random focusing along single Fermat paths produces intensity fluctuations bigger than Rayleigh. As p decreases below four this transition zone opens up and becomes a new saturated regime. The boundaries of this new regime can be found by studying the object $\frac{1}{2}\Omega(\mathbf{u})^2-\Phi f(\mathbf{u})$ of Eq. (6.12).

To see when the smaller scales can make multiple Fermat paths, imagine throwing out all scale sizes larger than λL where $1>\lambda>0$. The new scale length will be λL and Φ and Ω will be replaced by $\lambda^{p/2}\Phi$ and $\lambda^2\Omega$. The combination Φ/Ω becomes $\lambda^{(p-4)/2}\Phi/\Omega$ and is equal to unity when $\lambda=(\Phi/\Omega)^{2/(4-p)}$. Thus if $p<4$ the small scales can make multiple Fermat paths when $\Phi<\Omega$, i.e., before the large ones do at $\Phi=\Omega$. However, if these multiple paths are to be physically meaningful $\lambda^{p/2}\Phi$ must be greater than unity and the smallest permissible value of λ is $\Phi^{-2/p}$. Putting everything together, the small scales can make meaningful Fermat paths when $\Phi^{4/p}\Omega>1$. This is one boundary of the new saturated regime. To find the other boundary, we need to ask when the multiple Fermat paths can be separated by L. For a given λ the minima of $\frac{1}{2}\lambda^2\Omega\mathbf{u}^2-\lambda^{p/2}\Phi f(\mathbf{u})$ extend out to a maximum $|\mathbf{u}|$ which is the largest value of $|\mathbf{u}|$ for which the equation $\lambda^2\Omega\mathbf{u}-\lambda^{p/2}\Phi\nabla f(\mathbf{u})$ can be solved. The maximum $|\mathbf{u}|$ is $(\Phi/\Omega)\lambda^{(p-4)/2}$ and noting that $\mathbf{u}$ is distance in units of λL one sees that the Fermat paths can be separated by L when $(\Phi/\Omega)\lambda^{(p-2)/2}=1$. For $p>2$ the most separated paths are due to large scales with $\lambda=1$ and the other boundary of the new region is $\Phi=\Omega$. However, if $p<2$ the smaller scales produce the largest separation and taking the smallest permissible value $\Phi^{-2/p}$ for λ one sees that there can be Fermat paths separated by L when $\Phi^{2/p}/\Omega>1$. The regime where there are meaningful multiple Fermat paths all lying within L of each other will be called the *partially saturated regime*. The regime where the spacing between Fermat paths can be greater than L is analagous to the saturated regime of the single scale case and will be called the *fully saturated regime*. The boundaries of these regimes are summarized in Table I.

Although these boundaries have been obtained with a hueristic Fermat path argument they are in agreement with what one finds from more precise calculations. It is known that outside the saturated regimes the intensity fluctuations $(\langle I^2\rangle-\langle I\rangle^2)/\langle I\rangle^2$ are small, implying both the validity of the Rytov approximation and the absence of saturation. Inside the saturated regimes (as given by Table I) the intensity fluctuations as computed in the Rytov approximation are large, signaling the onset of saturation. The line between the fully and partially saturated regimes corresponds to the place where two pairs of paths, in the sense of Sec. 3, can be separated by more than L. When they are separated by more than L the pairs of paths are completely independent (full saturation) and Gaussian statistics for $\mathscr{E}$ follows immediately. If all pairs are within L of each other (partial saturation) then one expects that at least some statistics will not be Gaussian.

Nothing that was done in Sec. 2 or Appendix A depended in any essential way on the assumption of a single scale. The reader can verify that Eq. (1.14) for $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$ at equal frequencies continues to hold whenever the parabolic wave equation is valid. The only subtle point is that for $p<2$ the rms scattering angle is not well defined and, correspondingly, in Appendix A, Eq. (A6) cannot be approximated by Eq. (A9). However, a rather straightforward analysis of Eq. (A6) shows that the fractional error in Eq. (1.14) is of order $D(k^{-1},0)$ and it is known that $D(k^{-1},0)<1$ is the validity condition for the parabolic wave equation when $p<2$. Turning to coherences in frequency, there is however a significant defect in the theory if $p<2$. *When p is less than 2, the path integral in Eq.* (2.22) *cannot be approximated by that in Eq.* (2.24) *and Λ must be understood as a function defined by Eq.* (2.22) *whose evaluation would require a numerical calculation.*

In the fully saturated regime where pairs of paths can be separated by L or greater, the arguments of Sec. 3 proceed as before. One readily verifies that in the fully saturated regime the statistics of $\mathscr{E}$ are Gaussian and the discussion of Sec. 4 applies [except Eq. (4.7) which assumes Eq. (1.21) for Λ]. Equations (B12) and (B17) of Appendix B holds in the multiple scale case. The reader can then verify that for $p>2$, Eqs. (5.1), (5.2), (5.4), and (5.5) for the corrections to Gaussian statistics continue to hold in the fully saturated regime and that for $p<2$ these same equations hold if α is replaced by α' where

$$\alpha'=\frac{4(p+1)^{3/p}\Gamma(3/p)}{3^{3/2}\pi^{1/2}p}\frac{\Omega}{\Phi^{(6-2p)/p}}. \tag{7.3}$$

The situation for the fully saturated regime is summarized in Table II.

The higher order statistics in the partially saturated regime are more complicated. For the case $p<2$ everything can be worked out in detail and the results will be given in the next section. However, for $p>2$ the path integrals yield only qualitative information. It is summarized in Appendix D.

Finally, in multiple scale media the notion of multiple Fermat paths should be used with care. They exist but there are so many of them that they cannot, even in principle, be completely resolved. Nevertheless, the notion is useful in interpreting the path integral calculations and will continue to be employed.

8. THE PARTIALLY SATURATED REGIME FOR $p<2$

The partially saturated regime for $p<2$ is of considerable practical importance. Many atmospheric optics experiments lie in this region and, luckily, the complete statistics of

TABLE I. Boundaries of the saturated regimes.

	Partially Saturated Regime	Fully Saturated Regime
$2<p<4$	$\Phi>1, \Phi^{4/p}/\Omega>1, \quad \Phi/\Omega<1$	$\Phi>1, \quad \Phi/\Omega>1$
$1<p<2$	$\Phi>1, \Phi^{4/p}/\Omega>1, \quad \Phi^{2/p}/\Omega<1$	$\Phi>1, \quad \Phi^{2/p}/\Omega>1$

Table II. Changes needed to apply the formulas of Secs. 1–5 to the fully saturated regime.

	Boundaries	Limiting Statistics	Corrections to the Limiting Statistics
$2<p<4$	Unchanged	Unchanged	Unchanged
$1<p<2$	Replace $\Phi/\Omega>1$ by $\Phi^{2/p}/\Omega>1$	Unchanged except that Λ is not known explicitly	Replace α by α'

$\mathscr{E}$ can be worked out. There is a natural small parameter β defined by

$$\beta=(\Omega/\Phi^{4/p})^{2-p}. \tag{8.1}$$

For $p=\frac{5}{3}$, β is related to Tatarskii's[1] C_n by $\beta=1.19C_n^{4/5}R^{11/15}k^{7/15}$ and to the intensity fluctuations as computed in the Rytov approximation by $\langle(\ln I)^2\rangle-\langle\ln I\rangle^2|_{\text{Rytov}}=0.80\beta^{5/2}$. The signal statistics will be given through order β.

Partial saturation is due to the appearance of multiple Fermat paths all lying within L of each other. The larger scales ($\sim L$) will tend to correlate the locations of these paths leading in general to a complicated statistics. However, for $p<2$ the spectrum is so heavily weighted toward small scales that the locations of the Fermat paths turn out to be uncorrelated. This is not the case for $p>2$ where the multiple Fermat paths become correlated and the path integral yields over qualitative information (see Appendix D). Even for $p<2$ where the locations of the paths are uncorrelated the large scales can still correlate the phases along different Fermat paths. We will see this at the end of the section when coherences in frequency are studied.

Consider Eqs. (3.1) and (3.2) for $\langle I^2\rangle$ in the partially saturated regime with $p<2$. In the integration region (a) the separation between members of a pair of paths $\mathbf{v}_1(z)$ (using the notation of Appendix B) must be such that $d(|\mathbf{v}_1(z)|)\lesssim 1$, i.e., $|\mathbf{v}_1(z)|\lesssim L/\Phi^{2/p}$. The distance $\mathbf{v}_2(z)$ between pairs (again in the notation of Appendix B) will be limited by the oscillating terms in the path integral to values such that $\Omega|\mathbf{v}_2(z)|\,|\mathbf{v}_1(z)|\sim L^2$ or $|\mathbf{v}_2(z)|\lesssim L\Phi^{2/p}/\Omega$. Note that the ratio of the cutoff on $|\mathbf{v}_2|$ to that on $|\mathbf{v}_1|$ is $\Phi^{4/p}/\Omega$ and is large. Now both $|\mathbf{v}_1|$ and $|\mathbf{v}_2|$ are small compared to L and Eq. (7.2) can be used to evaluate M in Eqs. (3.2) or (B3). Taking account of the fact that $|\mathbf{v}_1|\ll|\mathbf{v}_2|$, the expression for M in Eq. (B3) of Appendix B then becomes [20]

$$M=\Phi^2R^{-1}\int_0^R\left|\frac{\mathbf{v}_1(z)}{L}\right|^p dz-\Phi^2R^{-1}p(p-1)$$

$$\times\int_0^R\left|\frac{\mathbf{v}_1(z)}{L}\right|^2\left|\frac{\mathbf{v}_2(z)}{L}\right|^{p-2}dz \tag{8.2}$$

and when $|\mathbf{v}_1/L|\sim\Phi^{2/p}$ and $|\mathbf{v}_2/L|\sim\Phi^{2/p}/\Omega$ the second term on the right-hand side of Eq. (8.2) is of order β and can be dropped. This is the same thing as saying that different pairs of paths in the path integral, or equivalently different Fermat paths in the sense of Sec. 6, are not correlated and $\langle I^2\rangle$ becomes $2\langle I\rangle^2$. What is happening is that for $p<2$ the fractional power behavior of d at small separations is making the arguments of Sec. 3 valid even though the different pairs are separated by less than L. Note that this will only happen for $p<2$. The generalization to $\langle I^n\rangle$ is straightforward and the result is a Rayleigh distribution with $\langle I^n\rangle=n!\langle I\rangle^n$.

The true test of the method comes when one evaluates the corrections to Rayleigh statistics. It is shown in Appendix C that to order β

$$\langle I^n\rangle=n!\langle I\rangle^n[1+\tfrac{1}{2}n(n-1)C(p)\beta], \tag{8.3}$$

where $C(p)$ is a constant which depends only on p. This constant is evaluated in Appendix C and $C(\frac{5}{3})=1.06$. The corrections are small for small β showing that the approximation scheme is consistent but there will be significant deviations from a Rayleigh distribution when

$$I/\langle I\rangle\gtrsim\sqrt{2/\beta C(p)}.$$

The statistics of $\mathscr{E}(\mathbf{r},\mathbf{r}_0)$ as a function of source and receiver locations can be investigated in a similar way. One finds that they are Gaussian and at equal times and frequencies the results of Secs. 3 and 4 hold in the limit $\beta=0$. There are coherence tails of order β. These are discussed in Appendix C.

In the fully saturated regime the dynamics of the medium enters only through $D(t)$. This is not always true in the partially saturated regime. It is true when the Taylor hypothesis is valid (a frozen field convected by a "wind") and the statistics in time can be obtained from the spatial statistics. However, one can consider a different kind of medium where the time dependence of μ is associated with linear waves whose dispersion relation is $\omega\sim k^{\delta/2}$. The Fourier transform of the second time derivative $\ddot{\rho}$ of ρ will then behave like

$$\tilde{\ddot{\rho}}(|\mathbf{q}|)=(\text{const})|\mathbf{q}|^{(2+p-\delta)} \tag{8.4}$$

at large $|\mathbf{q}|$. For the Taylor hypothesis Eq. (8.4) holds with $\delta=2$ and in general δ can be considered as being defined by Eq. (8.4). Assuming $p<2$, the statistics of $\mathscr{E}$ at unequal times are Gaussian in the partially saturated regime provided that $p-\delta<0$. This can be verified by explicitly computing the corrections. For $p<2$ and $p-\delta<0$ the corrections to Gaussian statistics are fractionally small for small β and the results of Secs. 3 and 4 continue to hold at unequal times. However for $p-\delta>0$, a direct calculation shows that the corrections to Gaussian statistics are *not* fractionally

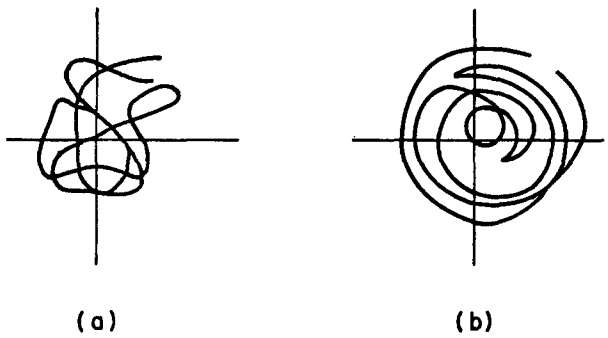

FIG. 5. (a) The schematic track of a signal satisfying Gaussian statistics in time. (b) The track of a signal which moves faster in phase than amplitude (phase wrapping).

small and therefore that the approximation scheme of Sec. 3 is not consistent at unequal times.

To see what is happening for $p-\delta>0$ one can compare the path integrals for $\langle I(t')I(t)\rangle$ and $\langle(\mathscr{E}^*(t'))^2(\mathscr{E}(t))^2\rangle$. The latter is sensitive to the time dependence of the phase of $\mathscr{E}$ while the former is not. A rather involved but straightforward calculation then shows that for $p-\delta>0$ the signal moves more rapidly in phase than in amplitude. This is to be contrasted with the case $p-\delta<0$ where the time statistics are Gaussian and according to Eq. (1.18) there is no tendency to move in phase as opposed to amplitude. As long as $p<2$ the signal has a Rayleigh distribution and over a long time the track of the signal will fill out a disk in the complex plane. The difference between $p-\delta<0$ and $p-\delta>0$ comes in how this disk is filled up. For $p-\delta<0$ the signal is Gaussian and it will make a track of the type shown in Fig. 5(a) which looks something like a random walk. However, for $p-\delta>0$ the track will wrap around in phase and slowly move in and out in amplitude as shown in Fig. 5(b).

These peculiar features of time statistics in the partially saturated regime can be understood in terms of Fermat paths. We know that $\mathscr{E}(t)$ is schematically $\Sigma_k A_k(t)\exp[i\phi_k(t)]$ where the locations of the paths are uncorrelated (for $p<2$) but the large scales may correlate the phases $\phi_k(t)$. The question of random walking vs. phase wrapping is equivalent to the question of whether or not the time derivatives $(d/dt)\phi_k=\dot{\phi}_k$ are correlated. For $p-\delta<0$, the time derivatives are sufficiently weighted towards small scales that the $\dot{\phi}_k$ are uncorrelated and the signal random walks. However for $p-\delta>0$, the effect of the large scales is strong enough to produce a correlated phase derivative common to all the Fermat paths.

Propagation of sound in the ocean is an example of a situation where $\mathscr{E}(t)$ phase wraps in the partially saturated regime.[7] For the ocean $p\approx 2,\delta\approx 0$, and in this special case it is possible to work out the detailed statistics of $\mathscr{E}(t)$.[7] However, for other combinations of p and δ it is not possible to compute fourth and higher moments of $\mathscr{E}(t)$ analytically, except when $p-\delta<0$.

Checking consistency, it was stated above that for $p<2$ the statistics of $\mathscr{E}(\mathbf{r},\mathbf{r}_0)$ as a function of $\mathbf{r}_0$ and $\mathbf{r}$ are Gaussian in the partially saturated regime. If the time derivatives on the right-hand side of Eq. (8.4) were replaced by spatial derivatives we would have $\delta=2$. Since $p-2<0$ for $p<2$ it is consistent that the statistics in $\mathbf{r}_0$ and $\mathbf{r}$ are Gaussian and that the statistics in time are Gaussian when the Taylor hypothesis (implying $\delta=2$) is valid.

For $p<2$ and $p-\delta<0$, the statistics of $\mathscr{E}(\mathbf{r},\mathbf{r}_0,t)$ in the partially saturated regime are essentially the same as in the fully saturated regime. The reader may therefore wonder what the basic distinction between the regimes is. The answer turns out to lie in the statistics in frequency.

Let us examine the path integral for $\langle\mathscr{E}^*(\omega_1)\mathscr{E}(\omega_2)\mathscr{E}^*(\omega_3)\mathscr{E}(\omega_4)\rangle$. Up to a normalization it is

$$\langle\mathscr{E}^*(\omega_1)\mathscr{E}(\omega_2)\mathscr{E}^*(\omega_3)\mathscr{E}(\omega_4)\rangle \sim\int d^4(\text{paths})\exp\left(\tfrac{1}{2}\sum_{j=1}^{4}(-1)^j\frac{\omega_j}{c}\int_0^R[\dot{\mathbf{r}}_j(z)]^2dz-N\right), \tag{8.5}$$

where with the Markov approximation, surpressing time t

$$N=\tfrac{1}{2}\sum_{i,j=1}^{4}(-1)^{i+j}\frac{\omega_i\omega_j}{c^2}\int_0^R\hat{\rho}(|\mathbf{r}_i(z)-\mathbf{r}_j(z)|)dz \tag{8.6}$$

and for simplicity the medium has been assumed to be nondispersive. There are the usual two important regions of path space (a) and (b). Let us concentrate on (a) where $|\mathbf{r}_1-\mathbf{r}_2|<L/\Phi^{2/p}$ and $|\mathbf{r}_3-\mathbf{r}_4|<L/\Phi^{2/p}$. First we will see how Gaussian statistics arise in the fully saturated regime and then see how the partially saturated case differs. In the fully saturated case typical values of, say, $|\mathbf{r}_1-\mathbf{r}_3|$ are large compared to L and $\hat{\rho}(|\mathbf{r}_1-\mathbf{r}_3|)$ can be set equal to zero. Ignoring correlations between the different pairs then yields

$$N\approx\tfrac{1}{2}\left(\sum_{i,j=1}^{2}+\sum_{i,j=3}^{4}\right)(-1)^{i+j}\frac{\omega_i\omega_j}{c^2}\int_0^R\hat{\rho}(|\mathbf{r}_i(z)-\mathbf{r}_j(z)|)dz \tag{8.7}$$

which is a sum of two terms one of which depends on ω_1 and ω_2 and the other on ω_3 and ω_4, and the result is Gaussian statistics. In the partially saturated case typical values of $|\mathbf{r}_1-\mathbf{r}_3|$ are small compared to L and $\hat{\rho}(|\mathbf{r}_1-\mathbf{r}_3|)$ is approximately equal to $\hat{\rho}(0)$. Now we have to set correlations between the different pairs of paths equal to $\hat{\rho}(0)$ rather than zero and N becomes

$$N\approx\tfrac{1}{2}\left(\sum_{j=1}^{4}(-1)^j\frac{\omega_j}{c}\right)^2R\hat{\rho}(0)+\tfrac{1}{2}\left(\sum_{i,j=1}^{2}+\sum_{i,j=3}^{4}\right)\times(-1)^{i+j}\frac{\omega_i\omega_j}{c^2}\int_0^R[\hat{\rho}(|\mathbf{r}_i(z)-\mathbf{r}_j(z)|)-\hat{\rho}(0)]dz. \tag{8.8}$$

The path integral again factors into a product of two double path integrals and is expressable in terms of Λ as defined in Eq. (2.22). Collecting the contribution from both regions (a) and (b) and supplying the correct normalization yields

$$\frac{\langle\mathscr{E}^*(\omega_1)\mathscr{E}(\omega_2)\mathscr{E}^*(\omega_3)\mathscr{E}(\omega_4)\rangle}{\mathscr{E}_0^*(\omega_1)\mathscr{E}_0(\omega_2)\mathscr{E}_0^*(\omega_3)\mathscr{E}_0(\omega_4)}$$

$$=\exp\left[-\tfrac{1}{2}\left(\sum_{j=1}^{4}(-1)^j\frac{\omega_j}{c}\right)^2 R\hat{\rho}(0)\right]$$

$$\times[\Lambda(\omega_2-\omega_1)\Lambda(\omega_4-\omega_3)+\Lambda(\omega_2-\omega_3)\Lambda(\omega_4-\omega_1)]. \tag{8.9}$$

Because of the common exponential factor in front of the two terms on the right-hand side this is not Gaussian statistics. What it corresponds to is an $\mathscr{E}$ of the form $\mathscr{E}(\omega)/\mathscr{E}_0(\omega)=e^{i\omega\psi}\chi(\omega)$ where ψ is a real Gaussian random variable with $\langle\psi\rangle=0$, $\langle\psi^2\rangle=R\hat{\rho}(0)c^{-2}$, and $\chi(\omega)$ is an independent complex Gaussian random variable with zero mean and covariances $\langle\chi(\omega)\chi(\omega')\rangle=\langle\chi^*(\omega)\chi^*(\omega')\rangle=0$ and $\langle\chi^*(\omega)\chi(\omega')\rangle=\Lambda(\omega'-\omega)$. It is straightforward to verify that this ansatz does in fact yield the correct $2n$th moment of $\mathscr{E}(\omega)$ in the partially saturated regime. In particular the second moment

$$\frac{\langle\mathscr{E}^*(\omega)\mathscr{E}(\omega')\rangle}{\mathscr{E}_0^*(\omega)\mathscr{E}_0(\omega')}=\langle\exp[i(\omega'-\omega)\psi]\chi^*(\omega)\chi(\omega')\rangle$$

$$=\langle\exp[i(\omega'-\omega)\psi]\rangle\langle\chi^*(\omega)\chi(\omega')\rangle$$

$$=\langle\exp[-\tfrac{1}{2}(\omega-\omega')^2R\hat{\rho}(0)c^{-2}]\Lambda(\omega'-\omega) \tag{8.10}$$

comes out right. For a dispersive medium $\langle\psi^2\rangle$ becomes ω_g^{-2} as in Eq. (1.20) and c_g rather than c appears in Λ.

Thus the fundamental distinction between the fully and partially saturated regimes is that in the former the statistics in frequency are Gaussian while in the latter they correspond to a phase times a Gaussian. Well inside the partially saturated regime ω_g is small compared to the width in ω of Λ. The phase $e^{i\omega\psi}$ then dominates the moments of $\mathscr{E}(\omega)$, except for correlations involving only $|\mathscr{E}(\omega)|^2$ where ψ cancels. As the boundary $\Phi^{2/p}/\Omega=1$ of the fully saturated regime is approached the width of $\Lambda(\omega)$ becomes comparable to ω_g and upon passing into the fully saturated regime Λ dominates the moments and the signal becomes Gaussian. In the terminology of Sec. 4, for partial saturation the spread is small compared to the wander. In pulse propagation $e^{i\omega\psi}$ represents a quasideterministic wander which dominates $\langle\mathscr{I}(\tau)\rangle$. The phase $e^{i\omega\psi}$ cancels out in the integral [Eq. (4.8)] for $P(\tau)$ and the spreading of a pulse is proportional to the inverse width of Λ.

In terms of Fermat paths $\mathscr{E}(\omega)=\Sigma_k A_k(\omega)\exp[i\phi_k(\omega)]$, the non-Gaussing statistics can be understood as follows. Each $\phi_k(\omega)$ can be written as $\omega\psi+\Delta\phi_k(\omega)$ where $\omega\psi$ is a common phase generated by the larger scales. The phase differences $\Delta\phi_k(\omega)$ are due to the small scales. They vary from path to path and are responsible for the Gaussian factor $\chi(\omega)$. Note that only correlations in frequency measure ψ directly. Correlations in space or time see only $\nabla\psi$ or $\dot{\psi}$ which for $p<2$ and $p-\delta<0$ are dominated by small rather than large scales, leading ultimately to Gaussian statistics. The phase wrapping in time for $p-\delta>0$ is a remnant of ψ.

The statistics of $\mathscr{E}$ in the partially saturated regime are summarized in Table III.

The reader may be curious as to what happens at $p=2$. The "small" parameter β is then equal to unity but according to Appendix C the coefficient $C(2)$ in Eq. (8.3) vanishes. A detailed investigation[7] then shows that the corrections to Rayleigh statistics in the partially saturated regime are of order $(\ln\Phi)^{-1}$. More generally, if $p=2$ and $\ln\Phi$ is large the statistics given in Table III apply with errors of order $(\ln\Phi)^{-1}$. At $p=2$ it is possible to compute Λ. It is given[7] by Eq. (1.21) with $\omega_0\alpha$ replaced by $\omega_0\alpha(\ln\Phi)^{-1}$. In general, a medium with $|p-2|\ln\Phi<1$ will act like one with $p=2$.

As mentioned before, the case of partial saturation for $p>2$ is discussed in Appendix D.

9. INHOMOGENEOUS AND ANISOTROPIC MEDIA

In practice, random media are only locally homogeneous and the covariance.

$$\rho(\mathbf{x}-\mathbf{x}',t-t';\bar{\mathbf{x}})=\langle\mu(\mathbf{x},t)\mu(\mathbf{x}',t')\rangle-\langle\mu(\mathbf{x},t)\rangle\langle\mu(\mathbf{x}',t')\rangle \tag{9.1}$$

depends on position $\bar{\mathbf{x}}\equiv\frac{1}{2}(\mathbf{x}+\mathbf{x}')$. It is always assumed that the variations of ρ in $\mathbf{x}-\mathbf{x}'$ are much more rapid than those in $\bar{\mathbf{x}}$ but over a long propagation path the dependence on $\bar{\mathbf{x}}$ cannot always be neglected. Also, in an inhomogeneous medium $\langle\mu(\mathbf{x})\rangle=\mu_0(\mathbf{x})$ will generally not be a constant and consequently cannot be absorbed in the definition $c=\omega/k$. Finally, the medium can be statistically anisotropic so that ρ depends on the orientation of $\mathbf{x}-\mathbf{x}'$ as well as its magnitude.

To obtain tractable path integrals in an inhomogeneous medium we will have to approximate the path dependence of $\bar{\mathbf{x}}$ in ρ by evaluating $\bar{\mathbf{x}}$ along some central path which will turn out to be an unperturbed ray. From Sec. 3 we know that paths are separated by $L\Phi/\Omega$ (the precise definitions of L, Φ and Ω for inhomogeneous anisotropic media will be given below) and the problem will be tractable if

(i) for changes in $\bar{\mathbf{x}}$ of order $L\Phi/\Omega$ the corresponding

TABLE III. The statistics of $\mathscr{E}$ in the partially saturated regime for $p<2$.

Intensity Distribution	Space	Variations in time		frequency
		$p-\delta<0$	$p-\delta>0$	
Rayleigh	Gaussian	Gaussian	phase wrapping	phase times a Gaussian

variations in ρ can be neglected. It will also turn out to be necessary to expand μ_0 in powers of distances between paths and we will have to require that

(ii) $\mu_0(\mathbf{x})$ is slowly varying over distances of order $L\Phi/\Omega$. The one other condition is that

(iii) the parabolic wave equation is a valid approximation.

When $\mu_0(\mathbf{x})$ is not a constant this requires that the normals to the wavefronts in the "unperturbed problem" where $\mu(\mathbf{x})=\mu_0(\mathbf{x})$ remain close to the z axis. If this is true locally but not globally, then solutions based on the parabolic approximation can be patched together in the obvious way.

When conditions (i), (ii), and (iii) are met it is reasonably straightforward to extend the path integral method to inhomogeneous and anisotropic media. It amounts to: (1) showing that with suitable definitions of Φ and D, $\langle\mathscr{E}\rangle$ remains $\exp[-\frac{1}{2}\Phi^2]$ and Eq. (1.14) continues to hold, (2) finding a suitable definition for Ω and then showing that the boundaries of the saturated regimes are still given by Table I, (3) showing that in the fully saturated regime the statistics of $\mathscr{E}$ are Gaussian and that in the partially saturated regime they are (for $p<2$) as given in Table III, (4) giving new formulas for the corrections to Gaussian statistics and coherence tails, and (5) giving a method for computing $\Lambda(\omega)$. These steps will be carried out in order. In doing so it will be assumed that a ray approximation is valid for the unperturbed problem with $\mu=\mu_0$.

A. The first and second moments

The path integral for $\langle\mathscr{E}\rangle$ will contain a factor

$$\exp\Big[-\frac{k^2}{2}\int_0^R dz\int_0^R dz'\,\rho(\mathbf{r}(z)-\mathbf{r}(z')+\mathbf{e}_z(z-z'),0;\tfrac{1}{2}[\mathbf{r}(z)+\mathbf{r}(z')]+\tfrac{1}{2}\mathbf{e}_z(z+z'))\Big]. \tag{9.2}$$

The path dependance of the third argument of ρ will be approximated by setting $\frac{1}{2}(\mathbf{r}(z)+\mathbf{r}(z'))=\mathbf{s}(\bar{z})$ where $\bar{z}=\frac{1}{2}(z+z')$ and $\mathbf{s}$ is the unperturbed ray satisfying

$$\mathbf{s}''(z)+\nabla\mu_0(\mathbf{s}(z)+\mathbf{e}_z z)=0. \tag{9.3}$$

Here $\mathbf{s}=(s_x,s_y)$ is a two-dimensional vector and $(\mathbf{s}(0),0)$ and $(\mathbf{s}(R),R)$ are the source and receiver coordinates. If there is more than one unperturbed ray connecting the source to the receiver, it is assumed that they are far enough apart that the path integral reduces to a sum of (statistically) independent terms coming from paths near each ray.[21] Defining a new path $\mathbf{u}(z)$ by $\mathbf{r}(z)=\mathbf{s}(z)+\mathbf{u}(z)$, the Markov approximation now amounts to setting

$$\mathbf{r}(z)-\mathbf{r}(z')+\mathbf{e}_z(z-z')\approx(\mathbf{s}'(\bar{z})+\mathbf{e}_z)(z-z'). \tag{9.4}$$

The essence of the approximation is neglecting $\mathbf{u}(z)-\mathbf{u}(z')$. By requirement (ii) the substitution $\mathbf{s}(z)-\mathbf{s}(z')\approx\mathbf{s}'(\bar{z})(z-z')$ is always valid. The reader will note that by (iii) $\mathbf{s}'$ is actually small compared to $\mathbf{e}_z$. However, in a sufficiently anisotropic medium $\mathbf{s}'$ cannot be neglected on the right-hand side of Eq. (9.4). Assuming for the moment the validity of the Markov approximation, the analog of $\hat{\rho}(0,0)$ will be $\hat{\rho}(\mathbf{0},0;z)$, where

$$\hat{\rho}(\mathbf{0},0;z)=\int_{-\infty}^{\infty}\rho([\mathbf{s}'(z)+\mathbf{e}_z]u,0;\mathbf{s}(z)+\mathbf{e}_z z)du \tag{9.5}$$

and the path integral for $\langle\mathscr{E}\rangle$, which is now trivial since ρ no longer contains the path $\mathbf{u}$, will yield $\langle\mathscr{E}\rangle=\mathscr{E}_0\exp[-\frac{1}{2}\Phi^2]$ where

$$\Phi^2=k^2\int_0^R\hat{\rho}(\mathbf{0},0;z)dz. \tag{9.6}$$

Continuing to assume the validity of the Markov approximation, the next thing to compute is $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$. There are two paths $\mathbf{r}_1=\mathbf{s}+\mathbf{w}_1$ and $\mathbf{r}_2=\mathbf{s}+\mathbf{w}_2$ where $\mathbf{s}$ satisfies Eq. (9.3) with the boundary conditions $\mathbf{s}(0)=\frac{1}{2}(\mathbf{r}_{01}+\mathbf{r}_{02})$ and $\mathbf{s}(R)=\frac{1}{2}(\mathbf{r}_1+\mathbf{r}_2)$ and the approximation is

$$\int_0^R dz\int_0^R dz'\,\rho(\mathbf{r}_i(z)-\mathbf{r}_j(z')+\mathbf{e}_z(z-z'),t_i-t_j;\tfrac{1}{2}[\mathbf{r}_i(z)+\mathbf{r}_j(z')]+\tfrac{1}{2}\mathbf{e}_z(z+z'))\approx\int_0^R\hat{\rho}(\mathbf{w}_i(z)-\mathbf{w}_j(z);t_i-t_j;z)dz \tag{9.7}$$

for $i,j=1,2$, where

$$\hat{\rho}(\mathbf{w},t;z)=\int_{-\infty}^{\infty}\rho(\mathbf{w}+[\mathbf{s}'(z)+\mathbf{e}_z]u,t;\mathbf{s}(z)+\mathbf{e}_z z)du. \tag{9.8}$$

The path integral for $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$ is then

$$\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle=\frac{1}{4k^2}\int d^2(\text{paths})\exp\Big[iS_0(\text{path }1)-iS_0(\text{path }2)-\int_0^R d(\mathbf{w}_1(z)-\mathbf{w}_2(z),t_1-t_2;z)dz\Big], \tag{9.9}$$

where

$$S_0=k\int_0^R[\tfrac{1}{2}(\mathbf{r}'(z))^2-\mu_0(\mathbf{r}(z)+\mathbf{e}_z z)]dz \tag{9.10}$$

and

$$d(\mathbf{w},t;z)=k^2[\hat{\rho}(\mathbf{0},0;z)-\hat{\rho}(\mathbf{w},t;z)]. \tag{9.11}$$

Introducing paths $\mathbf{u}=\frac{1}{2}(\mathbf{w}_1+\mathbf{w}_2)$ and $\mathbf{v}=\mathbf{w}_1-\mathbf{w}_2$ we can, according to (ii), expand $S_0(\text{path }1)-S_0(\text{path }2)$ in powers of $\mathbf{u}$ and $\mathbf{v}$ and keep only the leading terms which are quadratic. Proceeding in this way yields[22]

$$\frac{\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle}{\mathscr{E}_0^*(2)\mathscr{E}_0(1)}=|2k\mathscr{E}_0|^{-2}\int d^2(\text{paths})\exp\Big[ik\int_0^R(\mathbf{u}'(z)\cdot\mathbf{v}'(z)-\mathbf{u}_i(z)\mathbf{v}_j(z)\mu_{ij}(z))dz-\int_0^R d(\mathbf{v}(z),t_1-t_2;z)dz\Big], \tag{9.12}$$

where the two by two matrix (in $\mathbf{e}_x\mathbf{e}_y$ space) $\mu_{ij}(z)$ is

$$\mu_{ij}(z)=\frac{\partial^2}{\partial x_i\,\partial x_j}\mu_0(\mathbf{x})\big|_{\mathbf{x}=\mathbf{s}(z)+\mathbf{e}_z z}. \tag{9.13}$$

The path $\mathbf{u}$ now appears only as a linear factor in the exponential and integrating over it will produce a product of δ functions which force $\mathbf{v}(z)$ to be equal to the special path $\nu(z)$ which satisfies the differential equation and boundary conditions[22]

$$\nu_i''(z)+\mu_{ij}(z)\,\nu_j(z)=0,$$
$$\nu(0)=\mathbf{r}_{01}-\mathbf{r}_{02}, \tag{9.14}$$
$$\nu(R)=\mathbf{r}_1-\mathbf{r}_2.$$

Then setting $\mathbf{v}$ equal to ν in d the remaining path integral just produces[17] $|2k\mathscr{E}_0|^2$ and one finds Eq. (1.14)

$$\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle=\mathscr{E}_0^*(2)\,\mathscr{E}_0(1)\exp[-\tfrac{1}{2}D] \tag{1.14''}$$

with

$$D=2\int_R^0 d\,(\nu(z),t_1-t_2;z)dz. \tag{9.15}$$

The object D defined in Eq. (9.15) is just the phase structure function of first order geometric optics[1,7] for a general inhomogeneous anisotropic medium which satisfies (i), (ii), and (iii). Note that ν is always linear in $\mathbf{r}_{01}-\mathbf{r}_{02}$ and $\mathbf{r}_1-\mathbf{r}_2$. When μ_0 is a constant, $\nu(z)=(\mathbf{r}_{01}-\mathbf{r}_{02})(R-z)/R+(\mathbf{r}_1-\mathbf{r}_2)z/R$, and for a homogeneous isotropic medium Eq. (9.15) reduces to Eq. (1.16).

For an isotropic medium where ρ depends only on the magnitude of $\mathbf{x}-\mathbf{x}'$, the Markov approximation is valid whenever the parabolic wave equation is. The reason is the same as in Sec. 2. In Appendix E the formula for the first correction to the Markov approximation to $\langle\mathscr{E}^*\mathscr{E}\rangle$ is given. One can explicitly verify that the error is small when the parabolic wave equation is valid.

The situation for anisotropic media is more complicated. Consider an anisotropic but homogeneous medium with constant μ_0. Typical inhomogeneities will not be spherically symmetric and one needs to consider the three cases shown in Figs. 6. The asymmetric inhomogeneities introduce a new small angle θ_0, the ratio of the small dimension to the large one. Examining the error in the Markov approximation as given in Appendix E one finds that, for the case shown in Fig. 6(a), the Markov approximation fails when the rms multiple scattering angle is of order θ_0. For the case shown in Fig. 6(b), it fails when the rms multiple scattering angle is of order of the angle of incidence θ_i and for the situation in Fig. 6(c), it fails when the rms multiple scattering angle is of order unity, i.e., when the parabolic wave equation fails. Since θ_0 can be small compared to unity, the Markov approximation

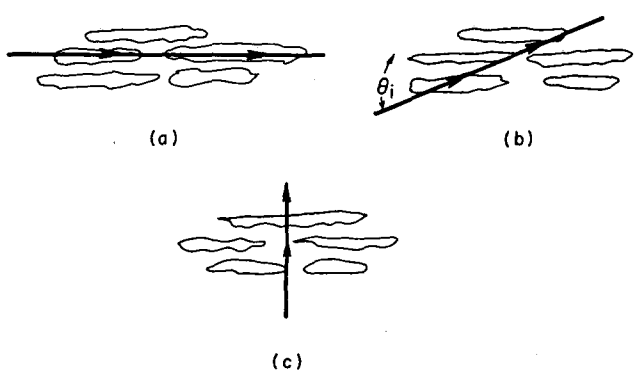

FIG. 6. Propagation through an anisotropic medium. The blobs are schematic inhomogeneities and the heavy directed lines are the unperturbed propagation path at various angles with respect to the long axis of the blobs.

can fail in an anisotropic medium before the parabolic wave equation does but only for some propagation paths. When it fails Eq. (1.14) is not valid and this represents a defect in the theory which is not easy to remove.

It should not be surprising that the Markov approximation can fail sooner in an anisotropic medium. The Markov approximation can be interpreted as the statement that the system has "no memory" in range, i.e., that scatterings at a given range point are independent of previous distant scatterings. In an isotropic medium this will be true as long as the rms multiple scattering angle is small and the wave keeps moving in the same direction. However, in an anisotropic medium, when the scattering by a given inhomogeneity can be highly dependent on the angle of incidence, a distant scattering which has deflected the wave only through a small angle will not be "forgotten." For the inhomogeneities shown in Fig. 6 the scattering is strongly dependent on angle of incidence (measured from the long axis of the inhomogeneities) when the angle is of order θ_0. When the incident wave is along the long axis as in Fig. 6(a), it begins to remember previous scatterings when the scattering angle builds up to θ_0 and the pieces of the wave have incidence angles greater than θ_0. For the case shown in Fig. 6(b) the past history of the wave becomes important when pieces of the wave have been deflected by θ_i and are incident along the long axis. When θ_i approaches $\pi/2$ as in Fig. 6(c) the process has no memory as long as the rms multiple scattering angle is less than unity.

Yet another way to understand the peculiarities of anisotropic media is to return to the remarks following Eqs. (6.8) and (6.9). For an isotropic medium the average of $\dot{\mu}$ integrated along a Fermat path $\mathbf{w}$ will be the same as the average of $\dot{\mu}$ integrated along the unperturbed ray $\mathbf{s}$ as long as the rms multiple scattering angle is small. However, in an anisotropic medium the average of $\dot{\mu}$ integrated along a path can be very sensitive to the local direction $\mathbf{w}'$ of the path. In fact, for the situation shown in Fig. 6(a), the average of $\dot{\mu}$ integrated along a Fermat path deviates from the average along an unperturbed ray as soon as $|\mathbf{w}'|\sim\theta_0$ and for the situation in Fig. 6(b) when $|\mathbf{w}'|\sim\theta_i$. This leads to the same criteria as before.

The cominbation of an anisotropic medium and a spatially varying $\mu_0(\mathbf{x})$ leads to a new set of complications. This will be illustrated for propagation in a channel where the unperturbed rays make loops as shown in Fig. 7 and where the long axis of the inhomogeneities is parallel to the channel axis. The medium will also be assumed to be statistically homogeneous in the direction of the channel axis but not necessarily in the transverse directions. (This is a prototype of the physical situation which occurs for sound propagation in the ocean.[7]) The scattering will be strongest when the tangent to the unperturbed ray is pointing along the long axis of the inhomogeneities, i.e., at the turning points. For small θ_0 one can in fact ignore all of the propagation path except for a set of discrete regions around turning point where the tangent to the ray is within θ_0 of the channel axis. Assuming that a Markov approximation is valid for propagation through one of these regions, it will also be valid for propagation through many turning points provided only that the average

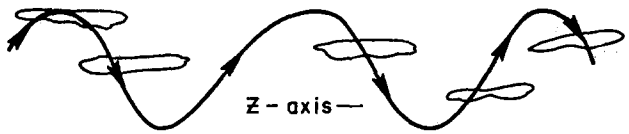

FIG. 7. Propagation in a channel. The channel axis (z axis) is parallel to the long axis of the inhomogeneities (blobs). The medium is assumed to be isotropic in the z direction and the unperturbed propagation path (heavy directed line) makes periodic loops.

scattering at a given turning point is at most weakly dependent on scattering at previous turning points. Assuming that the turning points are separated by more than a coherence length the effect of previous scatterings will be a random modulation of the range z_0 and (transverse) location $\mathbf{s}_0$ of a turning point. Now the average scattering around a turning point is dependent only on its location $\mathbf{s}_0$ in the channel and not on its range z_0. Thus the Markov approximation will be valid out to a range such that random variations in $\mathbf{s}_0$ are big enough to change the average scattering. This turns out to be a much longer range[7] than that for which the rms multiple scattering angle (which is dominated by variations in z_0) becomes of order θ_0. The extended validity of the Markov approximation can be demonstrated explicitly using the Fermat path formalism of Sec. 6. One works out the properties of Fermat paths which are randomly deflected at turning points and then compares averages of μ integrated along these paths to averages of μ integrated along the unperturbed ray. For a given channel one can then find out when the Markov approximation will break down. The result is just the criteria stated above.

B. The saturated regimes

It will temporarily be assumed that the medium has a single scale. Then in an anisotropic inhomogeneous medium the scale length L becomes a z dependent two-by-two matrix (in $\mathbf{e}_x\,\mathbf{e}_y$ space) defined by the expansion $\hat{\rho}$,[22]

$$\hat{\rho}(\mathbf{w},0;z)=\hat{\rho}(\mathbf{0},0;z)[1-\tfrac{1}{2}(L^{-2}(z))_{ij}w_iw_j+O(|\mathbf{w}|^3)]. \quad (9.16)$$

The first task in discussing the saturated regimes is to find the correct definition of Ω and establish their boundaries. The general definition of Ω will involve L and some geometric parameters associated with the unperturbed problem. From Secs. 3 and 7 one can see that Ω measures the rate at which the phase of the oscillating factor in the path integral varies as a path moves away from an upperturbed ray. To examine this in more detail consider paths that leave the source at $z=0$, go to the receiver at $z=R$, and at some point z_0 in between are separated from the unperturbed ray $\mathbf{s}(z_0)$ by $\mathbf{l}$. Let $\bar{S}(\mathbf{l},z_0)$ be the minimum of $S_0(\text{path})-S_0$ (unperturbed ray) taken over all paths of this class. The minimum is achieved for a path that follows an unperturbed ray from the source to $(\mathbf{s}(z_0)+\mathbf{l},z_0)$ and then another unperturbed ray from $(\mathbf{s}(z_0)+\mathbf{l},z_0)$ to the receiver. When μ_0 is a constant $\bar{S}(\mathbf{l},z_0)$ is simply

$$\bar{S}(\mathbf{l},z_0)=\tfrac{1}{2}(\mathbf{l})^2B(z_0), \quad (9.17)$$

where B which already appeared in Eq. (6.11) is

$$B(z_0)=\frac{kR}{z_0(R-z_0)}) \quad (9.18)$$

and Ω^{-1} is the average of $B^{-1}L^{-2}$, i.e.,

$$\Omega^{-1}=L^{-2}\overline{B^{-1}}$$

$$=L^{-2}\left[\frac{1}{R}\int_0^R B^{-1}(z_0)dz_0\right]. \quad (9.19)$$

Thus Ω is a measure of the phase change required to move a path a distance L away from the unperturbed ray. In general there is a two-by-two matrix B defined by the expansion for small $\mathbf{l}$,[22]

$$\bar{S}(\mathbf{l},z_0)=\tfrac{1}{2}l_il_jB_{ij}(z_0)+O(|\mathbf{l}|^3), \quad (9.20)$$

and Ω^{-1} will be an average of $L^{-2}B^{-1}$. It is convenient to weight the average by $\hat{\rho}(0,0;z)$ and Ω will be defined as[22]

$$\Omega^{-1}=\tfrac{1}{2}\overline{(L^{-2})_{ij}(B^{-1})_{ij}}$$

$$=\tfrac{1}{2}\frac{\int_0^R\hat{\rho}(\mathbf{0},0;z)(L^{-2}(z))_{ij}(B^{-1}(z))_{ij}dz}{\int_0^R\hat{\rho}(\mathbf{0},0;z)dz}. \quad (9.21)$$

With this definition of Ω one can follow through the arguments of Sec. 3 and verify that saturation and Gaussian statistics are expected when $\Phi>1$ and $\Phi/\Omega>1$. A more precise procedure is to compute $(\langle I^2\rangle-\langle I\rangle^2)/\langle I\rangle^2$ in the Rytov approximation to find the boundary of the saturated regime and then in the saturated regime compute the corrections to Gaussian statistics and verify that they are small. A straightforward evaluation of $(\langle I^2\rangle-\langle I\rangle^2)/\langle I\rangle^2|_{\text{Rytov}}$ shows that it does in fact exceed unity when $\Phi>1$ and $\Phi/\Omega>1$ indicating that the boundary is correct. Using the formulas of Appendix F one can verify that in the saturated regime the corrections to Gaussian statistics are indeed small.

With the appropriate change in the definition of $(L^{-2})_{ij}$ for $p<2$, the same procedure can be extended to media with multiple scales. The result is that with Ω defined as in Eq. (9.21) the boundaries given in Table I remain correct and that for $p<2$ the partially saturated statistics given in Table III also remain correct.

To actually calculate $B_{ij}(z)$ the following result is useful. Define a Green's function $g_{ij}(z,z')$ by[22]

$$\frac{\partial^2}{\partial z^2}g_{ij}(z,z')+\mu_{ik}(z)g_{kj}(z,z')=\delta_{ij}\delta(z-z'),$$
$$g_{ij}(0,z')=g_{ij}(R,z')=0. \quad (9.22)$$

Then it is straightforward to verify that

$$(B^{-1}(z))_{ij}=-g_{ij}(z,z). \quad (9.23)$$

C. Correlations in frequency

In general one can write

$$\frac{\langle\mathscr{E}^*(\omega')\mathscr{E}(\omega)\rangle}{\mathscr{E}_0^*(\omega')\mathscr{E}_0(\omega)}$$

$$=\exp\left[-\frac{1}{2}\left(\frac{\omega-\omega'}{\omega_g}\right)^2\right]\Lambda(\omega-\omega'), \tag{1.19'}$$

where the exponential factor comes from geometric optics and Λ is to be computed from the path integral. The geometric optics decorelation frequency ω_g is

$$\omega_g^{-2}=\left\langle\left(\frac{d}{d\omega}\int_0^R[\mu_\omega(\mathbf{s}_\omega(z)+\mathbf{e}_z z,t)-\langle\mu_\omega(\mathbf{s}_\omega(z)+\mathbf{e}_z z,t)\rangle]dz\right)^2\right\rangle \tag{9.24}$$

and for a general dispersive medium both μ and the unperturbed ray $\mathbf{s}$ will depend on ω. For a nondispersive medium ω_g is equal to $\bar{\omega}/\Phi$ where Φ is evaluated at the central frequency $\bar{\omega}$.

As before, the path integral for Λ is tractable only for media with $p>2$. The derivation proceeds as in Sec. 3 and after introducing scaled paths

$$\xi=\left(\frac{\bar{k}^2 c_g}{2(\omega-\omega')}\right)^{1/2}(\mathbf{v}-\mathbf{s}_{\bar{\omega}}),\ \text{where}\ \bar{\omega}=\tfrac{1}{2}(\omega+\omega')\ \text{and}\ c_g\ \text{is}$$

the group velocity at $\omega=\bar{\omega}$, the path integral for Λ gives

$$\Lambda(\omega)=\frac{K(\omega)}{K(0)}, \tag{9.25}$$

where[22]

$$K(\omega)=\int d(\text{paths})\exp\left\{-i\int_0^R[(\xi'(z))^2-\xi_i(z)\xi_j(z)(\mu_{ij}(z)+i\omega h_{ij}(z)]dz\right\} \tag{9.26}$$

with

$$h_{ij}(z)=c_g^{-1}\hat{\rho}(\mathbf{0},0;z)(L^{-2}(z))_{ij}. \tag{9.27}$$

If the path integral for K is written out in its finite form it becomes an ordinary integral of large dimension whose integrand is the exponential of a quadratic form. Such an integral is proportional to one over the square root of the determinant of the quadratic form and in particular Λ will be the square root of the ratio of two determinants. As the number of integration points goes to infinity the determinants becomes functional determinants. There are two equivalent methods[6] for computing the ratio of these functional determinants.

In the first method one has to find all the eigenvalues ω_n of the differential equation

$$\xi_i^{(n)''}(z)+\mu_{ij}(z)\xi_j^{(n)}(z)-\omega_n h_{ij}(z)\xi_j^{(n)}(z)=0 \tag{9.28}$$

subject to the boundary conditions $\xi^{(n)}(0)=\xi^{(n)}(R)=0$. Having done this $\Lambda(\omega)$ is

$$\Lambda(\omega)=\left(\prod_n\frac{1}{1+i\omega/\omega_n}\right)^{1/2}. \tag{9.29}$$

In the second method one defines a two-by-two matrix $M_{ij}(z,\omega)$ by the differential equation

$$M_{ij}''(z,\omega)+\mu_{ik}(z)M_{kj}(z,\omega)+i\omega h_{ik}(z)M_{kj}(z,\omega)=0 \tag{9.30}$$

and boundary conditions[22] in z

$$M_{ij}(0,\omega)=0,\quad M_{ij}'(0,\omega)=\delta_{ij}. \tag{9.31}$$

Then Λ is given by the ratio of determinants[23]

$$\Lambda(\omega)=\left(\frac{\det M(R,0)}{\det M(R,\omega)}\right)^{1/2}. \tag{9.32}$$

As an example of how Λ is computed consider a homogeneous isotropic medium where $\mu_{ij}=0$ and $h_{ij}=\delta\phi_{ij}c_g^{-1}\hat{\rho}(0,0)L^{-2}$. The eigenfunctions of the operator in Eq. (9.28) are then of the form $\delta_{i1}\sin(n_1\pi z/R)$ and $\delta_{j2}\sin(n_2\pi z/R)$ and the eigenvalues are $-n_1^2\pi^2\omega_1$ and $-n_2^2\pi^2\omega_1$ where $\omega_1=c_g L^{-2}/R^2\hat{\rho}(0,0)$ as in Sec. 4. The infinite product in Eq. (9.29) is then a product over two sets of integers

$$\Lambda(\omega)=\left[\prod_{n_1=1}^{\infty}\left(1-i\frac{\omega}{n_1^2\pi^2\omega_1}\right)^{-1}\times\prod_{n_2=1}^{\infty}\left(1-i\frac{\omega}{n_2^2\pi^2\omega_1}\right)^{-1}\right]^{1/2} \tag{9.33}$$

and the two equal factors just cancel the square root. The result is

$$\Lambda(\omega)=\prod_{n=1}^{\infty}\left(1-i\frac{\omega}{n^2\pi^2\omega_1}\right)^{1}$$

$$=\prod_{\substack{n=-\infty\\ n\neq 0}}^{\infty}\left[1+\left(\frac{i\omega}{\omega_1}\right)^{1/2}\frac{1}{n\pi}\right]^{-1}$$

$$=\frac{(i\omega/\omega_1)^{1/2}}{\sin(i\omega/\omega_1)^{1/2}} \tag{9.34}$$

and with $\omega_1=\omega_0\alpha/6$ Eq. (1.21) is reproduced. To compute Λ by the second method one finds immediately that

$$M_{ij}(z,\omega)=\delta_{ij}R\left(\frac{\omega_1}{i\omega}\right)^{1/2}\sin\left(\frac{z}{R}\left(\frac{i\omega}{\omega_1}\right)^{1/2}\right),$$

$$M_{ij}(z,0)=\delta_{ij}z,$$

and Eq. (9.32) yields the expected answer.

Once ω_g and Λ have been determined everything proceeds as in the homogeneous isotropic case. In particular $\mathscr{E}$ satisfies Gaussian statistics in the fuly saturated regime and in the partially saturated regime for $p<2$ (where Λ is unfortunately not known) it is a phase times a Gaussian. There is one new point worth mentioning. In the calculation of Sec. 4 $\langle\mathscr{I}(\tau)\rangle$ vanished for $\tau<0$ because Λ was analytic in the upper half-plane. When $\mu_{ij}(z)$ is nonzero there can be a finite number of positive eigenvalues ω_n. Then Λ is no longer analytic in the upper half plane and $\langle\mathscr{I}(\tau)\rangle$ is nonvanishing for

$\tau < 0$. This in fact happends for propagation of sound in the ocean.[7]

The transition to inhomogeneous anisotropic media has now been completed. The reader who is interested in seeing how the method works in detail for a realistic problem can consult the book of Flatte, *et al.*[7]

10. CONCLUSIONS

The path integral has turned out to be a powerful tool. It has provided a precise (very nearly), complete and global picture of what goes on in the saturated regimes. The unsaturated regime where the Rytov approximation is valid could also be treated by path integral methods. While this would lead to a more unified picture, in the end it would only amount to a rederivation of the Rytov approximation. A more fruitful endeavor would be to make an attack on the remaining unsolved problems in the saturated regimes. For situations where a scalar wave equation is sufficient and the (multiple) scattering angles are small the remaining problems are:

(a) How to compute (except numerically) the coherence in frequency, $\Lambda(\omega)$, for multiple scale media with $p < 2$.

(b) What are the detailed (beyond those given in Appendix D) statistics of $\mathscr{E}$ in the partially saturated regime for $4 > p > 2$?

(c) How to compute the second moment $\langle \mathscr{E}^*(2)\mathscr{E}(1)\rangle$ for those propagation paths in highly anisotropic media where the Markov approximation is not valid?

(d) What is the detailed behavior of $\mathscr{E}$ at the boundaries between the unsaturated and saturated regimes and between the fully and partially saturated regimes?

These are difficult problems which may not have any simple solution and, in particular, the path integral may not be the best method for attacking them. On the other hand, it is quite remarkable that the use of Feynman's path integral has reduced the problem to a few unknowns which occur only in special cases.

Among the other methods for treating wave propagation in random media, the most powerful ones uses the Markov approximation from the beginning. With the Markov approximation one can derive local partial differential equations for the moments of $\mathscr{E}$.[3-5] These equations have been stuided extensively, especially by the Russian school.[3,4] In the Markov approximation the path integrals for the moments are formal solutions to these partial differential equations. The equations for the first and second moments can be integrated analytically and correspondingly the path integrals can be done analytically. For the higher moments, the differential equations have yielded only some information about[4] $\langle I(1)I(2)\rangle$. The reason that this approach has not yielded more is that to determine the asymptotic (long-range) behavior of a function from its defining partial differential equation is highly nontrivial. The path integral has the advantage that it works on a global rather than local level, making it easier to determine the asymptotics.

The reader who is familar with Mercier's[12] treatment of the phase screen problem (an idealized case where all the scattering takes place on a thin sheet) will have noticed the similarity between his methods and those used here. The similarity is partly just the mathematics of manipulating integrals but there is also a physical reason. Any medium can be approximated by a (perhaps) large but finite number of phase screens. The wave field can then be expressed as a large but finite dimensional integral over the surfaces of the screens. But this is just the path integral in its finite form. Thus the path integral can be thought of as a scheme where one approximates the medium by n phase screens and then letting n go to infinity recovers the original problem.

APPENDIX A: CORRECTIONS TO THE MARKOV APPROXIMATION

The exact path integral for $\langle \mathscr{E}^*(2)\mathscr{E}(1)\rangle$ can be expanded as

$$\langle \mathscr{E}^*(2)\mathscr{E}(1)\rangle = \sum_{m=0}^{\infty} \frac{1}{m!}\frac{1}{4k^2}\int d^2(\text{paths})\exp\left(\frac{ik}{2}\int_0^R [(\dot{\mathbf{r}}_1(z))^2-(\dot{\mathbf{r}}_2(z))^2]dz - V_0\right)[V_0 - V]^m, \tag{A1}$$

where V is given by Eq. (2.10) and V_0 is the Markov approximation given by Eqs. (2.11). The $m=0$ term is the Markov approximation and the $m=1$ term will be computed below for the special case $\langle \mathscr{E}^*(1)\mathscr{E}(1)\rangle$.

For $\langle \mathscr{E}^*(1)\mathscr{E}(1)\rangle = \langle I\rangle$ the first correction contains $[V_0 - V]$ which can be replaced by $-V$ because, as may be seen from Sec. 2 the piece proportional to V_0 vanishes. We then have to first order

$$\langle I\rangle = \langle I\rangle_0 - \frac{1}{4k^2}\int d^2(\text{paths})\exp\left[ik\int_0^R \mathbf{v}'(z)\cdot\mathbf{u}'(z)dz - V_0\right]V, \tag{A2}$$

where paths $\mathbf{v}(z) = \mathbf{r}_1(z) - \mathbf{r}_2(z)$ and $\mathbf{u}(z) = \frac{1}{2}(\mathbf{r}_1(z) + \mathbf{r}_2(z))$ have been introduced. In terms of the Fourier transform $\tilde{\rho}$ of ρ [Eq. (5.3)] V is

$$V = 2k^2\int_0^R dz_1\int_0^R dz_2\int d^2q dq_z \tilde{\rho}(\mathbf{q}^2+q_z^2)^{1/2},0)\exp[iq_z(z_1-z_2)]\ \exp[i\mathbf{q}\cdot(\mathbf{u}(z_1)-\mathbf{u}(z_2))]\sin[\tfrac{1}{2}\mathbf{q}\cdot\mathbf{v}(z_1)]\sin[\tfrac{1}{2}\mathbf{q}\cdot v(z_2)], \tag{A3}$$

where $\mathbf{q} = (q_x, q_y)$ is a two-dimensional vector. Writing

$$\exp[i\mathbf{q}\cdot(\mathbf{u}(z_1)-\mathbf{u}(z_2))]=\exp\left\{i\mathbf{q}\cdot\int_0^R \mathbf{u}(z)[\delta(z-z_1)-\delta(z-z_2)]dz\right\}$$

and inserting V as given by Equation (A.3) into Eq. (A.2) one finds that since V_0 depends only on $\mathbf{v}$ the integral over $\mathbf{u}(z)$ can be done and that it leads to a δ-functional which forces $\mathbf{v}$ to satisfy the equation

$$\mathbf{v}''(z)=\frac{\mathbf{q}}{k}[\delta(z-z_1)-\delta(z-z_2)], \tag{A4}$$

with the boundary conditions $\mathbf{v}(0)=\mathbf{v}(R)=0$. In terms of the Greens function $g(z,z')$ which satisfies $(\partial^2/\partial z^2)g(z,z')=\delta(z-z')$ and $g(0,z')=g(R,z')=0$, $\mathbf{v}(z)$ is constrained to be $\mathbf{v}_0(z)$ where

$$\mathbf{v}_0(z)=\frac{\mathbf{q}}{k}(g(z,z_1)-g(z,z_2)). \tag{A5}$$

The path integral is then done by replacing $\mathbf{v}(z)$ by $\mathbf{v}_0(z)$ in V_0 and in the representation of V and the final result is

$$-\frac{\langle I\rangle-\langle I\rangle_0}{\langle I\rangle_0}=2k^2\int_0^R dz_1\int_0^R dz_2\int d^2q dq_z\tilde{\rho}(\sqrt{\mathbf{q}^2+q_z^2},0)\exp[iq_z(z_1-z_2)]$$

$$\times\sin\left[\frac{1}{2}\mathbf{q}\cdot\mathbf{v}_0(z_1)\right]\sin\left[\frac{1}{2}\mathbf{q}\cdot\mathbf{v}_0(z_2)\right]\exp\left[-\int_0^R d(|\mathbf{v}_0(z)|,0)dz\right]. \tag{A6}$$

To estimate the size of the integral in Eq. (A.6), one notes that $|z_1-z_2|$ will be restricted to order L or less and that for $|z_1-z_2|\sim L$, $\mathbf{q}\cdot\mathbf{v}_0(z)$ is of order $q^2L/k\sim q/k$ which is assumed to be small. The sines can then be expanded and using the identity

$$[g(z_1,z_1)-g(z_1,z_2)][g(z_2,z_1)-g(z_2,z_2)]=-\frac{(z_1-z_2)^2g(z_1,z_2)}{R} \tag{A7}$$

one finds

$$\frac{\langle I\rangle-\langle I\rangle_0}{\langle I\rangle_0}=\frac{1}{2R}\int_0^R dz_1\int_0^R dz_2\int d^2q dq_z\tilde{\rho}((\mathbf{q}^2+q_z^2)^{1/2},0)\mathbf{q}^4(z_1-z_2)^2$$

$$\times g(z_1,z_2)\exp\left[iq_z(z_1-z_2)-\int_0^R d(|\mathbf{v}_0(z)|,0)dz\right]. \tag{A8}$$

Ignoring the term $\int d(|\mathbf{v}_0(z)|,0)dz$ in the exponential which can only make the integral smaller, changing to variables $u=z_1-z_2$ and $\bar{z}=\frac{1}{2}(z_1+z_2)$, approximating their limits by $-\infty<u<\infty$ and $0<\bar{z}<R$, and setting $g(z_1,z_2)\approx g(\bar{z},\bar{z})$ then yields

$$\frac{\langle I\rangle-\langle I\rangle_0}{\langle I\rangle_0}=-\frac{R\hat{\rho}(0,0)}{3L^2}\sim-\frac{R\langle\mu^2\rangle}{3L}. \tag{A9}$$

The correction to Markov approximation to $\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle$ can be analyzed in the same way. It is fractionally small as long as $\langle\mu^2\rangle R/L$ is small.

APPENDIX B: CORRECTIONS TO GAUSSIAN STATISTICS

To begin with something simple, consider $\langle I^2\rangle$. It is given by the path integral in Eq. (3.1) with $t_i=t_j$ and the end point conditions $\mathbf{r}_i(0)=\mathbf{r}_i(R)='0$. Changing variables to $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{v}_1$, and $\mathbf{v}_2$ defined by

$$\begin{aligned}
\mathbf{r}_1(z)&=\mathbf{u}_1(z)+\tfrac{1}{2}\mathbf{v}_2(z)+\tfrac{1}{4}\mathbf{u}_2(z)+\tfrac{1}{2}\mathbf{v}_1(z),\\
\mathbf{r}_2(z)&=\mathbf{u}_1(z)+\tfrac{1}{2}\mathbf{v}_2(z)-\tfrac{1}{4}\mathbf{u}_2(z)-\tfrac{1}{2}\mathbf{v}_1(z),\\
\mathbf{r}_3(z)&=\mathbf{u}_1(z)-\tfrac{1}{2}\mathbf{v}_2(z)+\tfrac{1}{4}\mathbf{u}_2(z)-\tfrac{1}{2}\mathbf{v}_1(z),\\
\mathbf{r}_4(z)&=\mathbf{u}_1(z)-\tfrac{1}{2}\mathbf{v}_2(z)-\tfrac{1}{4}\mathbf{u}_2(z)+\tfrac{1}{2}\mathbf{v}_1(z),
\end{aligned} \tag{B1}$$

the integral over $\mathbf{u}_1(z)$ can be done and it produces a δ functional which forces $\mathbf{u}_2''(z)$ to vanish everywhere. With the end point conditions $\mathbf{u}_2(0)=\mathbf{u}_2(R)=0$, the only solution is $\mathbf{u}_2(z)=0$. The quadruple path integral then reduces to the double path integral over paths $\mathbf{v}_1$ and $\mathbf{v}_2$,

$$\frac{\langle I^2\rangle}{\langle I\rangle}=\frac{1}{4k^2}\int d^2(\text{paths})\exp\left[ik\int_0^R \mathbf{v}_1'(z)\cdot\mathbf{v}_2'(z)dz-M\right], \tag{B2}$$

where M was defined in Eq. (3.2) and for $\mathbf{u}_2=0$ it is explicitly

$$M=\int_0^R [2d(|\mathbf{v}_1(z)|,0)+2d(|\mathbf{v}_2(z)|,0)-d(|\mathbf{v}_1(z)+\mathbf{v}_2(z)|,0)-d(|\mathbf{v}_1(z)-\mathbf{v}_2(z)|,0)]dz. \tag{B3}$$

The two regions (a) and (b) discussed in Sec. 3 are $|\mathbf{v}_1(z)|<L/\Phi$ with $\mathbf{v}_2$ arbitrary and $|\mathbf{v}_2(z)|<L/\Phi$ with $\mathbf{v}_1$ arbitrary.

It was pointed out in the text that M is of order unity or smaller throughout regions (a) and (b). Actually, there is a further region [having some overlap with (a) and (b)] where M can be small. It is (c) $|\mathbf{v}_1(z)|<L/\sqrt{\Phi}$, $|\mathbf{v}_2(z)|<L/\sqrt{\Phi}$ and owes its existence to the fact that when both $|\mathbf{v}_1(z)|$ and $|\mathbf{v}_2(z)|$ are small, M is quartic in the $\mathbf{v}$'s. In all other regions of path space, e^{-M} is exponentially small.

Our first task is to dispose of the extra region (c) by showing that for small α the volume of path space occupied by this region is exponentially small compared to the volumes occupied by regions (a) and (b). An estimate of the volume of path space occupied by region (a) is

$$\frac{1}{4k^2}\int d^2(\text{paths})\exp\left(ik\int_0^R \mathbf{v}_1(z)\cdot\mathbf{v}_2'(z)dz-\frac{\Phi^2}{L^2R}\int_0^R(\mathbf{v}_1(z))^2dz\right)=1, \tag{B4}$$

where the integral is done by integrating over $\mathbf{v}_2(z)$ which produces a δ functional that forces $\mathbf{v}_1(z)$ to vanish. An estimate of the volume occupied by region (c) is

$$\frac{1}{4k^2}\int d^2(\text{paths})\exp\left(ik\int_0^R \mathbf{v}_1(z)\cdot\mathbf{v}_2'(z)dz-\frac{\Phi}{2RL^2}\int_0^R(\mathbf{v}_1^2(z)+\mathbf{v}_2^2(z))dz\right)=\frac{(6/\alpha)}{\sinh^2\sqrt{3/\alpha}+\sin^2\sqrt{3/\alpha}}\sim\frac{24}{\alpha}\exp(-2\sqrt{3/\alpha}), \tag{B5}$$

where the value of the path integral is taken from Ref. 6. For small α, the volume occupied by region (c) is therefore exponentially small compared to the volume occupied by regions (a) and (b). This result, which may surprise some readers, deserves an explanation. In region (a) where $|\mathbf{v}_1|$ is always less than L/Φ, the factor

$$\exp\left[ik\int_0^R \mathbf{v}_1(z)\cdot\mathbf{v}_2'(z)dz\right]=-\exp\left[ik\int_0^R \mathbf{v}_1(z)\cdot\mathbf{v}_2''(z)dz\right]$$

will restrict $|\mathbf{v}_2''|$ to values less than $\Phi/(kLR)$. Typical values of $|\mathbf{v}_2|$ will then be $R^2|\mathbf{v}_2''|/6\sim L/\alpha$. At a given range point z_0, the variables $\mathbf{v}_1(z_0)$ and $\mathbf{v}_2(z_0)$ span a four-dimensional space. In this space the volume occupied by paths in region (a) is roughly $L^4/(\Phi\alpha)^2$. At the same point the volume occupied by paths in region (c) is roughly L^4/Φ^2. Thus at each point z_0, the volume associated with region (c) is a factor of α^2 smaller than that occupied by region (a). To compute the total volume in path space, one has to multiply together the volumes at each range point z_0, taking into account the fact that the paths cannot bend too rapidly. The path integrals in Eqs. (B4) and (B5) do just this. The resulting exponential ratio of volumes should no longer be a surprise since at each range point the ratio is down by α^2.

It is therefore sufficient to consider only paths lying in regions (a) and (b). The fact that integrating separately over regions (a) and (b) leads to a slight over counting can also be ignored. The volume in path space where regions (a) and (b) overlap is even smaller than the volume occupied by region (c). Now as was pointed out in Sec. III, for most paths in region (a) $M\simeq M_0^{(a)}$, where

$$M_0^{(a)}=2\int_0^R d(|\mathbf{v}_1(z)|,0)dz \tag{B6}$$

and for most paths in region (b) $M\simeq M_0^{(b)}$ with

$$M_0^{(b)}=2\int_0^R d(|\mathbf{v}_2(z)|,0)dz. \tag{B7}$$

The path integral in Eq. (B2) can then be expanded according to

$$\int d^2(\text{paths})\exp\left[ik\int_0^R \mathbf{v}_1(z)\cdot\mathbf{v}_2'(z)dz-M\right]\sim\sum_{m=0}^{\infty}\int d^2(\text{paths})\exp\left[ik\int_0^R \mathbf{v}_1(z)\cdot\mathbf{v}_2'(z)dz-M_0^{(a)}\right]\frac{(M_0^{(a)}-M)^m}{m!}$$
$$+\sum_{m=0}^{\infty}\int d^2(\text{paths})\exp\left[ik\int_0^R \mathbf{v}_1(z)\cdot\mathbf{v}_2'(z)dz-M_0^{(b)}\right]\frac{(M_0^{(b)}-M)^m}{m!} \tag{B8}$$

which is an asymptotic series in α. It is not a convergent series because (exponentially small) contributions from region (c) and the overlap of regions (a) and (b) are not being treated correctly. The $m=0$ terms correspond to Rayleigh statistics and the $m=1$ terms are the first correction. They will be computed explicitly below.

Equations (B8) generalizes to an arbitrary correlation in the obvious way. For a 2nth order moment there are $n!$ important regions of path space. In each such region there is an M_0 given by the analog of Eq. (3.4) or (3.5). The generalization of Eq. (B8) is then a sum of $n!$ terms, each of which is a series of powers of the appropriate M_0-M.

The path integrals for the $m=1$ terms in Eq. (B8) can be evaluated by inserting a spectral representation for M_0-M. If $\tilde{\rho}$ is the three-dimensional Fourier transfer of ρ [see Eq. (5.3)], then in region (a)

$$M_0^{(a)}-M=4\pi k^2\int_0^R dz\int d^2q\,\tilde{\rho}(|\mathbf{q}|)\exp[i\mathbf{q}\cdot\mathbf{v}_2(z)]\,[1-\cos(\mathbf{q}\cdot\mathbf{v}_1(z))] \tag{B9}$$

and

$$\frac{1}{4k^2}\int d^2(\text{paths})\exp\left[ik\int_0^R \mathbf{v}_1'(z)\cdot\mathbf{v}_2'(z)dz-M_0^{(a)}\right][M_0^{(a)}-M\,]$$
$$=\pi\int_0^R dz'\int d^2q\tilde{\rho}(|\mathbf{q}|)\int d^2(\text{paths})\exp\left[-ik\int_0^R \mathbf{v}_2(z)\cdot\left(\mathbf{v}_1''(z)-\frac{\mathbf{q}}{k}\,\delta(z-z')\right)dz-M_0^{(a)}\right][1-\cos(\mathbf{q}\cdot\mathbf{v}_1(z'))]. \tag{B10}$$

Since M_0^a depends only on $\mathbf{v}_1$, the integral over $\mathbf{v}_2$ can be done and it produces a δ functional which forces $\mathbf{v}_1''(z)-(\mathbf{q}/k)\delta(z-z')$ to vanish. In terms of the Greens function $g(z,z')$ defined by

$$\frac{\partial^2}{\partial z^2}g(z,z')=\delta(z-z') \tag{B11}$$

with boundary conditions $g(0,z')=g(R,z')=0$, $\mathbf{v}_2(z)$ is constrained to be $(\mathbf{q}/k\,)g(z,z')$. In $M_0^{(a)}$ and $\cos(\mathbf{q}\cdot\mathbf{v}_1(z'))$ one can then set $\mathbf{v}_1(z)$ equal to $(\mathbf{q}/k\,)g(z,z')$ and the remaining path just gives $\langle I\rangle$. The calculation of the correction in region (b) is identical and to leading order in α,

$$\frac{\langle I^2\rangle-2\langle I\rangle^2}{\langle I\rangle^2}=4\pi k^2\int_0^R dz\int d^2q\,\tilde{\rho}(|\mathbf{q}|)Q(z,|\mathbf{q}|) \tag{B12}$$

with

$$Q(z,|\mathbf{q}|)=2\left[1-\cos\left(\frac{\mathbf{q}^2}{k}g(z,z)\right)\right]\exp\left[-2\int_0^R d\left(\frac{|\mathbf{q}|}{k}g(z,z'),0\right)dz'\right]. \tag{B13}$$

An examination of the integral on the right-hand side of Eq. (B12) shows that for small α,

(i) $d\left(\frac{|\mathbf{q}|}{k}g(z,z'),0\right)$

can be approximated by

$\frac{1}{2}k^2\hat{\rho}(0,0)\left(\frac{\mathbf{q}g(z,z')}{kL}\right)^2$,

(ii) $2\left[1-\cos\left(\frac{\mathbf{q}^2}{k}g(z,z)\right)\right]$ can be set equal to $\left(\frac{\mathbf{q}^2}{k}g(z,z)\right)^2$,

and

(iii) the dominant contribution comes from the regions $z\approx 0$ and $z\approx R$, where $g(z,z')\approx z(R-z')/R$ and $z'(R-z)/R$.

The contributions from the regions $z\approx 0$ and $z\approx R$ are the same and

$$\frac{\langle I^2\rangle-2\langle I\rangle^2}{\langle I\rangle^2}=8\pi k^2\int_0^\infty dz\int d^2q\tilde{\rho}(|\mathbf{q}|)\left(\frac{\mathbf{q}^2z}{k}\right)^2\exp\left[-\frac{R\hat{\rho}(0,0)z^2\mathbf{q}^2}{3L^2}\right]$$

$$=\alpha\left(\frac{3\pi}{4}\right)^{1/2}\frac{L\int_0^\infty q^2\tilde{\rho}(q,0)dq}{\int_0^\infty q\tilde{\rho}(q,0)dq}. \tag{B14}$$

The correction to $\langle I^n\rangle$ involves $n!$ regions of path space and in each of these regions there are $n(n-1)/2$ terms in M_0-M which differ only by permutations of the paths. The result is that the correction to $\langle I^n\rangle$ is $n!(n)\,(n-1)/4$ times the correction to $\langle I^2\rangle$.

Moving on to a more complicated object, consider $\langle I(1)I(2)\rangle$. It is given by the path integral in Eq. (B2) but the end point conditions on the paths are now $\mathbf{v}_1(0)=\mathbf{v}_1(R\,)=0$ and $\mathbf{v}_2(0)=\mathbf{r}_{01}-\mathbf{r}_{02}$, $\mathbf{v}_2(R\,)=\mathbf{r}_1-\mathbf{r}_2$ and now M is

$$M=\int_0^R [2d(|\mathbf{v}_1(z)|,0)+2d(|\mathbf{v}_2(z)|,t_1-t_2)-d(|\mathbf{v}_1(z)+\mathbf{v}_2(z)|,t_1-t_2)-d(|\mathbf{v}_1(z)-\mathbf{v}_2(z)|,t_1-t_2)]dz. \tag{B15}$$

The integration over region (a) gives $\langle I\rangle^2(1+\text{ corrections})$ while the integration over region (b) gives $\langle I\rangle^2 e^{-D(1,2)}(1+\text{ corrections})$. As indicated the corrections in region (b) are proportional to e^{-D} and are a small effect of no particular consequence. The corrections in region (a), on the other hand, are small but do not contain e^{-D} and hence fall much less rapidly. This leads to a coherence tail in $\langle I(1)I(2)\rangle$ which is not present in $\langle\mathscr{E}^*(1)\mathscr{E}(2)\rangle$. The interesting corrections in region (a) are computed by

changing variables from $\mathbf{v}_1(z)$ and $\mathbf{v}_2(z)$ *to* $\mathbf{v}_1(z)$ and $\mathbf{w}(z)=\mathbf{v}_2(z)-(z/R)(\mathbf{r}_1-\mathbf{r}_2)-[(R-z)/R](\mathbf{r}_{10}-\mathbf{r}_{02})$ and then proceeding in exactly the same way as before. The result is

$$\langle I(1)I(2)\rangle=\langle I\rangle^2[1+e^{-D(1,2)}+\gamma(\mathbf{r}_1-\mathbf{r}_2,\mathbf{r}_{01}-\mathbf{r}_{02},t_1-t_2)], \tag{B16}$$

where

$$\gamma(\mathbf{r},\mathbf{r}_0,t)=2\pi k^2\int_0^R dz\int d^2q\,\exp\left(i\frac{z\mathbf{q}\cdot\mathbf{r}}{R}+i\frac{(R-z)\mathbf{q}\cdot\mathbf{r}_0}{R}\right)\tilde{\rho}(|\mathbf{q}|,t)Q(z,|\mathbf{q}|) \tag{B17}$$

and Eq. (B12) is not reproduced for $I(1)=I(2)$ because a small term of order γe^{-D} from region (b) has been dropped. For small α this integral can be simplified in the same way that Eq. (B14) was obtained from Eq. (B12). It becomes

$$\gamma(\mathbf{r},\mathbf{r}_0,t)=\frac{\alpha\sqrt{3\pi}}{8}\,\frac{L\int_0^\infty q^2\tilde{\rho}(q,t)[J_0(q|\mathbf{r}|)+J_0(q|\mathbf{r}_0|)]dq}{\int_0^\infty d\tilde{\rho}(q,0)dq}. \tag{B18}$$

Corrections to more complicated correlations and terms of order α^2 or higher can also be computed—the only obstacle being the labor involved. The calculation of the general coherence tail involves only some combinatorics. It is

$$\frac{\langle\Pi_{k=1}^n(I(k))^{m_k}-\Pi_{k=1}^n\langle I(k)\rangle^{m_k}\rangle}{\Pi_{k=1}^n\langle I(k)^{m_k}\rangle}=\text{"Gaussian terms"}+\sum_{k,j=1}^n m_km_j\gamma(k-j), \tag{B19}$$

where $\gamma(k-j)=\gamma(\mathbf{r}_k-\mathbf{r}_j,\mathbf{r}_{0k}-\mathbf{r}_{0j},t_k-t_j)$, the "Gaussian terms" are what one would compute from the Gaussian distribution and all terms of order e^{-D} have been dropped.

APPENDIX C: CORRECTIONS TO GAUSSIAN STATISTICS FOR $p<2$

Equations (B12), (B13), (B16), and (B17) of Appendix B do not assume a single scale media and will be the starting point. For $p<2$, in either the fully or the partially saturated regime Q can be approximated by

$$Q(z,|\mathbf{q}|)\approx\left(\frac{\mathbf{q}^2}{k}g(z,z)\right)^2\exp\left(-\frac{\Phi^2}{p+1}\left|\frac{|\mathbf{q}|g(z,z)}{kL}\right|^p\right), \tag{C1}$$

where the cosine has been expanded, the short distance expansion for $\hat{\rho}$, Eq. (7.2), has been used and the identity

$$\int_0^R|g(z,z')|^p dz'=\frac{R}{p+1}|g(z,z)|^p \tag{C2}$$

has been employed.

In the fully saturated regime the main contribution to Eqs. (B12) and (B17) again comes from $z\approx 0$ and $z\approx R$. Using Eq. (C2) for Q then yields Eqs. (B14) and (B18) with α replaced by α' where α' is defined in Eq. (7.3).

In the partially saturated regime all values of z contribute to the integral but the dominant contribution comes from large $|q|$ where

$$\tilde{\rho}(q,t)=\frac{\hat{\rho}(0,t)2^p[\Gamma(1+\frac{1}{2}p)]^2\sin(\pi p/2)}{4\pi^3|L|^p|q|^{p+2}}. \tag{C3}$$

For $n=2$, Eq. (8.3), with

$$C(p)=\{(p+1)^{(4-p)/p}2^p[\Gamma(1+\tfrac{1}{2}p)\Gamma(p-1)]^2$$

$$\times\sin(\pi p/2)\Gamma((4-p)/p)/p\pi 6^{(2-p)}\Gamma(2p-2) \tag{C4}$$

is obtained by inserting Eqs. (C3) (with $t=0$) and (C1) in Eq. (B12). The extension to general n works in the same way as before. The coherence tails in the partially saturated regime are obtained by inserting Eqs. (C3) and (C1) into Eq. (B17). For $\mathbf{r},\mathbf{r}_0\neq 0$ this leads to integrals which cannot be done analytically.

APPENDIX D: THE PARTIALLY SATURATED REGIME FOR $p>2$

It is difficult to make quantitative statements about the partially saturated regime when $4>p>2$. There is however some qualitative information.

Equation (1.14) holds and $\langle\mathscr{E}\rangle$ is equal to $\mathscr{E}_0\exp(-\frac{1}{2}\Phi^2)$ in all regimes as long as the Markov approximation is valid. Furthermore $\langle\mathscr{E}^*(\omega')\mathscr{E}(\omega)\rangle$ continues to be given by Eqs. (1.19)–(1.21). The argument that any correlation involving an unequal number of $\mathscr{E}$'s and $\mathscr{E}^*$'s vanishes also goes through as before. Thus $\mathscr{E}$ is uniformly distributed in phase. The difficulty arises when one attempts to compute the nonvanishing higher moments.

The statistics are not Gaussian. This can be verified by assuming that they are and then computing the corrections. They are not small. Some information can be obtained however by comparing the path integral for $\langle(\mathscr{E}^*(2))^2(\mathscr{E}(1))^2\rangle$ with that for $\langle|\mathscr{E}(2)|^2|\mathscr{E}(1)|^2\rangle$. Upon doing this one finds that $\mathscr{E}$ always phase wraps as shown in Fig. 5(b). It turns out that the typical space time scales over which the phase and intensity change are those listed in Table DI [the parameter δ was defined in Eq. (8.4)]. Note that for partial saturation where $\Phi^{4/p}/\Omega>1$ but $\Phi/\Omega<1$ the rate at which the intensity changes is always small compared to the rate at which the phase changes. Examining more complicated correlations leads to the conclusion that at a fixed frequency $\mathscr{E}$ can be represented as

$$\mathscr{E}(j)=\mathscr{E}_0(j)\exp[i\phi(j)]\chi(j), \tag{D1}$$

where $\phi(j)$ is a real Gaussian random variable with $\langle\phi(j)\rangle=0$ and

$$\langle[\phi(i)-\phi(j)]^2\rangle=D(i,j) \tag{D2}$$

The other factor χ is an independent (of ϕ) complex random variable about which only three things are known:

(1) any correlation involving an unequal number of χ's and

TABLE DI. Time and space scales associated with phase and $\ln I$ in the partially saturated regime with $4>p>2$.

Phase	L/Φ		T/Φ
$\ln I$	$(L/\Phi)(\Omega/\Phi)^{(p-2)/(4-p)}$	$(T/\Phi)(\Omega/\Phi)^{2/(4-p)}$ $p-\delta>2$	$(T/\Phi)(\Omega/\Phi)^{(p-\delta)/(4-p)}$ $2>p-\delta>0$
	Scale Length		Scale Time

χ^*'s vanishes,

(2) $\langle|\chi|^2\rangle=1$,

and

(3) the decorrelation lengths and times for χ are those listed
under intensity in Table DI.

To see what the representation in Eq. (D1) means consider

$$\frac{\langle\mathscr{E}^*(2)\mathscr{E}(1)\rangle}{\mathscr{E}_0^*(2)\mathscr{E}_0(1)}=\langle\exp\{i[\phi(1)-\phi(2)]\}\rangle\langle\chi^*(2)\chi(1)\rangle$$

$$=\exp[-\tfrac{1}{2}D(1,2)]\langle|\chi|^2\rangle, \tag{D3}$$

where to get the second line one notes that $\langle\chi^*(2)\chi(1)\rangle$ will be approximately $\langle|\chi|^2\rangle$ for all space or time separations such that $\exp(-\frac{1}{2}D)$ is not vanishingly small. Thus, Eq. (1.19) is reproduced, as it should be. Similarly,

$$\frac{\langle(\mathscr{E}^*(2)^2(\mathscr{E}(1))^2\rangle}{(\mathscr{E}_0^*(2))^2(\mathscr{E}_0(1))^2}=\exp[-2D(1,2)]\langle|\chi|^4\rangle \tag{D4}$$

and this correlation is known up to a constant. However, all that is known about the intensity correlation

$$\langle I(2)I(1)\rangle=\langle|\chi(2)|^2|\chi(1)|^2\rangle \tag{D5}$$

are its space and time scales.

The extension to unequal frequencies is straightforward. At different frequencies $\langle(\phi(\omega)-\phi(\omega'))^2\rangle=[(\omega-\omega')/\omega_g]^2$ and $\langle\chi^*(\omega')\chi(\omega)\rangle=\Lambda(\omega-\omega')$. The higher order moments of χ are again non-Gaussian and unknown. However, their width in ω is large compared to ω_g. As in Sec. 8 this means that pulse propagation is dominated by wander rather than spreading.

As was mentioned in Sec. 8, there is a case where the non-Gaussian statistics of χ can be studied in detail. It is for correlations in time when $p=2$ and $\delta=0$ and is explained in Ref. 7.

The above results are most easily derived using the Fermat path formalism of Sec. 6. One can work out the joint probability that two paths will satisfy the perturbed ray equation. In the partially saturated regime with $p>2$ it turns out that the Fermat paths are highly correlated and tend to lie within $L(\Phi/\Omega)^{2/(4-p)}$ of each other. Studying averages of μ and μ' along correlated Fermat paths then leads to the above conclusions. The detailed calculations are relatively straightforward but tedious and will not be given here.

APPENDIX E: CORRECTIONS TO THE MARKOV APPROXIMATION FOR INHOMOGENEOUS ANISOTROPIC MEDIA

If the $\bar{\mathbf{x}}$ dependence of ρ is evaluated along the unperturbed ray then the first correction to the Markov approximation can be evaluated for a general homogeneous anisotropic medium. Let

$$\rho(\mathbf{x},t;\mathbf{s}(z)+\mathbf{e}_z z)=\int d^3l e^{i\mathbf{l}\cdot\mathbf{x}}\tilde{\rho}(\mathbf{l},t;z), \tag{E1}$$

then the generalization of Eq. (A6) is

$$-\frac{\langle I\rangle-\langle I\rangle_0}{\langle I\rangle_0}$$
$$=2k^2\int_0^R dz_1\int_0^R dz_2\int d^2q\,dq_z\tilde{\rho}(\mathbf{q}+\mathbf{e}_z q_z,0;\tfrac{1}{2}(z_1+z_2))$$
$$\times\exp[i(q_z+\mathbf{q}\cdot\mathbf{s}'(\tfrac{1}{2}(z_1+z_2))(z_1-z_2)]\sin[\tfrac{1}{2}\mathbf{q}\cdot\mathbf{v}_0(z_1)]$$
$$\times\sin[\tfrac{1}{2}\mathbf{q}\cdot\mathbf{v}_0(z_2)]\exp\left[-\int_0^R d(\mathbf{v}_0(z),0;z)dz\right], \tag{E2}$$

where $q=(q_x,q_y)$ is a two-dimensional vector, d is defined in Eq. (9.11), and

$$v_0(z)_i=\frac{q_j}{k}[g_{ij}(z,z_1)-g_{ij}(z,z_2)] \tag{E3}$$

with g_{ij} defined in Eq. (9.22).

For an isotropic medium Eq. (E2) can be analyzed in the same way as Eq. (A6) and one finds that $(\langle I\rangle-\langle I\rangle_0)/\langle I\rangle_0$ is of order of the rms multiple scattering angle.

It is also straightforward to analyze Eq. (E1) for a homogeneous but anisotropic medium. The result of doing this was stated in Sec. 9 A.

APPENDIX F: CORRECTIONS TO GAUSSIAN STATISTICS FOR INHOMOGENEOUS ANISOTROPIC MEDIA

When the approximation of Eq. (9.7) for the correlation between two paths is made, it is possible to compute the corrections to Gaussian statistics in the saturated regimes. The calculation is a fairly straightforward generalization of that done in Appendix B and only the final result will be given.

Define a function $\tilde{\rho}_1(\mathbf{q};z)$ where $\mathbf{q}=(q_x,q_y)$ by

$\tilde{\rho}_1(\mathbf{q},t;z)=\tilde{\rho}(\mathbf{q}-\mathbf{e}_z(\mathbf{s}'(z)\cdot\mathbf{q}),t;z)$ where $\tilde{\rho}$ is defined in Eq. (E1). Then the analog of Eq. (B12) is

$$\frac{\langle I^2\rangle-2\langle I\rangle^2}{\langle I\rangle^2}=4\pi k^2\int_0^R dz\int d^2q\tilde{\rho}_1(\mathbf{q},0;z)Q_1(z,\mathbf{q}), \tag{F1}$$

where

$$Q_1(z,\mathbf{q})=2[1-\cos q_i q_j g_{ij}(z,z)k^{-1}]$$

$$\times\exp\left[-2\int_0^R d(\mathbf{v}_1(z,z'),0;z')dz'\right] \tag{F2}$$

with

$$v_1(z,z')_i=k^{-1}q_j g_{ij}(z,z') \tag{F3}$$

and d and g_{ij} are defined in Eqs. (9.11) and (9.22).

Using Ω and Φ as defined by Eqs. (9.21) and (9.6) it is possible to show that the right-hand side of Eq. (F1) is small in the fully saturated regime and in the partially saturated regime for $p<2$. As before, in the fully saturated regime the dominant contribution comes from the regions $z\approx 0$ and $z\approx R$ and in the partially saturated regime $\tilde{\rho}_1$ can be approximated by its aymptotic form for large $|\mathbf{q}|$. Also, Q_1 can be simplified by expanding the cosine and replacing d by its expansion for small $\mathbf{v}_1$. The detailed calculation which is then fairly straightforward will be left to the reader.

The generalization to $\langle I^n\rangle$ works in the same way as in Appendix B.

The coherence tail is given by Eq. (B16) with

$$\gamma(\mathbf{r}_1-\mathbf{r}_2,\mathbf{r}_{01}-\mathbf{r}_{02},t)=2\pi k^2\int_0^R dz\int d^2q$$

$$\times\exp[i\mathbf{q}\cdot\boldsymbol{\epsilon}(z)]\tilde{\rho}_1(\mathbf{q},t;z)Q_1(\mathbf{q},z) \tag{F4}$$

where $\boldsymbol{\epsilon}$ is defined in Eq. (91.4). The approximations mentioned above can also be made in the integral for γ. Finally, Eq. (B19) holds with γ given by Eq. (F4).

[1] V. Tatarskii, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961).
[2] L. Chernov, *Wave Propagation in a Random Medium* (McGraw-Hill, New York, 1960).
[3] V. Tatarskii, *The Effects of Turbulent Atmospheres on Wave Propagation* (National Technical Information Services, Springfield, Virginia, 1971).
[4] A. Prokhorov, F. Bunkin, K. Gochelashvily, and V. Shishov, Proc. IEEE **63**, 790 (1975).
[5] Radio Sci. **10**, (1975), Special Issue: Waves in Random Media.
[6] R.P. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[7] S. Flatté, Ed., *Sound Transmission through a Fluctuating Ocean*, to be published by the Cambridge U. P.

[8] The technical meaning of "a single scale" is that the three-dimensional Fourier transform $\tilde{\rho}(\mathbf{q})$ of the covariance of μ should fall faster than $|\mathbf{q}|^{-6}$ at large $|\mathbf{q}|$. See Sec. 7.
[9] The assumption that ρ's at different times are jointly Gaussian, places restrictions on the dynamics of the medium. It is consistent with either the Taylor hypothesis (convection of a frozen field by a "wind") or time dependence due to linear wave motion.
[10] Boundary conditions corresponding to, say, a plane wave at $z=0$ can be obtained by superposition.
[11] In deriving this equation one neglects $(\partial^2/\partial z^2)\mathcal{A}$ relative to $2ik(\partial/\partial z)\mathcal{A}$, μ^2 relative to 2μ and time derivatives of $\mathcal{A}$. The latter requires the assumption that the medium does not change while a wave travels a distance L. This is true if condition (ii) (i.e., $kL\ll\omega T$) holds. Because R is always taken to be large compared to L, condition (iii) implies $\langle\mu^2\rangle^{1/2}\ll 1$.

[12] R. Mercier, Proc. Cambridge Philos. Soc. **58**, 382 (1962).

[13] The line $\Phi=\Omega$ is actually in a transition zone where the fluctuations are larger than Rayleilgh, see Sec. 7.
[14] Note that the rms multiple scattering angle is $6L/R\alpha$ so that α is restricted to be greater than $\sim 6L/R$. However, R is usually very large compared to L and there is no problem here.
[15] In the optics literature (Refs. 4 and 5) there is some controversy as to whether or not $P(I)$ is Rayleigh in the saturated regime. Asymptotically it is but the corrections may be substantial in some experiments, see Secs. 7 and 8.
[16] To make this rigorous, start with the finite form of the path integral in Eq. (2.1) and repeat the steps (including a summation by parts in the first term in the exponential) leading to Eq. (2.14) which is now an integral of finite dimension. Integrating over the discrete variables, $\mathbf{u}_k$ will produce a product of δ functions which forces the $\mathbf{v}_k$ to satisfy $\mathbf{v}_{k-1}-2\mathbf{v}_k+\mathbf{v}_{k+1}=0$ with $\mathbf{v}_0=\mathbf{r}_{01}-\mathbf{r}_{02}$ and $\mathbf{v}_n=\mathbf{r}_1-\mathbf{r}_2$. There is a unique solution and in the continuum limit Eq. (2.16) is reproduced.
[17] Up to a normalization it is the path integral for the unperturbed problem. The correct normalization is obtained by comparing both sides of the equation for the case where the fluctuations vanish and $d=0$.

[18] J. Lawson and G. Uhlenbeck, *Threshold Signals* (McGraw-Hill, New York 1950).

[19] It is also implicitly assumed that the signal is in the saturated regime for all important frequencies in f.
[20] This is M for region (a). In Appendix B it is shown that the integration over the center of gravity of all four paths produces a δ functional which forces the difference between members of each pair to be equal ($\mathbf{v}_1$).
[21] The case where there are many unperturbed rays is treated in detail in Ref. 7.
[22] When multiplying vectors and matrices the summation convention (repeated indices are summed over) is used.
[23] This result is equivalent to that of T.P. Williams, Proc. Roy. Soc. A **342**, 131 (1975). The author was not aware of his work at the time of writing.

Higgs mechanism and the inverse Einstein–Infeld–Hoffman problem

Ronald L. Mallett

Department of Physics, The University of Connecticut, Storrs, Connecticut 06268
(Received 21 August 1978)

The Guralnik–Hagen model for a self-coupled spin $-1/2$ field is minimally coupled to a gauge gravitational field. The corresponding free gravitational Lagrangian is not explicitly introduced. It is shown that, using spontaneous breakdown of Lorentz invariance and the generalized Gordon decomposition, the Higgs mechanism generates a gravitational Lagrangian which leads to the usual linearized Einstein gravitational field equations.

I. INTRODUCTION

A reformulation of the energy–momentum conservation laws consistent with the group of general coordinate transformations is well known to lead to the conclusion that the geodesic equations of motion of matter follow, as a necessary consequence, from the gravitational field equations. This result in one form or another may be called the Einstein–Infeld–Hoffman (EIH) problem.[1] One particularly simple formulation of this problem follows from the generally covariant zero divergence of the energy–momentum tensor which is assured by the field equations and the Bianchi identities.[2] By application of a suitable transformation to the covariant divergence a form is obtained which involves the ordinary divergence of the sum of the energy–momentum tensor of matter and pseudotensor of gravity. The symmetrical nature of this formulation of the conservation laws suggests the following question: To what extent are the gravitational field equations determined by the equations of motion of matter? This question may be called the *inverse EIH problem*.

In a previous paper,[3] we examined the consequences of the spontaneous breakdown of local Lorentz invariance for a self-coupled local spin-1 field interacting with a gauge gravitational field.[4] The corresponding gravitational Lagrangian was not explicitly introduced. It was found that this model produced a term in the Lagrangian via Higgs mechanism[5] which lead, in the static weak field approximation, to the Newtonian gravitational field equation. This model was particularly interesting because it suggested that, for the gravitational field, the gauge coupled matter field equations determined to some extent the structure of the gauge gravitational field equations of motion. In this paper we shall study the analogous problem for the Guralnik–Hagen model[6] of a self-coupled spinor field interacting with a gauge gravitational field. It is shown that, using spontaneous breakdown of Lorentz invariance and the generalized Gordon decomposition,[7] the Higgs mechanism generates a term in the Lagrangian which leads to the usual linearized Einstein field equations.

II. SELF-COUPLED LOCAL SPIN-$\frac{1}{2}$ FIELD AND THE GAUGE GRAVITATIONAL FIELD

We shall take as our model the Guralnik–Hagen Lagrangian[6] of a self-coupled spin-$\frac{1}{2}$ field and minimally couple it to a gauge gravitational field[4]

$$L = e\bar{\psi}\gamma^{\nu}(x)[\partial_{\nu} - \tfrac{1}{4}\sigma^{(\alpha\beta)}F_{\nu(\alpha\beta)}]\psi + P(\psi), \tag{1}$$

with

$$\gamma^{\nu}(x) = e^{\nu}_{(\alpha)}(x)\gamma^{(\alpha)}, \tag{2}$$

$$F_{\nu(\alpha\beta)} = e^{\tau}_{(\beta)}\nabla_{\nu}e_{(\alpha)\tau}, \tag{3}$$

$$\nabla_{\mu}e_{(\alpha)\nu} = \partial_{\mu}e_{(\alpha)\nu} - \Gamma^{\rho}_{\mu\nu}e_{(\alpha)\rho}, \tag{4}$$

$$\Gamma^{\rho}_{\mu\nu} = \tfrac{1}{2}g^{\rho\lambda}(\partial_{\nu}g_{\lambda\mu} + \partial_{\mu}g_{\nu\lambda} - \partial_{\lambda}g_{\mu\nu}), \tag{5}$$

$$g_{\mu\nu} = \delta_{(\alpha\beta)}e^{(\alpha)}{}_{\mu}e^{(\beta)}{}_{\nu}, \tag{6}$$

$$\delta_{(\alpha\beta)} = (-1,1,1,1), \tag{7}$$

$$\sigma^{(\alpha\beta)} = \tfrac{1}{2}(\gamma^{(\alpha)}\gamma^{(\beta)} - \gamma^{(\beta)}\gamma^{(\alpha)}), \tag{8}$$

$$e = \det(e_{(\alpha)\mu}), \tag{9}$$

$$P(\psi) = -\mu\bar{\psi}\psi + \tfrac{1}{2}\lambda(\bar{\psi}\psi)^2, \quad \mu,\lambda = \text{const} > 0, \tag{10}$$

where indices in parentheses indicate local Lorentz components. The vierbien field $e_{(\alpha)\mu}$ was originally introduced as a means of describing a spinor function ψ in a general Riemannian manifold where ψ transforms like a scalar under general coordinate transformations. The Lagrangian (1) is invariant under the combined space–time dependent Lorentz transformation of the local spinor fields ψ, $\bar{\psi}$ and gauge potentials $F_{\nu(\alpha\beta)}$ given by

$$\psi \to [1 + \tfrac{1}{4}\sigma^{(\alpha\beta)}\omega_{(\alpha\beta)}(x)]\psi, \tag{11}$$

$$\bar{\psi} \to \bar{\psi}[1 - \tfrac{1}{4}\sigma^{(\alpha\beta)}\omega_{(\alpha\beta)}(x)], \tag{12}$$

$$F_{\nu(\alpha\beta)} \to F_{\nu(\alpha\beta)} + \omega_{(\alpha\lambda)}(x)F_{\nu}{}^{(\lambda)}{}_{(\beta)} + \omega_{(\beta\lambda)}(x)F_{\nu(\alpha)}{}^{(\lambda)} + \partial_{\nu}\omega_{(\alpha\beta)}(x), \tag{13}$$

with

$$\gamma^{\nu}(x) \to \gamma^{\nu}(x) + \omega_{(\alpha\beta)}(x)e^{(\beta)\nu}\gamma^{(\alpha)}, \tag{14}$$

where

$$\omega_{(\alpha\beta)} = -\omega_{(\alpha\beta)} \quad \text{and} \quad F_{\nu(\alpha\beta)} = -F_{\nu(\beta\alpha)}.$$

III. GENERALIZED GORDON DECOMPOSITION OF L

The generalized Gordon decomposition[7] is a useful technique for separating the interaction density into a "convective" (i.e., γ matrix independent) part and an "internal" part. We will apply a partial form of this technique to the

 0022-2488/79/050921-02$01.00

Lagrangian density (1). Equation (1) may be rewritten as

$$L = e\bar{\psi}\gamma^{\nu}(x)\partial_{\nu}\psi - eS^{(\sigma\lambda\rho)}F_{(\sigma\lambda\rho)} + eP(\psi), \tag{15}$$

where $F_{(\sigma\lambda\rho)} = e_{(\sigma)}{}^{\nu}F_{\nu(\lambda\rho)}$ and the Dirac spin density $S^{(\sigma\lambda\rho)}$ is given by

$$S^{(\sigma\lambda\rho)} = \tfrac{1}{4}\bar{\psi}\gamma^{(\sigma)}\sigma^{(\lambda\rho)}\psi. \tag{16}$$

The equation of motion for the Dirac field is from Eq. (1),

$$\gamma^{\nu}(x)[\partial_{\nu} - \tfrac{1}{4}\sigma^{(\alpha\beta)}F_{\nu(\alpha\beta)}]\psi - \mu\psi + \lambda(\bar{\psi}\psi)\psi = 0. \tag{17}$$

Substituting Eq. (17) and its adjoint into Eq. (16), multiplying by $F_{(\sigma\lambda\rho)}$ and using $\gamma^{(\mu)}\gamma^{(\nu)} = \delta^{(\mu\nu)} + \sigma^{(\mu\nu)}$, we obtain

$$S^{(\sigma\lambda\rho)}F_{(\sigma\lambda\rho)} = S'^{(\sigma\lambda\rho)}F_{(\sigma\lambda\rho)} - (\tfrac{1}{32}\mu)\bar{\psi}\Gamma^{(\mu\alpha\beta\sigma\lambda\rho)}\psi \times F_{(\mu\alpha\beta)}F_{(\sigma\lambda\rho)} - (\tfrac{1}{32}\mu)\bar{\psi}\Gamma^{(\alpha\beta\mu\sigma\lambda\rho)}\psi F_{(\mu\alpha\beta)}F_{(\sigma\lambda\rho)}, \tag{18}$$

where

$$S'^{(\sigma\lambda\rho)} = (\tfrac{1}{8}\mu)\{\bar{\psi}\gamma^{(\sigma)}\sigma^{(\lambda\rho)}\gamma^{(\nu)}(x)\partial_{\nu}\psi - \partial_{\nu}[e\bar{\psi}\gamma^{\nu}(x)]\gamma^{(\sigma)}\sigma^{(\lambda\rho)}\psi + 2(\bar{\psi}\psi)\bar{\psi}\gamma^{(\sigma)}\sigma^{(\lambda\rho)}\psi\}, \tag{19}$$

and

$$\Gamma^{(\mu\alpha\beta\sigma\lambda\rho)} = \gamma^{(\mu)}\gamma^{(\alpha)}\gamma^{(\beta)}\gamma^{(\sigma)}\gamma^{(\lambda)}\gamma^{(\rho)}, \tag{20}$$

$$\Gamma^{(\alpha\beta\cdots\rho)} = \gamma^{(\alpha)}\gamma^{(\beta)}\cdots\gamma^{(\rho)}. \tag{21}$$

Applying the identity $\gamma^{(\mu)}\gamma^{(\nu)} = \delta^{(\mu\nu)} + \sigma^{(\mu\nu)}$ to the last two terms in Eq. (18), we obtain[8]

$$S^{(\sigma\lambda\rho)}F_{(\sigma\lambda\rho)} = S'^{(\sigma\lambda\rho)}F_{(\sigma\lambda\rho)} - (\tfrac{1}{32}\mu)\bar{\psi}\psi[F^{(\rho\lambda\sigma)}F_{(\sigma\lambda\rho)} - F^{(\mu)}{}_{(\mu)}{}^{(\rho)}F^{(\lambda)}{}_{(\lambda\rho)}] + \Sigma^{(\sigma\lambda\rho)}F_{(\sigma\lambda\rho)}, \tag{22}$$

where Σ contains all the terms $\bar{\psi}\Gamma\psi F$ with an explicit γ matrix dependence.

IV. SYMMETRY BREAKING AND THE LINEARIZED EINSTEIN EQUATIONS

The Lorentz invariance of (1) is spontaneously broken by imposing the condition

$$\langle 0|\psi|0\rangle = \sigma \neq 0, \tag{23}$$

where σ is a constant local spinor. Calculating

$$\left.\frac{\partial P(\psi)}{\partial\bar{\psi}}\right|_{\psi = \langle\psi\rangle} = 0, \tag{24}$$

we have from Eqs. (10), (23), and (24),

$$\bar{\sigma}\sigma = \mu/\lambda. \tag{25}$$

Using Eqs. (15) and (22) and expressing the Lagrangian in terms of fields $\psi' = \psi - \sigma$ possessing a zero vacuum expectation value, Eq. (1) becomes

$$L = L_{\psi'} + L_F + L_I, \tag{26}$$

where

$$L_{\psi'} = e\bar{\psi}'\gamma^{\nu}(x)\partial_{\nu}\psi', \tag{27}$$

$$L_F = (e/32\lambda)(F^{(\rho\lambda\sigma)}F_{(\sigma\lambda\rho)} - F^{(\mu)}{}_{(\mu)}{}^{(\lambda)}F^{(\sigma)}{}_{(\sigma\lambda)}) - \tfrac{1}{2}e\mu^2/\lambda, \tag{28}$$

with L_I denoting the remaining terms.[9] To interpret L_F we linearize the gauge field F with

$$e_{\mu(\alpha)} = \delta_{\mu\alpha} + h_{\mu\alpha}. \tag{29}$$

Keeping terms bilinear in the derivatives of $h_{\mu\alpha}$ and neglecting the cosmologicallike term $(\mu^2/\lambda)(1 + h)$, we have from Eq. (28),

$$L_h = (1/32\lambda)(\partial^{\lambda}h^{\sigma\rho}\partial_{\lambda}h_{\sigma\rho} + 2\partial_{\lambda}h\partial_{\mu}h^{\lambda\mu} - \partial^{\lambda}h\partial_{\lambda}h - \partial_{\mu}h^{\lambda\mu}\partial_{\sigma}h^{\sigma}_{\lambda} - \partial^{\lambda}h^{\rho\sigma}\partial_{\rho}h_{\lambda\sigma}), \tag{30}$$

where $h = h^{\sigma}_{\sigma}$. Taking the variation of Eq. (30) yields

$$\frac{\delta L_h}{\delta h_{\alpha\beta}} = (1/16\lambda)[\Box^2 h^{\alpha\beta} - (\partial_{\lambda}\partial^{\alpha}h^{\beta\lambda} + \partial_{\lambda}\partial^{\beta}h^{\alpha\lambda}) + \partial^{\beta}\partial^{\alpha}h + \delta^{\alpha\beta}(\partial_{\lambda}\partial^{\mu}h^{\lambda}_{\mu} - \Box^2 h)]. \tag{31}$$

Equation (31) is the usual linearized Einstein gravitational field equation.[10]

ACKNOWLEDGMENTS

The author wishes to thank Dr. G.A. Peterson of United Technologies Research Center for his continuing encouragement.

[1] S. Bazanski, *Recent Developments in General Relativity* (Pergamon, New York, 1962), p. 13.
[2] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1975), p. 351.
[3] R.L. Mallett, Phys. Rev. D. **13**, 2691 (1976).
[4] R. Utiyama, Phys. Rev. **101**, 1597 (1956); T.W. Kibble, J. Math. Phys. **2**, 212 (1961).
[5] For a discussion of symmetry breaking and the Higgs mechanism see, e.g., the review article by E.S. Abers and B.W. Lee, Phys. Rev. C **9**, 1 (1973).
[6] G.S. Guralnik and C.R. Hagen, Nuovo Cimento **45**, 959 (1966).
[7] R.L. Mallett, Phys. Rev. D **16**, 295 (1977).
[8] Symbolically, the contractions leading to Eq. (22) may be represented by $\bar{\psi}\Gamma^{(\mu\alpha\beta\sigma\lambda\rho)}\psi + \bar{\psi}\Gamma^{(\alpha\beta\mu\sigma\lambda\rho)}\psi$. Other contractions either lead to equivalent terms or reintroduce explicit γ matrix dependence.
[9] Equation (27) suggests the existence of Goldstone fermions. However, there may be difficulties in this interpretation see, e.g., A. Salam and J. Strathdee, Phys. Lett. B **49**, 465 (1974).
[10] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 435.

Zorn algebra in general relativity

Colber G. Oliveira

Departamento de Física, Universidade de Brasília, Brazil 70.000 Brasília-DF-Brazil

Marcos D. Maia

Departamento de Matemática, Universidade de Brasília, 70.000 Brasília-DF-Brazil
(Received 8 November 1977; revised manuscript received 18 June 1978)

The covariant differential properties of the split Cayley subalgebra of local real quaternion tetrads is considered. Referred to this local quaternion tetrad several geometrical objects are given in terms of Zorn–Weyl matrices. Associated with a pair of real null vectors we define two-component spinor fields over the curved space and the associated Zorn–Weyl matrices which satisfy the Dirac equation written in terms of the Zorn algebra. The formalism is generalized by considering a field of complex tetrads defining a Hermitian second rank tensor. The real part of this tensor describes the gravitational potentials and the imaginary part the electromagnetic potentials in the Lorentz gauge. The motion of a charged spin zero test body is considered. The Zorn–Weyl algebra associated with this generalized formalism has elements belonging to the full octonion algebra.

1. INTRODUCTION

In most applications of Cayley algebras to relativity theory a modification of the algebraic structure is required so as to make it compatible with the indefinite metric structure of the spacetime. Such modification can be obtained by using a suitable redefinition of the algebra over the complex field. In this case the algebra loses its division property and becomes a split algebra. A well-known example is given by the associative subalgebra of complex quaternions (or biquaternions, or split quaternions) or the Cayley algebra in special relativity, where a Weyl representation by 2×2 complex matrices is used.[1]

Recently the split quaternion subalgebra of the Zorn algebra has been applied to the Maxwell and Yang—Mills fields in special relativity.[2] Further applications to particle physics are also known in the literature.[3—5] The purpose of this paper is to investigate the application of the Zorn algebra to the study of the relativistic wave equations in curved space. It is found that by using the conventional tetrad formalism, which connects the tangent space to the pseudo-Riemannian spacetime, it is possible to construct an algebraic tetrad structure belonging to a split quaternion subalgebra of the octonion algebra, where the algebra of octonions is here reperesented by a modified version of the Zorn matrices.[6]

It follows that the role played by the second-rank Hermitian matrices of the two-component spinor formalism are taken over by four Zorn—Weyl matrices which are associated to each local Weyl representation in the curved space. Since the "internal" symmetry group is the local $SL_2(C)$ we have two different Weyl representations, which define two sets ${}^{(a)}\mathcal{H}_\mu$ $(a=1,2)$ of the above matrices. The analogy between the ${}^{(a)}\mathcal{H}_\mu$ and the second-rank Hermitian matrices again indicates that the present formalism is equivalent to a local quaternion subalgebra of the full Cayley algebra. This property follows from the fact that the ${}^{(a)}\mathcal{H}_\mu$ are really 4×4 matrices as compared with the 2×2 matrices σ_μ, τ_μ. We also show that this local quaternion tetrad written as Zorn—Weyl matrices acts as projection operators which associate to each geometrical object (tensor or spinor) a well defined Zorn—Weyl matrix. The flat spacetime limit is then easily obtained and coincides with known results.

Our present Zorn—Weyl representation of octonions (and of quaternions) may also be thought of as a 4×4 matrix (similarly as a γ-matrix). Such matrices, and their nonassociative law of product were already considered in the literature.[5] However, we mention that for our covariant treatment involving the "internal" group $SL_2(C)$ such analogy is not relevant. Indeed, it is not possible to associate to each ${}^{(a)}\mathcal{H}_\mu$ a γ-matrix since for each value of a we have only a type of Weyl basis, and as is known a γ-matrix contains the two-Weyl basis. Due to this we interpret the ${}^{(a)}\mathcal{H}_\mu$ as Zorn matrices referred to a Weyl basis, the nonassociative product being defined locally by introducing "scalar" and "wedge" products of the quaternion basis.

In the following sections we consider the definition of the differential operator in flat spaces and determine the Maxwell equations in the Zorn—Weyl formalism for the Lorentz gauge. Then we determine the Zorn—Weyl covariant derivative and apply the formalism for the relativistic spin-$\frac{1}{2}$ wave equation in curved space.

Finally the Zorn—Weyl formalism is extended for the case of complex tetrads which generate a Hermitian second-rank tensor field that plays the role of a generalized "metric." The symmetric (or real) part of this tensor describes gravitation according to general relativity and the antisymmetric (or imaginary) part describes the electromagnetic potentials in the Lorentz gauge. It is shown that the algebraic structure of the complex tetrad contains elements belonging to the split octonion algebra. The covariant differential properties are extended to this formalism, and as an application the problem of the motion of a charged spin-zero test body in this generalized geometry is considered.

The conventions and notations which will be used throughout this paper are the following: The four-dimensional space of general relativity is assumed to have metric signature +2. Greek indices running from 1 to 4 denote tensor degrees of freedom. Latin indices indicate spacelike degrees of freedom and run from 1 to

 0022-2488/79/050923-08$01.00

3. Bracketed indices are used to indicate tetrad indices. Capital dotted, or undotted, italic indices are reserved for two component spinor degrees of freedom and run from 1 to 2. Finally, boldface roman indices are used for algebraic elements, running from 1–7 for the capital indices and from 1–3 for the other indices. Summation convention is used throughout and applied to all kinds of indices.

2. THE SPLIT CAYLEY ALGEBRA IN THE ZORN REPRESENTATION

Let $\{e_{\bar{A}}\}$ be a basis in a seven-dimensional real vector space with an inner product. The real Cayley algebra, or octonion algebra, $\mathcal{O}$ is the linear algebra constructed in the above space, with the product operation defined by

$$e_{\bar{A}}e_{\bar{B}} = \epsilon_{\bar{A}\bar{B}\bar{C}}\, e_{\bar{C}} - \delta_{\bar{A}\bar{B}}\, e_{\bar{0}}, \tag{2.1}$$

where $\epsilon_{\bar{A}\bar{B}\bar{C}}$ is totally antisymmetric and satisfies $\epsilon_{\bar{A}\bar{B}\bar{C}} = 1$ when $\bar{A}, \bar{B}, \bar{C}$ assume the values $(\bar{1}, \bar{2}, \bar{3})$, $(\bar{5}, \bar{1}, \bar{6})$, $(\bar{6}, \bar{2}, \bar{4})$, $(\bar{4}, \bar{3}, \bar{5})$, $(\bar{6}, \bar{7}, \bar{3})$, $(\bar{4}, \bar{7}, \bar{1})$, and $(\bar{5}, \bar{7}, \bar{2})$. For all other cases $\epsilon_{\bar{A}\bar{B}\bar{C}}$ vanishes. The identity element of the algebra is $e_{\bar{0}}$.

It follows immediately from (2.1) that if the indices vary on each one of the seven above triads $(\bar{A}, \bar{B}, \bar{C})$ a quaternion subalgebra is obtained. Thus, the real Cayley algebra contains seven quaternion subalgebras. In the basis $\{e_{\bar{0}}, e_{\bar{A}}\}$ a general real Cayley number is expressed by $A = x_{\bar{0}}e_{\bar{0}} + x_{\bar{A}}e_{\bar{A}}$, $x_{\bar{0}}, x_{\bar{A}} \in R$. The multiplication table implies that the product operation is in general nonassociative. Furthermore it follows that a real Cayley algebra is a division algebra.

Now we consider the algebra of complex Cayley numbers $\mathcal{O}/\mathbb{C}$ which may be taken as the set of elements of the form

$$A = z_{\bar{0}}\, e_{\bar{0}} + z_{\bar{A}}e_{\bar{A}}, \quad z_{\bar{0}}, z_{\bar{A}} \in \mathbb{C}. \tag{2.2}$$

The complex conjugation applied to the components $z_{\bar{0}}, z_{\bar{A}}$ gives a new Cayley number, $A^* = z^*_{\bar{0}}e_{\bar{0}} + z^*_{\bar{A}}e_{\bar{A}}$, and this operation commutes with the Cayley number conjugation $(\bar{A})^* = (\overline{A^*})$. In particular consider the following complex Cayley numbers

$$u_{\bar{0}} = \tfrac{1}{2}(e_{\bar{0}} + ie_{\bar{7}}), \quad u_{\bar{i}} = \tfrac{1}{2}(e_{\bar{i}} + ie_{\bar{i}+3})$$

and their complex conjugates. From the multiplication table (2.1) it follows that the above set of complex Cayley numbers together with their complex conjugates form a basis for $\mathcal{O}/\mathbb{C}$. The product between these basis elements is given by equations

$$u_{\bar{i}}u_{\bar{j}} = \epsilon_{\bar{i}\bar{j}\bar{k}}u^*_{\bar{k}}, \quad u_{\bar{i}}u^*_{\bar{j}} = -\delta_{\bar{i}\bar{j}}u_{\bar{0}},$$

$$u_{\bar{i}}u_{\bar{0}} = 0, \quad u_{\bar{i}}u^*_{\bar{0}} = u_{\bar{i}}, \quad u_{\bar{0}}u_{\bar{i}} = u_{\bar{i}},$$

$$u^*_{\bar{0}}u_{\bar{i}} = 0, \quad u_{\bar{0}}u_{\bar{0}} = u_{\bar{0}}, \quad u_{\bar{0}}u^*_{\bar{0}} = 0,$$

and their complex conjugates.

A general complex Cayley number in this basis assumes the form

$$A = au^*_{\bar{0}} + bu_{\bar{0}} + x_{\bar{i}}u^*_{\bar{i}} + y_{\bar{i}}u_{\bar{i}}, \tag{2.3}$$

where the coefficients $a, b, x_{\bar{i}}$, and $y_{\bar{i}}$ are in general complex numbers. In particular they can be real, with A still a complex Cayley number.

In order to introduce a representation of the complex Cayley algebra we consider an application Z from $\mathcal{O}/\mathbb{C}$ assuming values on the set $M_{2\times 2/H}$ of 2×2 matrices defined over the quaternion field H. Such application is defined by

$$Z(u^*_{\bar{0}}) = \begin{pmatrix} e_{\bar{0}} & 0 \\ 0 & 0 \end{pmatrix}, \quad Z(u_{\bar{0}}) = \begin{pmatrix} 0 & 0 \\ 0 & e_{\bar{0}} \end{pmatrix}, \quad Z(u_{\bar{i}}) = \begin{pmatrix} 0 & 0 \\ e_{\bar{i}} & 0 \end{pmatrix},$$

$$Z(u^*_{\bar{i}}) = \begin{pmatrix} 0 & -e_{\bar{i}} \\ 0 & 0 \end{pmatrix}. \tag{2.4}$$

Defining the sum of two such matrices and multiplication by $\mathbb{C}$ in the usual way, it follows that the application Z is linear in $\mathcal{O}/\mathbb{C}$ and from (2.3) and (2.4) we have

$$Z(A) = \begin{pmatrix} ae_{\bar{0}} & -x_{\bar{i}}e_{\bar{i}} \\ y_{\bar{i}}e_{\bar{i}} & be_{\bar{0}} \end{pmatrix} = \begin{pmatrix} a & -x \\ y & b \end{pmatrix}, \tag{2.5}$$

where we have denoted $x_{\bar{i}}e_{\bar{i}}$ by x, the same for y.

The set of matrices of the form (2.5) may define a representation of $\mathcal{O}/\mathbb{C}$ in $M_{2\times 2/H}$ provided a product between such matrices is defined in such a way that the application Z is a homomorphism. In this case the matrices (2.5) are called Zorn matrices.[6]

In order to introduce the definition of the Zorn product for matrices of the form (2.5) we define the scalar and wedge product of quaternions as

$$e_{\bar{i}} * e_{\bar{j}} = -\tfrac{1}{2}(e_{\bar{i}}e_{\bar{j}} + e_{\bar{j}}e_{\bar{i}}) = \delta_{\bar{i}\bar{j}}e_{\bar{0}}, \tag{2.6}$$

$$e_{\bar{i}} \wedge e_{\bar{j}} = \tfrac{1}{2}(e_{\bar{i}}e_{\bar{j}} - e_{\bar{j}}e_{\bar{i}}) = \epsilon_{\bar{i}\bar{j}\bar{k}}e_{\bar{k}}. \tag{2.7}$$

The Zorn product between Zorn matrices is now defined in such a way that it reproduces the multiplication table of the complex Cayley basis

$$Z(AB) = Z(A)\odot Z(B) = \begin{pmatrix} ac - x * w & -az - dx - y \wedge w \\ cy + bw + x \wedge z & bd - y * z \end{pmatrix} \tag{2.8}$$

for

$$Z(A) = \begin{pmatrix} a & -x \\ y & b \end{pmatrix}, \quad Z(B) = \begin{pmatrix} c & -z \\ w & d \end{pmatrix}.$$

The unit element of the resulting Zorn matrix algebra is

$$1 = \begin{pmatrix} e_{\bar{0}} & 0 \\ 0 & e_{\bar{0}} \end{pmatrix} = Z(u^*_{\bar{0}}) + Z(u_{\bar{0}}). \tag{2.9}$$

As can be seen, a complex octonion like (2.3) reduces, in general, to a complex quaternion when $a = b, x = y$. This quaternion belongs to the quaternion subalgebra (1, 2, 3) of the octonion algebra. The Zorn matrix associated to this quaternion is

$$Z(A) = \begin{pmatrix} a & -x \\ x & a \end{pmatrix}. \tag{2.10}$$

The conjugation operation induces on the complex basis the transformation $u_{\bar{0}}^{*}$, $u_{\bar{i}} \to u_{\bar{i}}$. Thus, a general complex Cayley number B such as (2.3) which in the Zorn representation reads as (2.5), transforms under conjugation to

$$Z(\bar{B}) = \begin{pmatrix} b & x \\ -y & a \end{pmatrix}.$$

The norm of this octonion is given by

$$Z(B\bar{B}) = Z(B)\,\bar{\circ}\,Z(\bar{B}) = (ab + x_{\bar{i}}y_{\bar{i}})\mathbb{I}.$$

Here we are interested in the situation where the octinion B reduces to a quaternion of the form (2.10). In this case the norm of this quaternion in the Zorn representation will be $(a^2 + x_{\bar{i}}x_{\bar{i}})\mathbb{I}$. If we take a as an imaginary number, $a = ix_{\bar{o}}$, and $x_{\bar{o}}, x_{\bar{i}}$ as reals, it follows that

$$Z(A) \cdot Z(\bar{A}) = Q(A) \cdot 1 = -x_{\bar{o}}^2 + x_{\bar{i}}x_{\bar{i}}. \tag{2.11}$$

Therefore, the Zorn matrix (2.10) with $a = ix_{\bar{0}}$ may be thought as representing a 4-vector in Minkowski spacetime (in this case the algebraic indices become world indices). Likewise given an general octonion like (2.5) with $a \neq b$, $x \neq y$ but with a, b imaginary numbers, $a = ix_{\bar{o}}$, $b = iy_{\bar{o}}$, and all $x_{\bar{o}}, x_{\bar{i}}, y_{\bar{o}}, y_{\bar{i}}$ reals, we can associate to this octonion a pair of 4-vectors in Minkowski spacetime. From now on we will consider only this particular type of octonions and quaternions and their Zorn matrices. The corresponding quaternion subalgebra of the complex Cayley algebra given by the set of matrices of the form (2.10), with the law of product given by (2.8), and with norm given by (2.11) is a split algebra. This corresponds to the property that the Minkowski spacetime contains isotropic vectors. Similarly we have a split octonion algebra which corresponds to the property that the Minkowski spacetime contains pairs of orthogonal 4-vectors.

3. EXTENSION OF THE METHOD TO CURVED SPACETIMES IN THE TETRAD REPRESENTATION

In this section we apply the previous algebraic methods to a curved four-dimensional spacetime. Such a type of formalism is an extension of previous works which apply these algebraic methods to special relativity.[2] The formalism which will be developed in this section corresponds to the use of only a part of the Cayley algebra, namely the quaternion subalgebra of the complex Cayley algebra. As in the previous section we will use this algebra with reference to the Zorn matrices defined in a quaternion basis. In applications to relativity it is of interest to use the quaternion basis in terms of the three Pauli matrices and the 2×2 identity matrix, that is, in terms of a Weyl representation of the quaternion algebra.[7] From the algebraic point of view this Weyl representation is obtained by the application W: $H/\mathbb{C} \to M_{2\times2/C}$ which may be defined either by

$$W(e_{\bar{o}}) = {}^{0}\sigma_{\bar{0}}, \quad W(e_{\bar{i}}) = {}^{0}\sigma_{\bar{i}}, \tag{3.1}$$

or by

$$W(e_{\bar{o}}) = {}^{0}\sigma_{\bar{0}}, \quad W(e_{\bar{i}}) = \left(\frac{1}{i}\right){}^{0}\sigma_{\bar{i}}. \tag{3.2}$$

The symbols ${}^{0}\sigma_{\bar{i}}$, ${}^{0}\sigma_{\bar{0}}$ denote the Pauli matrices and the 2×2 identity matrix, respectively. In the first case they satisfy the usual law of multiplication of the Pauli matrices together with the 2×2 identity matrix, and in the second case they satisfy the same law of product as the $e_{\bar{o}}, e_{\bar{i}}$, namely,

$$\begin{aligned} &W(e_{\bar{i}})\ W(e_{\bar{j}}) = -\delta_{\bar{i}\bar{j}}\ W(e_{\bar{o}}) + \epsilon_{\bar{i}\bar{j}\bar{k}}\ W(e_{\bar{k}}), \\ &W(e_{\bar{i}})\ W(e_{\bar{o}}) = W(e_{\bar{o}})\ W(e_{\bar{i}}) = W(e_{\bar{i}}). \end{aligned} \tag{3.3}$$

We will use the second alternative. It should be observed that the Zorn representation of the quaternion algebra treated on the last section is distinct from the usual Weyl representation of quaternions. However, since the Zorn matrices associated to quaternions are defined over the quaternion field, a combined Zorn—Weyl representation of quaternions can be obtained by considering the Weyl representation of elements of the Zorn matrices. Denoting the resulting composition by ZW, we have for a Zorn matrix like (2.5) for $a = ix_{\bar{o}}$, which gives the Zorn representation of the split quaternion subalgebra of the complex Cayley algebra.

$$ZW(A) = \begin{pmatrix} W(a) & -W(x) \\ W(x) & W(a) \end{pmatrix} = \begin{pmatrix} ix_{\bar{0}}W(e_{\bar{0}}) & -x_{\bar{i}}W(e_{\bar{i}}) \\ x_{\bar{i}}W(e_{\bar{0}}) & ix_{\bar{0}}W(e_{\bar{0}}) \end{pmatrix},$$

which, from (3.2) takes the form

$$ZW(A) = \begin{pmatrix} ix_{\bar{0}}{}^{0}\sigma_{\bar{0}} & -\frac{1}{i}x_{i}{}^{0}\sigma_{\bar{i}} \\ \frac{1}{i}x_{\bar{i}}{}^{0}\sigma_{\bar{i}} & ix_{\bar{0}}{}^{0}\sigma_{\bar{0}} \end{pmatrix}. \tag{3.4}$$

Presently we have to adapt this notation to our problem of a curved spacetime with a Riemannian structure. For this purpose we consider only the local properties of this spacetime translated in terms of the Zorn—Weyl algebra. With this in mind we consider the local tangent space at each point of the Riemannian spacetime, and the set of four local tetrad vectors $h_{\alpha} = (h_{(\lambda)\alpha})$. All algebraic quantities, with indices $\bar{o}, \bar{i}, \bar{j}$, etc. now become quantities defined on the local tangent space with indices (0), (*i*), (*j*), etc. The metric $g_{\mu\nu}$ is related to the Minkowski metric $\eta_{\alpha\beta}$ by means of the local tetrad field

$$g_{\mu\nu} = h_{\mu}^{(\alpha)}h_{\nu}^{(\beta)}\eta_{\alpha\beta}.$$

The matrices ${}^{0}\sigma_{(0)}$, ${}^{0}\sigma_{(i)}$, σ_{i}, σ_{j} have degrees of freedom given by the indices K and M which run from 1 to 2, both indices taken as contravariant indices. The matrices denoted by τ have both indices as covariant indices, also of the type KM. The matrix denoted by I is the 2×2 identity matrix of the form (δ_{S}^{K}). In the local tangent space which here is taken with signature +2 we can define three types of identity matrices: ${}^{0}\sigma^{(0)}$, ${}^{0}\tau_{(0)}$, and I. The matrix ${}^{0}\tau_{(0)}$ is defined by ${}^{0}\tau_{(0)} = \epsilon\,{}^{0}\sigma^{*}_{(0)}\epsilon = I$ and is numerically identical to the identity 2×2 matrix, but has covariant indices of the type KM. Here ϵ is the matrix

$$\epsilon = (\epsilon^{KM}) = (\epsilon_{KM}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Since local indices are raised and lowered

respectively by $\eta^{\alpha\beta}$ and $\eta_{\alpha\beta}$, we have ${}^{o}\sigma_{(o)} = -I$. The covariant matrix product is defined only between matrices of the type σ and τ. This implies that the algebraic formulas defining the quaternion structure like (3.3), and the formulas defining the ZW representation of quaternions have to be translated in this covariant notation. With this in mind we define the following Weyl representations on the local tangent space, which are associated to $SL_2(C)$:

$$W_1(e_{(i)}) = \frac{1}{i}\,{}^{o}\sigma_{(i)}, \quad W_1(e_{(o)}) = {}^{o}\sigma^{(o)},$$
$$W_2(e_{(i)}) = \frac{1}{i}\,{}^{o}\tau_{(i)}, \quad W_2(e_{(o)}) = {}^{o}\tau_{(o)}. \tag{3.5}$$

In the limit where we consider only the action of $SU_2(C)$ on the spinor degrees of freedom they degenerate in the representation given by (3.2) for the spacelike degrees of freedom, $W_1(e_{(i)}) \to W_2(e_{(i)})$. Besides this we also define

$$W_3(e_{(o)}) = I = (\delta^K_S).$$

It can be shown that the covariant law of product for the Weyl representations (3.5) has the form

$$W_1(e_{(i)})W_2(e_{(j)}) = -\delta_{(i)(j)}W_3(e_{(o)}) + \epsilon_{(i)(j)(k)}W_1(e_{(k)})W_2(e_{(o)}), \tag{3.6}$$

$$W_1(e_{(o)})W_2(e_{(o)}) = W_3(e_{(o)}). \tag{3.7}$$

These formulas presently substitute the formulas (3.3).

Given the Zorn matrix associated to a quaternion we can write it in the Weyl representation of the type W_1 as

$$ZW_1(A) = \begin{pmatrix} ie_{(o)}W_1(e_{(o)}) & -a_{(i)}W_1(e_{(i)}) \\ a_{(i)}W_1(e_{(i)}) & ia_{(o)}W_1(e_{(o)}) \end{pmatrix}. \tag{3.8}$$

For the same quaternion, or in general for any other quaternion, we can write the ZW matrix of the type W_2 by replacing the subscripts 1 by 2.

On the remainder of this section we will use these matrices in place of the matrices (3.4).

The product of these matrices is defined similarly as before [see Eq. (2.10)]:

$$ZW_1(A) \odot ZW_2(B) = ZW_3(C) = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{3.9}$$

where

$$\alpha = \delta = -a_{(o)}b_{(o)}W_3(e_{(o)}) - a_{(i)}b_{(j)}W_1(e_{(i)}) * W_2(e_{(j)})$$
$$\beta = -\gamma = -(ia_{(o)}b_{(i)} + ia_{(i)}b_{(o)}W_1(e_{(i)})W_2(e_{(o)}) - a_{(i)}b_{(j)}W_1(e_{(i)}) \wedge W_2(e_{(j)}).$$

From (3.6) we define by analogy with (2.6) and (2.7),

$$W_1(e_{(i)}) * W_2(e_{(j)}) = -\tfrac{1}{2}S_{(i,j)}(W_1(e_{(i)})W_2(e_{(j)})) = \delta_{(i)(j)}W_3(e_{(o)}). \tag{3.10}$$

$$W_1(e_{(i)}) \wedge W_2(e_{(j)}) = \tfrac{1}{2}E_{(i,j)}(W_1(e_{(i)})W_2(e_{(j)})) = \tfrac{1}{2}\epsilon_{(i)(j)(k)}[W_1(e_{(k)})W_2(e_{(o)}) + W_1(e_{(o)})W_2(e_{(k)})]$$
$$= \epsilon_{(i)(j)(k)}W_1(e_{(o)})W_2(e_{(k)}) = \epsilon_{(i)(j)(k)}W_1(e_{(k)})W_2(e_{(o)}), \tag{3.11}$$

where

$$S_{(i,j)}T_{(i)(j)} \text{ and } E_{(i,j)}T_{(i)(j)},$$

for any $T_{(i)(j)}$ mean

$$S_{(i,j)}T_{(i)(j)} = T_{(i)(j)} + T_{(j)(i)}, \quad E_{(i,j)}T_{(i)(j)} = T_{(i)(j)} - T_{(j)(i)}.$$

Of fundamental importance are the ZW matrices associated to the tetrad field, which are defined by

$${}^{(1)}\!/\!/_{\mu} = ZW_1(H_\mu) = \begin{pmatrix} ih_{\mu(0)}W_1(e_{(o)}) & -h_{\mu(i)}W_1(e_{(i)}) \\ h_{\mu(i)}W_1(e_{(i)}) & ih_{\mu(0)}W_1(e_{(o)}) \end{pmatrix}, \tag{3.12}$$

$${}^{(2)}\!/\!/_{\mu} = ZW_2(H_\mu) = \begin{pmatrix} ih_{\mu(0)}W_2(e_{(o)}) & -h_{\mu(i)}W_2(e_{(i)}) \\ h_{\mu(i)}W_2(e_{(i)}) & ih_{\mu(0)}W_2(e_{(o)}) \end{pmatrix}, \tag{3.13}$$

where $H_\mu = h_{\mu(o)}e_{(o)} + h_{\mu(i)}e_{(i)}$.

From these definitions it follows that the metric of the Riemannian spacetime is given in terms of the ZW matrices (3.12) and (3.13) as

$$2g_{\mu\nu}ZW_3(e_{(0)}) = ZW_1(H_\mu) \odot ZW_2(\overline{H}_\nu) + ZW_1(H_\nu) \odot ZW_2(\overline{H}_\mu), \tag{3.14}$$

where

$$ZW_3(e_{(0)}) = \begin{pmatrix} W_3(e_{(0)}) & 0 \\ 0 & W_3(e_{(0)}) \end{pmatrix} = \mathbb{I}.$$

The world indices labeling the several elements of the algebraic quantities given by (3.12) and (3.13) are raised by the metric field $g^{\mu\nu}$. The process of raising (lowering) world indices is presently equivalent to a sum of terms representing the multiplication of Zorn–Weyl scalars, the metric components, by Zorn–Weyl matrices which display free world indices,

$${}^{(1)}\!/\!/^{\mu} = g^{\mu\nu}\,{}^{(1)}\!/\!/_{\nu}, \quad {}^{(2)}\!/\!/^{\mu} = g^{\mu\nu}\,{}^{(2)}\!/\!/_{\nu}.$$

This process is extended to any other ZW matrix possessing free world indices.

4. THE ZORN-WEYL DIFFERENTIAL OPERATOR IN FLAT SPACETIME

We define a flat spacetime Zorn–Weyl differential operator in the quaternion representation as

$$D_a = \begin{pmatrix} iW_a(e_0)\partial_0 & -W_a(e_j)\partial_j \\ W_a(e_j)\partial_j & -iW_a(e_0)\partial_0 \end{pmatrix} = ZW_a(\partial_\mu), \tag{4.1}$$

where a takes the values 1 or 2. Here ∂_o, ∂_j denote the usual partial derivatives. The operator (4.1) acts on a Zorn–Weyl matrix as $D_a(A) = D_a \odot A$. It follows that

$$D_1 \odot \overline{D}_2 = \overline{D}_1 \odot D_2 = D_2 \odot D_1 = \overline{D}_2 \odot D_1 = \Box \cdot \mathbb{I}. \tag{4.2}$$

where

$$\Box = \eta^{\alpha\beta}\partial_\alpha\partial_\beta, \quad \eta^{\alpha\beta} = \mathrm{diag}(-1, +1, +1, +1).$$

If A_μ is the electromagnetic potential its associated Zorn–Weyl matrix is given by $A_a = ZW_a(A_\mu)$. Using

(4.1) it is possible to show that the Maxwell equations in the Lorentz gauge assume the form[2]

$$D_2 \circ (\bar{D}_1 \circ \mathcal{A}_2) = -\bar{J}_2, \tag{4.3}$$

where $J_a = ZW_a(j_\mu)$ are the two Zorn–Weyl representations of the current 4-vector.

5. ZORN-WEYL MATRICES ASSOCIATED TO GEOMETRICAL OBJECTS

The Zorn–Weyl matrices associated to the four vectors of the tetrad $H_\mu = (h_{\mu(0)}, h_{\mu(1)}, h_{\mu(2)}, h_{\mu(3)})$ are given by ${}^{(a)}\mathcal{H}_\mu = ZW_a(H_\mu)$. Then the Zorn–Weyl matrices associated to a vector B_μ are defined by

$${}^{(a)}B = {}^{(a)}\mathcal{H}^\mu B_\mu = {}^{(a)}\mathcal{H}^\mu \circ \mathbb{1} B_\mu. \tag{5.1}$$

Algebraic Zorn–Weyl objects may also be associated to tensors, spinors, and mixed geometrical objects. To $B_{\sigma\mu}$, a tensor of rank two, we can associate the Zorn–Weyl matrices

$${}^{(a)}C_\lambda = {}^{(a)}\mathcal{H}^\mu B_{(\lambda)\mu}, \quad a = 1, 2, \tag{5.2}$$

where $B_{(\lambda)\mu} = h^\sigma_{(\lambda)} B_{\sigma\mu}$.

It is also possible to associate with $B_{\sigma\mu}$ a further Zorn–Weyl matrix given by

$$C = ({}^{(1)}\mathcal{H}^\mu \circ {}^{(2)}\mathcal{H}^\nu) B_{\mu\nu}. \tag{5.3}$$

If $B_{\mu\nu}$ is symmetric the expression (5.3) becomes $C = g^{\mu\nu} B_{\mu\nu} \cdot \mathbb{1}$. If $B_{\mu\nu}$ is antisymmetric C contains only nondiagonal "matrix elements."

Now we consider the problem of associating ZW matrices to spinor fields in curved spaces. This correspondence is obtained by recalling that two-component spinors are related to tensors through well-known formulas. Here we are mainly interested in two-component spinor fields of the type χ_A, $\omega^{\dot{A}}$, since we want to obtain the Dirac equation for a massive spin-$\frac{1}{2}$ particle in terms of the Zorn algebra. With this in mind we consider a pair of real null vectors $V_\mu(x)$ and $W_\mu(x)$. Then

$$V_\mu(x) = \tfrac{1}{2}\sigma_\mu^{A\dot{B}}(x)\chi_A(x)\chi_{\dot{B}}^*(x), \tag{5.4}$$

$$W_\mu(x) = \tfrac{1}{2}\sigma_\mu^{A\dot{B}}(x)\omega_A(x)\omega_{\dot{B}}^*(x) = \tfrac{1}{2}\sigma_{\mu A\dot{B}}(x)\omega^A(x)\omega^{\dot{B}}(x), \tag{5.5}$$

where $\sigma^{A\dot{B}}_{\mu(x)} = h_{\mu(\alpha)}(x)\sigma^{(\alpha)A\dot{B}}$. In matrix notation we have

$$\chi = (\chi_A) = \begin{pmatrix}\chi_1\\ \chi_2\end{pmatrix}, \quad \chi^\dagger = (\chi_{\dot{A}}^*) = (\chi_1^*\chi_2^*), \quad \sigma_\mu = (\sigma_\mu^{A\dot{B}}) = \sigma_\mu^\dagger.$$

Similarly we denote

$$\tau_\mu = \epsilon \cdot \sigma_\mu^* \cdot \epsilon = -(\sigma_\mu^{*\ \dot{A}B}) = \tau_\mu^\dagger, \quad \Omega = (\omega^A) = \begin{pmatrix}\omega^1\\ \omega^2\end{pmatrix},$$

$$\Omega^\dagger = (\omega^{\dot{A}}) = (\omega^{\dot{1}}, \omega^{\dot{2}}).$$

Raising (lowering) of spinor indices is obtained by the use of the spinors ϵ_{AB}, ϵ^{AB}, as usual, $\psi^A = \epsilon^{AK}\psi_K$, $\psi_A = \psi^K \epsilon_{KA}$. Now Eq. (5.5) can be written as $W_\mu = \frac{1}{2}\Omega^\dagger \rho_\mu \Omega$, where $\rho_\mu = (\sigma_{\mu A\dot{B}}) = \rho_\mu^\dagger$. Since W_μ is a real vector field,

$$W_\mu^* = W_\mu = \tfrac{1}{2}\Omega^T \rho_\mu^* \Omega^* = -\tfrac{1}{2}\Omega^T \tau_\mu \Omega^*.$$

Similarly Eq. (5.4) gives

$$V_\mu = \tfrac{1}{2}\chi^T \sigma_\mu \chi.$$

Therefore, we get the Zorn–Weyl matrices associated to the spinor χ and Ω:

$${}^{(a)}B = {}^{(a)}\mathcal{H}^\mu V_\mu = \tfrac{1}{2}{}^{(a)}\mathcal{H}^\mu \circ \mathbb{1}\chi^\dagger \sigma_\mu \chi. \tag{5.6}$$

$${}^{(a)}C = {}^{(a)}\mathcal{H}^\mu W_\mu = -\tfrac{1}{2}{}^{(a)}\mathcal{H}^\mu \circ \mathbb{1}\Omega^T \tau_\mu \Omega^*. \tag{5.7}$$

Zorn–Weyl matrices associated with higher rank spinor fields may also be constructed. For example, if $\phi_{A\dot{B}C\dot{D}}$ is a spinor associated to a second rank tensor field,

$$B_{\mu\nu}(x) = \tfrac{1}{4}\sigma^{A\dot{B}}_\mu(x)\sigma^{C\dot{D}}_\nu(x)\phi_{A\dot{B}C\dot{D}}(x).$$

Then Zorn–Weyl matrices of the type (5.2) or (5.3) may be obtained. Finally we may also construct Zorn–Weyl matrices associated to mixed spin-tensor objects. For example, for the spin-tensor field $\psi^A_\alpha(x)$, the associated Zorn–Weyl matrices are

$${}^{(a)}\mathcal{J}^A = {}^{(a)}\mathcal{H}^\mu \psi^A_\mu = {}^{(a)}\mathcal{H}^\mu \circ \mathbb{1} \cdot \psi^A_\mu. \tag{5.8}$$

6. THE COVARIANT ZORN-WEYL DIFFERENTIAL OPERATOR AND FIELD EQUATIONS

Now we consider the problem of forming higher order tensors, spinors, or mixed objects by taking covariant derivatives in the Zorn–Weyl formulation of these objects. For that purpose we introduce an affine connection and define the differential operator

$$\mathcal{D}_\mu = \mathbb{1} \cdot \partial_\mu + \Gamma_\mu, \tag{6.1}$$

and, using (5.1), its associated Zorn–Weyl covariant differential operators are constructed as

$${}^{(a)}\mathbb{D} = g_{\mu\nu}\,{}^{(a)}\mathcal{H}^\mu \mathcal{D}^\nu = {}^{(a)}\mathcal{H}_\mu \mathcal{D}^\mu. \tag{6.2}$$

The definition of Γ_μ will depend on the space where ${}^{(a)}\mathbb{D}$ operates. We can write $\Gamma_\mu = (\Gamma_\mu{}^{\bar{\alpha}}{}_{\bar{\beta}})$ where the indices $\bar{\alpha}$, $\bar{\beta}$ are to be taken as world indices, or spinor indices. Thus if ${}^{(a)}\mathbb{D}$ operates on a Zorn–Weyl matrix associated to a world vector, then $\Gamma_\mu{}^{\bar{\alpha}}{}_{\bar{\beta}}$ is given by the Christoffel symbols $\{\mu^\alpha{}_\beta\}$. On the other hand, if ${}^{(a)}\mathbb{D}$ operates on a Zorn–Weyl matrix associated with a spinor field the $\Gamma_\mu{}^{\bar{\alpha}}{}_{\bar{\beta}}$ are the components of the spinor connection. We may also consider ${}^{(a)}\mathbb{D}$ acting on Zorn–Weyl matrices which are associated with mixed objects displaying vector and spinor indices, in this case $\Gamma_\mu{}^{\bar{\alpha}}{}_{\bar{\beta}}$ is a more complicated object where the indices $\bar{\alpha}$, $\bar{\beta}$ take on the values of spinor and tensor indices. In this case $(\Gamma_\mu{}^{\bar{\alpha}}{}_{\bar{\beta}})$ are represented by a sum of terms involving the Christoffel symbols and the spinor affinities. In Eq. (6.1) 1 denotes the identity element with the same index structure as the term in Γ_μ. For example, considering the mixed object ψ^A_μ of (5.8) we have

$$\mathcal{D}_\mu \psi^A_\nu = \left(\delta^A_D \delta^\rho_\nu \partial_\mu - \begin{Bmatrix}\rho\\ \mu\nu\end{Bmatrix}\delta^A_D + \Gamma_\mu{}^A{}_D \delta^\rho_\nu\right)\psi^D_\rho.$$

Therefore, in this case 1 is represented by $\delta^A_D \delta^\rho_\nu$ and

$$(\Gamma^{\bar{\alpha}}_\mu{}_{\bar{\beta}}) \to \left(-\begin{Bmatrix}\rho\\ \mu\nu\end{Bmatrix}\delta^A_D + \Gamma_\mu{}^A{}_D \delta^\rho_\nu\right).$$

Here $\Gamma_\mu{}^A{}_D$ represents the spinor affinity associated to local unimodular transformations of the spinor indices. In the flat space limit, in Cartesian coordinates, $h^\mu_{(\beta)} \to \delta^\mu_\beta$, $h_{\mu(\beta)} \to \eta_{\mu\beta}$, and $\Gamma_\mu \to 0$, so that ${}^{(a)}\mathbb{D} \to D_a$. Now using the covariant operator (6.2) we may construct covariant wave equations involving tensors spinors or mixed objects. As a first example consider the expression

$${}^{(1)}\mathbb{D} \odot {}^{(2)}B = {}^{(1)}\mathcal{H}^\mu \mathcal{D}_\mu \odot {}^{(2)}\mathcal{H}^\nu B_\nu = {}^{(1)}\mathcal{H}^\mu \odot {}^{(2)}\mathcal{H}^\nu \mathcal{D}_\mu B_\nu. \tag{6.3}$$

where $B_\nu(x)$ is an arbitrary vector field. Using the properties

$$\partial_\mu h^\alpha_{(\nu)} = 0, \quad \partial_\mu \mathbb{1} = 0 \tag{6.4}$$

and assuming that $B_\mu = A_\nu$ (the electromagnetic potential) in (6.3), then ${}^{(1)}\overline{D} \odot {}^{(2)}B$ gives the Zorn–Weyl matrix associated to the electromagnetic field $F_{\mu\nu}$ in curved spacetime. A straightforward calculation shows that in the flat limit this expression reduces to the expression $\overline{D}_1 \odot A_2$ of (4.3). Now consider the Zorn covariant derivative of a real null vector associated to the spinor χ [see Eq. (5.6)]. Denoting

$$\Lambda = {}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{B} \tag{6.5}$$

and using unities such that $c = \hbar = 1$ and usual spinor connection condition $\partial_\mu \sigma_\lambda = 0$, we have that the diagonal matrix elements in (6.5) are given by $\frac{1}{2}(\chi^\dagger_{;\alpha}\sigma^\alpha\chi + \chi^\dagger\sigma^\alpha\chi_{;\alpha})\mathbb{1}$.

Defining for any Zorn–Weyl matrix

$$S^{(a)}(N) = {}^{(a)}N + {}^{(a)}\overline{N} = \mathrm{tr}^{(a)}N \begin{pmatrix} W_a(e_{(0)}) & 0 \\ 0 & W_a(e_{(0)}) \end{pmatrix},$$

we have for (6.5)

$$S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{B}) = (\chi^\dagger_{;\alpha}\sigma^\alpha\chi + \chi^\dagger\sigma^\alpha\chi_{;\alpha})\mathbb{1}. \tag{6.6}$$

Similarly for the matrix (5.7) we have

$$S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{C}) = -(\Omega^T_{;\alpha}\tau^\alpha\Omega^* + \Omega^T\tau^\alpha\Omega^*_{;\alpha})\mathbb{1}. \tag{6.7}$$

Introducing

$$\mathcal{L} = (s_1\Omega^T \cdot \chi + s_2\chi^\dagger \cdot \Omega^*)\mathbb{1}, \quad \mathcal{K} = (s_3\chi^\dagger\Omega^* + s_4\Omega^T \cdot \chi)\mathbb{1}, \tag{6.8}$$

where s_1, s_2, s_3, and s_4 are constant numbers to be determined, we find

$$\begin{aligned} S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{B}) + \mathcal{L} &= \{[\chi^\dagger_{;\alpha}\sigma^\alpha + s_1\Omega^T]\chi \\ &\quad + \chi^\dagger[\sigma^\alpha\chi_{;\alpha} + s_2\Omega^*]\}\mathbb{1}, \end{aligned} \tag{6.9}$$

$$\begin{aligned} S({}^{(2)}\mathbb{D} \odot \overline{{}^{(1)}C}) + \mathcal{K} &= \{\Omega^T[-\tau^\alpha\Omega^*_{;\alpha} + s_4\chi] \\ &\quad + [-\Omega^T_{;\alpha}\tau^\alpha + s_3\chi^\dagger]\Omega^*\}\mathbb{1}. \end{aligned} \tag{6.10}$$

For the choice $s_1 = s_2 = -m$, $s_3 = s_4 = m$, the terms between brackets in the right-hand side of (6.9) and (6.10) give the left-hand side of the Dirac equation written in terms of two-component spinors.[8] Here m is the rest mass of the spin-$\frac{1}{2}$ particle. According to our method we may present the Dirac equation in the Zorn algebra on a curved space as

$$\begin{aligned} &S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{B}}) - 2m\,\mathrm{Re}(\Omega^T \cdot \chi)\mathrm{ZW}_3(e_{(0)}) = 0, \\ &S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathrm{C}}) + 2m\,\mathrm{Re}(\Omega^T \cdot \chi)\mathrm{ZW}_3(e_{(0)}) = 0. \end{aligned} \tag{6.11}$$

Note that from the right-hand side of (6.9) and (6.10) the Dirac equation is written as $\gamma^\alpha\psi_{;\alpha} - im\psi = 0$, for

$$\gamma^\alpha = \begin{pmatrix} 0 & -{}^*i\sigma^\alpha_{\dot{B}A} \\ i\sigma^{\alpha\dot{B}A} & 0 \end{pmatrix} = \begin{pmatrix} 0 & i\tau^\alpha \\ i\sigma^\alpha & 0 \end{pmatrix}, \quad \psi = \begin{pmatrix} \chi_A \\ \omega^{\dot{A}} \end{pmatrix} = \begin{pmatrix} \chi \\ \Omega^* \end{pmatrix}. \tag{6.12}$$

From our previous definitions it follows that

$$\gamma_{(\alpha}\gamma_{\beta)} = -2g_{\alpha\beta} \cdot 1.$$

It is also possible to derive a direct analog of the left-hand side of the Dirac equation without the problem presented by (6.11) which is quadratic in the spinor χ, Ω. For obtaining such direct analogy we recall the definition (5.8) and rewrite the relations (5.6) and (5.7) as ${}^{(a)}B = \chi_{\dot{A}}\,{}^{(a)}\mathcal{M}^{\dot{A}}$, ${}^{(a)}C = \omega^A\,{}^{(a)}\mathcal{N}_A$, where

$${}^{(a)}\mathcal{M}^{\dot{A}} = \tfrac{1}{2}\,{}^{(a)}\mathcal{H}^\mu\sigma_\mu^{\dot{A}B}\chi_B = \tfrac{1}{2}\,{}^{(a)}\mathcal{H}^\mu\xi_\mu^{\dot{A}}, \tag{6.13}$$

$${}^{(a)}\mathcal{N}_A = \tfrac{1}{2}\,{}^{(a)}\mathcal{H}^\mu\sigma^*_{\mu\dot{B}A}\omega^{\dot{B}} = \tfrac{1}{2}\,{}^{(a)}\mathcal{H}^\mu\eta_{\mu A}. \tag{6.14}$$

These expressions are linear in the spinors χ and Ω, and a straightforward calculation gives

$$S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{M}}^{\dot{A}}) = \sigma^{\nu\dot{A}B}\chi_{B;\nu}\mathbb{1}. \tag{6.15}$$

Similarly

$$S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{N}}_A) = {}^*\sigma^\mu_{B\dot{A}}\omega^{\dot{B}}_{;\mu}\mathbb{1}. \tag{6.16}$$

Define $\mathcal{P}^{\dot{A}} = -m\omega^{\dot{A}}\mathbb{1}$, $\mathcal{J}_A = m\chi_A\mathbb{1}$. Then, from (6.15) and (6.16) we have

$$S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{M}}^{\dot{A}}) + \mathcal{P}^{\dot{A}}) + \mathcal{P}^{\dot{A}} = |\sigma^{\nu\dot{A}B}\chi_{B;\nu} - m\omega^{\dot{A}}|, \tag{6.17}$$

$$S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{N}}_A) + \mathcal{J}_A = [{}^*\sigma^\mu_{B\dot{A}}\omega^{\dot{B}}_{;\mu} + m\chi_A]. \tag{6.18}$$

Therefore, the Dirac equation in the Zorn algebra may also be directly obtained from (6.17) and (6.18) as

$$S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{M}}^{\dot{A}}) + \mathcal{P}^{\dot{A}} = 0, \quad S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{N}}_A) + \mathcal{J}_A = 0. \tag{6.19}$$

Now we derive the Zorn–Weyl version of the Klein–Gordon equation. Defining the spinor operators (or Cartan matrices associated with the covariant derivative) $\sigma^*_{\dot{C}A}\partial_\mu$ and $\sigma^{\mu\,\dot{C}B}\partial_\mu$, we can form the ZW elements

$$\hat{0}^{(2)}_{\dot{C}A} = \mathrm{ZW}_3(e_{(0)})\sigma^*_{\dot{C}A}\,\partial_\mu, \quad 0^{(1)\dot{C}B} = \mathrm{ZW}_3(e_{(0)})\sigma^{\mu\,\dot{C}B}\partial_\mu.$$

From (6.15) we find

$$\hat{0}^{(2)}_{\dot{C}A} \odot S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{M}}^{\dot{A}}) = \mathbb{1} \circ \sigma^*_{\dot{C}A}\,\sigma^{\nu\dot{A}B}\chi_{B;\nu;\mu}, \tag{6.20}$$

$$\hat{0}^{(1)A\dot{B}} \odot S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{N}}_A) = \mathbb{1} \circ \sigma^{\mu\,A\dot{B}}\,\sigma^*_{C\dot{A}}\omega^{\dot{C}}_{;\nu;\mu}. \tag{6.21}$$

We have

$$\begin{aligned} \chi_{B;\nu;\mu} &= \tfrac{1}{2}(\chi_{B;\nu;\mu} + \chi_{B;\mu;\nu}) + \tfrac{1}{2}(\chi_{B;\nu;\mu} - \chi_{B;\mu,\nu}) \\ &= \tfrac{1}{2}\chi_{B(;\nu;\mu)} + \tfrac{1}{2}P^R_{\nu\mu B}\chi_R, \end{aligned} \tag{6.22}$$

where $P_{\nu\mu}$ is the curvature 2-spinor. Similarly

$$\omega^{\dot{C}}_{;\nu;\mu} = \tfrac{1}{2}\omega^{\dot{C}}_{(;\nu;\mu)} + \tfrac{1}{2}P^{\dot{C}}_{\nu\mu\dot{R}}\omega^{\dot{R}}. \tag{6.23}$$

Substitution of (6.22) into (6.20) and (6.23) into (6.21) gives

$$\begin{aligned} \hat{0}^{(2)}_{\dot{C}A} \odot S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{M}}^{\dot{A}}) &= -ZW_3(e_{(0)})\chi_{C;\mu;\nu}g^{\mu\nu} \\ &\quad + \tfrac{1}{4}ZW_3(e_{(0)})P^R_{\nu\mu B}\textstyle\sum^{\mu\nu B}C\chi_R, \end{aligned} \tag{6.24}$$

$$\begin{aligned} \hat{O}^{(1)A\dot{B}} \odot S({}^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{N}}_A) &= -ZW_3(e_{(0)})\omega^{\dot{B}}{}_{;\mu;\nu}g^{\mu\nu} \\ &\quad - \tfrac{1}{4}ZW_3(e_{(0)})P^{\dot{C}}_{\nu\mu\dot{R}}\textstyle\sum^{\mu\nu\dot{B}}\dot{C}\omega^{\dot{R}}, \end{aligned} \tag{6.25}$$

where

$$\textstyle\sum^{\mu\nu B}C = {}^*\sigma^\mu_{\dot{C}A}\sigma^{\nu\dot{A}B} - {}^*\sigma^\mu_{\dot{C}A}\sigma^{\mu\,\dot{A}B}.$$

From (6.18) it follows that

$$\hat{0}^{(2)}_{\dot{C}A} \odot \mathcal{P}^{\dot{A}} = -m\mathbb{1}\sigma^{*\mu}_{\dot{C}A}\omega^A_{;\mu} = m^2\,\mathbb{1}\chi_C.$$

Similarly from (6.17) and (6.19)

$$\hat{O}^{(1)A\dot{B}} \odot \mathcal{J}_A = m\,\mathbb{1}\,\sigma^{\mu\,A\dot{B}}\chi_{A;\mu} = m^2\,\mathbb{1}\,\omega^{\dot{B}}.$$

Then, the Klein–Gordon equation for each component of χ and Ω^* has the form

$$\hat{0}^{(2)}_{CA} \odot \{\S(^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{N}}\dot{A}) + p^{\dot{A}}\}$$
$$= \mathbb{1}\{-\Box\chi_C + \tfrac{1}{4}P^R_{\nu\mu B}\Sigma^{\mu\nu B}{}_C\chi_R + m^2\chi_C\} = 0, \tag{6.26}$$

$$\hat{0}^{(1)A\dot{B}} \odot \{\S(^{(2)}\mathbb{D} \odot {}^{(1)}\overline{\mathcal{N}}_A) + \mathcal{T}_A\}$$
$$= \mathbb{1}\{-\Box\omega^{\dot{B}} - \tfrac{1}{4}\Sigma^{\mu\nu\dot{B}}{}_{\dot{C}}P_{\nu\mu}{}^{\dot{C}}{}_{\dot{R}}\omega^{\dot{R}} + m^2\omega^{\dot{B}}\} = 0, \tag{6.27}$$

where $\Box$ represents the covariant D'Alembert operator. The equations (6.26) and (6.27) in the limit of flat space reduce to the correct Klein–Gordon equation, for the signature (+2), in special relativity.

7. SYMMETRIC-ANTISYMMETRIC THEORY IN A COMPLEX TETRAD FORMALISM

As was seen in the previous sections, the geometry of the four-dimensional Riemannian space, described locally by the tetrad field, is algebraically described as a split quaternion subalgebra of the Cayley algebra. In this section we look for a generalization of this geometry in such a way that part of its algebraic description is contained in the full Cayley algebra. With this in mind we consider a general second-rank tensor field $G^{\mu\nu}(x^\alpha)$ given in terms of a complex tetrad as

$$G^{\mu\nu} = h^\mu_{(\alpha)}h^{*\nu}_{(\beta)}\eta^{\alpha\beta}, \tag{7.1}$$

Here $\eta^{\alpha\beta}$ indicates the Minkowski tensor with signature (+2). The matrix $G = (G^{\mu\nu})$ is Hermitian, $G^{*\nu\mu} = G^{\mu\nu}$. The symmetric and antisymmetric parts of this matrix are given by

$$G^{(\mu\nu)} = \tfrac{1}{2}(G^{\mu\nu} + G^{\nu\mu}) = \mathrm{Re}(G^{\mu\nu}),$$
$$G^{[\mu\nu]} = \tfrac{1}{2}(G^{\mu\nu} - G^{\nu\mu}) = i\mathrm{Im}(G^{\mu\nu}). \tag{7.2}$$

Denoting the matrices associated with the symmetric and antisymmetric parts of $G^{\mu\nu}$ by g and if, we have $G = g + if$. The matrices g and f are supposed to be nonsingular, and the matrix g is used for raising four-dimensional indices (and g^{-1} for lowering these indices),

$$A_{\mu\cdots} = G_{(\mu\nu)}A^{\nu\cdots}, \quad A^{\mu\cdots} = G^{(\mu\nu)}A_{\nu\cdots}, \quad G_{(\mu\nu)}G^{(\nu\sigma)} = \delta^\sigma_\mu.$$

The use of complex tetrads is known in the literature,[9] and our present formalism giving the Hermitian tensor $G^{\mu\nu}$ in terms of a complex tetrad is a condensed notation for a formalism due to Smith.[10]

From (7.1) we have

$$G_{\mu\nu} = h_{\mu(\alpha)}h^*_{\nu(\beta)}\eta^{\alpha\beta}, \quad h_{\mu(\alpha)} = G_{(\mu\rho)}h^\rho_{(\alpha)}. \tag{7.3}$$

In matrix notation this takes the form $K = (G_{\mu\nu})$, $K = g^{-1} + ig^{-1}\cdot f\cdot g^{-1} = K^\dagger$.

Associated with the field of complex tetrads we define in each Zorn–Weyl basis the set of four split octonion elements (for each of the two values of a)

$${}^{(a)}\mathcal{K}^\mu = \begin{pmatrix} ih^\mu_{(0)}W_a(e_{(0)}) & -h^{*\mu}_{(k)}W_a(e_{(k)}) \\ h^\mu_{(k)}W_a(e_{(k)}) & ih^{*\mu}_{(0)}W_a(e_{(0)}) \end{pmatrix}. \tag{7.4}$$

which may be written as ${}^{(a)}\mathcal{K}^\mu = ZW_a(K^\mu)$, where

$$K^\mu = ih^\mu_{(o)}u^*_{(o)} + ih^{*\mu}_{(o)}u_{(o)} + h^\mu_{(k)}u_{(k)} + h^{*\mu}_{(k)}u^*_{(k)}.$$

In the limit $\mathrm{Im}(h^\mu_{(\alpha)}) \to 0$ the Cayley numbers K^μ degenerate in elements of the split quaternion subalgebra of the octonion algebra. A straightforward calculation gives

$$ZW_1(K^{(\mu}) \odot ZW_2(\bar{K}^{\nu)}) = G^{(\mu\nu)}ZW_3(e_{(0)}), \tag{7.5}$$

where for any quantities A^μ, B^μ

$$A^{(\mu}B^{\nu)} = \tfrac{1}{2}(A^\mu B^\nu + A^\nu B^\mu).$$

Therefore, the ZW elements ${}^{(a)}\mathcal{K}^\mu$ are associated to the symmetric part of the Hermitian tensor $G^{\mu\nu}$.

It is also possible to introduce ZW elements belonging to the split quaternion subalgebra associated to the complex tetrad,

$${}^{(a)}\mathcal{L}^\mu = \begin{pmatrix} ih^\mu_{(0)}W_a(e_{(o)}) & -h^\mu_{(s)}W_a(e_{(s)}) \\ h\mu_{(s)}W_a(e_{(s)}) & ih^\mu_{(0)}W_a(e_{(o)}) \end{pmatrix}.$$

Since these objects are 4×4 matrices we may introduce their Hermitian conjugates

$${}^{(a)}\mathcal{L}^{\dagger\mu} = \begin{pmatrix} -ih^{*\mu}_{(0)}W^\dagger_a(e_{(0)}) & h^{*\mu}_{(s)}W^\dagger_a(e_{(s)}) \\ -h^{*\mu}_{(s)}W^\dagger_a(e_{(s)}) & -ih^{*\mu}_{(0)}W^\dagger_a(e_{(o)}) \end{pmatrix}.$$

In this equation we have to use that $W^\dagger_a(e_{(o)}) = W_a(e_{(o)})$ and $W^\dagger_a(e_{(s)}) = -W_a(e_{(s)})$. Defining for any ZW element the operation $\S({}^{(a)}M) = {}^{(a)}M = {}^{(a)}\overline{M}$, we find by a direct calculation

$$\S({}^{(1)}\mathcal{L}^{[\mu} \odot {}^{(2)}\mathcal{L}^{\dagger\nu]}) = -2G^{[\mu\nu]}ZW_3(e_{(o)}).$$

This is a relation involving product in the ZW algebra which generates the antisymmetric part of $G^{\mu\nu}$.

In the formalism presently considered the real part of $G^{\mu\nu}$ plays the role of metric of a Riemannian geometry with affinity $\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$ (the Christoffel symbols). Thus, only one kind of covariant differentiation is used, namely the usual covariant differentiation used in general relativity. Therefore, all previous conventions regarding covariant differentiation in the ZW algebra apply here. The antisymmetric part of the Hermitian tensor $G^{\mu\nu}$ is related to the electromagnetic potentials by the definition[10]

$$A^\mu = \lambda G^{[\mu\nu]}_\nu, \tag{7.6}$$

where λ is a constant. The potentials A^μ satisfy the covariant Lorentz condition $A^\mu_{;\mu} = -\lambda R_{\mu\nu}G^{[\mu\nu]} = 0$, where $R_{\mu\nu} = R^\alpha_{\mu\nu\alpha}$ is the Ricci tensor of the Riemannian geometry.

The operator of covariant differentiation in the ZW algebra is here defined similarly as before by

$${}^{(a)}\mathbb{D} = {}^{(a)}\mathcal{K}^\mu\partial_\mu.$$

As an application of the present formalism we consider the motion of a charged spin-zero massive test body under the action of gravitation and electromagnetism, described by the corresponding covariant Klein–Gordon equation. We take unities such that $c = \hbar = 1$. The equation of motion takes the form

$$\left(G^{(\mu\nu)}\frac{D}{Dx^\mu}\frac{D}{Dx^\nu} + m^2\right)\Psi = 0, \tag{7.7}$$

$$\frac{D}{Dx^\mu} = \frac{1}{i}\partial_\mu\Psi - eA_\mu\Psi. \tag{7.8}$$

Using the notation $P_\mu\Psi = (1/i)\partial_\mu\Psi$, Eq. (7.7) takes the form (in the Lorentz gauge)

$$G^{(\mu\nu)}P_\mu P_\nu\psi - 2\lambda G^{[\mu\nu]}_{;\nu}P_\mu\psi = -(m^2 + \lambda^2 e^2 G^{[\mu\nu]}_{;\nu}G^{\alpha}_{[\mu\alpha]})\psi. \tag{7.9}$$

Writing ${}^{(a)}B = {}^{(a)}K^\mu P_\mu\psi$ we have

$${}^{(1)}\overline{\mathbb{D}}\odot{}^{(2)}B = {}^{(1)}\overline{K}^\mu\partial_\mu\odot{}^{(2)}K^\nu P_\nu\psi = i\,{}^{(1)}\overline{K}^\mu\odot{}^{(2)}K^\nu P_\mu P_\nu\psi.$$

From (7.5) we get

$${}^{(1)}\overline{\mathbb{D}}\odot{}^{(2)}B = iG^{(\mu\nu)}P_\mu P_\nu\psi = iG^{\mu\nu}P_\mu P_\nu\psi. \tag{7.10}$$

A similar operation may be extended for the vector field

$$\frac{D\psi}{Dx^\mu} = P_\mu\psi - \lambda e G_{[\mu\alpha]}{}^{;\alpha}\psi.$$

Defining

$${}^{(a)}\Pi = {}^{(a)}K^\mu\frac{D}{Dx^\mu} = \frac{1}{i}\,{}^{(a)}\mathbb{D} - e\,{}^{(a)}A, \quad {}^{(a)}R = {}^{(a)}K^\mu\frac{D\psi}{Dx^\mu},$$

we have

$${}^{(1)}\overline{\Pi}\odot{}^{(2)}R = {}^{(1)}\overline{K}^\mu\odot{}^{(2)}K^\nu\frac{D}{Dx^\mu}\frac{D}{Dx^\nu}\psi.$$

Then,

$$\mathcal{S}({}^{(1)}\overline{\Pi}\odot{}^{(2)}R) = \mathcal{S}({}^{(1)}\overline{K}^\mu\odot{}^{(2)}K^\nu)\frac{D}{Dx^\mu}\frac{D}{Dx^\nu}\Psi$$

$$= 2G^{(\mu\nu)}\cdot\mathbb{1}\cdot\frac{D}{Dx^\mu}\frac{D}{Dx^\nu}\psi.$$

Therefore, the Klein–Gordon equation takes the simpler form

$$\mathcal{S}({}^{(1)}\overline{\Pi}\odot{}^{(2)}R) + 2m^2\,\mathbb{1}\cdot\psi = 0.$$

The first term on the left-hand side of this equation is the Zorn–Weyl gauge invariant covariant "D'Alembertian" (divided by a factor $\frac{1}{2}$).

8. CONCLUSION

In the Zorn algebra formulation presented in the previous sections the spacetime remains four-dimensional. Therefore, the Zorn–Weyl differential operator is written in Sec. 7

$${}^{(a)}\mathbb{D} = {}^{(a)}K^\mu\partial_\mu,$$

has octonion coefficients ${}^{(a)}K^\mu$ but the covariant differential operator ∂_μ acts on coordinates of the four-dimensional spacetime. This is in part related to the fact that exist four octonions ${}^{(a)}K^\mu$ for each of the two values of the "Weyl index" (a). It should be possible to think of an operator which involves four extra coordinates as

$$\mathbb{D} = {}^{(a)}K^\mu D_{(a)\mu}, \quad a = 1,2,$$

such that

$$\overline{\mathbb{D}}\odot\mathbb{D} = \sum_{\substack{a,b\\ a\neq b}}{}^{(a)}\overline{K}^\mu\odot{}^{(b)}K^\nu\partial_{a\mu}\partial_{b\nu}.$$

This operator would be called a generalized "D'Alembertian." However, it can not be interpreted in a usual way since there are no coordinates $X^{(a)\mu}$ in the conventional relativistic formalism. One possible interpretation should be to take $X^{(a)\mu} = {}^{(a)}\sigma^\mu$, where ${}^{(1)}\sigma^\mu = (\sigma^{\mu AB})$, ${}^{(2)}\sigma^\mu = (\sigma^\mu_{AB})$. Such a formalism is quite different than the conventional relatvistic theory, since here "coordinates" are the field quantities ${}^{(a)}\sigma^\mu(x)$ of the usual theory, and the transition of this formalism to the conventional theory is not straightforward. Since such a generalization is not directly reducible to the conventional relativistic theory, we have not considered this extension in the present paper.

Finally it is worth noticing that in the example given in Sec. 7, where an octonion algebra appears, the physical interpretation of the extra four components of the complex tetrad, or the octonion, is to incorporate the electromagnetic interaction on a massive charged test body moving in gravitational and electromagnetic fields. Thus it can be said that, in our present formalism, the effect of introduction of octonions (complex tetrads) is to absorb the electromagnetic interaction of the test particle.

[1]P. A. M. Dirac, Proc. Royal Irish Academy Sec. A **50**, 261 (1945).
[2]A. M. Buoncristiani, J. Math. Phys. **14**, 849 (1973).
[3]M. Gnaydin and F. Gürsey, J. Math. Phys. **14**, 1651 (1973).
[4]R. Penney, Nuovo Cimento B **3**, 95 (1971); Am. J. Phys. **36**, 871 (1968).
[5]M. Gnaydin, J. Math. Phys. **17**, 1875 (1976).
[6]M. Zorn, Abh. Math. Sem. Univ. Hamburg **8**, 123 (1931); G. B. Seligman, Trans. Ann. Math. Soc. **97**, 286 (1960).
[7]R. Brauer and H. Weyl, Am. J. Maths. **57**, 425 (1935).
[8]E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Blackie, London, 1954), p. 10.
[9]D. W. Sciama, Nuovo Cimento **8**, 417 (1958).
[10]P. D. P. Smith, Nuovo Cimento B **24**, 255 (1974).

The Yang–Mills system in compactified Minkowski space; Invariance conditions and SU(2) invariant solutions

J. Harnad

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada

S. Shnider

Department of Mathematics, McGill University, Montréal, Québec, Canada

Luc Vinet

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada
(Received 25 May 1978)

In this work the SU(2) Yang–Mills equations are studied in compactified Minkowski space. The manifold is identified with that of the Lie group $U(1)\times SU(2)$ and a classification is made of all SU(2) principal bundles over this base space in terms of homotopy classes of mappings $f:S^3\to S^3$. Invariance of gauge fields under transformation groups is defined in terms of bundle mappings and the case of invariance under SU(2) translations is shown to imply a trivial bundle structure. All solutions to the field equations invariant under $U(1)\times SU(2)$ translations are obtained as well as all (anti-) self-dual solutions invariant under SU(2) translations.

1. INTRODUCTION

The interpretation of classical gauge fields as connection forms on principal fibre bundles has led to a deepened understanding of the underlying geometric and topological structure.[1–4] Solutions to the SU(2) Yang–Mills equations in Euclidean space have been extensively studied in the recent literature and particular attention has been paid to those which arise as a pullback of some solution on the space S^4 under a compactification mapping. A complete algebraic geometric classification has been given by Atiyah *et al.*[5] for the self-dual Yang–Mills fields of this type, and numerous families of explicit solutions have been obtained.[6–10]

The Yang–Mills system in Minkowski space has received much less attention, partly because of the more difficult problems associated with hyperbolic differential equations. No systematic classification has been attempted, although certain particular solutions with interesting properties have been obtained.[11–19] Again, a compactification of the space is useful on the one hand in order to have a global realization of certain group actions leaving the field equations invariant[20] and is necessary on the other hand in order to give a meaningful interpretation to such topological invariants as the Chern class (instanton) number. A well-known procedure[20–23] leads to a homogeneous space of the Minkowskian conformal group [locally isomorphic to O(4,2)] factorized by the isotropy group at the origin of Minkowski space. The resulting manifold, referred to as the conformally compactified Minkowski space, is most easily realized as the projective cone of null vectors in $\mathbb{R}^6$ under a quadratic form of signature (2,4) and is diffeomorphic to $(S^1\times S^3)/\mathbb{Z}_2$, the manifold of the Lie group U(2). For simplicity, we shall mainly be working with the twofold covering space $S^1\times S^3$, identifiable as the manifold of the group $U(1)\times SU(2)$, which may alternatively be regarded as the compactification of the universal covering space $\mathbb{R}\times S^3$.

A simplifying assumption which has proved useful in the determination of particular solutions to the gauge field equations has been the requirement that these be invariant under some relatively large transformation group, in particular, subgroups of the conformal group of space–time. For the case of ordinary tensor fields, such a characterization is unambiguous and the form of such fields may be determined in a straightforward fashion.[23] The notion of invariance of a gauge field, however, has been somewhat loosely treated in the literature, since in certain individual cases a seemingly arbitrary gauge transformation which accompanies the group action must first be picked.[6,7,14] An attempt has been made by Bergmann and Flaherty[24] to sharpen this notion of invariance at the level of vector bundles. This was limited, however to infinitesimal invariance under one-dimensional Lie groups and involved the replacement of the finite gauge transformation by a seemingly equally arbitrary infinitesimal one. In fact, the precise definition of invariance, valid for finite group actions of any dimension, is easily interpreted at the level of connection forms on the principal bundle and will be given in Sec. 3.

In Sec. 2 a summary is given regarding the coordinate systems and reference frames used in the subsequent calculations and the relevant transformation groups. Of particular interest will be the action of $U(1)\times SU(2)$ on itself by left or right translations. A convenient set of coframes is provided by the canonical left or right invariant forms of Maurer–Cartan. The corresponding metric under which these frames are orthonormal is the natural $O(2)\times O(4)$ invariant one, conformal to the Minkowskian metric. In Sec. 3, the conditions for invariance of gauge fields under a transformation group are studied, the "arbitrary" gauge transformations interpreted as transition functions between local sections in the principal bundle and a criterion given (Proposition 3.1)

[a]Research supported in part by the National Research Council of Canada.

 0022-2488/79/050931-12$01.00

for when these transformations may be eliminated by a suitable choice of sections. The construction of all possible SU(2) bundles over $S^1\times S^3$ and their classification by homotopy classes of maps $f: S^3\to SU(2)$ is given (Proposition 3.2) and the Chern class number is shown to be equal to the degree of f (Proposition 3.3). It is proved (Proposition 3.4) that no nontrivial bundle exists which admits an SU(2) group action projecting onto left or right translation in $S^1\times S^3$, and moreover (Proposition 3.5), that the existence on the base manifold of a local connection form (gauge field) which is invariant under SU(2) translations and which satisfies the Yang–Mills equations is sufficient to imply a trivial bundle structure. Thus, all SU(2) invariant solutions have vanishing instanton number. In Sec. 4, we consider those gauge fields which are invariant under left or right translations by $U(1)\times SU(2)$. Since the group acts transitively, the fields are determined through invariance by their value at any one point, and the field equations reduce to a set of algebraic ones which may be solved. Among the solutions are certain new ones which are essentially complex, in the sense that no gauge transformation may cast them into a real form. An interpretation of these necessarily involves a complexification of the Lie algebra and hence an extension of the structure group to $SL(2,\mathbb{C})$. The details of such a complexification are not analyzed here; instead, we limit ourselves to deriving all the $U(1)\times SU(2)$ invariant complex solutions to the field equations regarded as forms on the base manifold. In Sec. 5, all self-dual and anti-self-dual fields invariant under SU(2) translations are obtained. The analysis for both Sec. 4 and 5 involves the canonical forms of complex symmetric 3×3 matrices. The results of these sections are summarized in Propositions (4.2) and (5.1)–(5.3). Finally, in Sec. 6, there is a brief discussion of gauge fields invariant under the group SO(4) formed from left and right SU(2) translations and the implementation of cyclic boundary conditions.

2. THE MANIFOLD AND GROUP ACTIONS

Let us identify each point in the space $S^1\times S^3$ with an element $(e^{i\psi},v)$ of the group $U(1)\times SU(2)$. Introducing six-dimensional coordinates as

$$e^{i\psi}=u_5+iu_0,\quad v=u^4-iu^i\sigma_i,$$
$$u_0^2+u_5^2=u_1^2+u_2^2+u_3^2+u_4^2=1 \tag{2.1}$$

(when i appears as a coefficient it means $\sqrt{-1}$), where σ_i are the Pauli matrices, the points $\{(u^a)_{a=0,\ldots,5}\in\mathbb{R}^6\}$ constitute an orbit under $O(2)\times O(4)$ on the cone

$$C^5=\{(\eta^a)\in\mathbb{R}^6 \mid \eta_0^2-\eta_1^2-\eta_2^2-\eta_3^2-\eta_4^2+\eta_5^2=0\}. \tag{2.2}$$

[Raising and lowering of six-dimensional indices is always to be understood as defined by the metric diag $(+1,-1,-1,-1,-1,+1)$.] Since this orbit intersects each ray on C^5 in two points (u^a) and $(-u^a)$, the projective cone may be identified with the space $(S^1\times S^3)/\mathbb{Z}_2$. At the group level, this may be realized by the homomorphism $j: U(1)\times SU(2)\to U(2)$ defined by

$$j:(e^{i\psi},v)\to u\equiv e^{-i\psi}v\in U(2) \tag{2.3}$$

which has the two element kernel $\{(1,\mathbb{1}),(-1,-\mathbb{1})\}$. We may note however that while $S^1\times S^3$ is thus identified as a twofold covering of $(S^1\times S^3)/\mathbb{Z}_2$, the two spaces are actually diffeomorphic under the identification

$$U(2)\ni\rho v\leftrightarrow(\rho^2,v_\rho)\in U(1)\times SU(2), \tag{2.4}$$

where $\rho\in U(1)$; v, $v_\rho\in SU(2)$ and

$$v_\rho\equiv v\begin{pmatrix}\rho & 0\\ 0 & \rho^{-1}\end{pmatrix}.$$

Note also that although (2.4) does not define a group isomorphism, it does preserve the action of left translation under SU(2) and a similar diffeomorphism may be defined which preserves right translation.

The compactification of Minkowski space M may be realized by identifying the point with Cartesian coordinates $(x^\mu)_{\mu=0,\ldots,3}$ with the Hermitian matrix,

$$h=x^\mu\sigma_\mu\quad(\sigma_0\equiv\mathbb{1}) \tag{2.5}$$

and using the Cayley transform[21]

$$C: h\to u=(1-ih)(1+ih)^{-1}\in U(2). \tag{2.6}$$

In terms of coordinate components, this gives the usual relations:

$$u^\mu=\pm\frac{x^\mu}{\tau},\quad u^4=\pm\frac{1+\mathbf{x}^2}{2\tau},\quad u^5=\pm\frac{1-\mathbf{x}^2}{2\tau}, \tag{2.7}$$

where $\tau=[x_0^2+\frac{1}{4}(1-\mathbf{x}^2)^2]^{1/2}$ and $\mathbf{x}^2=x_0^2-x_1^2-x_2^2-x_3^2$, and their (singular) inverse

$$x^\mu=\frac{u^\mu}{u^4+u^5}. \tag{2.8}$$

We have the following natural group actions on $S^1\times S^3$:

Left and right translations under SU(2):

$$SU(2)_L\quad L_w:(e^{i\psi},v)\to(e^{i\psi},wv), \tag{2.9a}$$
$$SU(2)_R\quad R_w:(e^{i\psi},v)\to(e^{i\psi},vw), \tag{2.9b}$$

where

$$w=\begin{pmatrix}a+ib & c+id\\ -c+id & a-ib\end{pmatrix}\in SU(2), \tag{2.10}$$

with $a^2+b^2+c^2+d^2=1$.

Left and right translations under U(1):

$$L_\phi(e^{i\psi},v)=R_\phi(e^{i\psi},v)\to(e^{i(\psi+\phi)},v). \tag{2.11}$$

Left action of the product $SU(2)_L\times SU(2)_R$

$$L_{(w,w')}:(e^{i\psi},v)\to(e^{i\psi},wvw'^{-1}),\quad w,w'\in SU(2). \tag{2.12}$$

The diagonal SU(2) *subgroup* $[SU(2)_L\otimes SU(2)_R]_D$ [which is conjugation of the SU(2) group on itself]:

$$D_w:(e^{i\psi},v)\to(e^{i\psi},wv\,w^{-1}). \tag{2.13}$$

From (2.1), (2.7), and (2.9)–(2.11) we see that left translation by $(e^{i\phi},w)$ and right translations by $(e^{-i\phi},w^{-1})$ are conjugate to each other under the map

$$\mathcal{J}_{PT}:(u^0,u^1,u^2,u^3,u^4,u^5)\to(-u^0,-u^1,-u^2,-u^3,u^4,u^5) \tag{2.14}$$

corresponding to space–time inversion in Minkowski space.

The product group $SU(2)_L \otimes SU(2)_R$ in this representation may be identified with the group SO(4) acting on the (u^1,u^2,u^3,u^4) subspace. [More specifically, we have a homomorphism: $SU(2)_L \otimes SU(2)_R \to SO(4)$ of which the two-element kernel $\{(\mathbb{1},\mathbb{1}),(-\mathbb{1},-\mathbb{1})\}$ acts as the identity.] Furthermore, U(1) translations on S^1 may be identified with the action of the group SO(2) in the (u^0,u^5) subspace. The diagonal subgroup $[SU(2)_L \otimes SU(2)_R]_D$ acts on the (u^1,u^2,u^3) subspace and may be identified (again, factorizing first by the kernel) as the SO(3) subgroup defining proper rotations in Minkowski space.

Define a basis $\{t_\mu\}$ for the $u(1) \oplus su(2)$ Lie algebra as

$$t_0 = i\mathbb{1}, \quad t_i = \frac{\sigma_i}{2i}, \quad i = 1,2,3, \tag{2.15}$$

$$[t_0,t_i] = 0, \quad [t_i,t_j] = \epsilon_{ijk}t_k.$$

A corresponding basis for the cotangent space $T^*(S^1 \times S^3)$ is provided by the canonical (Maurer–Cartan) left-invariant forms[25] on $U(1) \times SU(2)$:

$$\overline{\omega}_L = v^{-1}dv + id\psi = \omega_L^i t_i + \omega_L^0 t_0, \tag{2.16}$$

where

$$\omega_L^k = -2\eta_{\alpha\beta}^k u^\alpha du^\beta, \quad \alpha,\beta = 1,2,3,4, \tag{2.17a}$$

$$\omega_L^0 = d\psi = u^5 du^0 - u^0 du^5. \tag{2.17b}$$

The symbol $\eta_{\alpha\beta}^k$ is the one defined by 't Hooft,[26]

$$\eta_{\alpha\beta}^k = -\eta_{\alpha\beta}^k, \quad \eta_{4i}^k = -\delta_i^k, \quad \eta_{ij}^k = \epsilon^{ijk}. \tag{2.18}$$

The canonical forms satisfy the Maurer–Cartan structure equations:

$$d\omega_L^i + \tfrac{1}{2}\epsilon^{ijk}\omega_L^j \wedge \omega_L^k = 0, \tag{2.19a}$$

$$d\omega_L^0 = 0. \tag{2.19b}$$

Similarly, we may define a basis of right invariant forms $\overline{\omega}_R$ which are related to the left-invariant ones by space–time inversion,

$$\overline{\omega}_L = -\mathscr{I}_{PT}^* \overline{\omega}_R. \tag{2.20}$$

In terms of the above forms, we may express the natural $O(2) \times O(4)$ invariant Lorentzian metric for $S^1 \times S^3$ as

$$\begin{aligned} g &= du_0^2 - du_1^2 - du_2^2 - du_3^2 - du_4^2 + du_5^2 \\ &= \omega_L^0 \otimes \omega_L^0 - \tfrac{1}{4}\omega_L^i \otimes \omega_L^i = \omega_R^0 \otimes \omega_R^0 - \tfrac{1}{4}\omega_R^i \otimes \omega_R^i \\ &= \frac{1}{\tau^2} g_m, \end{aligned} \tag{2.21}$$

which, is conformal to the Minkowski metric g_m in M. The sets $\{\omega_i^0, \tfrac{1}{2}\omega_L^i\}$ and $\{\omega_R^0, \tfrac{1}{2}\omega_R^i\}$ define fields of orthonormal (nonholonomic) coframes for $S^1 \times S^3$. The corresponding volume element is thus

$$d\Omega = \tfrac{1}{8}\omega_L^0 \wedge \omega_L^1 \wedge \omega_L^2 \wedge \omega_L^3 = \tfrac{1}{8}\omega_R^0 \wedge \omega_R^1 \wedge \omega_R^2 \wedge \omega_R^3, \tag{2.22}$$

which, under the pullback to M gives $1/\tau^4$ times the Minkowskian volume element.

The dual $*F$ of a 2-form F is unchanged under a conformal change of metric, and may be simply expressed in terms of the basis for $\Lambda^2 T^*(S^1 \times S^3)$ provided by the above coframes. In view of their orthonormality, we have

$$*\omega^0 \wedge \omega^i = \tfrac{1}{4}\epsilon^{ijk}\omega^j \wedge \omega^k \tag{2.23a}$$

and

$$*\omega^i \wedge \omega^j = -2\epsilon^{ijk}\omega^0 \wedge \omega^k \tag{2.23b}$$

(where ω^μ represents either ω_L^μ or ω_R^μ.)

The change of relative sign in (2.23) is due to the Lorentzian type of metric which also implies that an eigenvector of the * operator may only have eigenvalue $\pm i$ since

$$**F = -F. \tag{2.24}$$

3. THE PRINCIPAL BUNDLE AND INVARIANT CONNECTIONS

We shall now investigate the restrictions that invariance conditions for a connection form place on the bundle on which the connection is defined. Let H be a Lie group, $\mathfrak{H}$ its Lie algebra, $\{U_\alpha\}$ an open covering of a manifold M and ω_α an $\mathfrak{H}$-valued 1-form on U_α. Let G be a Lie transformation group acting on M on the left,

$$f_g : M \to M, \quad g \in G \tag{3.1}$$

and suppose that the U_α are G invariant,

$$f_g U_\alpha \subset U_\alpha, \quad \forall\ g \in G. \tag{3.2}$$

The condition that $f_g^* \omega_\alpha$ be gauge equivalent to ω_α is

$$f_g^* \omega_\alpha = \mathrm{Ad}\rho_\alpha^{-1}(g,p)\omega_\alpha + \rho_\alpha^{-1}(g,p)d_p\rho_\alpha(g,p), \quad p \in U_\alpha, \quad \forall\ g \in G, \tag{3.3}$$

where

$$\rho_\alpha : G \times U_\alpha \to H \tag{3.4}$$

defines a local gauge transformation and satisfies, for consistency of the composition law in G,

$$\rho_\alpha(g',p)\rho_\alpha(g, f_{g'}(p)) = \rho_\alpha(gg',p). \tag{3.5}$$

Furthermore, if the local forms ω_α are related by

$$\omega_\alpha = \mathrm{Ad}(k_{\alpha\beta}^{-1})\omega_\beta + k_{\alpha\beta}^{-1} dk_{\alpha\beta} \quad \text{on } U_\alpha \cap U_\beta \tag{3.6}$$

for some gauge function

$$k_{\alpha\beta} : U_\alpha \cap U_\beta \to H \tag{3.7}$$

we must have, for consistency

$$\rho_\beta(g,p)k_{\alpha\beta}(f_g(p)) = k_{\alpha\beta}(p)\rho_\alpha(g,p) \tag{3.8}$$

and

$$k_{\alpha\beta}^{-1}(p)k_{\beta\gamma}^{-1}(p) = k_{\alpha\gamma}^{-1}(p), \quad \forall\ p \in U_\alpha \cap U_\beta \cap U_\gamma. \tag{3.9}$$

Using the function $k_{\alpha\beta}$ as transition functions, we can construct a principal H bundle $\mathscr{P}(M,H)$ over M which is locally trivial over each U_α,

$$\tau_\alpha : \mathscr{P}|_{U_\alpha} \to U_\alpha \times H, \tag{3.10}$$

where the maps τ_α defining the trivialization satisfy

$$\tau_\beta \tau_\alpha^{-1}(p,h) = (p, k_{\alpha\beta}(p)h), \quad \forall\ p \in U_\alpha \cap U_\beta, \quad h \in H. \tag{3.11}$$

A G-action $\tilde{f}$ on $\mathscr{P}$ projecting to the action f on M may then be defined by

$$\tilde{f}_g(\tau_\alpha^{-1}(p,h)) = \tau_\alpha^{-1}(f_g(p),\rho_\alpha^{-1}(g,p)h), \tag{3.12}$$

which is a consistent definition by virtue of Eq. (3.8). A connection $\tilde{\omega}$ on $\mathscr{P}$ may now be defined by

$$\tilde{\omega}_{\tau_\alpha(p,h)} = \tau_\alpha^*(\mathrm{Ad}h^{-1}(\omega_\alpha) + h^{-1}dh), \tag{3.13}$$

where $h^{-1}dh$ is the Maurer–Cartan from on $T(H) \subset T(U_\alpha \times H)$. Equation (3.6) shows that this definition is unambiguous and Eq. (3.3) implies that the connection form $\tilde{\omega}$ is G invariant under the action $\tilde{f}: \mathscr{P} \to \mathscr{P}$. Thus the information in Eqs. (3.3)–(3.9) is equivalent to the existence of a principal H bundle with an action of G as bundle transformations, together with a G-invariant connection. The local connection forms are the pullbacks of ω under the sections σ_α given by

$$\sigma_\alpha(p) = \tau_\alpha^{-1}(p,e) \tag{3.14}$$

and the functions $\rho_\alpha(g,p)$ are just the transition functions relating σ_α and its image under G,

$$(\tilde{f}_g \circ \sigma_\alpha \circ f_g^{-1})(p) = R_{\rho_\alpha^{-1}(g,f_g(p))}\sigma_\alpha(p). \tag{3.15}$$

When determining the local forms ω_α corresponding to a G-invariant connection, it is helpful to know whether a suitable choice of sections ρ_α may be made such that the functions ρ_α take the simplest possible form. In particular, we may inquire whether a choice exists for which the ρ_α take the identity element in G as their only value. The following proposition gives the conditions for this to hold.

Proposition 3.1: If for each U_α there is a smooth cross section S_α of the orbits of G in M such that

(i) $G \times S_\alpha \to U_\alpha$ has constant rank, and

(ii) for all $s \in S_\alpha$ and $g \in G_s$ (the isotropy group of G at s), we have $\rho_\alpha(g,s) = e$, then there exists a G-invariant section $\hat{\sigma}_\alpha$ of $\mathscr{P}|_{U_\alpha}$ and a form $\hat{\omega}_\alpha = \hat{\sigma}_\alpha^* \tilde{\omega}$ which is gauge equivalent to ω_α and satisfies

$$f_g^* \hat{\omega}_\alpha = \hat{\omega}_\alpha. \tag{3.16}$$

Proof: By the implicit function theorem, for all $p \in U_\alpha$, there is an open set $V \subset U_\alpha$ containing p and smooth functions:

$$g_{\alpha,v}: V \to G, \quad s_{\alpha,v}: V \to S_\alpha,$$

such that

$$f_{g_{\alpha,v}(q)}(s_{\alpha,v}(q)) = q, \quad q \in V.$$

The map

$$h_{\alpha,v}: q \to \rho_\alpha(g_{\alpha,v}(q), s_{\alpha,v}(q))$$

defines a smooth function $h_{\alpha,v}: V \to H$ and it is immediate that for $q \in V_1 \cap V_2$,

$$\rho_\alpha(g_{\alpha,v_1}(q), s_{\alpha,v_1}(q)) = \rho_\alpha(g_{\alpha,v_2}(q), s_{\alpha,v_2}(q)).$$

Thus, we have a smooth function

$$h_\alpha: U_\alpha \to H$$

and may verify directly that the section defined by

$$\hat{\sigma}_\alpha(p) = \tau_\alpha^{-1}(p, h_\alpha(p))$$

is G invariant, that is,

$$\tilde{f}_g^{-1} \circ \hat{\sigma}_\alpha \circ f_g = \hat{\sigma}_\alpha,$$

implying, by Eq. (3.15) that

$$\hat{\omega}_\alpha = \hat{\sigma}_\alpha^* \omega = \mathrm{Ad}h_\alpha^{-1}\omega_\alpha + h_\alpha^{-1}dh_\alpha$$

satisfies (3.16).

A particular case for which the proposition applies is when the group action satisfies (i) and is free; that is, the isotropy group at all points is the identity element. This certainly holds for any closed subgroup of a Lie group acting by left or right translations, which will be our main concern in Sec. 4 and 5.

We now turn to the classification of principal SU(2) bundles over $S^1 \times S^3$. Corresponding to any smooth map $f: S^3 \to \mathrm{SU}(2)$, we may define such a bundle by considering the space

$$E = (0,\tfrac{3}{2}) \times S^3 \times \mathrm{SU}(2), \tag{3.17}$$

regarded as the trivial SU(2) bundle over the product of the open interval $(0,\frac{3}{2})$ with S^3, and factorizing by the equivalence relation

$$(t,p,h) \underset{f}{\sim} (1 + t,p,f(p)h), \tag{3.18}$$
$$\forall t \in (0,\tfrac{1}{2}), \quad p \in S^3, \quad h \in \mathrm{SU}(2).$$

Let E_f denote the quotient of E by this relation, regarded as a principal bundle over $S^1 \times S^3$.

Proposition 3.2: Any SU(2) bundle over $S^1 \times S^3$ is equivalent to E_f for some f. The bundles E_f and E_g are isomorphic if and only if f is homotopic to g.

The proof involves certain ideas in the homotopy theory of fibre bundles which may be found in any standard text,[27,28] to which the reader is referred for further background.

Proof: First note that by using the covering of S^3 by contractible neighborhoods of the hemispheres, we find that for any Lie group H, the isomorphism class of a principal H bundle over S^3 is determined by a homotopy class in $\pi_2(H)$. Since $\pi_2(\mathrm{SU}(2)) = 0$, any SU(2) bundle over S^3 is trivial. Further, $(0,\frac{3}{2}) \times S^3$ being homotopy equivalent to S^3, the same triviality holds true for any SU(2) bundle over $(0,\frac{3}{2}) \times S^3$. Given an SU(2) bundle B over $S^1 \times S^3$, let $\pi^* B$ be the pullback of B to $(0,\frac{3}{2}) \times S^3$ under the map $\pi: (0,\frac{3}{2}) \times S^3 \to S^1 \times S^3$ defined by

$$\pi(t,p) = (e^{2\pi i t},p). \tag{3.19}$$

Letting $\hat{\pi}: (\pi^* B) \to B$ denote the corresponding bundle map, an isomorphism $\phi: E \to \pi^* B$ may be chosen under which

$$\hat{\pi}\phi(t,p,h) = \hat{\pi}\phi(1 + t,p,f(p)h), \quad \forall\, t \in (0,\tfrac{1}{2}) \tag{3.20}$$

for some smooth map $f: S^3 \to \mathrm{SU}(2)$. Generally any isomorphism $\tilde{\phi}: E \to \pi^* B$ defines a homotopy of maps $f_t(p)$, $t \in (0,\frac{1}{2})$,

but may be composed with a suitable automorphism which replaces these by a fixed $f(p)$ in the same class. The map $\hat{\pi}\hat{\phi} : E \to B$, in view of (3.20), passes to the quotient under the equivalence (3.18), thereby defining an isomorphism $\hat{\phi} : E_f \to B$.

Suppose now that $f,g \to SU(2)$ are homotopic by a homotopy h_t satisfying $h_t \equiv g, t \in (0,\frac{1}{2})$ and $h_t \equiv f$ for $t \in (1,\frac{3}{2})$. Then the automorphism of E,

$$(t,p,h) \to (t,p,h_t^{-1}(p)h),$$

takes points which are equivalent under f into points which are equivalent under g and hence determines an isomorphism $E_f \cong E_g$. Conversely, to prove that if E_f is isomorphic to E_g then f is homotopic to g, we apply Hopf's theorem, stating that two maps $f,g : S^n \to S^n$ are homotopic if and only if they have the same degree. Suppose there exists an isomorphism $\sigma : E_f \to E_g$. Then there is an automorphism $\hat{\sigma}$ of E which takes f equivalent points into g equivalent ones. Writing

$$\hat{\sigma}(t,p,h) = (t,p,\sigma_t(p)h)$$

we see that

$$g(p)\sigma_t(p) = \sigma_{1+t}(p)f(p)$$

and therefore

$$\deg g + \deg\sigma_t = \deg f + \deg\sigma_{1+t}$$

which follows most simply by replacing f, g, σ_t, and σ_{1+t} by the appropriate power map $p \to p^n$, $p \in SU(2)$ of degree n, to which they are homotopic. Since σ_t and σ_{1+t} are homotopic, their degrees are equal and hence so are $\deg f$ and $\deg g$, implying f and g are homotopic. Next we prove:

Proposition 3.3: The second Chern class number of E_f equals the degree of f; that is,

$$C_2 = -\frac{1}{16\pi^2}\int_{S^1\times S^3} B(\Omega\wedge\Omega) = \deg f, \tag{3.21}$$

where Ω is the curvature of any connection $\tilde{\omega}$ on E_f pulled back to $S^1\times S^3$ through any choice of local sections over an open covering and B is the Killing form:

$$B(X,Y) = \mathrm{Tr}\,\mathrm{ad}X\,\mathrm{ad}Y, \quad X,Y \in su(2). \tag{3.22}$$

Proof: Any connection on E may be written as

$$\tilde{\omega}_{(t,p,h)} = h^{-1}dh + \mathrm{Ad}h^{-1}\omega_{(t,p)}, \tag{3.23}$$

where $h^{-1}dh$ is the Maurer–Cartan form on SU(2) and $\omega_{(t,p)}$ is an su(2)-valued 1-form on $(0,\frac{3}{2})\times S^3$. In order that $\tilde{\omega}$ define a connection on E_f, we must have

$$\omega_{(t,p)} = \mathrm{Ad}f^{-1}\omega_{(1+t,p)} + f^{-1}df, \quad \forall\, t \in (0,\tfrac{1}{2}). \tag{3.24}$$

If ϕ is chosen as a C^∞ real-valued function on $(0,\frac{3}{2})$ with values $\phi \equiv 1$ on $(0,\frac{1}{2})$ and $\phi \equiv 0$ on $(1,\frac{3}{2})$, then

$$\omega_{(t,p)} \equiv \phi(t)f^{-1}df \tag{3.25}$$

satisfies (3.24). The curvature $\tilde{\Omega}$ corresponding to this choice of connection on the open set $(0,1)\times S^3$ in $S^1\times S^3$ [where (0,1) is regarded as S^1 minus a point], pulled back by the section $\sigma(t,p) \to (t,p,e)$ is:

$$\begin{aligned}\Omega &= d\omega + \tfrac{1}{2}[\omega,\omega]\\ &= \dot{\phi}dt\wedge f^{-1}df + \tfrac{1}{2}(\phi^2-\phi)[f^{-1}df f^{-1}df],\\ B(\Omega\wedge\Omega) &= \tfrac{1}{2}\dot{\phi}(\phi^2-\phi)dt\wedge B(f^{-1}df\wedge[f^{-1}df f^{-1}df]).\end{aligned} \tag{3.26}$$

In the notation of the preceding section, we have

$$B(h^{-1}dh\wedge[h^{-1}dh, h^{-1}dh]) = -12\omega_L^1\wedge\omega_L^2\wedge\omega_L^3. \tag{3.27}$$

The S^1 integral in (3.21) gives

$$\tfrac{1}{2}\int_0^1 \dot{\phi}(\phi^2-\phi)dt = \tfrac{1}{2}\int_1^0 (\phi^2-\phi)d\phi = \frac{1}{12}. \tag{3.28}$$

Let $f^{-1}df_p = A_{1,p}\omega^1 + A_{2,p}\omega^2 + A_{3,p}\omega^3$ for $p \in S^3$ and define a linear transformation A_p by $A_p(t_i) = A_{i,p}$. then by (3.27) and (3.28)

$$\begin{aligned}\int_{S^1} B(\Omega\wedge\Omega) &= -\int_{S^1}(\det A)\omega_L^1\wedge\omega_L^2\wedge\omega_L^3\\ &= -16\pi^2\deg f.\end{aligned} \tag{3.29}$$

The last equality follows from Brouwer's degree theorem.[29] The following proposition shows that it is impossible to introduce a left SU(2) action on a principal SU(2) bundle over $S^1\times S^3$ which projects onto the left-translations (2.9a) on the base unless the bundle is trivial.

Proposition 3.4: E_f admits a left SU(2) action compatible with the action $L_g(t,p) = (t,gp)$ $[g,p \in SU(2)]$ if and only if $\deg f = 0$.

Proof: Let $\tilde{L}_g$ represent the action of $g \in SU(2)$ on E_f. Then $\tilde{L}_g$ induces a map $\hat{L}_g$ on E. If we write $\hat{L}_g(t,p,h) = (t,gp,\rho_t(p,g)h)$, then we must have

$$\rho_{1+t}(g,p)f(p) = f(gp)\rho_t(g,p). \tag{3.30}$$

Fix a point p_0. Then the map $g \to f(gp_0)$ has the same degree as f. The two maps $g \to \rho_t(g,p_0)$ and $g \to \rho_{1+t}(g,p_0)$ being homotopic have equal degrees. Since the degrees are additive under group multiplication, (3.30) implies that $\deg f = 0$.

Even if we ignore the possibility of a group action on the bundle E_f, an invariance condition interpreted entirely on the base imposes strong restrictions on f. If the local connection form of a connection on E_f pulled back to the base is invariant under left SU(2) translations, it may be expressed as

$$\omega_{(t,p)} = A_t\omega_L + B_t\,dt, \tag{3.31}$$

where $A_t \in \mathrm{End}(su(2))$, $B_t \in su(2)$ and ω_L is the S^3 part of the Maurer–Cartan form $\bar{\omega}_L$ on $S^1\times S^3$ [identified with $U(1)\times SU(2)$]. A differential equation for ω, such as the Yang–Mills equation, imposes certain conditions on A_t and B_t, which then imply a trivial bundle structure.

Proposition 3.5: Suppose ω is the local connection form of a connection on E_f which satisfies $L_g^*\omega = \omega$ for all $g \in SU(2)$, and hence has the form (3.31). If ω satisfies a differential equation of the type

$$A_t^{(n)} = F(t,A_t,A_t^{(1)},\ldots,A_t^{(n-1)},B_t,B_t^{(1)},\ldots,B_t^{(m)}) \tag{3.32}$$

such that F is smooth and $F(t,0,0,\ldots,0) = 0$, then $\deg f = 0$.

Proof: Assuming $\deg f \neq 0$, we shall arrive at a contradiction. The condition (3.24) for the connection to be consistently defined on E_f takes the form:

$$\mathrm{Ad}f^{-1}A_{1+t}(\omega_L) + f^{-1}df = A_t(\omega_L), \tag{3.33a}$$

$$\mathrm{Ad}f^{-1}B_{1+t} = B_t, \qquad t\in(0,\tfrac{1}{2}). \tag{3.33b}$$

Since $\deg f \neq 0$, the image of f covers SU(2) and we conclude from (3.33b) that $B_{1+t} \in$ center su(2) $= \{0\}$ and hence $B_{1+t} = B_t = 0$.

Taking the exterior derivative of (3.33a) at point p, substituting, using the Maurer–Cartan equations and translating to the origin e we obtain

$$\mathrm{Ad}f^{-1}(p)(A_1[\omega_L,\omega_L] - [A_1\omega_L, A_1\omega_L])_e = A_0[\omega_L,\omega_L]_e - [A_0\omega_L, A_0\omega_L]_e. \tag{3.34}$$

Also, by repeated application of the exterior derivative we find

$$\mathrm{Ad}f^{-1}A_1^{(n)}\omega_L = A_0^{(n)}\omega_L, \quad n = 1,2,\cdots, \tag{3.35}$$

and conclude that $A_1^{(n)} = A_0^{(n)} = 0$.

Due to the uniqueness of solution of (3.32) under these boundary conditions, A_t must therefore be constant. Substituting $A_0 = A_1$ in Eq. (3.33a) at $t = 0$, we have

$$f^{-1}df = A_1(\omega_L) - \mathrm{Ad}f^{-1}(\omega_L)$$

$$= (I - \mathrm{Ad}f^{-1})A_1(\omega_L).$$

Since

$$\det(I - \mathrm{Ad}f^{-1}) = 0$$

the argument at the end of Proposition 3.3 shows $\deg f = 0$, contradicting the hypothesis.

4. THE FIELD EQUATIONS AND U(1)×SU(2) INVARIANT SOLUTIONS

The Yang–Mills potential ω is defined, subject to a choice of section in the principal SU(2) bundle over $S^1\times S^3$, as an su(2)-valued 1-form on $S^1\times S^3$. The corresponding field is the curvature 2-form,

$$F = d\omega + \tfrac{1}{2}[\omega,\omega] \tag{4.1}$$

and the field equations are

$$D{*F} = d{*F} + [\omega,{*F}] = 0, \tag{4.2}$$

where the dual $*F$ is taken with respect to the metric (2.21). In view of Eq. (2.24), the Lorentzian analog of the (anti) self-dual equations is

$${*F} = \mp iF, \tag{4.3}$$

implying that such fields are necessarily complex.

We shall be interested in determining solutions to (4.2) and (4.3) that are invariant under the group actions defined in Sec. 2. Since left and right translations by U(1)×SU(2) or SU(2) satisfy the conditions of Proposition 3.1, we may without loss of generality choose the value of the function ρ_α as the identity, giving the simple requirements

$$L_g^*\omega = \omega \tag{4.4a}$$

or

$$R_g^*\omega = \omega \tag{4.4b}$$

$\forall g\in$U(1)×SU(2) or SU(2).

Consider now the gauge potentials ω satisfying (4.4a). These may be characterized most easily in terms of the left-invariant canonical form ω_L,

$$\omega = (A^i{}_j\omega_L^j + B^i\omega_L^0)t_i, \tag{4.5}$$

where the components $A^i{}_j$ and B^i will be constant for $\mathrm{U}(1)\times\mathrm{SU}(2)_L$ invariant forms and, more generally, functions of the S^1 angle ψ for $\mathrm{SU}(2)_L$-invariant ones. For right-invariant gauge fields, satisfying (4.4b), we simply replace ω_L in (4.5) by ω_R. However, since the field equations are invariant under the space–time inversion $\mathcal{I}_{PT}$, we may obtain all right-invariant solutions from the left-invariant ones by making the replacements:

$$A^i{}_j \to -A^i{}_j(-\psi), \quad B^i(\psi)\to -B^i(-\psi),$$
$$\psi\to -\psi \quad \text{and} \quad \omega_L^\mu\to\omega_R^\mu. \tag{4.6}$$

It is therefore sufficient to consider the left-invariant case (4.5) only. The field strength may be conveniently expressed in terms of the basis $\omega_L^\mu\wedge\omega_L^\nu$ for $\Lambda^2T^*(S^1\times S^3)$ as:

$$F = [\tfrac{1}{2}F^i{}_{\mu\nu}\omega_L^\mu\wedge\omega_L^\nu]t_i, \tag{4.7}$$

where

$$F^i_{0k} = \frac{\partial A^i{}_k}{\partial\psi} + \epsilon_{imn}B^mA^n{}_k \tag{4.8a}$$

and

$$F^i_{lk} = -\epsilon_{lkm}A^i{}_m + \epsilon_{imn}A^m_lA^n_k. \tag{4.8b}$$

The components of $*F$ are easily obtained using (2.23). The field equations (4.2), expressed in components, take the form

$$\frac{1}{2}\frac{d^2A^i{}_k}{d\psi^2} + \tfrac{1}{2}\epsilon_{imn}\frac{dB^m}{d\psi}A^n{}_k + \epsilon_{imn}B^m\frac{dA^n{}_k}{d\psi} + 2A^i{}_k - 3\epsilon_{imn}\epsilon_{kpq}A^m{}_pA^n{}_q + 2(A^i{}_kA^l{}_mA^l{}_m - A^i{}_mA^l{}_mA^l{}_k) + \tfrac{1}{2}(B^iB^mA^m{}_k - A^i{}_kB^mB^m) = 0 \tag{4.9a}$$

and

$$\epsilon_{ijk}A^j_l\frac{dA^k{}_l}{d\psi} - A^i{}_jA^m_lB^m + A^m_lA^m_lB^i = 0 \tag{4.9b}$$

while (4.3) becomes

$$\frac{dA^i{}_l}{d\psi} \mp 2iA^i{}_l + \epsilon_{ijk}(B^jA^k_l \pm i\epsilon_{lmn}A^j{}_mA^k{}_n) = 0. \tag{4.10}$$

We shall be considering complex solutions to these equations, hence ω takes values in the complexified gauge algebra su(2) $\otimes\,\mathbb{C}$ = sl(2,$\mathbb{C}$).

For the $\mathrm{U}(1)\times\mathrm{SU}(2)_L$ invariant case the derivative terms in (4.9) vanish, leaving a third degree set of algebraic equations which we may solve completely. In doing this, it will be helpful to make use of the fact that these equations are invariant not only under the complex gauge transformations:

$$A^i{}_k \to R^iA^i{}_k \tag{4.11a}$$

$$B^i \to R^i{}_j B^j \tag{4.11b}$$

$\forall (R^i{}_j) \in SO(3,\mathbb{C})$,

but also under the following transformations, corresponding to right-SU(2) translations:

$$A^i{}_k \to A^i{}_j R^j{}_k, \tag{4.12a}$$

$$B^i \to B^i, \tag{4.12b}$$

$\forall (R^i{}_j) \in SO(3,\mathbb{C})$.

Let $A \in sl(2,\mathbb{C}) \otimes sl(2,\mathbb{C})^*$ denote the $sl(2,\mathbb{C})$ endomorphism whose components in the $\{t_i \otimes t^*_j\}$ basis are $A^i{}_j$, and B the $sl(2,\mathbb{C})$ element with components B^i. Multiplying (4.9a) by $A^j{}_k$ and summing over k, we obtain the matrix equation

$$X(1 + \mathrm{tr}X) \mp 3(\det X)^{1/2}\mathbb{1} - X^2 + \tfrac{1}{4}(B \otimes XB - B^2 X) = 0, \tag{4.13a}$$

where

$$X \equiv AA^T \tag{4.14}$$

and

$$B^2 = \sum_i B^i B^i, \tag{4.15}$$

while (4.9b) may be written

$$XB = (\mathrm{tr}X)B. \tag{4.13b}$$

The gauge invariance condition (4.11) now becomes the invariance of (4.13) under the transformations:

$$X \to RXR^T, \quad B \to RB, \tag{4.16}$$

where R is the $SO(3,\mathbb{C})$ matrix with components $R^i{}_j$. It is possible to solve (4.13) by using these transformations to standardize the forms of X and B. The result may be summarized as follows:

Lemma 4.1: Up to a gauge transformation (4.16), the solutions of (4.13) are:

(i) $B = 0$ and X has one of the forms:

$$X = \mu\mathbb{1}, \quad \mu = \tfrac{1}{4} \text{ or } 1, \tag{4.17a}$$

$$X = \begin{pmatrix} -1 & & \\ & -1 & \\ & & 9 \end{pmatrix}, \tag{4.17b}$$

$$X = \begin{pmatrix} -1 & & \\ & -1 & \\ & & 0 \end{pmatrix}, \tag{4.17c}$$

(ii)

$$X = \begin{pmatrix} y & iy & 0 \\ iy & -y & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 \\ 0 \\ \pm 2 \end{pmatrix}, \quad y \neq 0. \tag{4.18}$$

(iii) $X = 0$ and B is arbitrary. (4.19)

Proof: (i) If $B = 0$, then X satisfies the polynomial equation

$$X^2 - X(1 + \mathrm{tr}X) \mp 3\det X^{1/2}\mathbb{1} = 0. \tag{4.20}$$

Since this is of degree 2, X must have one of the following Jordan forms:

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}.$$

In either case, there is a two-dimensional subspace of eigenvectors with equal eigenvalues, and hence X has at least one eigenvector of nonzero length. It follows that by a transformation of the type (4.16), X may be cast in the form

$$X = \begin{pmatrix} x + y & iy & 0 \\ iy & x - y & 0 \\ 0 & 0 & z \end{pmatrix}.$$

Substitution in (4.20) yields the equations:

$$x(1 + x + z \pm 3z^{1/2}) = 0,$$

$$y(1 + z) = 0,$$

$$z(1 + 2x) \pm 3xz^{1/2} = 0,$$

whose solution leads to the expressions (4.17) or $X = 0$. If $B \neq 0$, applying the LHS of (4.13a) to it and using (4.13b) gives

$$\mathrm{Tr}X = \pm 3(\det X)^{1/2} \tag{4.21a}$$

while taking the trace of (4.13a) gives

$$\mathrm{Tr}X + (\mathrm{Tr}X)^2 - \mathrm{Tr}X^2 = \mp 9(\det X)^{1/2}. \tag{4.21b}$$

(ii) If $B^2 \neq 0$, B may be rotated into the form $(0,0,b)$ by the transformation (4.16). Equation (4.13b) then implies that X may, after a suitable rotation in the (1,2) subspace, be put in one of the forms:

$$X = \begin{pmatrix} y & iy & 0 \\ iy & -y & 0 \\ 0 & 0 & z \end{pmatrix},$$

$$\mathrm{Tr}X = z, \quad \mathrm{Tr}X^2 = z^2, \quad \det X = 0$$

if it has only one eigenvector of nonzero length, or

$$X = \begin{pmatrix} y & & \\ & -y & \\ & & z \end{pmatrix}$$

$$\mathrm{tr}X = z, \quad \mathrm{Tr}X^2 = z^2 + 2y^2, \quad \det X = -y^2 z$$

otherwise. In either case, substitution in (4.21) yields $z = 0$ and hence Eq. (4.13a) becomes

$$X(1 - \tfrac{1}{4}b^2) = X^2,$$

implying that either X, B are of the form (4.18) or $X = 0$.

(iii) If $B^2 = 0$, $B \neq 0$ it may be rotated into the form $(1,i,0)$ by the transformation (4.16). Equation (4.13b) then implies that X is of the form

$$X = \begin{pmatrix} x + y & iy & z \\ iy & x - y & iz \\ z & iz & -x \end{pmatrix},$$

$$\mathrm{Tr}X = x, \quad \mathrm{Tr}X^2 = 3x^2, \quad \det X = -x^3.$$

Substitution in (4.21) yields $x = 0$ and hence Eq. (4.13a) becomes

$$X = X^2,$$

implying $y = z = 0$ and hence $X = 0$.

Using the results of this lemma, we may obtain all constant [U(1)×SU(2) invariant] solutions to (4.9), the result being:

Proposition 4.2: Up to a constant gauge transformation, the U(1)×SU(2) invariant solutions to (4.9) are of the form:

$$\text{I. } A = \alpha \otimes \gamma, \quad B = \beta, \tag{4.22}$$

where $\alpha, \beta \in \mathrm{sl}(2,\mathbb{C})$, $\gamma \in \mathrm{sl}(2,\mathbb{C})^*$ satisfy

$$\alpha \cdot \beta = 0, \quad \beta^2 = 4 \tag{4.23a}$$

and either

$$\alpha^2 = 0 \quad \text{or} \quad \gamma^2 = 0 \tag{4.23b}$$

[the inner product on sl(2,ℂ) and sl(2,ℂ)* being defined relative to $\{t_i\}$ and $\{t_i^*\}$ as orthonormal bases].

$$\text{II. } A = \begin{pmatrix} i & & \\ & i & \\ & & 3 \end{pmatrix} R, \quad B = 0, \ R \in \mathrm{SO}(3,\mathbb{C}). \tag{4.24}$$

$$\text{III. } A = \lambda \mathbb{1}, \quad B = 0, \tag{4.25}$$

where $\lambda = \frac{1}{2}$ or 1.

$$\text{IV. } A = 0, \quad B \text{ arbitrary} \tag{4.26}$$

Proof: Consider first case (iii) of the lemma. Since X vanishes, Eq. (4.14) implies that the lengths and inner products of the rows of A vanish. It follows that they are all proportional to a single zero length vector γ:

$$A = \alpha \otimes \gamma, \quad \gamma^2 = 0, \quad \alpha \in \mathrm{sl}(2,\mathbb{C}), \ \gamma \in \mathrm{sl}(2,\mathbb{C})^*.$$

Substitution in (4.9a) gives

$$\alpha(1 - \tfrac{1}{4}B^2) + \tfrac{1}{4}B(\alpha \cdot B) = 0 \quad \text{if } A \neq 0.$$

Taking the inner product with B, we obtain

$$\alpha \cdot B = 0$$

and hence $B^2 = 4$ or $A = 0$. In either case, Eq. (4.9b) is also satisfied, yielding solutions of type I or IV.

Turning next to case (ii) of the lemma, Eqs. (4.14) and (4.18) imply that, up to a suitable transformations of type (4.11) and (4.12), A and B have the form:

$$A = \begin{pmatrix} 0 & 0 & a \\ c & i\epsilon c & ia \\ d & i\epsilon d & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ \pm 2 \end{pmatrix}$$

$$a^2 = y \neq 0, \quad \epsilon = \pm 1, \quad c \text{ and } d \text{ arbitrary}.$$

Substitution into Eq. (4.9a) shows that $c = d = 0$, and therefore $A = \bar{\alpha} \otimes \bar{\gamma}$, where

$$\bar{\alpha} = t_1 + it_2, \quad \bar{\gamma} = at_3^*,$$

applying the transformations (4.11) and (4.12), we obtain, in general:

$$A = \alpha \otimes \gamma, \quad B = \beta,$$

where $\alpha, \beta \in \mathrm{sl}(2,\mathbb{C})$, $\gamma \in \mathrm{sl}(2,\mathbb{C})^*$ satisfy:

$$\alpha \cdot \beta = 0, \quad \alpha^2 = 0, \quad \beta^2 = 4,$$

corresponding to the other case of solution I.

Solution II follows directly from the case (4.17b) of the lemma while solution III follows from (4.17a). A rotation of type (4.12) need not be included in parametrizing the latter case, since it becomes equivalent to a gauge transformation (4.11). The last case (4.17c) corresponds, up to transformation (4.12), to A of the form

$$A = \begin{pmatrix} i & & \\ & \pm i & \\ & & 0 \end{pmatrix}.$$

which does not satisfy Eq. (4.9a). This exhausts all the cases of Lemma 4.1 and therefore case I–IV represent all U(1)×SU(2) invariant solutions of Eqs. (4.9).

Denoting the $T^*(S^3)$ part of $\bar{\omega}_L$ as ω_L, we may write the gauge potential ω and field F corresponding to the solutions (4.22) as

$$\omega = \alpha(\gamma, \omega_L) + \beta\omega_L^0 \tag{4.27}$$

and

$$F = -\tfrac{1}{2}\alpha(\gamma, [\omega_L, \omega_L]) + [\alpha, \beta](\gamma, \omega_L) \wedge \omega_L^0. \tag{4.28}$$

The condition (4.23) imply, in the case $\alpha^2 = 0$ that

$$[\alpha, \beta] = \pm 2i\alpha \tag{4.29}$$

and hence (4.28) is an (anti) self-dual field; the only one, in fact, among those listed in Proposition (4.2). Furthermore, since

$$\alpha^2 = \tfrac{1}{4}([\alpha, \beta])^2 \quad \text{and} \quad \alpha \cdot [\alpha, \beta] = 0 \tag{4.30}$$

the integrands $B(F \wedge F)$ and $B(F \wedge {}^*F)$ defining, respectively, the class number and the action integral both vanish.

The gauge potential and field for solutions (4.24) may, up to a suitable gauge transformation, be expressed as

$$\omega = (3 - i)\hat{\alpha}(\hat{\alpha}^*, \omega_L) + i\omega_L, \tag{4.31}$$

$$F = 2i[\omega_L, \omega_L] + 2(1 - i)\hat{\alpha}(\hat{\alpha}^*, [\omega_L, \omega_L]), \tag{4.32}$$

where $\hat{\alpha} \in \mathrm{sl}(2,\mathbb{C})$ is any unit vector and $\hat{\alpha}^* \in \mathrm{sl}(2,\mathbb{C})^*$ its dual, defined by the inner product on sl(2,ℂ). The action integral

$$A = -\frac{1}{4g^2} \int_{(S^1 \times S^3)/\mathbb{Z}_2} F^i \wedge {}^*F^i \tag{4.33}$$

takes the value $64\pi^2/g^2$, while the Chern number, consistent with Proposition (3.5) again vanishes. The case $R = \mathbb{1}$ of solution (4.24) has been found independently by Howe and Tucker.[19] Solution (4.25) with $\lambda = \frac{1}{2}$ is the SO(4)×SO(2) invariant one studied by de Alfaro *et al.*,[11,13] while the case $\lambda = 1$ of (4.25) and solution (4.26) represent pure gauge potentials corresponding to vanishing fields.

5. SELF-DUAL SU(2)$_L$ -INVARIANT SOLUTIONS

The self-duality equation (4.10) for SU(2)$_L$ -invariant fields may be cast in a simple vectorial form by identifying the three sl(2,ℂ) elements $\{A_i\}$ whose components are the columns of the matrix A. In terms of these and the sl(2,ℂ) vector B, the gauge potential (4.5) is

$$\omega = A_i\omega_L^i + B\omega_L^0 \tag{5.1}$$

and Eq. (4.10) becomes

$$\dot{A}_i \mp 2iA_i + [B, A_i] \pm i\epsilon_{ijk}[A_j, A_k] = 0. \tag{5.2}$$

Defining new sl(2,ℂ) vectors $\{\tilde{A}_i\}$, $\tilde{B}$ by

$$A_i \equiv -\tilde{A}_i e^{\pm 2i\psi}, \quad B \equiv \pm 2i\tilde{B}e^{\pm 2i\psi}, \tag{5.3}$$

and using three-dimensional cross products to replace the Lie brackets, we have

$$\tilde{A}_i \pm ie^{\pm 2i\psi}\{2\tilde{B}\times\tilde{A}_i - \epsilon_{ijk}\tilde{A}_j\times\tilde{A}_k\} = 0. \tag{5.4}$$

Finally, changing variables to

$$w \equiv e^{\pm 2i\psi} \tag{5.5}$$

we obtain the equation

$$\tilde{A}_i' + \tilde{B}\times\tilde{A}_i = \tfrac{1}{2}\epsilon_{ijk}\tilde{A}_j\times\tilde{A}_k, \tag{5.6}$$

where $\tilde{A}_i'$ denotes the derivative of $\tilde{A}_i$ with respect to w.

The solutions to (5.6) may be systematically found by separately considering the cases when the vectors $\{\tilde{A}_i\}$ are colinear, when they span a (complex) two-dimensional space and when they are all linearly independent. In the latter case, we may further distinguish between three classes depending upon the eigenvectors of the complex symmetric matrix Y formed from the dot products

$$Y_{ij} \equiv \tilde{A}_i\cdot\tilde{A}_j. \tag{5.7}$$

The results may be summarized as follows:

Proposition 5.1: If the three vectors $\tilde{A}_i$ satisfying Eq. (5.6) are colinear, then up to a gauge transformation they are constant and $\tilde{B}$ vanishes. The gauge potential ω is thus of the form

$$\omega = e^{\pm 2i\psi}\alpha(\gamma,\omega_L), \tag{5.8}$$

where $\alpha\in$sl(2,$\mathbb{C}$) and $\gamma\in$sl(2,$\mathbb{C}$)* are arbitrary complex vectors.

Proposition 5.2: If the vectors $\tilde{A}_i$ span a complex two-dimensional space, then up to a gauge transformation and a cyclic permutation of the labels (1,2,3), the most general solution to (5.6) is of the form:

$$\tilde{A}_1 = \alpha, \qquad \tilde{A}_2 = \beta,$$
$$\tilde{A}_3 = i(\alpha\cos\theta + \beta\sin\theta), \qquad \tilde{B} = i(-\alpha\sin\theta + \beta\cos\theta), \tag{5.9}$$

where $\alpha,\beta\in$sl(2,$\mathbb{C}$) are arbitrary noncolinear, complex vectors and θ is any complex number.

Proposition 5.3: If the vectors $\tilde{A}_i$ are linearly independent, then three classes of solutions to Eq. (5.6) exist, depending upon whether the matrix Y has two (and hence three) eigenvectors of nonzero length, only one, or none at all. Denoting by $\tilde{A}$ the matrix whose columns are $\tilde{A}_1$, $\tilde{A}_2$, and $\tilde{A}_3$, then up to a gauge transformation, B vanishes and

(i) If Y has three eigenvectors of nonzero length,

$$\tilde{A} = \begin{pmatrix} p & & \\ & q & \\ & & r \end{pmatrix} R, \tag{5.10}$$

where

$$p = b\,\mathrm{ds}(b(w-w_0)\mid m),$$
$$q = b\,\mathrm{ns}(b(w-w_0)\mid m), \tag{5.11}$$
$$r = b\,\mathrm{cs}(b(w-w_0)\mid m),$$

ds, ns, and cs are Jacobi–Glaisher functions,[10] b, m, $w_0\in\mathbb{C}$, and $R\in$SO(3,$\mathbb{C}$) are arbitrary.

(ii) If Y has only one eigenvector of nonzero length,

$$\tilde{A} = \begin{pmatrix} p+q & iq & 0 \\ iq & p-q & 0 \\ 0 & 0 & r \end{pmatrix} R, \tag{5.12}$$

where

$$p = \frac{a}{\sinh[a(w-w_0)]},$$
$$q = \frac{b\sinh[a(w-w_0)]}{a}, \tag{5.13}$$
$$r = \frac{b\coth[a(w-w_0)]}{a},$$

with a, b, $w_0\in\mathbb{C}$, $b\neq 0$, and $R\in$SO(3,$\mathbb{C}$).

(iii) If Y has no eigenvectors of nonzero length,

$$\tilde{A} = \begin{pmatrix} p+q & iq & r \\ iq & p-q & ir \\ r & ir & p \end{pmatrix} R, \tag{5.14}$$

where

$$p = -\frac{1}{w-w_0},$$
$$q = -\tfrac{1}{2}(w-w_0)[a^2(w-w_0)^2 - b], \tag{5.15}$$
$$r = -a(w-w_0),$$

with a, b, $w_0\in\mathbb{C}$, $a\neq 0$, and $R\in$SO(3,$\mathbb{C}$).

We may remark that Eqs. (5.12)–(5.13) with $b = 0$ still define a solution, however this is a degenerate case of class (i) with only two distinct eigenvalues. Similarly, Eqs. (5.14) and (5.15) with $a = 0$ define a degenerate case belonging to class (ii) with only one eigenvalue. The solutions (5.10)–(5.11) with $R = \mathbb{1}$ and $w_0 = 0$, have previously been obtained by Howe and Tucker[19]; the others given here are all new. It is relevant to note that gauge transformations within the class of SU(2)-invariant solutions alter the vectors $\tilde{A}_i$ only by a complex rotation and hence leave the matrix Y of inner products invariant. Therefore, the various classes of solutions given above are really distinct. The vector B, on the other hand, is transformed into $\mathrm{Ad}g^{-1}(B + \dot{g}g^{-1})$ under a transformation which preserves the SU(2) symmetry (the gauge function g depending upon the S^1 angle ψ only) and hence can always be made to vanish by solving a linear, first-order differential equation for g. However, this does not always give rise to the simplest form for the solutions to the field equations.

Proof of Proposition 5.1: Since the $\tilde{A}_i$'s are colinear, Eq. (5.6) becomes

$$\tilde{A}_i' + \tilde{B}\times\tilde{A}_i = 0.$$

Taking the inner product with $\tilde{A}_j$ shows that the dot products $(\tilde{A}_i\cdot\tilde{A}_j)$ are constant, and hence the ψ dependence is generated by a rotation. There exists therefore an SU(2)-valued function g on S^1 such that

$$\tilde{A}_i = \mathrm{Ad}g\alpha_i,$$

where α_i are constant sl(2,$\mathbb{C}$) elements which are colinear,

$$\alpha_i = C_i\alpha.$$

Applying the gauge transformation g, $\tilde{A}_i$ becomes α_i and the field equation becomes

$$\beta \times \alpha = 0,$$

where β is the transformed value of B. Therefore, β is proportional to α and we have

$$\omega = e^{\pm 2i\psi}(\alpha(\gamma,\omega_L) + b\alpha\omega_L^0),$$

where $\gamma \in \mathrm{sl}(2,\mathbb{C})^*$ has components C_i and b is some complex function on S^1. Applying the gauge transformation defined by $\exp(-\alpha \int b\, d\psi)$ gives the expression (5.8).

Proof of Proposition 5.2: Since the $\tilde{A}_i$'s span a two-dimensional space,

$$\det\tilde{A} = \tilde{A}_1 \cdot (\tilde{A}_2 \times \tilde{A}_3) = 0.$$

In the gauge where $B = 0$, taking the inner product of (5.6) with $\tilde{A}_l$ again shows that the ψ dependence is determined by a rotation acting upon fixed vectors α_i. We apply, as above, the appropriate gauge transformation to make the $\tilde{A}_i$'s take these constant values, which gives rise in general to a nonvanishing B. Within a permutation of indices, we may define:

$$\tilde{A}_1 = \alpha, \quad \tilde{A}_2 = \beta, \quad \tilde{A}_3 = a\alpha + b\beta,$$

for some constant vectors α, β and numbers a, b. Substitution into Eq. (5.6) then shows that

$$a^2 + b^2 = -1$$

and

$$\tilde{B} = (-b\alpha + a\beta).$$

Therefore, defining

$$a = i\cos\theta \quad \text{and} \quad b = i\sin\theta$$

we obtain the result (5.9).

Proof of Proposition 5.3: Preliminary to the proof, we remark that any complex, symmetric 3×3 matrix C may, after conjugation by a suitable complex rotation, be cast in one of three canonical forms:

$$C^{(1)} = \begin{pmatrix} x & & \\ & y & \\ & & z \end{pmatrix}, \tag{5.16a}$$

$$C^{(2)} = \begin{pmatrix} x+y & iy & 0 \\ iy & x-y & 0 \\ 0 & 0 & z \end{pmatrix}, \quad y \neq 0, \tag{5.16b}$$

$$C^{(3)} = \begin{pmatrix} x+y & iy & z \\ iy & x-y & iz \\ z & iz & x \end{pmatrix}, \quad z \neq 0, \tag{5.16c}$$

corresponding, respectively, to three, one, or no eigenvectors of nonzero length.

Now choosing a gauge in which B vanishes, taking the inner product of (5.6) with $\tilde{A}_l$ and symmetrizing in i, l gives

$$Y' = 2\det\tilde{A}\cdot\mathbb{1}. \tag{5.17}$$

Integrating this, we have

$$Y = C + g\mathbb{1}, \tag{5.18}$$

where C is a symmetric, constant matrix and g is a complex function on S^1 satisfying

$$g' = 2\det\tilde{A}. \tag{5.19}$$

Since the second term in (5.18) is unchanged by a rotation, we may always standardize Y by a constant rotation:

$$\tilde{A}_i \to \tilde{A}_i R, \quad Y \to R^{-1} Y R, \tag{5.20}$$

to one of the three forms:

$$Y^{(1)} = \begin{pmatrix} a^2 + g_1 & & \\ & b^2 + g_1 & \\ & & g_1 \end{pmatrix}, \tag{5.21a}$$

$$Y^{(2)} = \begin{pmatrix} g_2 + 2b & 2ib & 0 \\ 2ib & g_2 - 2b & 0 \\ 0 & 0 & g_2 + a^2 \end{pmatrix}, \quad b \neq 0, \tag{5.21b}$$

$$Y^{(3)} = \begin{pmatrix} g_3 + b & ib & 2a \\ ib & g_3 - b & 2ia \\ 2a & 2ia & g_3 \end{pmatrix}, \quad a \neq 0, \tag{5.21c}$$

where $a, b \in \mathbb{C}$ are arbitrary. (The squares are introduced to simplify the final result and do not imply any restrictions on the parameters involved since a, b are complex.) Substituting each of these into (5.17) and using

$$\det Y = (\det\tilde{A})^2 \tag{5.22}$$

gives rise to the three differential equations:

$$(g_1')^2 = 4g_1(g_1 + a^2)(g_1 + b^2), \tag{5.23a}$$

$$(g_2')^2 = 4g_2^2(g_2 + a^2), \tag{5.23b}$$

$$(g_3')^2 = 4g_3^3. \tag{5.23c}$$

Integrating, we obtain:

$$g_1 = b^2\,\mathrm{cs}^2(b(w - w_0) \mid m) \quad \text{with} \quad m = 1 - a^2/b^2, \tag{5.24a}$$

$$g_2 = \frac{a^2}{\sinh^2[a(w - w_0)]}, \tag{5.24b}$$

$$g_3 = \frac{1}{(w - w_0)^2}, \tag{5.24c}$$

where $w_0 \in \mathbb{C}$ is arbitrary. Since

$$Y = \tilde{A}^T\tilde{A} \tag{5.25}$$

this just determines $\tilde{A}$ up to a transformation,

$$\tilde{A} \to R\tilde{A}, \quad R \in \mathrm{O}(3,\mathbb{C}) \tag{5.26}$$

which is a gauge transformation only if R is constant on S^1 and $\det R = 1$. (A nonconstant gauge transformation must also change the value of B.) Moreover, since a symmetrization was involved in passing from Eq. (5.6) to (5.17), not all $\tilde{A}$'s satisfying Eq. (5.25) solve (5.6). However, if $\tilde{A}$ is itself chosen to be symmetric, then (5.18) and (5.25) determine a solution of (5.6) (up to a sign), since A is invertible. Its eigenvectors in that case are the same as those of Y. Therefore, $\tilde{A}$ will correspondingly take one of the three forms:

$$\tilde{A}^{(1)} = \begin{pmatrix} p & & \\ & q & \\ & & r \end{pmatrix}, \tag{5.27a}$$

$$\tilde{A}^{(2)} = \begin{pmatrix} p+q & iq & 0 \\ iq & p-q & 0 \\ 0 & 0 & r \end{pmatrix}, \tag{5.27b}$$

$$\tilde{A}^{(3)} = \begin{pmatrix} p+q & iq & r \\ iq & p-q & ir \\ r & ir & p \end{pmatrix}. \tag{5.27c}$$

Solving (5.25) to determine p, q, r yields the expressions of Eqs. (5.11), (5.13), and (5.15) up to a sign, which is then determined by returning to Eq. (5.6). Applying a transformation of type (5.26) to these solutions and substituting in Eq. (5.6) shows that the rotation R must be constant and have determinant 1. This is therefore only a gauge transformation. However, the diagonalization of Y was achieved by applying the transformation (5.20) which leaves Eqs. (5.6) invariant. Therefore, the general solution is obtained by applying an arbitrary transformation of this type to the solutions (5.27), thereby giving the results stated in Eqs. (5.10)–(5.15) and completing our proof.

Among the various solutions given above, the only one which is gauge equivalent to one in the previous section is (5.8) for the case when α^2 vanishes. In this case, there exists an element $\beta \in \mathrm{sl}(2,\mathbb{C})$ satisfying Eq. (4.23a) and hence applying a gauge transformation defined by the transition function $\exp(\pm i\psi\beta)$ transforms solution (5.8) into the form (4.27). An interesting feature of all the solutions of this section is that they are periodic in ψ with period π. This implies that the gauge fields are well defined as forms on $(S^1 \times S^3)/\mathbb{Z}_2$ and that the Chern class number (and hence also the action integral) vanishes (provided the integration constants are shown so that the solutions are regular), in consistency with the results of Sec. 3.

As a further remark regarding rank 1 solutions of the type (5.8), we note that in this case the field equations (4.9) become linear. Therefore, any linear combination of the self-dual and anti-self-dual solutions

$$\omega = \alpha[\cos 2\psi(\gamma_1, \omega_L) + \sin 2\psi(\gamma_2, \omega_L)], \tag{5.28}$$

where $\alpha \in \mathrm{sl}(2,\mathbb{C})$, $\gamma_1, \gamma_2 \in \mathrm{sl}(2,\mathbb{C})^*$

is also a solution to (4.9), though not self-dual. In fact, it is easily verified that (5.28) is, up to a gauge transformation, the most general form for rank 1 solutions.

6. DISCUSSION OF SO(4) INVARIANT SOLUTIONS

Solutions to the field equations (4.2) (in $\mathbb{R} \times S^3$) have been obtained by Lüscher[14] and Schechter[15] under the requirement of invariance under the group SO(4) formed from left and right SU(2) translations. In this case the isotropy group at the origin is not the identity but the group SO(3) identified locally with $[\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R]_D$. The conditions of Proposition (3.1) are therefore not necessarily satisfied and there need not exist a G-invariant section in the principal bundle under which the gauge field represents a pullback of the connection form. It follows that the gauge function $\rho_\alpha(g,p)$ of Eq. (3.3) may not necessarily be reduced to the identity element by a gauge transformation. The choice made in Refs. 14 and 15 corresponds to a gauge function which is *constant* on the manifold, taking as value the $\mathrm{SU}(2)_R$ component of any SO(4) element. The resulting form is therefore necessarily an SU(2)-invariant scalar multiple of the left-invariant Maurer–Cartan form on S^3. It should be mentioned that such a choice for ρ_α is not the only one permissible. Left invariance could, for instance, be replaced by (inverse) right invariance. However, this merely amounts to making a gauge transformation with the map $v : S^3 \to \mathrm{SU}(2)$ defined by Eq. (2.1) as transition function. An inequivalent invariance criterion results from choosing the value of ρ_α as the identity element; that is, considering gauge fields which are strictly invariant as forms on the manifold, not only up to a gauge transformation. However, the resulting forms are in $T^*(S^1)$ and hence closed, giving a vanishing field.

Up to gauge equivalence, these are actually the only possibilities. This follows from the results of Sec. 3 applied to the inclusion $\mathrm{SU}(2)_L \to \mathrm{SO}(4)$, which allow us to assume that $\rho(g,p) = e$ for $g \in \mathrm{SU}(2)_L$. Furthermore, since $\mathrm{SU}(2)_L$ and $\mathrm{SU}(2)_R$ commute, it follows from the composition rule (3.5) that $\rho(g,p)$ is independent of p for $g \in \mathrm{SU}(2)_R$ and therefore defines a homomorphism,

$$\rho : \mathrm{SU}(2)_R \to \mathrm{SU}(2) \tag{6.1}$$

into the gauge group. Since SU(2) is simple, this may only be an inner automorphism or the constant map onto the identity element. The former case is gauge equivalent to $\rho(g) = g$ for $g \in \mathrm{SU}(2)_R$, the choice of Refs. 14 and 15.

If one wishes to interpret the SO(4) invariant solutions in $(S^1 \times S^3)/\mathbb{Z}_2$, rather than $\mathbb{R} \times S^3$ (which is necessary to give meaning to the Chern class number), then suitable boundary conditions must be satisfied. In the notation of Sec. 3, we must have, in a neighborhood of the point $\psi = 2\pi$,

$$\omega_{(1+t,p)} = \mathrm{Ad} f_t^{-1} \omega_{(t,p)} + f_t^{-1} df_t, \qquad -\epsilon < t < \epsilon, \quad t \equiv \psi/2\pi \tag{6.2}$$

for sufficiently small ϵ, where f_t is a homotopy of maps $f_t : S^3 \to \mathrm{SU}(2)$ (which may be chosen as constant in t) belonging to the same class as the map $f : S^3 \to \mathrm{SU}(2)$ defining the bundle. Since, as shown in Proposition 3.5, the existence of an SU(2)-invariant connection implies that $\deg f$ vanishes, f_t may be chosen as the constant map onto the identity element, in which case (6.2) simply becomes the condition that ω be periodic,

$$\omega_{(1+t,p)} = \omega_{(t,p)}. \tag{6.2'}$$

Over the space $(S^1 \times S^3)/\mathbb{Z}_2$, moreover, the period for SU(2) invariant forms must be π rather than 2π. This severely restricts the permissible SO(4)-invariant solutions. In fact, the only such case for which the components in the Maurer–Cartan basis are ψ dependent is the one studied by Rebbi,[16]

$$\omega = \frac{1}{2} \frac{\exp[\pm i(\psi - \psi_0)]}{\cos(\psi - \psi_0)} \omega_L, \tag{6.3}$$

which arises as a degenerate case of the solutions of Refs. 14 and 15 analytically continued and is also a limiting case for each of the solutions (5.11)–(5.15).

Naturally, the constant $\mathrm{SO}(4) \times \mathrm{SO}(2)$ invariant solutions of de Alfaro *et al.*[11] also satisfy the periodicity require-

ment and therefore are well defined on the compact manifold.

7. FURTHER DEVELOPMENTS

In this paper we have been concerned with the characterization of gauge fields invariant under transformation groups and in particular the solutions to the Yang–Mills equations which are invariant under SU(2) [and $U(1)\times SU(2)$] translations. In order to obtain solutions with nonzero class number (and hence, in the self-dual case, nonvanishing action integral), it is necessary, in view of Proposition 3.5 to abandon this invariance requirement. An alternative may be to consider solutions which are SO(3) invariant, a case which has been shown by Witten[8] to lead in the Euclidean case to multi-instanton solutions of any class number.

Further developments relating to the present work also suggest themselves; for example: (i) extension of the methods of this paper to other gauge groups; (ii) a study, under the same invariance conditions, of the combined system of gauge fields coupled to scalars or spinors; (iii) an analysis of the present solutions pulled back to Minkowski space or continued into the Euclidean domain; (iv) determination of semiclassical expansions about these solutions for quantum amplitudes. We plan to address ourselves to these and related questions in future articles.

ACKNOWLEDGMENTS

The authors would like to thank M. Perroud and P. Winternitz for numerous helpful discussions and comments. A National Research Council of Canada scholarship is gratefully acknowledge by one of us (L.V.).

[1]A. Trautman, Rep. Math. Phys. **1**, 29 (1970); see also "Infinitesimal Connections in Physics," Proc. Internat. Symposium on New Mathematical Methods in Physics, edited by K. Bleuler and A. Reetz, Bonn, 1973.
[2]R. Hermann, *Vector Bundles in Mathematical Physics* (Benjamin, New York, 1970), Vol. 1.
[3]J.P. Harnad and R.B. Pettitt, J. Math. Phys. **17**, 1827 (1976).
[4]W. Drechsler and M.E. Mayer, *Fiber Bundle Techniques in Gauge Theories*, Vol. 67 of Lecture Notes in Physics (Springer-Verlag, New York, 1977).
[5]M.F. Atiyah and R.S. Ward, Commun. Math. Phys. **55**, 117 (1977); M.F. Atiyah, N.J. Hitchin, and I.M. Singer, Proc. Nat. Acad. Sci. U.S. **74**, 2662 (1977).
[6]A.M. Polyakov, Phys. Lett. B **59**, 82 (1975); A. Belavin, A. Polyakov, A. Schwartz, and Y. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[7]R. Jackiw and C. Rebbi, Phys. Rev. D **14**, 517 (1976).
[8]E. Witten, Phys. Rev. Lett. **38**, 121 (1977).
[9]R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
[10]E.F. Corrigan, D.B. Fairlie, P. Goddard, and R.G. Yates, Commun. Math. Phys. **58**, 223 (1978).
[11]V. de Alfaro, S. Fubini, and G. Furlan, Phys. Lett. B **65**, 163 (1976).
[12]J. Cervero, L. Jacobs, and C. Nohl, Phys. Lett. B **69**, 351 (1977).
[13]V. de Alfaro, S. Fubini, and G. Furlan, Phys. Lett. B **72**, 203 (1977).
[14]M. Lüscher, Phys. Lett. B **70**, 321 (1977).
[15]B. Schechter, Phys. Rev. D **16**, 3015 (1977).
[16]C. Rebbi, Phys. Rev. D **17**, 483 (1978).
[17]R. Jackiw, C. Nohl, and C. Rebbi, "Classical and Semi-Classical Solutions of Yang–Mills Theory," MIT preprint CTP 675 (1977).
[18]J. Harnad and L. Vinet, Phys. Lett. B **76**, 589 (1978).
[19]P.S. Howe and R.W. Tucker, Nucl. Phys. B **69**, 73 (1978); R.W. Tucker and W.J. Zakrzewski, Nucl. Phys. B **143**, 424 (1978).
[20]G. Mack and A. Salam, Ann. Phys. **53**, 174 (1969).
[21]I.E. Segal, *Mathematical Cosmology and Extragalactic Astronomy* (Academic, New York, 1976).
[22]T.H. Go, H.A. Kastrup, and D. Mayer, Rep. Math. Phys. **6**, 395 (1976).
[23]J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).
[24]P.G. Bergmann and E.J. Flaherty Jr., J. Math. Phys. **19**, 212 (1978).
[25]J. Dieudonné, *Eléments d'Analyse* (Gauthier-Villars, Paris, 1971), t. IV [English translation, *Treatise on analysis* (Academic, New York, 1970), Vol. IV].
[26]G. 't Hooft, Phys. Rev. D **14**, 3432 (1976).
[27]D. Husemoller, *Fibre Bundles* (Springer-Verlag, New York, 1975), 2nd ed.
[28]M.W. Hirsch, *Differential Topology* (Springer-Verlag, New York, 1976).
[29]V.W. Guillemin and A. Pollack, *Differential Topology* (Prentice Hall, Englewood Cliffs, N.J., 1974).
[30]P. du Val, *Elliptic Functions and Elliptic Curves* (Cambridge U.P., Cambridge, 1973).

The spectral properties of many-electron atomic Hamiltonians and the method of configuration interaction. II. Compactness proof associated with an infinite system of linear equations for two-electron atoms

M. H. Choudhury

Department of Applied Mathematics, The University, Hull, England
(Received 21 June 1977; revised manuscript received 28 December 1977)

The Schrödinger equation for a two-electron atomic system is reduced to an infinite system of linear equations. The linear operator defined by this system of equations is then shown to be compact in a region of the complex energy plane which excludes the various bound state and multiparticle scattering cuts (i.e., the eseential spectrum of the Hamiltonian of the two-electron atomic system). This permits one to truncate the infinite system of equations with the assurance that the N energy eigenvalues obtained from the $N\times N$ truncated system will uniformly approximate the lowest N energy eigenvalues of the original infinite system.

1. INTRODUCTION

In the previous paper it was shown that the N energy eigenvalues and eigenvectors obtained from the truncated system of equations

$$\sum_{m=1}^{N} \langle\phi_n|H|\phi_m\rangle\langle\phi_m|\,|\Psi\rangle = E\,\langle\phi_n|\Psi\rangle$$

will not converge uniformly (as the size of the truncated matrix is increased), to the lowest N energy eigenvalues and eigenvectors of the original infinite system

$$\sum_{m=1}^{\infty} \langle\phi_n|H|\phi_m\rangle\langle\phi_m|\,|\Psi\rangle = E\,\langle\phi_n|\Psi\rangle.$$

In this paper, an infinite system of linear equations will be derived from the schrödinger equation of a two-electron atomic system which will precisely accomplish this. The basic idea behind this formulation is to derive an infinite system of linear equations which defines a compact linear operator in a suitable region of the complex energy plane. This ensures that the N energy eigenvalues obtained from the $N\times N$ truncated system of equations will uniformly approximate the lowest N energy eigenvalues of the original infinite system[1] of this formulation.

In Sec. 2 we derive from the schrödinger equation of a two-electron atomic system an infinite system of linear equations. In Sec. 3 we show that the linear operator defined by this infinite system of equations is compact in a region of the complex energy plane which excludes the various multiparticle or continuum cuts. These are

(i) the bound state scattering cuts starting at the Hydrogenic bound state energy E_n, $n = 1,2,...$, and extending to $+\infty$.

(ii) the multiparticle cut starting at $E = 0$ and extending to $+\infty$.

The region of compactness, in our case, also excludes the "spurious" point $\{E_{n_1} + E_{n_2}\}$ $n_1,n_2 = 1,2,....$ We call these points spurious, as they do not belong to the essential spectrum of H.

Finally, Sec. 4 states our conclusions.

2. REDUCTION OF THE SCHRÖDINGER EQUATION TO AN INFINITE SYSTEM OF LINEAR EQUATIONS

The schrödinger equation for a two-electron atomic system can be written

$$H\,|\Psi\rangle = (H_{01} + H_{02} + 1/r_{12})|\Psi\rangle + E\,|\Psi\rangle, \qquad (2.1)$$

where $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$, $\mathbf{r}_1$ and $\mathbf{r}_2$ being the position operators of electrons 1 and 2, respectively, and H_{01} and H_{02} are the hydrogenic Hamiltonian operators associated with electrons 1 and 2, which in the coordinate representation are given by

$$H_{0i} = -\tfrac{1}{2}\nabla_i^2 - \frac{Z}{r_i}, \quad r_i = |\mathbf{r}_i|, \quad i = 1,2, \qquad (2.2)$$

with

$$H_{0i}|\mathbf{n}_i\rangle = E_{n_i}|\mathbf{n}_i\rangle, \quad i = 1,2, \qquad (2.3)$$

and

$$H_{0i}|\mathbf{k}_{Hi}\rangle = E_{k_i}|\mathbf{k}_{Hi}\rangle, \quad i = 1,2. \qquad (2.4)$$

Here $|\mathbf{n}\rangle = |nlm\rangle$ are the bound states of the hydrogenic atom with corresponding energy eigenvalues $E_n = -Z^2/(2n^2)$, and $|\mathbf{K}_H\rangle = |k_H lm\rangle$ the continuum states, with corresponding energies E_k. In the coordinate representation the bound and continuum states can be written as[2]

$$\langle\mathbf{r}|\mathbf{n}\rangle = \langle\mathbf{r}|nlm\rangle = R_{nl}(r)Y_{lm}(\theta,\phi), \qquad (2.5)$$

and

$$\langle\mathbf{r}|\mathbf{k}_H\rangle = \langle\mathbf{r}|k_H lm\rangle = R_l(kr)Y_{lm}(\theta,\phi), \qquad (2.6)$$

where $Y_{lm}(\theta,\phi)$ are the usual spherical Harmonics, and $R_{nl}(r)$, $R_l(kl)$ are the radial bound and continuum state eigenfunctions, respectively.

Let $\mathscr{H}_1$ and $\mathscr{H}_2$ be the space of states associated with electrons 1 and 2, respectively. The resolutions of the identity in these two spaces are given by

$$\sum_{n_i=1}^{\infty}\sum_{l_i=0}^{n_i-1}\sum_{m_i=-l_i}^{+l_i} |n_i l_i m_i\rangle\langle n_i l_i m_i|$$

 0022-2488/79/050943-10$01.00

$$+\sum_{l_i=0}^{\infty}\sum_{m_i=-l_i}^{+l_i}\int_0^{\infty} k_i^2\,dk_i\,|k_{Hi}l_im_i\rangle\langle k_{Hi}l_im_i| = I_{Hi},$$
$$i=1,2, \tag{2.7}$$

which can be written

$$\sum_{\mathbf{n}_i}|\mathbf{n}_i\rangle\langle\mathbf{n}_i| + \left(\int\sum_{\mathbf{K}_{Hi}}\right)|\mathbf{k}_{Hi}\rangle\langle\mathbf{k}_{Hi}| = I_{Hi}, \quad i=1,2, \tag{2.7a}$$

or in a still more abridged form as

$$\left(\int\sum_{\nu_i}\right)|\nu_i\rangle\langle\nu_i| = I_{Hi}, \quad i=1,2. \tag{2.7b}$$

Here the sign $(\int\Sigma)$ is used to mean a summation over the discrete quantum numbers and an integration over the continuous set as appropriate. In the product space $\mathscr{H}=\mathscr{H}_1\otimes\mathscr{H}_2$, the identity operator is given by

$$I_H = I_{H1}\otimes I_{H2}\cdots \tag{2.8}$$

so that the resolution of the identity I_H using (2.7a) is

$$\sum_{\mathbf{n}_1,\mathbf{n}_2}|\mathbf{n}_1,\mathbf{n}_2\rangle\langle\mathbf{n}_1,\mathbf{n}_2| + \sum_{\mathbf{n}_1}\left(\int\sum_{\mathbf{k}_{H2}}\right)|\mathbf{n}_1,\mathbf{k}_{H2}\rangle\langle\mathbf{n}_1,\mathbf{k}_{H2}|$$
$$+\left(\int\sum_{\mathbf{k}_{H1}}\right)\sum_{\mathbf{n}_2}|\mathbf{k}_{H1},\mathbf{n}_2\rangle\langle\mathbf{k}_{H1},\mathbf{n}_2|$$
$$+\left(\int\sum_{\mathbf{k}_{H1}}\right)\left(\int\sum_{\mathbf{k}_{H2}}\right)|\mathbf{k}_{H1},\mathbf{k}_{H2}\rangle\langle\mathbf{k}_{H1},\mathbf{k}_{H2}| = I_H, \tag{2.9}$$

which can also be written in the abridged form

$$\left(\int\sum_{\nu_1}\right)\left(\int\sum_{\nu_2}\right)|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2| = I_H. \tag{2.9a}$$

It is not difficult to write down a resolution of the identity involving only symmetrized or antisymmetrized (spatial) states. The use of such states in our investigation would only introduce tedious complications without adding anything qualitatively new or different to the results regarding the spectrum of the Hamiltonian operator for two-electron atomic systems.

Equation (2.1) can be inverted and written in the form

$$|\Psi\rangle = (E-H_{01}-H_{02})^{-1}\frac{1}{r_{12}}|\Psi\rangle. \tag{2.10}$$

Let $\{|\alpha_1,\alpha_2\rangle\}$ denote a (complete) discrete basis belonging to the domain of $\mathscr{H}$. The resolution of the identity with respect to this basis is

$$\sum_{\alpha_1,\alpha_2}|\alpha_1,\alpha_2\rangle\langle\alpha_1,\alpha_2| = I_H. \tag{2.11}$$

Taking the inner product on both sides of (2.10) with respect to $\langle\alpha_1,\alpha_2|$ and using (2.11) we obtain the infinite system of equations

$$\langle\alpha_1,\alpha_2|\Psi\rangle = \sum_{\alpha_1',\alpha_2'}\langle\alpha_1,\alpha_2|(E-H_{01}-H_{02})^{-1}$$
$$\times\frac{1}{r_{12}}|\alpha_1',\alpha_2'\rangle\langle\alpha_1',\alpha_2'|\Psi\rangle. \tag{2.12}$$

In the next section we shall show that the operator $(E-H_{01}-H_{02})^{-1}(1/r_{12})$ is compact in a suitable region of the complex energy plane excluding the essential spectrum of $\mathscr{H}$. This permits one to truncate (2.12) with the assurance that the N energy eigenvalues obtained from the $N\times N$ truncated matrix uniformly approximates the lowest N energy eigenvalues of the original infinite matrix.

3. PROOF OF COMPACTNESS

In this section it will be shown that the operator

$$K(E) = (E-H_{01}-H_{02})^{-1}\frac{1}{r_{12}}. \tag{3.1}$$

defined by the system of equations (2.12) is compact in a region D_E (to be specified later) of the complex energy plane. Let us first note that the domain of $1/r_{12}$ is a subset D_V of $\mathbb{R}^3\times\mathbb{R}^3=\mathbb{R}^6$ defined by

$$D_V = \{(\mathbf{r}_1,\mathbf{r}_2):(\mathbf{r}_1,\mathbf{r}_2)\in\mathbb{R}^6, |\mathbf{r}_1-\mathbf{r}_2|\geqslant\epsilon\}. \tag{3.2}$$

We have

$$V(|\mathbf{r}_1-\mathbf{r}_2|) = \frac{1}{r_{12}} = \sum_{k=0}^{\infty}F_k(r_1,r_2)P_k(\cos\theta_{12}), \tag{3.3}$$

where θ_{12} is the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$ and

$$F_k(r_1,r_2) = \frac{1}{r_1}\left(\frac{r_2}{r_1}\right)^k\theta(r_1-r_2) + \frac{1}{r_2}\left(\frac{r_1}{r_2}\right)^k\theta(r_1-r_2), \tag{3.4}$$

where $\theta(r)$ is the step function defined by

$$\theta(r) = \begin{cases}1, & r>0,\\ 0, & r<0.\end{cases}$$

Next we define the sequence of potentials

$$V_n = \sum_{k=0}^{n}F_k(r_1,r_2)P_k(\cos\theta_{12}). \tag{3.5}$$

If now we can show that

(i) the sequence of operators

$$K_n(E) = (E-H_{01}-H_{02})^{-1}V_n \tag{3.6}$$

are compact whenever $E\in D_E$ and

(ii) $\|K-K_n\|\to 0$ as $n\to\infty$

in the uniform topology of the operator norm, then the operator $K(E)$ is compact for $E\in D_E$. Conditions (i) and (ii) stated above for the compactness or $K(E)$ are just the statement of the results[3] that if an operator K can be approximated uniformly in the norm by a sequence of compact operators, then the operator itself is compact.

To demonstrate (i), it is sufficient to show that for some positive integer m, $[K_n(E)]^m$ belongs to the class of Hilbert–Schmidt operators[4] for $E\in D_E$. In fact, we shall show that $[K_n(E)]^2$ belongs to the Hilbert–Schmidt class for $E\in D_E$.

To show that $[K_n(E)]^2$ is of the Hilbert–Schmidt type whenever $E\in D_E$, we must show that

$$\int\sum_{\nu_1,\,1,2}\left(\int\sum_{\nu_2}\right)|\langle\nu_1,\nu_2|[K_n(E)]^2|\nu_1',\nu_2'\rangle|^2<\infty,$$
$$E\in D_E. \tag{3.7}$$

Using (2.9) we have

$$\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|\,[K_n(E)]^2\,|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle$$

$$= \sum_{\mathbf{n}_1}\sum_{\mathbf{n}_2} \langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|(E-H_{01}-H_{02})^{-1}V_n|\mathbf{n}_1,\mathbf{n}_2\rangle$$
$$\times(E-E_{n_1}-E_{n_2})^{-1}\langle \mathbf{n}_1,\mathbf{n}_2|\,V_n\,|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle$$
$$+\sum_{\mathbf{n}_1}\left(\int\sum_{\mathbf{k}_{H2}}\right)\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|(E-H_{01}-H_{02})^{-1}V_n|\mathbf{n}_1,\mathbf{k}_{H2}\rangle$$
$$\times(E-E_{n_1}-E_{k_2})^{-1}\langle \mathbf{n}_1,\mathbf{k}_{H2}|\,V_n\,|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle$$
$$+\left(\int\sum_{\mathbf{k}_{H1}}\right)\sum_{\mathbf{n}_2}\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|(E-H_{01}-H_{02})^{-1}V_n|\mathbf{k}_{H1},\mathbf{n}_2\rangle$$
$$\times(E-E_{k_1}-E_{n_2})^{-1}\langle \mathbf{k}_{H1},\mathbf{n}_2|\,V_n\,|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle$$
$$+\left(\int\sum_{\mathbf{k}_{H1}}\right)\left(\int\sum_{\mathbf{k}_{H2}}\right)$$
$$\times\,\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|(E-H_{01}-H_{02})^{-1}V_n|\mathbf{k}_{H1},\mathbf{k}_{H2}\rangle$$
$$\times(E-E_{k_1}-E_{k_2})^{-1}\langle \mathbf{k}_{H1},\mathbf{k}_{H2}|\,V_n\,|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle. \tag{3.8}$$

We consider each of the four terms on the right-hand side of (3.8) and make the following observations:

(i) $|(E-E_{n_1}-E_{n_2})^{-1}|\leqslant$const (3.9)

for all values of E in the complex energy plane for which $E\neq E_{n_1}+E_{n_2}$ $(n_1,n_2=1,2,3,....)$ We denote this domain by D_{E1}.

(ii) $|(E-E_{n_1}-E_{k_2})^{-1}|\leqslant$const (3.10)

in a domain D_{E2} which consists of all points in the complex energy plane excluding those which make up the branch cuts starting at the points E_{n_1} $(n_1=1,2,\cdots)$ and extending to $+\infty$. Note that because of the integration over $\mathbf{k}_{H2}$, the pole at E_{n_1} is converted into a cut along the line $[-|E_{n_1}|,\infty]$. These are the so-called bound state scattering cuts.[5]

(iii) same as in (ii) with

$|(E-E_{k_1}-E_{n_2})^{-1}|\leqslant$const. (3.11)

(iv) $|(E-E_{k_1}-E_{k_2})^{-1}|\leqslant$const (3.12)

in a domain D_{E4} which consists of all the points in the complex energy plane excluding those which make up the branch cut starting $E=0$ and extending to $+\infty$.

Denoting by D_E the intersections of the domain D_{Ei} $(i=1,2,3,4)$, that is,

$$D_E=D_{E1}\cap D_{E2}\cap D_{E3}\cap D_{E4} \tag{3.13}$$

it is clear that the inequalities (3.9)–(3.12) are simultaneously valid in the domain D_E. We find it convenient to express this fact in the abridged form

$$|(E-E_{\nu_1}-E_{\nu_2})^{-1}|\leqslant\text{const},\quad E\in D_E. \tag{3.14}$$

Using the result (A1) (proved in Appendix A) with

$$\langle\Phi_1|=\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|(E-H_{01}-H_{02})^{-1}V_n$$

and

$$|\Phi_2\rangle=V_n|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle,$$

we find that

$$|\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|\,[K_n(E)]^2\,|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle|$$
$$\leqslant\text{const}\,|\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|(E-E_{\nu_1}-E_{\nu_2})^{-1}V_n^2|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle|.$$

Hence the compactness condition (3.7) remains satisfied if

$$\left(\int\sum_{\nu_1}\right)\left(\int\sum_{\nu_2}\right)|\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|(E-E_{\nu_1}-E_{\nu_2})^{-1}V_n^2|\boldsymbol{\nu}_1',\boldsymbol{\nu}_2'\rangle|^2$$
$$=\left(\int\sum_{\nu_1}\right)|(E-E_{\nu_1}-E_{\nu_2})^{-1}|^2$$
$$\times\langle \boldsymbol{\nu}_1,\boldsymbol{\nu}_2|\,[V_n]^4\,|\boldsymbol{\nu}_1,\boldsymbol{\nu}_2\rangle<\infty. \tag{3.15}$$

Noting (2.9), we can write (3.15) as the sum of four terms. Further, using the addition theorem of spherical harmonics

$$\sum_{m=-l}^{+l}\overline{Y_{lm}(\theta,\phi)}\,Y_{lm}(\theta,\phi)=\frac{(2l+1)}{4\pi}$$

to sum over the m_i, $i=1,2$, one obtains

$$\sum_{n_i,l_i}|(E-E_{n_1}-E_{n_2})^{-1}|^2(4\pi)^{-2}(2l_1+1)(2l_2+1)$$
$$\times\int d^3r_1\int d^3r_2[R_{n_1l_1}(r_1)R_{n_2l_2}(r_2)]^2[V_n]^4$$
$$+\sum_{n_1}\sum_{l_1=0}^{n_1-1}\sum_{l_2=0}^{\infty}\int_0^\infty k_2^2\,dk_2\,\frac{(2l_1+1)(2l_2+1)}{(4\pi)^2}$$
$$\times\frac{\int d^3r_1\int d^3r_2[R_{n_1l_1}(r_1)R_{l_2}(k_2r_2)]^2[V_n]^4}{|E-E_{n_1}-\frac{1}{2}k_2^2|^2}$$
$$+\text{(Similar Term)}$$
$$+\sum_{l=0}^{\infty}\int_0^\infty k_1^2\,dk_1\int_0^\infty dk_2\,\frac{(2l_1+1)(2l_2+1)}{(4\pi)^2}$$
$$\frac{\int d^3r_1\int d^3r_2[R_{l_1}(k_1r_1)R_{l_2}(k_2r_2)]^2[V_n]^4}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}<\infty,$$
$$E\in D_E. \tag{3.16}$$

Denote the four terms on the left-hand side of this inequality by I_i, $i=1,2,3,4$, respectively. For compactness, it is therefore sufficient to show that

$$I_i<\infty,\quad i=1,2,3,4\ \text{ for }\ E\in D_E. \tag{3.17}$$

Using the inequality

$$\left(\sum_{k=0}^{n}C_k\right)^2\leqslant\text{const}\sum_{k=0}^{n}C_k^2,\quad C_k\in\mathbb{R}$$

twice, one obtains

$$\left[V_n^4\leqslant\text{const}\sum_{k=0}^{n}[F_k(r_1,r_2)]^4\right][P_k(\cos\theta_{12})]^4.$$

Further, the angular integrations on the first term I_1 of (3.16) yields a constant independent of n_i, l_i, so that

$$I_1\leqslant\text{const}\sum_{n_i,l_i}|(E-E_{n_1}-E_{n_2})^{-1}|^2(2l_1+1)(2l_2+1)$$

$$\times\Big[\sum_{k=0}^{n}\int_0^\infty r_1^2\,dr_1\int_0^\infty r_2^2\,dr_2[R_{n_1l_1}(r_1)R_{n_2l_2}(r_2)]^2$$

$$\times[F_k(r_1,r_2)]^4\Big],\quad E\in D_E. \tag{3.18}$$

Substitution for F_k from (3.4) in the radial integral in (3.18) yields

$$\int_0^\infty r_1^2\,dr_1[R_{n_1l_1}(r_1)]^2\frac{1}{r_1^{4k+4}}\int_0^{r_1} r_2^2\,dr_2[R_{n_2l_2}(r_2)]^2r_2^{4k}$$
$$+\int_0^\infty r_2^2\,dr_2[R_{n_2l_2}(r_2)]^2\frac{1}{r_2^{4k+4}}$$
$$\times\int_0^{r_2} r_1^2\,dr_1[R_{n_1l_1}(r_1)]^2r_1^{4k}$$
$$=\int_0^\infty dr_1\int_0^1 t^{4k}[R_{n_1l_1}(r_1)r_1]^2\frac{1}{r_1^3}[R_{n_2l_2}(r_1t)r_1t]^2$$
$$+\int_0^\infty dr_2\int_0^1 t^{4k}[R_{n_2l_2}(r_2)r_2]^2\frac{1}{r_2^3}[R_{n_1l_1}(r_2t)r_2t]^2, \tag{3.19}$$

where, to obtain the last step, the substitution $(r_2/r_1)=t$ was made in the first of the integrals on the left-hand side of (3.19) and the substitution $(r_1/r_2)=t$ in the second of the integrals. Noting that the two integrals on the right-hand side of (3.19) are precisely of the same type, the inequality (3.18) can now be written

$$I_1\leqslant\text{const}\sum_{n_2l_2}|(E-E_{n_1}-E_{n_2})^{-1}|^2(2l_1+1)(2l_2+1)$$
$$\times\sum_{k=0}^{n}R^{(k)}(n_1l_1,n_2l_2),\quad E\in D_E, \tag{3.20}$$

where

$$R^{(k)}(n_1l_1,n_2l_2)=\int_0^\infty dr\int_0^1 t^{4k}\,dt\,[R_{n_1l_1}(r)r]^2$$
$$\times\frac{1}{r^3}[R_{n_2l_2}(rt)rt]^2. \tag{3.21}$$

Similarly, one obtains

$$I_2\leqslant\text{const}\sum_{n_1l_1}\sum_{l_2=0}^{\infty}(2l_1+1)(2l_2+1)$$
$$\times\int_0^\infty k_2^2\,dk_2\frac{[\Sigma_{k=0}^{n}R^{(k)}(n_1l_1,k_2l_2)]}{|E-E_{n_1}-\frac{1}{2}k_2^2|^2},\quad E\in D_E, \tag{3.22}$$

$$I_3\leqslant\text{const}\sum_{n_2l_2}\sum_{l_1=0}^{\infty}(2l_1+1)(2l_2+1)$$
$$\times\int_0^\infty k_1^2\,dk_1\frac{[\Sigma_{k=0}^{n}R^{(k)}(k_1l_1,n_2l_2)]}{|E-\frac{1}{2}k_1^2-E_{n_2}|^2},\quad E\in D_E, \tag{3.23}$$

$$I_4\leqslant\text{const}\sum_{l_1=0}^{\infty}(2l_1+1)(2l_2+1)$$
$$\times\int_0^\infty k_1^2\,dk_1\int_0^\infty k_2^2\,dk_2\frac{[\Sigma_{k=0}^{n}R^{(k)}(k_1l_1,k_2l_2)]}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2},$$
$$E\in D_E, \tag{3.24}$$

where

$$R^{(k)}(n_1l_1,k_2l_2)=\int_0^\infty dr\int_0^1 t^{4k}\,dt\,[R_{n_1l_1}(r)r]^2$$
$$\times\frac{1}{r^3}|R_{l_2}(k_2rt)rt|^2, \tag{3.25}$$

a similar term for $R^{(k)}(k_1l_1,n_2l_2)$, and

$$R^{(k)}(k_1l_1,k_2l_2)=\int_0^\infty dr\int_0^1 dtt^{4k}|R_{l_1}(k_1r)r|^2$$
$$\times\frac{1}{r^3}|R_{l_2}(k_2rt)rt|^2. \tag{3.26}$$

Considering I_1 first, the inequality (3.20) can be written

$$I_1\leqslant\text{const}\Big[\sum_{n_1=2}^{\infty}\sum_{l_1=1}^{n_1-1}|(E-E_{n_1}-E_{n_2})^{-1}|^2(2l_1+1)(2l_2+1)$$
$$\times\sum_{k=0}^{n}R^{(k)}(n_1l_1,n_2l_2)$$
$$+\sum_{n_1=1}^{\infty}\sum_{n_2=2}^{\infty}\sum_{l_2=1}^{n_2-1}|(E-E_{n_1}-E_{n_2})^{-1}|^2(2l_2+1)$$
$$\times\sum_{k=0}^{n}R^{(k)}(n_10,n_2l_2)$$
$$+\sum_{n_1=2}^{\infty}\sum_{l_1=1}^{n_1-1}\sum_{n_2=1}^{\infty}|(E-E_{n_1}-E_{n_2})^{-1}|^2(2l_1+1)$$
$$\times\sum_{k=0}^{n}k^{(k)}(n_1l_1,n_20)$$
$$+\sum_{n_1=1}^{\infty}|(E-E_{n_1}-E_{n_2})^{-1}|^2\sum_{k=0}^{n}R^{(k)}(n_10,n_20)\Big]<\infty,$$
$$E\in D_E. \tag{3.27}$$

Use of the inequality (B1) (see Appendix B) on (3.21) yields

$$R^{(k)}(n_1l_1,n_2l_2)$$
$$\leqslant\frac{\text{const}}{(n_1n_2)^{1/2}}\int_0^1 t^{4k}\,dt\int_0^\infty dr\left|\frac{R_{n_1l_1}(r)r}{r^{3/2}}\cdot\frac{R_{n_2l_2}(rt)rt}{r^{3/2}}\right|.$$

Use of the Cauchy–Schwarz inequality in the form

$$\int_0^\infty|f(r)g(r)|\,dr\leqslant\Big(\int_0^\infty|f(r)|^2\,dr\Big)^{1/2}\Big(\int_0^\infty|g(r)|^2\,dr\Big)^{1/2} \tag{3.28}$$

then gives

$$R^{(k)}(n_1l_1,n_2l_2)$$
$$\leqslant\frac{\text{const}}{(n_1n_2)^{1/2}}\int_0^1 t^{4k}\,dt\Big(\int_0^\infty dr\frac{[R_{n_1l_1}(r)r]^2}{r^3}\Big)^{1/2}$$

$$\times\left(\cdot\int_0^\infty dr\,\frac{[R_{n_2l_2}(rt)rt]}{r^3}\right)^{1/2}. \tag{3.29}$$

Noting that[6]

$$\int_0^\infty dr\,\frac{[R_{n_1l_1}(r)r]^2}{r^3} = \frac{Z^3}{n^3(l+1)(l+\frac{1}{2})l}, \tag{3.30}$$

the substitution $\rho = rt$ yields

$$\int_0^\infty dr\,\frac{[R_{n_1l_1}(r)r]^2}{r^3} = t^2\int_0^\infty d\rho\,\frac{[R_{nl}(\rho)\rho]^2}{\rho^3}$$

$$= \frac{t^2Z^3}{n^3(l+1)(l+\frac{1}{2})l}, \tag{3.31}$$

so that (3.29) reduces to

$$R^{(k)}(n_1l_1,n_2l_2)\leqslant \text{const}(n_1n_2)^{-2}(l_1l_2)^{-3/2}. \tag{3.32}$$

Again, use of the inequalities (B1) and (B2) in (3.21) with $l_1 = 0$ gives

$$R^{(k)}(n_10,n_2l_2)$$

$$\leqslant \frac{\text{const}}{n_1^{3/2}n_2^{1/2}}\int_0^1 t^{4k}\,dt\int_0^\infty dr\left|\frac{R_{n_10}(r)r}{r^{1/2}}\cdot\frac{R_{n_2l_2}(rt)rt}{r^{3/2}}\right|.$$

Use of the Cauchy–Schwarz inequality (3.28) yields

$$R^{(k)}(n_10,n_2l_2)$$

$$\leqslant \frac{\text{const}}{n_1^{3/2}n_2^{1/2}}\int_0^1 t^{4k}\,dt\left(\int_0^\infty dr\,\frac{[R_{n_10}(r)r]^2}{r}\right)^{1/2}$$

$$\times\left(\int_0^\infty dr\,\frac{[R_{n_2l_2}(rt)rt]^2}{r^3}\right)^{1/2}. \tag{3.33}$$

Use of (3.31) and the formula[7]

$$\int_0^\infty dr\,\frac{[R_{nl}(r)r]^2}{r} = \frac{Z}{n^2} \tag{3.34}$$

reduces (3.33) to

$$R^{(k)}(n_10,n_2l_2)\leqslant \text{const}n_1^{-5/2}n_2^{-2}l_2^{-3/2}. \tag{3.35}$$

Similarly one finds that

$$R^{(k)}(n_1l_1,n_20)\leqslant \text{const}n_1^{-2}l_1^{-3/2}n_2^{-5/2}. \tag{3.36}$$

and

$$R^{(k)}(n_10,n_20)\leqslant \text{const}n_1^{-5/2}n_2^{-5/2}. \tag{3.37}$$

Substituting the estimates (3.32), and (3.35)–(3.37) into (3.27) and using the fact that

$$\sum_{l=1}^{n} l^{-1/2}\leqslant \text{const}n^{1/2}, \tag{3.38}$$

yields

$$I_1\leqslant \text{const}\Bigg[\sum_{n_1=2}^\infty (n_1n_2)^{-3/2} + \sum_{n_1=1}^\infty n_1^{-5/2}\sum_{n_2=2}^\infty n_2^{-3/2}$$

$$+\sum_{n_1=2}^\infty n_1^{-3/2}\sum_{n_2=1}^\infty n_2^{-5/2} + \sum_{n_i=1}^\infty (n_1n_2)^{-5/2}\Bigg] < \infty, \tag{3.39}$$

for $E\in D_E$.

It is clear that had we not split up (3.20) into four parts as in (3.27), the summation over l_i, $i = 1, 2$, which starts from zero would have caused difficulties due to the use which we have made of the formulae (3.30) and (3.31). We next consider I_2 and I_3. These are exactly similar and can be obtained, one from the other, by an interchange of the subscripts 1 and 2. It will therefore be sufficient to consider I_2.

The consideration which prompted us to split (3.2) into four parts now leads us to split (3.22) into two parts and write it in the form

$$I_2\leqslant \text{const}\Bigg[\sum_{n_1=2}^\infty\sum_{l_1=1}^{n_1-1}\sum_{l_2=0}^\infty (2l_1+1)(2l_2+1)$$

$$\times\int_0^\infty k_2^2\,dk_2\,\frac{[\Sigma_{k=0}^n R^{(k)}(n_1l_1,k_2l_2)]}{|E-E_{n_1}-\frac{1}{2}k_2^2|^2}$$

$$+\sum_{n_1=1}^\infty\sum_{l_2=0}^\infty (2l_2+1)\int_0^\infty k_2^2\,dk_2$$

$$\times\frac{[\Sigma_{k=0}^n R^{(k)}(n_10,k_2l_2)]}{|E-E_{n_1}-\frac{1}{2}k_2^2|^2}\Bigg]. \tag{3.40}$$

The inequality (B1) applied to (3.25) followed by a use of the Cauchy–Schwarz inequality yields

$$R^{(k)}(n_1l_1,k_2l_2)$$

$$\leqslant \frac{\text{const}}{n_1^{1/2}}\int_0^1 t^{4k}\,dt\left(\int_0^\infty dr\,\frac{[R_{n_1l_1}(r)r]^2}{r^3}\right)^{1/2}$$

$$\times\left(\int_0^\infty dr\,\frac{|R_{l_2}(k_2rt)rt|^4}{r^3}\right)^{1/2}.$$

Using (3.30), this can be written in the form

$$R^{(k)}(n_1l_1,k_2l_2)\leqslant \frac{\text{const}}{n_1^2l_1^{3/2}}\left[\int_0^\infty |R_{l_2}(k_2r)|^4 r\,dr\right]^{1/2}. \tag{3.41}$$

Note that, since the radial function $R_{l_2}(k_2r)$ is finite at $r = 0$ and[8]

$$R_{l_2}(k_2r)\underset{r\to\infty}{\sim}(2/\pi)^{1/2}(k_2r)^{-1}$$

$$\times\sin\left(k_2r-\tfrac{1}{2}l_2\pi+\eta_{l_2}+\frac{Z}{k_2}\log 2k_2r\right), \tag{3.42}$$

with

$$\eta_{l_2} = \arg\Gamma\left(l_2+1-i\frac{Z}{k_2}\right),$$

we have

$$\int_0^\infty r|R_{l_i}(k_2 r)^4|\,dr<\infty. \tag{3.43}$$

Use of the result (C1) (see Appendix C) in (3.41) yields

$$R^{(k)}(n_1 l_1,k_2 l_2) \leqslant \frac{\text{const}}{n_1^2 l_1^{3/2}}\left|R_{l_2}\left(\frac{k_2 u_0}{1-u_0}\right)\right|^2\left[\frac{u_0}{(1-u_0)^3}\right]^{1/2}, \quad u_0\in[0,1]. \tag{3.44}$$

Similarly, one finds that

$$R^{(k)}(n_1 0,k_2 l_2) \leqslant \frac{\text{cons}}{n_1^{5/2}}\left|R_{l_2}\left(\frac{k_2 u_0}{1-u_0}\right)\right|^2\left[\frac{u_0}{(1-u_0)^3}\right]^{1/2}, \quad u_0\in[0,1]. \tag{3.45}$$

The estimates (3.44) and (3.45) when substituted into (3.40) yield

$$I_2\leqslant\text{const}\left[\left(\sum_{n_1=2}^{\infty}\sum_{l_1=1}^{n_1-1} n_1^{-2} l_1^{-3/2}+\sum_{n_1=1}^{\infty} n_1^{-5/2}\right)\times\left(\sum_{l_2=0}^{\infty}(2l_2+1)\int_0^\infty k_2^2\,dk_2\left|R_{l_2}\left(\frac{k_2 u_0}{1-u_0}\right)\right|^2\left[\frac{u_0}{(1-u_0)^3}\right]^{1/2}\times\left|E-E_{n_1}-\tfrac{1}{2}k_2^2\right|^{-2}\right)\right]. \tag{3.46}$$

Note that $|R_{l_2}(k_2u_0/(1-u_0))|^2$ is bounded as a function of k_2 in the interval $[0,\infty)$. Let $k_{2m}\in[0,\infty)$ be the value of k_2 at which $|R_{l_2}(k_2u_0/(1-u_0))|^2$ attains its maximum value. We have

$$\int_0^\infty k_2^2\,dk_2\left|R_{l_2}\left(\frac{k_2u_0}{1-u_0}\right)\right|^2\left[\frac{u_0}{(1-u_0)^3}\right]^{1/2}|E-E_{n_1}-\tfrac{1}{2}k_2^2|^{-2} \leqslant\left|R_{l_2}\left(\frac{k_{2m}u_0}{1-u_0}\right)\right|^2\left[\frac{u_0}{(1-u_0)^3}\right]^{1/2}\int_0^\infty\frac{k_2^2\,dk_2}{|E-E_{n_1}-\tfrac{1}{2}k_2^2|^2} \leqslant\text{const}\left|R_{l_2}\left(\frac{k_{2m}u_0}{1-u_0}\right)\right|^2\left[\frac{u_0}{(1-u_0)^3}\right]^{1/2}, \quad \text{for } E\in D_E.$$

Using the above inequality and (3.38), we find that (3.46) reduces to

$$I_2\leqslant\text{const}\sum_{l_2=0}^{\infty}(2l_2+1)\left|R_{l_2}\left(\frac{k_{2m}u_0}{1-u_0}\right)\right|^2\times\left[\frac{u_0}{(1-u_0)^3}\right]^{1/2}<\infty,$$

where in the last step we have used the result (C3)

Finally we consider I_4. An application of the Cauchy–Schwarz inequality (3.28) to the expression (3.26) followed by a change of variable $\rho=rt$ in one of the integrals yields

$$R^{(k)}(k_1 l_1,k_2 l_2)\leqslant\text{const}\left[\int_0^\infty r|R_{l_1}(k_1 r)|^4\,dr\right]^{1/2}\times\left[\int_0^\infty|R_{l_2}(k_2\rho)|^4\,d\rho\right]^{1/2}.$$

Using the result (C1), this reduces to

$$R^{(k)}(k_1 l_1,k_2 l_2)\leqslant\text{const}\left|R_{l_1}\left(\frac{k_1u_0}{1-u_0}\right)\right|^2\times\left|R_{l_2}\left(\frac{k_2u_0}{1-u_0}\right)\right|^2\frac{u_0}{(1-u_0)^3}.$$

Substitution of this estimate into the inequality (3.24) for I_4 yields

$$I_4\leqslant\text{const}\sum_{l_i=0}^{\infty}(2l_1+1)(2l_2+1)\int_0^\infty k_1^2\,dk_1\int_0^\infty k_2^2\,dk_2\times\left|R_{l_1}\left(\frac{k_1u_0}{1-u_0}\right)\right|^2\left|R_{l_2}\left(\frac{k_2u_0}{1-u_0}\right)\right|^2\frac{u_0}{(1-u_0)^3}\times|E-\tfrac{1}{2}k_1^2-\tfrac{1}{2}k_2^2|^2 \quad \text{for } E\in D_E. \tag{3.47}$$

Transforming into polar coordinates $k_1=\rho\cos\theta$, $k_2=\rho\sin\theta$ and noting that $\rho^4/|E-\frac{1}{2}\rho^2|^2$ is bounded for $E\in D_E$, the double integral in (3.47) satisfies the inequality

$$\begin{aligned}&\int_0^\infty k_1^2\,dk_1\int_0^\infty k_2^2\,dk_2\times\left|R_{l_1}\left(\frac{k_1u_0}{1-u_0}\right)\right|^2\left|R_{l_2}\left(\frac{k_2u_0}{1-u_0}\right)\right|^2\times|E-\tfrac{1}{2}k_1^2-\tfrac{1}{2}k_2^2|^2\\ &=\int_0^\infty\rho^5\,d\rho\int_0^{\pi/2}\sin^2\theta\cos^2\theta\,d\theta\times\left|R_{l_1}\left(\frac{\rho u_0\cos\theta}{1-u_0}\right)\right|^2\times\left|R_{l_2}\left(\frac{\rho u_0\sin\theta}{1-u_0}\right)\right|^2|E-\tfrac{1}{2}\rho^2|^{-2}\\ &\leqslant\text{const}\int_0^\infty\rho\left|R_{l_1}\left(\frac{\rho u_0\cos\theta_0}{1-u_0}\right)\right|^2\times\left|R_{l_2}\left(\frac{\rho u_0\sin\theta_0}{1-u_0}\right)\right|^2d\rho,\quad\theta_0\in\left[0,\frac{\pi}{2}\right],\quad E\in D_E,\end{aligned} \tag{3.48}$$

where, the last step has been obtained by using the mean value theorem of the integral calculus applied to the integration over θ in the interval $[0,\pi/2]$. Using the result (C2) on the last integral in (3.48) and substituting the resulting estimate of the

double integral into (3.47), one obtains

$$I_4 \leqslant \text{const}\left[\sum_{l_1=0}^{\infty}(2l_1+1)\left|R_{l_1}\left(\frac{u_0\nu_0\cos\theta_0}{(1-u_0)(1-\nu_0)}\right)\right|^2\right]$$

$$\times\left[\sum_{l_2=0}^{\infty}(2l_2+1)\left|R_{l_2}\left(\frac{u_0\nu_0\sin\theta_0}{(1-u_0)(1-\nu_0)}\right)\right|^2\right]$$

$$\times\left(\frac{u_0\nu_0}{[(1-u_0)(1-\nu_0)]^3}\right)<\infty, \quad E\in D_E, \tag{3.49}$$

the last step being made possible by the result (C3).

This completes the proof of compactness, as desired, of the sequence of operators (3.6). To show that the operator

$$K(E)=(E-H_{01}-H_{02})^{-1}V, \quad E\in D_E \tag{3.1}$$

where V is given by (3.3) is compact, all we need to do is to show that $\|K-K_n\|\to 0$ as $n\to\infty$. Noting that V and V_n have domain D_V defined by (3.2), we observe that $(V-V_n)\in L^{\infty}(D_V)$, where the norm on the Banach space $L^{\infty}(D_V)$ is defined by

$$\|f\|=\sup_{(\mathbf{r}_1,\mathbf{r}_2)\in D_V}|f(\mathbf{r}_1,\mathbf{r}_2)|. \tag{3.50}$$

Since $\|V-V_n\|\to 0$ as $n\to\infty$

and $(E-H_{01}-H_{02})^{-1}$ is a bounded operator for $E\in D_E$, we have

$$\|K-K_n\|=\|(E-H_{01}-H_{02})^{-1}(V-V_n)\|$$

$$\leqslant\|(E-H_{01}-H_{02})^{-1}\|\|(V-V_n)\|\to 0$$

$$\text{as } n\to\infty, \quad E\in D_E.$$

Hence, the operator $K(E)$ is compact for $E\in D_E$.

4. CONCLUSIONS

We have shown that the linear operator defined by the infinite system of linear equations (2.12) is compact in a region of the complex energy plane which excludes

(i) the bound state scattering cuts starting at the hydrogenic bound state energies E_n, $n=1,2,\cdots$, and extending to $+\infty$.

(ii) the multiparticle cut starting at $E=0$ and extending to $+\infty$.

Our region of compactness also excludes the spurious points $\{E_{n_1}+E_{n_2}\}$, $n_1,n_2=1,2,\cdots$. These points do not belong to the essential spectrum of H.

The above results permit us to truncate the infinite system of equations (2.12) with the assurance that the N energy eigenvalues obtained from the $N\times N$ truncated equations will uniformly converge to the lowest N eigenvalues of the original infinite system. Questions regarding the choice of the complete basis sets $\{|\alpha_1,\alpha_2\rangle\}$ have been discussed in I. Here we merely remark that they must belong to the domain of H and be discrete. One could, for instance, choose a Hartree–Fock (RHF) basis set for $\{|\alpha_1,\alpha_2\rangle\}$.

APPENDIX A

We prove the result that if $|\Phi_1\rangle$, $|\Phi_2\rangle\in\mathscr{H}$ and $\langle\Phi_1|\Phi_2\rangle\neq 0$, then

$$\left|\left(\int\sum_{\nu_1}\right)\left(\int\sum_{\nu_1}\right)\langle\Phi_1|\nu_1,\nu_2\rangle(E-E_{\nu_1}-E_{\nu_2})^{-1}\langle\nu_1,\nu_2|\Phi_2\rangle\right|$$
$$\leqslant C\,|\langle\Phi_1|\Phi_2\rangle|, \quad E\in D_E, \tag{A1}$$

where C is a positive constant.

Proof: Let

$$\text{Re}[(E-E_{\nu_1}-E_{\nu_2})^{-1}]=L(E,\nu_1,\nu_2), \tag{A2a}$$

and

$$\text{Im}[(E-E_{\nu_1}-E_{\nu_2})^{-1}]=M(E,\nu_1,\nu_2). \tag{A2b}$$

We have, using (3.14)

$$|L(E;\nu_1,\nu_2)|\leqslant|(E-E_{\nu_1}-E_{\nu_2})^{-1}|\leqslant A \quad \text{for } E\in D_E, \tag{A3}$$

where A is a positive constant. Similarly

$$|M(E;\nu_1,\nu_2)|\leqslant A \quad \text{for } E\in D_E. \tag{A4}$$

If $|\Phi_1\rangle=|\Phi_2\rangle=|\Phi\rangle$ say, then using (A3) we have

$$\left(\int\sum_{\nu_i}\right)\langle\Phi|\nu_1,\nu_2\rangle L(E;\nu_1,\nu_2)\langle\nu_1,\nu_2|\Phi\rangle$$
$$\leqslant\left(\int\sum_{\nu_i}\right)\langle\Phi|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2|\Phi\rangle|L(E;\nu_1,\nu_2)|$$
$$\leqslant A\left(\int\sum_{\nu_i}\right)\langle\Phi|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2|\Phi\rangle$$
$$=\langle\Phi|\Phi\rangle, \quad E\in D_E.$$

We can therefore choose a real constant A_1 which may be either positive or negative and for which

$$\left(\int\sum_{\nu_i}\right)\langle\Phi|\nu_1,\nu_2\rangle L(E;\nu_1,\nu_2)\langle\nu_1,\nu_2|\Phi\rangle$$
$$=A_1\langle\Phi|\Phi\rangle, \quad E\in D_E. \tag{A5}$$

Similarly, one obtains

$$\left(\int\sum_{\nu_i}\right)\langle\Phi|\nu_1,\nu_2\rangle M(E;\nu_1,\nu_2)\langle\nu_1,\nu_2|\Phi\rangle$$
$$=A_2\langle\Phi|\Phi\rangle, \quad E\in D_E. \tag{A6}$$

Using $|\Phi_1+\Phi_2\rangle$ in (A5) instead of $|\Phi\rangle$ we get

$$\left(\int\sum_{\nu_i}\right)[\langle\Phi_1|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2|\Phi_1\rangle+\langle\Phi_2|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2|\Phi\rangle$$
$$+\langle\Phi_1|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2|\Phi_2\rangle+\langle\Phi_2|\nu_1,\nu_2\rangle$$
$$\times\langle\nu_1,\nu_2|\Phi_1\rangle]L(E;\nu_1,\nu_2)$$
$$=A_1[\langle\Phi_1|\Phi_1\rangle+\langle\Phi_2|\Phi_2\rangle+\langle\Phi_1|\Phi_2\rangle+\langle\Phi_2|\Phi_1\rangle],$$
$$E\in D_E,$$

which by (A5) implies that

$$\left(\int\sum_{\nu_i}\right)[\langle\Phi_1|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2|\Phi_2\rangle+\langle\Phi_2|\nu_1,\nu_2\rangle\langle\nu_1,\nu_2|\Phi_1\rangle]$$

$$= A_1[\langle \Phi_1|\Phi_2\rangle + \langle \Phi_2|\Phi_1\rangle].$$

This can be written in the form

$$\mathrm{Re}\left[\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle L(E;\nu_1,\nu_2)\right] = A_1\,\mathrm{Re}[\langle \Phi_1|\Phi_2\rangle], \quad E\in D_E. \tag{A7}$$

Similarly, using (A6) one obtains

$$\mathrm{Re}\left[\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle M(E;\nu_1,\nu_2)\right] = A_2\,\mathrm{Re}[\langle \Phi_1|\Phi_2\rangle], \quad E\in D_E. \tag{A8}$$

In a similar manner, using $|\Phi_1 - i\Phi_2\rangle$ instead of $|\Phi\rangle$ in (A5) and (A6) one obtains

$$\mathrm{Im}\left[\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle L(E;\nu_1,\nu_2)\right] = A_1\,\mathrm{Im}[\langle \Phi_1|\Phi_2\rangle], \quad E\in D_E, \tag{A9}$$

and

$$\mathrm{Im}\left[\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle M(E;\nu_1,\nu_2)\right] = A_2\,\mathrm{Im}[\langle \Phi_1|\Phi_2\rangle], \quad E\in D_E. \tag{A10}$$

respectively. Also, (A7) and (A9) together imply that

$$\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle L(E;\nu_1,\nu_2) = A_1\langle \Phi_1|\Phi_2\rangle, \quad E\in D_E. \tag{A11}$$

Similarly, (A8) and (A10) imply that

$$\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle M(E;\nu_1,\nu_2) = A_2\langle \Phi_1|\Phi_2\rangle, \quad E\in D_E. \tag{A12}$$

We have, using (A11) and (A12)

$$\begin{aligned}\left|\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle (E-E_{\nu_1}-E_{\nu_2})^{-1}\right| &\leqslant \left|\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle L(E;\nu_1,\nu_2)\right| \\ &\quad + \left|\left(\int \sum_{\nu_i}\right)\langle \Phi_1|\nu_1,\nu_2\rangle\langle \nu_1,\nu_2|\Phi_2\rangle M(E;\nu_1,\nu_2)\right| \\ &= |(|A_1| + |A_2|)|\,|\langle \Phi_1|\Phi_2\rangle| = C\,|\langle \Phi_1|\Phi_2\rangle|.\end{aligned}$$

This completes the proof.

APPENDIX B

It will be shown that

$$|R_{nl}(r)r| \leqslant \frac{\text{const}}{n^{1/2}} \tag{B1}$$

and

$$|R_{n0}(r)| \leqslant \frac{\text{const}}{n^{3/2}}, \tag{B2}$$

where

$$R_{nl}(r) = \left[\left(\frac{2Z}{n}\right)^3 \frac{(n-l-1)!}{2n(n+l)!}\right]^{1/2}\left(\frac{2Zr}{n}\right)^l \times L_{n-l-1}^{2l+1}\left(\frac{2Zr}{n}\right)e^{-Zr/n}. \tag{B3}$$

The following recurrence relations will be used[9]:

$$xL_n^{\alpha+1}(x) = (n+\alpha+1)L_n^{\alpha}(x) - (n+1)L_{n+1}^{\alpha}(x), \tag{B4}$$

$$L_n^{\alpha-1}(x) = L_n^{\alpha}(x) - L_{n-1}^{\alpha}(x), \tag{B5}$$

$$xL_n^{\alpha+1}(x) = \alpha L_n^{\alpha}(x) - (n+1)L_{n+1}^{\alpha-1}(x), \tag{B6}$$

$$\sum_{m=0}^{n} L_m^{\alpha}(x) = L_n^{\alpha+1}(x), \quad L_0^{\alpha}(x) = 1. \tag{B7}$$

The third of these relations is obtained from the first two. Also, the following inequality will be required[10]:

$$|L_n(x)e^{-x/2}| \leqslant 1. \tag{B8}$$

The inequality (B1) will be proved by induction. From (B3), we have for $l = 0$

$$R_{n0}(r) = (Z/n^3)^{1/2}\left(\frac{2Zr}{n}\right)L_{n-1}^{1}\left(\frac{2Zr}{n}\right)e^{-Zr/n}. \tag{B9}$$

Using (B4), we find that (B9) yields the inequality

$$|R_{n0}(r)r| \leqslant (Z/n)^{1/2}\left[\left|L_{n-1}\left(\frac{2Zr}{n}\right)e^{-Zr/n}\right| + \left|L_n\left(\frac{2Zr}{n}\right)e^{-Zr/n}\right|\right] \leqslant \frac{\text{const}}{n^{1/2}},$$

where, the last step is obtained by using the inequality (B8). Assume now that the statement is true for l:

$$|R_{nl}(r)r| \leqslant \frac{\text{const}}{n^{1/2}}. \tag{B10}$$

The statement will be proved for $l+1$. Now

$$R_{n,l+1}(r)r = \left[\frac{Z}{n^2}\frac{(n-l-2)!}{(n+l+1)!}\right]^{1/2}\left(\frac{2Zr}{n}\right)^{l+2} \times L_{n-l-2}^{2l+3}\left(\frac{2Zr}{n}\right)e^{-Zr/n}. \tag{B11}$$

Using (B6) and (B7), we have

$$\left(\frac{2Zr}{n}\right)L_{n-l-2}^{2l+3}\left(\frac{2Zr}{n}\right) = -(n-l-1)L_{n-l-1}^{2l+1}\left(\frac{2Zr}{n}\right) + (2l+2)\sum_{s=1}^{n-l-1} L_{n-l-1-s}^{2l+1}\left(\frac{2Zr}{n}\right).$$

Hence (B11) satisfies the inequality

$$|R_{n,l+1}(r)r| \leqslant \left|\left[\frac{Z}{n^2}\frac{(n-l-2)!}{(n+l+1)!}\right]^{1/2}(n-l-1)\left(\frac{2Zr}{n}\right)^{l+1}\right.$$

$$\times L_{n-l-1}^{2l+1}\left(\frac{2Zr}{n}\right)e^{-Zr/n}\Bigg|$$

$$+(2l+2)\sum_{s=1}^{n-l-1}\left|\left[\frac{Z}{n^2}\frac{(n-l-2)!}{(n+l+1)!}\right]^{1/2}\left(\frac{2Zr}{n}\right)^{l+1}\right.$$

$$\left.\times L_{(n-s)-l-1}^{2l+1}\left(\frac{2Zr}{n}\right)e^{-Zr/n}\right| \tag{B12}$$

$$=\Psi_{nl}(r)+\Phi_{nl}(r), \text{ say.} \tag{B13}$$

Consider the first term on the right-hand side of the inequality (B12). Using the hypothesis (B10) we have

$$\Psi_{nl}(r)=\left[\frac{(n-l-1)}{(n+l+1)}\right]^{1/2}\left|\left[\frac{Z}{n^2}\frac{(n-l-1)!}{(n+l)!}\right]^{1/2}\right.$$

$$\left.\times\left(\frac{2Zr}{n}\right)^{l+1}L_{n-l-1}^{2l+1}\left(\frac{2Zr}{n}\right)e^{-Zr/n}\right|$$

$$\leqslant\left[\frac{(n-l-1)}{(n+l+1)}\right]^{1/2}\frac{\text{const}}{n^{1/2}},$$

which implies

$$\Psi_{nl}(r)\leqslant\frac{\text{const}}{n^{1/2}}. \tag{B14}$$

We now deal with the second term in (B12). We can write

$$\Phi_{nl}(r)$$

$$=(2l+1)\sum_{s=2}^{n-l-1}\left[\frac{(n-s)^2(n-l-s)\cdots(n-l-2)}{n^2(n-s+l+1)\cdots(n+l+1)}\right]^{1/2}$$

$$\times|R_{n-s,l}(k_{n,s}r)k_{n,s}r|$$

$$+(2l+2)\left[\frac{(n-1)^2}{n^2}\frac{1}{(n+l)(n+l+1)}\right]^{1/2}$$

$$\times|R_{n-1,l}(k_{n,1}r)k_{n,1}r|, \tag{B15}$$

where $k_{n,s}=(n-s)/n$. Note that $0<k_{n,s}<1$ so that $0\leqslant k_{n,s}r<\infty$. Since

$$|R_{n-s,l}(k_{n,s}r)k_{n,s}r|\leqslant\frac{\text{const}}{(n-s)^{1/2}},$$

We find from (B15) that $\Phi_{nl}(r)$ satisfies the inequality

$$\Phi_{nl}(r)\leqslant\text{const}n^{-1}\sum_{s=1}^{n-l-1}(n-s)^{-1/2}$$

$$\leqslant\text{const}n^{-1}\sum_{s=1}^{n-1}s^{-1/2}\leqslant\text{const}n^{-1/2}, \tag{B16}$$

where we have used (3.38) to obtain the last step. Noting (B14) and (B16), the result follows.

To obtain (B2), we have, using (B7),

$$|R_{n0}(r)|\leqslant(Z/n^3)^{1/2}(2Z/n)\sum_{m=0}^{n-1}\left|L_m\left(\frac{2Zr}{n}\right)e^{-Zr/n}\right|$$

$$\leqslant(Z/n^3)^{1/2}(2Z/n)(n-1), \quad \text{by use of (B8)}$$

$$\leqslant\frac{\text{const}}{n^{3/2}},$$

which is the required result.

APPENDIX C

Finally, we prove the following results involving the radial functions for the continuum states of the hydrogen atom:

$$\int_0^\infty|R_l(kr)|^4\,dr=\left|R_l\left(\frac{ku_0}{1-u_0}\right)\right|^4\frac{u_0}{(1-u_0)^3}, \tag{C1}$$

$$u_0\in[0,1],$$

$$\int_0^\infty\rho|R_{l_1}(\rho r_1)|^2|R_{l_2}(\rho r_2)|^2\,d\rho$$

$$=\left|R_{l_1}\left(\frac{v_0r_1}{1-v_0}\right)\right|^2\left|R_{l_2}\left(\frac{v_0r_2}{1-v_0}\right)\right|^2\frac{v_0}{(1-v_0)^3}, \tag{C2}$$

$$v_0\in[0,1],$$

$$\sum_{l=0}^{\infty}(2l+1)|R_l(kr)|^2<\infty. \tag{C3}$$

To prove (C1), we transform the infinite integral into one over the finite interval [0,1] by the change of variable

$$u=r/(1+r), \quad \text{so that } r=u/(1-u). \tag{C4}$$

This gives

$$\int_0^\infty|R_l(kr)|^4\,dr=\int_0^1\left|R_l\left(\frac{ku}{1-u}\right)\right|^4\frac{u}{(1-u)^3}\,du$$

$$=\int_0^1 f_l(u)\,du, \quad \text{say.} \tag{C5}$$

Note that $f_l(0)=0$. Also using (3.42) we have

$$f_l(1)=\lim_{u\to1}\left|R_l\left(\frac{ku}{1-u}\right)\right|^4\frac{u}{(1-u)^3}$$

$$=\lim_{u\to1}\left(\frac{2}{\pi}\right)^2\frac{u}{(1-u)^3}$$

$$\times\frac{\sin^4(ku/(1-u)-\frac{1}{2}l\pi+\eta_l+(Z/k)\log[2ku/(1-u)}{[ku/(1-u)]^4}$$

$$=\lim_{u\to1}(2/\pi)^2(1-u)\frac{\sin^4(\cdots)}{k^4u^3}=0,$$

so that $f_l(u)$ is defined at $u=0$ and $u=1$.

We can therefore apply the mean value theorem of the integral calculus to the finite integral on the right hand side of (C5) to obtain

$$\int_0^\infty|R_l(kr|^4\,dr=\left|R_l\left(\frac{ku_0}{1-u_0}\right)\right|^4\frac{u_0}{(1-u_0)^3}, \quad u_0\in[0,1].$$

The result (C2) is obtained in exactly the same manner as (C1). Note that[11] $R_l(kr)\to0$ as $k\to0$ and that the asymptotic form of $R_l(kr)$ for large k is of the same form as that for large r [see (3.42)].

To prove (C3), we have[12]

$$\psi(r,\theta) = e^{Z\pi/2k}\Gamma(1 - iZ/k)e^{ikZ}$$

$$\times F_1(-iZ/k,1;ikr(1-\cos\theta))$$

$$= \sum_{l=0}^{\infty} (2l+1)\exp[i(\eta_l + \tfrac{1}{2}l\pi)]\, R_l(kr)P_l(\cos\theta)$$

so that

$$\int_0^{\pi} |\psi(r,\theta)|^2 \sin\theta\, d\theta$$

$$= \sum_{l,l'=0}^{\infty} (2l+1)(2l'+1)$$

$$\times \exp[i(\eta_l - \eta_{l'} + \tfrac{1}{2}l\pi - \tfrac{1}{2}l'\pi)]$$

$$\times R_l(kr)\overline{R_{l'}(kr)} \int_0^{\pi} P_l(\cos\theta)P_{l'}(\cos\theta)\sin\theta\, d\theta.$$

Using the orthogonality relations for the Legendre polynomials we obtain immediately

$$\sum_{l=0}^{\infty} (2l+1)|R_l(kr)|^2 = \int_0^{\pi} |\psi(r,\theta)|^2 \sin\theta\, d\theta < \infty.$$

[1]N. Dunford and J. Schwartz, *Linear Operators* (Interscience, New York, 1963), Part II, p. 1091.
[2]H.A. Bethe and E.E. Salpeter, *Quantum Mechanics of One and Two-electron Atoms* (Springer-Verlag, Berlin, 1957), Chap. I.
[3]J. Dieudonne, *Foundations of Modern Analysis* (Academic, New York, 1960), p. 314.
[4]R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960), p. 18.
[5]C. Lovelace, Three particle systems and unstable particles in *Strong Interactions and High Energy Physics* (Oliver and Boyd, Edinburgh and London, 1964), R.G. Moorhourse, Ed.
[6]See Ref. 2, p. 17.
[7]See Ref. 2, p. 17.
[8]N.F. Mott and H.S.W. Massey, *The theory of Atomic Collisions* (Oxford U. P., 1949), 2nd ed., p. 46.
[9]*Higher Transcendental Functions, Bateman Manuscript Project* (Erdelyi, Ed. (McGraw-Hill, New York, 1953), Vol. II, pp. 190, 192.
[10]See Ref. 9, p. 205.
[11]L.D. Landau and E.M. Lifshitz, *Quantum Mechanics* (Pergamon, London, 1965), 2nd ed., p. 122.
[12]See Ref. 8, p. 46.

Geometry of spacetime founded on spacelike metric

Angel Montesinos

Departamento de Topología y Geometría, Facultad de Matemáticas, Universidad Complutense, Madrid-3, Spain
(Received 6 September 1977; revised manuscript received 22 May 1978)

The first part of this paper contains new mathematical techniques for describing a spacetime anisotropy as suggested by the violation of parity conservation. Geometric measures of spacetime involve both the laboratory doing them and the events upon which these measures are done. The time form c and the spacelike length γ are the basic issues of those measures. Both depend on events and also on the timelike direction of the laboratory. Relativity tells that the field $\gamma - c \otimes c$ depends, on the contrary, on events only; in this sense, relativistic spacetime is isotropic. If γ and c do not have that property, the manifold where the observable geometry takes place must be the set of timelike directions. The geometric structure of this manifold given by c and γ is studied in detail. The second part of the paper contains the study of a line of thought opposite to chronogeometry: Building the geometry from lengths instead of times. The datum is γ; through the conditions of stationary spacelike volume and of stationary proper time, a class of time forms and a gauge are obtained under some weak restrictions. Newtonian and relativistic spacelike metrics fulfill these restrictions. Standard connections are induced; they define the absolute derivative of physical fields and the geometric structure of the manifold of timelike directions. The paper ends with some comments about the remaining problem: to suggest and justify field equations.

I. MATHEMATICAL TECHNIQUES FOR SPACETIME PHYSICS

1. INTRODUCTION

We will consider spacetime as an n-dimensional differentiable manifold M, whose underlying set is the set of events. We call timelike the nonvanishing vectors tangent to the possible world line of particles. Let $\widetilde{\mathcal{T}}M$ be the set of timelike vectors. Then, if $\bar{x}\in\widetilde{\mathcal{T}}M$ and $\alpha>0$, we have $\alpha\bar{x}\in\widetilde{\mathcal{T}}M$. Also we admit that $\widetilde{\mathcal{T}}M$ is an open subset of T_0^1M, $\pi_0^1:T_0^1M\to M$ being the tangent bundle over M. In ordinary language, this assumption corresponds to the following experimental fact: Given a particle, it is possible to have particles whose relative movement (with respect to the former) has arbitrary direction. We put $\tilde{\pi}:\widetilde{\mathcal{T}}M\to M$, where $\tilde{\pi}=\pi_0^1|\widetilde{\mathcal{T}}M$, and suppose $\widetilde{\mathcal{T}}_m=\tilde{\pi}^{-1}(m)$ to be nonempty for each $m\in M$.

We emphasize that $\widetilde{\mathcal{T}}M$ is not related here to a Lorentz metric, because we are looking for a wider mathematical ground than the relativistic one.

Any physical quantity must be measured from some laboratory, and every physical experiment must be devised referring it to several instruments. These instruments constitute the laboratory, and they are built by particles following their respective world lines. Let U be the spacetime neighborhood where the experiment takes place. Then, we can provide a rough description of the laboratory as a cross section $\bar{r}$ of $\tilde{\pi}$ on U, where $\bar{r}_m$ stands for the tangent to the world line of the particle (belonging to the involved instruments) at $m\in U$. Thus, one could expect the result to be a function of $\bar{r}$ and other parameters. Obviously, this happens in practice: for example, the Doppler effect of a signal received in earth from a satellite.

However, this experiment and others like it are too far from our geometric goal. So, we shall fix our attention upon the measurement of geometric features of spacetime: (a) time elapsed between two events, as measured by clocks (laboratories) following different world lines connecting both events; (b) spacelike distance between two events as measured by different meter sticks (laboratories), such that both events occur on each meter stick. In both cases, the resulting quantity depends on the laboratory, i.e., on the local cross section of $\tilde{\pi}$ attached to each clock or meter stick.

The wondrous thing would be that one could find, from that type of measures, a magnitude depending on events of spacetime only, and not also or $\bar{r}$. If this did occur, we could say that spacetime geometry was isotropic, since it did not depend on the timelike directions of the laboratories measuring it. Einstein's standard relativity is, of course, the best example.

But spacetime is not isotropic in its mass or charge distribution, at least on local scale. Moreover, the violation of parity conservation suggests an anisotropic spacetime at the microscopic level, as it has been explained by Horváth.[1] So, *one could regard general relativity as a first approximation that neglects anisotropy, and consider the manifold of timelike directions as the proper ground for the measurable spacetime geometry*. We say directions instead of vectors because $\bar{r}$ and $a\bar{r}$ do represent the same laboratory if $a:M\to\mathbb{R}$ is a positive function. Thus, the true manifold must be $\mathcal{T}M$, the quotient of $\widetilde{\mathcal{T}}M$ under the equivalence relation given by homotheties.

Now, what could one expect to find out as measurable quantities? Of course, the same we are obtaining until now, that is, ordinary numbers, vectors, or tensors. Thus, our physical fields will be maps from $\mathcal{T}M$ to $\mathbb{R}$ (scalar fields), or to T^r_sM (tensor fields).

The goal of Part I is to develop a suitable mathematical formalism for the treatment of these "mixed" fields also depending on directions. It provides a common geometric framework for the study and comparison of different spacetime theories (Newtonian and relativistic for instance). As far as I am aware, it constitutes a new mathematical technique; however, for the sake of brevity, we shall restrict our-

 0022-2488/79/050953-13$01.00

selves to the concepts we will directly use in spacetime theory; the risk of such restriction is to conceal somewhat the mathematical reasons for giving certain definitions. Anyway, our paper (Montesinos[2]) could serve as an introduction to those techniques.

Through Part II, my own physical theory is developed under the formalism of Part I. The fundamental field will be the spacelike metric. From it, we build simultaneity and, partially, time length; in this sense, my theory is somewhat new, since it no longer takes time or light signals as fundamental. It could be looked at as the opposite veiwpoint of chronogeometry.

Besides this Introduction, Part I has nine sections. In Sec. 2. we briefly describe the notation and some mathematical notations which we will use.

As for Sec. 3, let $\mathcal{F}M$ be the quotient manifold of $\tilde{\mathcal{F}}M$ under the equivalence relation given by positive homotheties. If $\pi:\mathcal{F}M\to M$ is the induced projection, then the manifold of timelike directions, $\mathcal{F}M$, becomes an open submanifold of the sphere bundle over M. Physical fields are maps as $h:\mathcal{F}M\to T^r_sM$, satisfying $\pi^r_s\circ h=\pi$, where $\pi^r_s:T^r_sM\to M$ is the tangent tensor bundle of type (r,s). This condition tells us that a physical field assigns to each timelike direction r_m a tensor lying in the tensor space tangent to M at m, the event where that timelike direction lies. We can consider physical fields as included in the algebra of Finsler tensor fields over $\tilde{\mathcal{F}}M$ because there is a one-to-one correspondence with homogeneous degree zero Finsler fields.

This material serves for describing the basic geometric features of spacetime, namely the time function $\tilde{f}$, the time form c, and the spacelike metric γ (Sec. 4). We discuss the physical meaning of these fields and give two examples, Newtonian and relativistic spacetimes, clearing up the wide range of spacetime models where this scheme applies.

The mixed nature of physical fields makes a direct treatment difficult. So, *we shall submit it to the techniques for usual fields over $\mathcal{F}M$. Besides the physical motivations for my viewpoints, that is the main objective of this part.* Thus, in Sec. 5 we define horizontal and vertical homomorphisms from the module of physical vector fields to that of ordinary vector fields over $\mathcal{F}M$.

In Sec. 6, these homomorphisms are extended to be graded tensor algebra homomorphisms (lifts) from ΠM, the algebra of physical fields, to $V\mathcal{F}M$, the algebra of ordinary tensor fields over $\mathcal{F}M$. Each lift has a unique lowering that is its transpose map. We define crossed pairs of lift lowerings. They induce the horizontal and vertical projectors. The main result of this section tells that a pair of horizontal and vertical homomorphisms, in the sense of Sec. 5, do define a unique pair of crossed lifts.

In Sec. 7 we define and interpret several types of connections we will use later, namely horizontal and vertical connections on ΠM, physical connections, and the j-connection D, an important mathematical tool. We interpret $\eta=Dc$ as the rate of time retardation when the relative speed increases. This field plays an important role in the existence problem for connections.

Section 8 is devoted to the definition, explanation, existence, and uniqueness of horizontal and vertical torsionless metric connections. They have a suggestive meaning: The vertical connection measures the absolute directional dependence of physical fields; the horizontal one, the absolute along spacetime dependence.

In Sec. 9, we lift the pair of these horizontal and vertical connections for having a unique physical connection. It defines the absolute dependence of physical fields along the time manifold. This connection is also lifted for having the linear connection **D**, that yields the final geometric structure of the time manifold itself. These results are briefly resumed in the conclusion (Sec. 10).

2. NOTATION

M, n-dimensional Hausdorff second countable real C^∞ manifold, briefly manifold. It stands for spacetime.

$\pi^r_s:T^r_sM\to M$, tangent tensor bundle over M of type (r,s); M_m, tangent space at $m\in M$.

V^0_0M, the ring of C^∞ real functions on M; V^r_sM, the V^0_0M-module of C^∞ cross sections of π^r_s; $VM=\oplus V^r_sM$, tensor $\mathbb{R}$ algebra, graded by the indexes (r,s).

$\tilde{\mathcal{F}}M$, the set of timelike vectors, is an open submanifold of T^1_0M; $\tilde{\pi}:\tilde{\mathcal{F}}M\to M$ is defined by $\tilde{\pi}=\pi^1_0|\tilde{\mathcal{F}}M$. We suppose that $0\notin\tilde{\mathcal{F}}_m=\tilde{\pi}^{-1}(m)\neq\emptyset$ for every $m\in M$. In addition, we require that if $\bar{x}\in\tilde{\mathcal{F}}M$, then $\alpha\bar{x}\in\tilde{\mathcal{F}}M$ for every $0<\alpha\in\mathbb{R}$.

Since $\tilde{\mathcal{F}}M$ is itself a manifold, we use $V^r_s\tilde{\mathcal{F}}M$ and $V\tilde{\mathcal{F}}M$ to denote the module of ordinary tensor fields of type (r,s) over $\tilde{\mathcal{F}}M$, and the respective graded tensor $\mathbb{R}$ algebra.

$\tilde{\Pi}^0_0M=V^0_0\tilde{\mathcal{F}}M$, the ring of real C^∞ functions on $\tilde{\mathcal{F}}M$; $\tilde{\Pi}^r_sM$ is the $\tilde{\Pi}^0_0M$ module of Finsler tensor fields of type (r,s), that is C^∞ maps $\tilde{h}:\tilde{\mathcal{F}}M\to T^r_sM$ satisfying $\pi^r_s\circ\tilde{h}=\tilde{\pi}$; $\tilde{\Pi}M=\oplus\tilde{\Pi}^r_sM$, graded tensor $\mathbb{R}$ algebra of Finsler fields. We say that a Finsler tensor field $\tilde{h}$ is homogeneous of degree $\alpha\in\mathbb{R}$ if $\tilde{h}_{q\bar{x}}=q^\alpha\tilde{h}_{\bar{x}}$ for every $0<q\in\mathbb{R}$ and $\bar{x}\in\tilde{\mathcal{F}}M$. That property will be denoted $h(\alpha)$.

$\tilde{u}:\tilde{\mathcal{F}}M\to T^1_0M$, the canonic Finsler vector field, is defined as the inclusion. Hence, $\tilde{u}$ is $h(1)$.

$i:\tilde{\Pi}^1_0M\to V^1_0\tilde{\mathcal{F}}M$, the vertical injection. That is, if $\bar{v}\in M_m$ and $\bar{x}\in\tilde{\mathcal{F}}_m$, then $i_{\bar{x}}(\bar{v})$ is the tangent at $t=0$ to the curve $\sigma:t\to\bar{x}+\bar{v}t$. Since $\tilde{\mathcal{F}}M$ is open in T^1_0M, then $\tilde{\mathcal{F}}_m$ is open in M_m; therefore, for some $\epsilon>0$ that curve lies in $\tilde{\mathcal{F}}_m\subset\tilde{\mathcal{F}}M$ if $-\epsilon<t<\epsilon$. Thus, $\sigma(t)$ is a curve on $\tilde{\mathcal{F}}M$, whence its tangent $i_{\bar{x}}(\bar{v})$ at $t=0$ is a vector belonging to $(\tilde{\mathcal{F}}M)_{\bar{x}}$. Hence, if $\tilde{v}:\tilde{\mathcal{F}}M\to T^1_0M$ is a Finsler vector field, we define $i\tilde{v}\in V^1_0\tilde{\mathcal{F}}M$ by means of $(i\tilde{v})_{\bar{x}}=i_{\bar{x}}(\tilde{v}_{\bar{x}})$.

$\langle s,v\rangle$, the contraction of the 1-form s (belonging to V^0_1M, $V^0_1\tilde{\mathcal{F}}M$, $\tilde{\Pi}^0_1M$, etc.) with the vector field v (belonging to V^1_0M, $V^1_0\tilde{\mathcal{F}}M$, $\tilde{\Pi}^1_0M$, etc., respectively).

3. THE TIME MANIFOLD. PHYSICAL FIELDS

On $\tilde{\mathcal{F}}M$ we define an equivalence relation $\sim$ by means of $\bar{x}\sim\bar{y}$ if $\tilde{\pi}(\bar{x})=\tilde{\pi}(\bar{y})$ and $\bar{x}=\alpha\bar{y}$ for some $\alpha>0$. Let $\mathcal{F}M$ be

the set of equivalence classes, and $p:\tilde{\mathcal{F}}M\to\mathcal{F}M$ the natural projection, which applies an element $\bar{x}\in\tilde{\mathcal{F}}M$ into its class $p\bar{x}$. Then, $\mathcal{F}M$ can be given a unique differentiable manifold structure making p a submersion. We will call $\mathcal{F}M$, with that structure, the *time manifold*. It represents the manifold of timelike directions. The map $\pi:\mathcal{F}M\to M$, where $\pi\circ p=\tilde{\pi}$, defines the *time bundle*. Note that a cross section r of π on $U\subset M$ can be looked at as a laboratory whose instruments have at $m\in U$ a particle with speed r_m.

Let Π^0_0M be the ring of C^∞ real functions on $\mathcal{F}M$. We use Π^r_sM to denote the set of *physical fields* of type (r,s), i.e., C^∞ maps $h:\mathcal{F}M\to T^r_sM$ satisfying $\pi^r_s\circ h=\pi$. Then, if for example $v\in\Pi^1_0M$, its value v_{r_m} at $r_m\in\mathcal{F}_m=\pi^{-1}(m)$ is a vector of M_m, the tangent space to M at m. Thus, Π^r_sM becomes a Π^0_0M-module, and we can build the graded tensor algebra ΠM of physical fields.

If $h\in\Pi^r_sM$, we put $e_ah=h\circ p:\tilde{\mathcal{F}}M\to T^r_sM$. Then $\pi^r_s\circ e_ah=\pi^r_s\circ h\circ p=\pi\circ p=\tilde{\pi}$; therefore, e_ah is a Finsler tensor field of type (r,s). Since $p(\alpha\bar{x})=p\bar{x}$ for every $\alpha>0$ and $\bar{x}\in\tilde{\mathcal{F}}M$, we conclude that e_ah is h (0). Hence, $e_a:\Pi M\to\tilde{\Pi}M$ is a graded $\mathbb{R}$ algebra homomorphism mapping ΠM onto the graded subalgebra of h (0) Finsler tensor fields. Conversely, if $\tilde{h}\in\tilde{\Pi}M$ is h (0), it defines $e_b\tilde{h}\in\Pi M$ by means of $e_b\tilde{h}\circ p=\tilde{h}$. Thus, $e_b\circ e_a=\mathrm{id}$ on ΠM, and $e_a\circ e_b=\mathrm{id}$ on the subalgebra of h (0) Finsler tensor fields. So we have bridge between Finsler techniques and those we present here.

4. SPACE AND TIME FORMS

Spacetime geometry involves two main concepts, spacelike and timelike length, and a link between them: synchronization. This last is the troubling point because since Einstein's relativity, light signals came in. The trouble is: timelike length defines by itself a synchronization, as we shall see at once; spacelike length also does that (see Part II). So, what do light signals do in all this matter? This question is purposely bold, but I think it is not merely rhetorical. It aims to raise doubts about the role light signals must play on spacetime geometry, and to make more plausible the viewpoint of this paper. In fact, my methodological way is the following: to look at space and timelike length as the basic (related between them or not) geometric data of spacetime, and to consider gravitational or electromagnetic phenomena (light signals among them) as desirable *dynamical* issues from the *static* (geometric) description. So, *in this paper light signals do not play any direct role among the basic geometric features of spacetime.* Of course, electromagnetic signals are the best practical tool for the study of spacetime in several areas. I simply say they are unnecessary for our theoretical purposes.

Let us consider time length first. As it has been pointed out by chronogeometry, time length must be defined by a h (1) function $\tilde{f}\in\tilde{\Pi}^0_0M$, such that if $\sigma:[a,b]\to M$ is the world line of an atomic clock, then $\int_a^b\tilde{f}_\sigma dt$ is the time measured by that clock between $\sigma(a)$ and $\sigma(b)$. The function $\tilde{f}$ must be h (1) for time elapsed could be invariant under parametrization changes of σ. We will call $\tilde{f}$ the *time function*.

A synchronization is given by a *time form*, that is a field $c\in\Pi^0_1M$ such that $\langle\tilde{c},\tilde{u}\rangle$ is everywhere nonzero (we always will put $\tilde{c}=e_ac$). Briefly, if $\bar{x}\in\tilde{\mathcal{F}}M$, then the hyperplane of M_m spanned by the vectors $\bar{v}\in M_m$ satisfying $\langle\tilde{c}_{\bar{x}},\bar{v}\rangle=0$ defines the simultaneity relative to the timelike direction $p\bar{x}$. Note that $\langle\tilde{c},\tilde{u}\rangle_{\bar{x}}=\langle c_{p\bar{x}},\bar{x}\rangle\neq0$ because $\bar{x}$ stands for a tangent to the world line of the particle defining the simultaneity $c_{p\bar{x}}$; since that line is timelike, different events on it cannot be simultaneous. Note also that if c is multiplied by any nonvanishing function $q\in\Pi^0_0M$, then qc defines the same simultaneity than c.

Let us relate $\tilde{f}$ and c. If $\tilde{f}$ and c are given, then c can be multiplied by some function $q\in\Pi^0_0M$ such that $\langle e_a(qc),\tilde{u}\rangle=\tilde{f}$, since it is enough to take $q=e_b(\tilde{f}/\langle\tilde{c},\tilde{u}\rangle)$. Thus, *an arbitrary given time function can be defined on this way from an arbitrary simultaneity.* The choice of a "length" for a simultaneity c (the multiplication by q) fixes a time scale on each synchronized laboratory. That is, if r_m is a timelike direction at m, then c_{r_m} stratifies on equitime hyperplanes the affine tangent space M_m. Thus, if $\bar{v}\in M_m$, then $\langle c_{r_m},\bar{v}\rangle$ stands for the time shift between the tail and the head events determining $\bar{v}$. This time shift depends on the inclination (synchronization) of c_{r_m}, and also depends on the separation of equitime hyperplanes (the lenght of c_{r_m}). Now, if $\sigma:[a,b]\to M$ is a world line and $r:M\to\mathcal{F}M$ is a laboratory, then $\int_a^b\langle c_{r\circ\sigma},\dot{\sigma}\rangle dt$ is the time inverted by the particle σ from $\sigma(a)$ to $\sigma(b)$, as measured by the laboratory r. If r is the particle itself, that is $r\circ\sigma=p\dot{\sigma}$, and $\langle\tilde{c},\tilde{u}\rangle=\tilde{f}$, then $\int_a^b\langle c_{r\circ\sigma},\dot{\sigma}\rangle dt=\int_a^b\langle\tilde{c},\tilde{u}\rangle_\sigma dt=\int_a^b\tilde{f}_\sigma dt$. In other words, the condition $\langle\tilde{c},\tilde{u}\rangle=\tilde{f}$ means that *we have picked for the synchronized laboratories the same time scale which measures proper times by means of* $\tilde{f}$.

Let us consider the inverse problem: Given $\tilde{f}$, find out a time form c such that $\langle\tilde{c},\tilde{u}\rangle=\tilde{f}$. A solution is the element of Π^0_1M defined through $\langle\tilde{c},\tilde{v}\rangle=i\tilde{v}(\tilde{f})$ for every $\tilde{v}\in\tilde{\Pi}^1_0M$. In fact we have $\langle\tilde{c},\tilde{u}\rangle=i\tilde{u}(\tilde{f})=\tilde{f}$ because $\tilde{f}$ is h (1); also $\tilde{c}$ is h (0) because $\tilde{u}$ is h (1). Therefore, $e_b\tilde{c}=c$ is a solution. Now, if $b\in\Pi^0_1M$ satisfies $\langle e_ab,\tilde{u}\rangle=0$, then $c+b$ is another solution. But only the first one has a decisive property: *The simultaneity if furnishes corresponds to that of infinitely slow clock transport.* In fact, we will see in Part II Sec. 4 that this correspondence is characterized by the property $i\tilde{v}(\langle\tilde{c},\tilde{u}\rangle)=\langle\tilde{c},\tilde{v}\rangle$ for every $\tilde{v}\in\tilde{\Pi}^1_0M$. So, we can say that *a time function* $\tilde{f}$ *gives raise to a unique compatible time form* c, the one satisfying $\langle\tilde{c},\tilde{u}\rangle=\tilde{f}$, $i\tilde{v}(\langle\tilde{c},\tilde{u}\rangle)=\langle\tilde{c},\tilde{v}\rangle$. Due to this, in the following we will use time forms instead of time functions.

As for spacelike length, it is given by a field $\gamma\in\Pi^0_2M$, symmetric, of signature $(0,+,\ldots,+)$, and such that $\tilde{\gamma}(\tilde{u},)=0$, where $\tilde{\gamma}=e_a\gamma$ Along Part II we will justify this assertion and see in what manner γ defines a time form. So, we shall then reach another puzzling point: the compatibility of the time forms obtained from time functions or from spacelike metrics (II.1). Until then, we will leave this question and go on to describe two typical examples under this formalism.

Let $\bar{g}\in V^0_2M$ be a Lorentz metric. Then it defines the time form $\tilde{c}=-\tilde{g}(\tilde{u},)/(-\tilde{g}(\tilde{u},\tilde{u}))^{1/2}$ where $\tilde{g}=\bar{g}\circ\tilde{\pi}$, and the spacelike metric $\gamma=\bar{g}\circ\pi+c\otimes c$. This is the *relativistic model.*

As for a *generalized Newtonian spacetime* (locally absolute time and length), let M admit a symmetric field $\bar{q}\in V_2^0M$ of signature $(0,+,\ldots,+)$, and a field $\bar{b}\in V_1^0M$, everywhere nonzero, such that if $0\neq\bar{v}\in M_m$ and $\langle\bar{b}_m,\bar{v}\rangle=0$, then $\bar{q}_m(\bar{v},\bar{v})>0$. Thus, $\bar{b}$ defines the local absolute time and $\bar{q}$ the local absolute length. We put $\tilde{\mathcal{F}}M=\{\bar{x}_{\underline{m}}\in T_0^1M:\langle\bar{b}_m,\bar{x}_m\rangle\neq0\}$. Then, the time form is given by $c=\bar{b}\circ\pi$ and the spacelike metric by

$$e_a\gamma=\tilde{q}+\frac{\tilde{q}(\tilde{u},\tilde{u})}{\langle\tilde{c},\tilde{u}\rangle^2}\tilde{c}\otimes\tilde{c}-\frac{\tilde{q}(\tilde{u},)\otimes\tilde{c}}{\langle\tilde{c},\tilde{u}\rangle}-\frac{\tilde{c}\otimes\tilde{q}(\tilde{u},)}{\langle\tilde{c},\tilde{u}\rangle},$$

where $\tilde{q}=\bar{q}\circ\tilde{\pi}$.

See also Ref. 3.

5. VERTICAL AND HORIZONTAL HOMOMORPHISMS

A vector field $\mathbf{v}\in V_0^1\mathcal{F}M$ is said to be vertical if $\mathbf{v}(\bar{a}\circ\pi)=0$ for every $\bar{a}\in V_0^0M$. That is, vertical vector fields are tangent to the fibres $\pi^{-1}(m)$. The set of vertical vector fields is a $V_0^0\mathcal{F}M$-module, locally $(n-1)$-dimensional, for it is the annihilator of the $V_0^0\mathcal{F}M$-module spanned by the elements $d(\bar{a}\circ\pi)\in V_1^0\mathcal{F}M$, and this last module is clearly n-dimensional [take for example $\bar{a}=\bar{x}^i$, where $\{\bar{x}^i\}$ is a coordinate system on $U\subset M$, and note that $\mathcal{F}M$ is $(2n-1)$-dimensional].

Suppose that a time form c is given. Then, it defines in a natural way a homomorphism $j:\Pi_0^1M\to V_0^1\mathcal{F}M$ such that its image, $j(\Pi_0^1M)$, equals the module of vertical vector fields (in this sense we say that j is a *vertical homomorphism*). In face, let $v\in\Pi_0^1M$; then $\tilde{v}=e_av\in\tilde{\Pi}_0^1M$, and $\langle\tilde{c},\tilde{u}\rangle i\tilde{v}$ is a vertical field of $V_0^1\tilde{\mathcal{F}}M$. By its own definition, $(\langle\tilde{c},\tilde{u}\rangle i\tilde{v})_{\alpha\bar{x}}$ is the tangent, at $t=0$, to the curve $t\to\alpha\bar{x}+\langle\tilde{c}_{\alpha\bar{x}},\alpha\bar{x}\rangle t\tilde{v}_{\alpha\bar{x}}$. Now, because the factor $\langle\tilde{c}_{\alpha\bar{x}},\alpha\bar{x}\rangle=\alpha\langle\tilde{c}_{\bar{x}},\bar{x}\rangle$, p projects all these curves (varying the number α) upon the same curve $p(\bar{x}+\langle c_{p\bar{x}},\bar{x}\rangle tv_{p\bar{x}})$, whose tangent at $t=0$ defines $j_{p\bar{x}}(v_{p\bar{x}})$. Thus, we put $(jv)_{p\bar{x}}=j_{p\bar{x}}(v_{p\bar{x}})$. Hence we have $(jv)\circ p=p_*\circ(\langle\tilde{c},\tilde{u}\rangle ie_av)$, where p_* stands for the derived map of p. If $a\in\Pi_0^0M=V_0^0\mathcal{F}M$, then $jv(a)$ defines a derivation along the fibres; that is, *jv measures the dependence of functions on directions, not on events of spacetime.* We have that $\ker j$ is spanned by $k=e_b(\tilde{u}/\langle\tilde{c},\tilde{u}\rangle)$, because $p_*\circ i\tilde{u}=0$.

Now, let $\Lambda:\Pi_0^1M\to V_0^1\mathcal{F}M$ be a homomorphism. Then we say it is *horizontal* if $(\Lambda v)_r(\bar{a}\circ\pi)=v_r(\bar{a})$ for every $r\in\mathcal{F}M$ and $\bar{a}\in V_0^0M$. The definition tells that Λ is injective. Note that our condition is equivalent to $\langle d(\bar{a}\circ\pi),\Lambda v\rangle=\langle(d\bar{a})\circ\pi,v\rangle$. Let us give an interpretation of Λ. We have that the elements $v\in\Pi_0^1M$ can be locally written as $v^i(\partial/\partial\bar{x}^i\circ\pi)$, where $v^i\in\Pi_0^0M$ and $\{\bar{x}^i\}$ is a coordinate system on $U\subset M$. Since Λ is Π_0^0M-linear, we shall only give the interpretation of Λ upon associated fields, that is such as $\bar{v}\circ\pi$, with $\bar{v}\in V_0^1M$.. A horizontal homomorphism Λ is an assignment of a field $\Lambda v\in V_0^1\mathcal{F}M$ to the field $v=\bar{v}\circ\pi$ such that Λv projects upon v under the map π_*. In other words, integral curves of Λv are projectecd by π on integral curves of v. Or roughly speaking, *a horizontal homomorphism is an interpretation of derivatives along M as derivatives along $\mathcal{F}M$.*

6. LIFTS AND LOWERINGS

Our purpose is now to extend a pair of vertical and horizontal homomorphisms to vertical and horizontal lifts for arbitrary fields of ΠM.

The map $A:\Pi M\to V\mathcal{F}M$ is called a *horizontal (vertical) lift* if: (a) $A\,|\,\Pi_0^1M$ is horizontal (vertical) homomorphism; (b) A is a type preserving graded $\mathbb{R}$ algebra homomorphism; (c) if v is in annihilator of $\ker A\,|\,\Pi_1^0M$ and $s\in\Pi_1^0M$, then $\langle As,Av\rangle=\langle s,v\rangle$; and if s is in annihilator of $\ker A\,|\,\Pi_0^1M$ and $v\in\Pi_0^1M$, then $\langle As,Av\rangle=\langle s,v\rangle$.

Note that if A is horizontal, then $A\,|\,\Pi_0^1M$ is injective; thus every $s\in\Pi_1^0M$ belongs to the annihilator of $\ker A\,|\,\Pi_0^1M$; hence, if A is horizontal, condition (c) tells us that $\langle As,Av\rangle=\langle s,v\rangle$ for every s,v. Note also that for every lift we have that $Aa=a$ if $a\in\Pi_0^0M$.

The *lowering* B of a lift A is its transpose map $B:V\mathcal{F}M\to\Pi M$. In other words, B is the graded $\mathbb{R}$ algebra homomorphism such that $Ba=a$, $\langle B\mathbf{s},v\rangle=\langle\mathbf{s},A\mathbf{v}\rangle$, $\langle s,B\mathbf{v}\rangle=\langle As,\mathbf{v}\rangle$.

If A is horizontal, then $\langle s,BAv\rangle=\langle As,Av\rangle=\langle s,v\rangle=\langle BAs,v\rangle$, whence $BA=\mathrm{id}$. If A is vertical we have $BAB=B$, $ABA=A$. In fact, if $v\in\Pi_0^1M$ and $\mathbf{s}\in V_1^0\mathcal{F}M$, then $\langle\mathbf{s},ABAv\rangle=\langle B\mathbf{s},BAv\rangle$. But if $z\in\ker A\,|\,\Pi_0^1M$, then $\langle B\mathbf{s},z\rangle=\langle\mathbf{s},Az\rangle=0$. Hence $B\mathbf{s}$ belongs to the annihilator of $\ker A\,|\,\Pi_0^1M$. Therefore, $\langle B\mathbf{s},BAv\rangle=\langle AB\mathbf{s},Av\rangle=\langle B\mathbf{s},v\rangle=\langle\mathbf{s},Av\rangle$. Since $\mathbf{s}$ is arbitrary we have $ABA=A$ on Π_0^1M; in the same way $ABA=A$ on Π_1^0M; therefore, this relation holds on the whole ΠM. The proof for $BAB=B$ is similar.

The maps A,B have a local character, as is easily proved as customary. This means that if $v_r=w_r$, then $(Av)_r=(Aw)_r$ and so on.

The following definition will be useful for our purposes. Let A_1 be horizontal, A_2 vertical, and B_1, B_2 their respective lowerings. Then we say they form a *crossed lift pair* if $B_1A_2=B_2A_1=0$ on Π_s^rM for $(r,s)\neq(0,0)$ (on Π_0^0M, these homomorphisms are always the identity), and $A_1B_1+A_2B_2=\mathrm{id}$ on $V_0^1\mathcal{F}M$ and on $V_1^0\mathcal{F}M$.

Then we shall put $\mathbf{H}=A_1B_1$, $\mathbf{V}=A_2B_2$. Thus we have $\mathbf{H}^2=\mathbf{H}$, $\mathbf{V}^2=\mathbf{V}$, $\mathbf{HV}=\mathbf{VH}=0$ on $V_s^r\mathcal{F}M$ with $(r,s)\neq(0,0)$. Thus, $\mathbf{H}$ and $\mathbf{V}$ project fields of $V\mathcal{F}M$ into their horizontal and vertical components. These components sum the given field if it is a vector field or a 1-form because then $\mathbf{H}+\mathbf{V}=\mathrm{id}$.

Now we reach the fundamental result of this section:

Theorem: Given the vertical and horizontal homomorphisms j and Λ, they define a unique crossed lift pair A_1, B_1, A_2, B_2 satisfying $A_1|\Pi_0^1M=\Lambda$, $A_2|\Pi_0^1M=j$, $\langle c,B_2\mathbf{v}\rangle=0$ for every $\mathbf{v}\in V_0^1\mathcal{F}M$. Moreover, then $B_1|V_0^1\mathcal{F}M=\pi_*$.

Proof: First we prove $j(\Pi^1_0 M) \oplus \Lambda(\Pi^1_0 M) = V^1_0 \mathcal{F}M$. In fact, if $\mathbf{v} = jv = \Lambda w$, then for every $\bar{a} \in V^0_0 M$ we have $\mathbf{v}_r(\bar{a} \circ \pi) = (jv)_r(\bar{a} \circ \pi) = 0 = (\Lambda w)_r(\bar{a} \circ \pi) = w_r(\bar{a})$. Hence $w_r = 0$ and $\mathbf{v}_r = 0$. Thus, the intersection of those submodules is zero. Now, Λ is injective and the image of j equals the submodule of vertical vector fields. Therefore, the maps Λ_r, j_r defined at each $r \in \mathcal{F}M$ by $\Lambda_r v_r = (\Lambda v)_r$, $j_r v_r = (jv)_r$ have rank n and $n - 1$, respectively. Hence $j_r(M_{\pi r}) \oplus \Lambda_r(M_{\pi r}) = (\mathcal{F}M)_r$ because $(\mathcal{F}M)_r$ is $(2n - 1)$-dimensional. Now, it is a simple matter to extend this direct sum globally for having our first claim. As a consequence, if $\mathbf{v} \in V^1_0 \mathcal{F}M$, it can be written in a unique manner as $\mathbf{v} = \Lambda v_1 + jv_2$, where $v_1, v_2 \in \Pi^1_0 M$ and $\langle c, v_2 \rangle = 0$ (note that $\ker j$ is spanned by k, and $\langle c, k \rangle = 1$). We put $\langle A_1 s, \mathbf{v} \rangle = \langle s, v_1 \rangle$, $\langle A_2 s, \mathbf{v} \rangle = \langle s, v_2 \rangle$. These maps, together with Λ and j, in fact define the whole lifts A_1, A_2 satisfying our requirements. The proof is rather mechanical and is left to the reader. As for the assertion $B_1 | V^1_0 \mathcal{F}M = \pi_*$, we have $\langle s, B_1 \mathbf{v} \rangle = \langle A_1 s, \mathbf{v} \rangle = \langle s, v_1 \rangle$ if $\mathbf{v} = \Lambda v_1 + jv_2$. Then $\pi_* \mathbf{v} = \pi_* \circ \Lambda v_1 + \pi_* \circ jv_2 = \pi_* \circ \Lambda v_1 = v_1$ as we have seen in our interpretation of horizontal homomorphisms. Therefore, $B_1 \mathbf{v} = \pi_* \circ \mathbf{v}$.

Note that if $v \in \Pi^1_0 M$, it can be written as $v = (v - \langle c, v \rangle k) + \langle c, v \rangle k$. Thus, $A_2 v = j(v - \langle c, v \rangle k)$ and $\langle s, B_2 A_2 v \rangle = \langle A_2 s, A_2 v \rangle = \langle A_2 s, j(v - \langle c, v \rangle k) \rangle$ $= \langle s, v - \langle c, v \rangle k \rangle$ because $\langle c, v - \langle c, v \rangle k \rangle = 0$. Hence, $B_2 A_2$ is the identity on annihilator of c. On a similar way, $B_2 A_2$ is the identity on the annihilator of k.

7. CONNECTIONS

The map $\nabla : (v, h) \in \Pi^1_0 M \times \Pi M \to \nabla_v h \in \Pi M$ is called a *horizontal (vertical) connection on* ΠM if: (a) $\nabla_v : \Pi M \to \Pi M$ is a derivation of degree zero on the graded $\mathbb{R}$ algebra ΠM; (b) $\nabla_v a = Av(a)$, $A : \Pi^1_0 M \to V^1_0 \mathcal{F}M$ being a horizontal (vertical) homomorphism and $a \in \Pi^0_0 M$; (c) it is $\Pi^0_0 M$-linear in v, that is $\nabla_{av + bw} = a \nabla_v + b \nabla_w$; (d) if $s \in \Pi^0_1 M$ and $w \in \Pi^1_0 M$, then $\nabla_v \langle s, w \rangle = \langle \nabla_v s, w \rangle + \langle s, \nabla_v w \rangle$.

The map $\Delta : (\mathbf{w}, h) \in V^1_0 \mathcal{F}M \times \Pi M \to \Delta_{\mathbf{w}} h \in \Pi M$ is called a *physical connection* if: (a) $\Delta_{\mathbf{w}}$ is a derivation of degree zero on ΠM; (b) $\Delta_{\mathbf{w}} a = \mathbf{w}(a)$ for $a \in \Pi^0_0 M$; (c) it is $V^0_0 \mathcal{F}M$-linear in $\mathbf{w}$; (d) $\Delta_{\mathbf{w}} \langle s, v \rangle = \langle \Delta_{\mathbf{w}} s, v \rangle + \langle s, \Delta_{\mathbf{w}} v \rangle$.

From a geometric and physical viewpoint, physical connections are more natural than connections on ΠM, but these are easier to handle. We will use them as a tool for finding physical metric connections. However, both types have a physical significance. The meaning of physical connections is that they give the covariant derivatives of physical fields along the directions $\mathbf{w}$, that is, along curves on $\mathcal{F}M$; in other words, when we move from a point m at which the laboratory has direction r_m, to a point m' where the laboratory has direction $r_{m'}$, in such a manner than the points r_m and $r_{m'}$ of $\mathcal{F}M$ are detached between them by the vector $\mathbf{w}$ (roughly speaking).

Now, as another useful tool, we build the j connection D, which is a vertical connection on ΠM. It is defined by $D_v a = jv(a)$ if $a \in \Pi^0_0 M$, and $D_v(\bar{h} \circ \pi) = 0$ if $\bar{h} \in VM$. It is not difficult to prove the consistency of this definition. We have:

Theorem: If c is a time form such that $i\tilde{v}(\langle \tilde{c}, \tilde{v} \rangle) = \langle \tilde{c}, \tilde{v} \rangle$ (see Sec. 4), then $\langle D_k c, w \rangle = \langle D_w c, k \rangle = 0$, and $\langle D_v c, w \rangle = \langle D_w c, v \rangle$ for $v, w \in \Pi^1_0 M$.

Proof: Since these expressions are $\Pi^0_0 M$-linear in v, w, we can suppose that they are associated fields, that is $Dv = Dw = 0$. Thus, $\langle D_k c, w \rangle = D_k \langle c, w \rangle - \langle c, D_k w \rangle$ $= D_k \langle c, w \rangle = jk(\langle c, w \rangle) = 0$ because $jk = 0$. Now $(D_w \langle \mathrm{c}, \mathrm{v} \rangle) \circ p = \langle \tilde{c}, \tilde{u} \rangle i\tilde{w}(i\tilde{v}(\langle \tilde{c}, \tilde{u} \rangle)) = \langle \tilde{c}, \tilde{u} \rangle i\tilde{v}(i\tilde{w}(\langle \tilde{c}, \tilde{u} \rangle))$ because v, w are associated fields and iv, iw are ordinary derivatives (in the same sense used in $\mathbb{R}^n$) on the fibres of $\tilde{\pi}$. Hence $0 = D_w \langle c, v \rangle - D_v \langle c, w \rangle = \langle D_w c, v \rangle - \langle D_v c, w \rangle$. Therefore, $\langle D_w c, k \rangle = \langle D_k c, w \rangle = 0$.

This theorem tells us that $\eta = Dc$ defines a symmetric element of $\Pi^0_2 M$ such that $\eta(k,) = 0$. This field gives the rate of time retardation when the relative speed increases. In fact, let $\sigma(t)$ be the world line of a particle, and r a cross section of π, that is a laboratory. If $\dot{\sigma}(t)$ is the tangent to σ, we can roughly think of $\dot{\sigma}(t)$ as a vector joining two events in the world line, namely $\sigma(t)$ and $\sigma(t) + \dot{\sigma}(t)$. Then $\tau_r = \langle c_{r \circ \sigma(t)}, \dot{\sigma}(t) \rangle$ is the time interval, measured by the synchronized laboratory r, for the track of that particle between $\sigma(t)$ and $\sigma(t) + \dot{\sigma}(t)$. Thus, if σ remains fixed, this time interval depends on r only. Thus, $D_v \tau_r$ is the rate of variation of τ_r with respect to s, at $s = 0$, when we take laboratories $p(\bar{r} + \langle c_{r \circ \sigma(t)}, \bar{r} \rangle s\bar{v})$ measuring it (see Sec. 5), where we suppose $p\bar{r} = r \circ \sigma(t)$ and $\bar{v} = v_{r \circ \sigma(t)}$. That is, $D_v \tau_r$ is the rate of variation of τ_r when the speed of the laboratory changes towards the $\bar{v}$ direction. But $D_v \tau_r = \langle (D_v c)_{r \circ \sigma(t)}, \dot{\sigma}(t) \rangle$ $= \eta_{r \circ \sigma(t)}(\bar{v}, \dot{\sigma}(t))$. If $\bar{v}$ is a positive multiple of $\dot{\sigma}(t)$, this means we are approaching the laboratory speed to that of the particle because $\langle c_{r \circ \sigma(t)}, \bar{r} \rangle$ (proper time) is supposed to be positive. Then, if $\eta_{r \circ \sigma(t)}(\dot{\sigma}(t), \dot{\sigma}(t)) < 0$, we have that clocks relatively retard with respect to each other when their relative speed increases (as a thinking guide, bear in mind special relativity).

In relativity we have $\gamma + \eta = 0$. In Newtonian spacetime $\eta = 0$.

8. METRIC CONNECTIONS

If ${}^1\nabla, {}^2\nabla$ are horizontal and vertical connections on ΠM, respectively, then we have that ${}^1T(v, w) = {}^1\nabla_v w - {}^1\nabla_w v - B_1[A_1 v, A_1 w]$ and ${}^2T(v, w) = B_2 A_2({}^2\nabla_v w - {}^2\nabla_w v - B_2[A_2 v, A_2 w])$ are $\Pi^0_0 M$-bilinear operators, where A_1, B_1, A_2, B_2 is the crossed lift pair defined through the Theorem in Sec. 6 from the homomorphisms associated to these connections. These operators define elements ${}^1T, {}^2T \in \Pi^1_2 M$, called the *horizontal and vertical torsion*, respectively.

Thus, we say that ${}^1\nabla$ (${}^2\nabla$) is a *horizontal (vertical) metric connection* if ${}^1\nabla c = 0$, ${}^1\nabla \gamma = 0$, ${}^1T = 0$ (${}^2\nabla_c = 0$, ${}^2\nabla \gamma = 0$, ${}^2T = 0$).

Then, since $\gamma(k,) = 0$, we have that ${}^1\nabla_v \gamma(k\ ,)$ $= ({}^1\nabla_v \gamma)(k,) + \gamma({}^1\nabla_v k,) = \gamma({}^1\nabla_v k,) = 0$. Hence, ${}^1\nabla_v k$ must be a multiple of k; but $\langle c, k \rangle = 1$ and ${}^1\nabla c = 0$. Therefore, ${}^1\nabla k = 0$, and in the same way we can prove ${}^2\nabla k = 0$.

The problem of existence for horizontal metric connections is rather difficult; in Appendix A is proved that *if the signature of* γ *is* $(0, +, \ldots, +)$, *the signature of* η *is* $(0, -, \ldots, -)$ *and* c *satisfies* $i\tilde{v}(\langle\tilde{c},\tilde{u}\rangle) = \langle\tilde{c},\tilde{v}\rangle$, *then there is a unique horizontal metric connection on* IIM. The root of the difficulty is that we do not know *a priori* the horizontal homomorphism associated with that connection. It must be determined from our requirements together with the action of ${}^1\nabla$ upon vector fields.

Now, the increasing half-lifes of particles has been verified for many speeds and directions. Thus, we have an experimental reason for taking $(0, -, \ldots, -)$ as the signature of η. Assuming this for granted, there is a unique horizontal metric connection. In Newtonian spacetime, $\eta = 0$ and that connection, if it exists, is not unique; the existence condition is that $\bar{b}$ be an exact 1-form. This means a universal absolute time. The proof of that assertion is too long for bringing it here.

The physical meaning of ${}^1\nabla$ is the following: *It defines the absolute derivative of physical fields along spacetime* (cf. the interpretation of homomorphisms given in Sec. 5) *from a laboratory whose particles are each other at relative rest* (at the limit when these particles are close to the event where the derivative is taken). The reason for this last remark is that ${}^1\nabla k = 0$, and k could be looked at, in some respects, as the laboratory field. An account for this interpretation is given in Ref. 4.

As for the vertical metric connection, it defines the absolute derivative of fields along the fibres of π, having γ and c as an absolute measure for the directional dependence of fields. That vertical metric connection also is uniquely determined, and given by

$$
\begin{aligned}
{}^2\nabla_v w = D_v w + \tfrac{1}{2} g^{-1}((D_v g)(w,) + (D_w g)(v,) - (Dg)(v,w),) \\
- g(v,w)k - \langle c,w\rangle v,
\end{aligned}
$$

where we have put $g = \gamma - c \otimes c$ (see Appendix B).

9. LIFTING CONNECTIONS

If Δ is a physical connection and A,B is a lift lowering, then $\nabla_v w = \Delta_{A_v} w$ defines a connection on IIM. The following assertion justifies our use of connections on IIM:

Given the horizontal and vertical metric connections ${}^1\nabla$, ${}^2\nabla$ *there is a unique physical connection* Δ *giving* ${}^1\nabla$ *and* ${}^2\nabla$ *through the above process. It is metric in the sense that* $\Delta c = 0$ *and* $\Delta\gamma = 0$.

For if Δ satisfies that condition, then $\Delta_w h = \Delta_{\mathbf{H}w} h + \Delta_{\mathbf{V}w} h = {}^1\nabla_{B_1 w} h + {}^2\nabla_{B_2 w} h$ for every $h \in \mathit{IIM}$, $w \in V^1_0\mathcal{F}M$. Now it is a trivial matter to prove this formula effectively gets a physical connection. Moreover $\Delta\gamma = 0$ and $\Delta c = 0$ because ${}^2\nabla_\gamma = {}^1\nabla_\gamma = 0$ and ${}^1\nabla c = {}^2\nabla c = 0$. Also we have $\Delta k = 0$.

The formula giving Δ is rather striking: It manifests itself our way to get it. It splits in two terms, corresponding to the horizontal and vertical components of $\mathbf{w}$, that is, of the tangent to the curve on $\mathcal{F}M$ along which we compute the derivative. Thus, it does not require a more detailed explanation.

Unfortunately, physical connections are awkward to handle because of their mixed nature. Due to this, we shall lift Δ for having an ordinary linear connection on $V\mathcal{F}M$. The process is the following.

Let A_1, B_1 be the horizontal lift lowering given by the metric connections ${}^1\nabla$, ${}^2\nabla$. Then, if $\mathbf{D}$ is a linear connection on $V\mathcal{F}M$, we have that $\Delta_w h = B_1 \mathbf{D}_w A_1 h$ defines a physical connection. We shall demand that Δ should be *the* physical metric connection we have just defined.

As for A_2, B_2, the formula $B_2\mathbf{D}_w A_2 h$ does not define a physical connection because B_2A_2 is not the identity. But $B_2A_2B_2 = B_2$, whence the preceding formula defines a physical connection on $B_2(V\mathcal{F}M)$ that is a subalgebra of IIM. But our physical metric connection also is a connection on this subalgebra, because $B_2(V^1_0\mathcal{F}M)$ is the annihilator of c, and $B_2(V^0_1\mathcal{F}M)$ is the annihilator of k. For if $\langle s,k\rangle = 0$, then $\langle \Delta_w s,k\rangle = -\langle s,\Delta_w k\rangle = 0$; also, if $\langle c,v\rangle = 0$, then $\langle c,\Delta_w v\rangle = 0$. So we shall demand that $\Delta_w B_2\mathbf{h} = B_2\mathbf{D}_w A_2 B_2\mathbf{h}$ for every $\mathbf{h} \in V\mathcal{F}M$. In addition, we demand that the parallel displacement given by $\mathbf{D}$ should apply horizontal vectors into horizontal vectors; in other words, that $\mathbf{DH} = 0$.

Theorem: There is a unique linear connection $\mathbf{D}$ on $V\mathcal{F}M$ such that
$\Delta_w h = B_1\mathbf{D}_w A_1 h$, $\Delta_w B_2\mathbf{h} = B_2\mathbf{D}_w A_2 B_2\mathbf{h}$, $\mathbf{DH} = 0$.

Proof: Note first that $\mathbf{H}$ linearly applies $V^r_s\mathcal{F}M$ into $V^r_s\mathcal{F}M$; hence, each restriction $\mathbf{H}|V^r_s\mathcal{F}M$ can be looked at as a tensor field of type $(r+s, r+s)$; in this sense, $\mathbf{DH}$ has a definite meaning. If $\mathbf{D}$ is the required connection, then
$\mathbf{D}_w\mathbf{v} = +\mathbf{D}_w\mathbf{H}\mathbf{v} + \mathbf{D}_w\mathbf{V}\mathbf{v} = \mathbf{D}_w\mathbf{H}^2\mathbf{v} + \mathbf{D}_w\mathbf{V}^2\mathbf{v}$
$= \mathbf{H}\mathbf{D}_w\mathbf{H}\mathbf{v} + \mathbf{V}\mathbf{D}_w\mathbf{V}\mathbf{v}$ because $\mathbf{H} + \mathbf{V} = \mathrm{id}$ on $V^1_0\mathcal{F}M$ and as a consequence $\mathbf{DV} = 0$. Thus $\mathbf{D}_w\mathbf{v}$
$= A_1B_1\mathbf{D}_w A_1B_1\mathbf{v} + A_2B_2\mathbf{D}_w A_2B_2\mathbf{v} = A_1\Delta_w B_1\mathbf{v} + A_2\Delta_w B_2\mathbf{v}$.
Hence, if such a linear connection exists, it is unique and given by the above formula (valid for elements of $V^1_0\mathcal{F}M$ and $V^0_1\mathcal{F}M$; for other tensor types, the expression is more complicated). Now it is a trivial exercise to prove that formula fulfill our demands.

10. GEOMETRY ON THE TIME BUNDLE

We look at γ and c as the primordial geometric features of spacetime. From them, we build unique vertical and horizontal metric connections, and they define the physical metric connection, which describes the absolute derivative of physical fields along the time manifold. Also we have the linear connection $\mathbf{D}$ that could be regarded as getting *the geometry of the time bundle itself*; in fact, the torsion of $\mathbf{D}$, its curvature and Ricci fields, Bianchi identities, etc., can now be computed as customary. Thus, our goal has been reached: we have translated the problem of spacetime geometry to the geometry of the time manifold, the manifold where the observable physics takes place. This lifting process has the advantage of recovering the usual techniques of differential geometry.

However, to tell the truth, I have some doubts about this process, in the following sense. One could also say that η

measures the increasing relative energy when the relative speed increases, because time retardation and relative energy are directly related in relativity and quantum mechanics. Therefore, η would stand for the *vertical potential* in the manner as in relativity $\bar{g}$ stands for the gravitational (horizontal) potential. Then, this symmetry lends some strength to the definition of vertical metric connections through ${}^2\nabla c = 0$, ${}^2\nabla \eta = 0$, ${}^2T = 0$. On this assumption the lifting process for connections becomes the same because we also have that $\eta(k,) = 0$; but then $\Delta\gamma$ and $\Delta\eta$ are in general different from zero. Thus, what is the appropriate field, γ or η, to be used for defining a metric on the fibres of π? Relativity is not an aid because then $\gamma + \eta = 0$, whence the choice does not matter. But in Newtonian spacetime, $\eta = 0$; thus, no vertical distance among velocities?, no relative energy?, no inertia? These strange outcomes and the nonmetric character of Δ compel me to prefer γ instead η.

Disregarding these doubts, I believe this process is not merely a desperate issue from an unnecessarily puzzled starpoint; on the contrary, it seems to me more natural than the relativistic one, because it allows a step by step construction of different models of spacetime, clearing up the different options one must take for having different theories.

II. SPACELIKE LENGTH AND SPACETIME

1. INTRODUCTION

Until now, we have considered as independent data the time form and the spacelike metric. But are they independent magnitudes? In relativity the answer is no, because then $\gamma + Dc = \gamma + \eta = 0$, and there is experimental evidence favoring some link between γ and η—the Michelson–Morley experiment for instance.

Let us accept that link, but suppose that $\tilde{f}$ is a general time function, perhaps not a relativistic one. From $\tilde{f}$ we build c and $Dc = \eta$. Suppose the signature of η is everywhere $(0, -, \ldots, -)$. Increasing half-lifes is the experimental support for this assumption. Then, it seems a suggestive attitude to postulate that the relation between γ and η is the same as the relativistic one, i.e., $\gamma + \eta = 0$. In other words, we are *defining* the spacelike metric as $\gamma = -Dc$. From this point, we could apply the techniques of Part I for reaching a geometry of the time manifold. That would be the track of a *pure chronogeometry: to reject meters, adopt clocks and build lengths from times.* Classical chronogeometry in additon postulates that $\eta + c \otimes c = \bar{g} \circ \pi$, with $\bar{g} \in V_2^0 M$, that is we can mix these magnitudes for having a Lorentz metric.

So far I do not know examples of the opposite viewpoint: *to reject clocks, adopt meters, and build times from lengths.* My own position is the construction of a very general spacetime geometry from the datum of a spacelike metric. At least I judge this task convenient, as complementary with respect to chronogeometry. Moreover, I find some physical arguments favoring my position. First, η and γ have very different physical meanings: η stands for the rate of time retardation, and γ for spacelike length as measured by metersticks; thus, the relation $\gamma + \eta = 0$ seems rather accidental. Second, I think of time as a more dynamic feature than spacelike length, whence also a more secondary datum from our methodological viewpoint (see Part I, Sec. 4); I believe that is in the same line of thought as the way in which superspace theories are going; that is, spacelike length carries information about time, but can we say that time carries information about space? Third, Pythagoras' theorem, on which our theory leans, has always been verified at the macroscopic level, and always supposed at the microscopic one.

Thus, our departure point is a spacelike metric, that is a field $\gamma \in \Pi_2^0 M$, symmetric, of signature $(0, +, \ldots, +)$, and such that $\tilde{\gamma}(\tilde{u},) = 0$. This field describes Pythagoras' theorem at each laboratory (Sec. 2).

The key point of the paper is Sec. 3. On it, we define a simultaneity from γ through the criterion of stationary spacelike volume. It is a generalization of the oldest definition of simultaneity, that given by a person saying: "I cannot be in two places at the same time!" He signifies that he cannot reduce the distance (relative to him) between two events happening at different places if they are simultaneous. We will take volume instead of distance, but the basic point is the same: to take spacelike measures instead of interchanging signals for defining the simultaneity. The criterion of spacelike volume gives a time form c under a multiplicative function.

Each choice of that function defines a time function; we demand that time function to be consistent with the time form by means of infinitely slow clock transport (Sec. 4). However, this requirement does not entirely determine the time form; the equivalence among these consistent time forms gives raise to a gauge (Sec. 5).

In Sec. 6 we characterize our geometric model of space time in terms of a nonsingular symmetric field $g \in \Pi_2^0 M$. Some examples are shown.

Gauge invariance makes the definitions of metric connections on ΠM more difficult. Along the study of this problem, a field $\phi \in \Pi_1^0 M$ arises (Sec. 7). It determines the horizontal metric connection, and perhaps could be interpreted as the electromagnetic potential.

In Sec. 8 we apply the techniques of Part I for lifting connections, and so reach a physical metric connection which is gauge invariant, and a linear connection on $V\mathcal{F}M$ giving the geometry of the time bundle.

Section 9 contains some comments about our results.

2. PYTHAGOREAN SPACELIKE METRIC

For a better understanding, we will translate back and forth our constructive process from the *special* to the *general* case, in a similar manner to that of special and general relativity.

In the special case, spacetime is considered as the four-dimensional affine space. Geometric features of spacetime, that is γ or c, are supposed to be independent of events; they could perhaps depend on laboratory directions. As in special relativity or classical mechanics, if no forces act upon a particle, its world line is straight. An inertial laboratory is now a

set of solidary particles, i.e., whose world lines are parallel straight lines. Any nonzero vector tangent to them must be timelike by definition. So, each laboratory is characterized by a timelike vector $\bar{x}$ or by any nonvanishing multiple of it. Let us consider ourselves traveling with the laboratory $\bar{x}$. Given an arbitrary event, it appears *located* at a well defined point of our laboratory. In spacetime language, location is the world line of the particle of our laboratory whose history contains the given event. Events which happened at the same point of our laboratory must have the same location, no matter the time elapsed among them. If two events are given, we can measure the distance between their locations by means of a meter stick *at rest in our laboratory*. Obviously, this is the ordinary method of spacelike length measurements among events: The bottle carrying the help message was found *four thousand miles* away from the wreckage····

The resulting quantity depends on the vector $\bar{y}$ joining both events (4-vector of spacetime). But it is clear that it also depends on the selected laboratory, that is on $\bar{x}$. Now, we assume that Pythagoras' theorem holds at each laboratory. In other words, the spacelike length of $\bar{y}$ at $\bar{x}$ is given by $\tilde{\gamma}_{\bar{x}}(\bar{y},\bar{y})$, where $\tilde{\gamma}_{\bar{x}}$ is a quadratic form that depends on $\bar{x}$, but not (in the special case) on the events of spacetime. Obvious properties of this field $\tilde{\gamma}:\bar{x}\to\tilde{\gamma}_{\bar{x}}$ are: (a) $\tilde{\gamma}_{\bar{x}}(\alpha\bar{x},\ ) = 0$ for every $\alpha\in\mathbb{R}$, because $\alpha\bar{x}$ stands for the vector joining two events having the same location at the laboratory $\bar{x}$; (b) $\tilde{\gamma}_{\bar{x}} = \tilde{\gamma}_{\alpha\bar{x}}$ for $\alpha > 0$ since $\bar{x}$ and $\alpha\bar{x}$ stand for the same laboratory (hence we say that $\tilde{\gamma}$ is homogeneous of degree zero); (c) $\tilde{\gamma}_{\bar{x}}(\bar{y},\bar{y}) > 0$ if $\bar{y}$ is not a multiple of $\bar{x}$.

By a standard generalization, in the general case a laboratory will be a local cross section of $\pi:\mathcal{F}M\to M$, the time bundle, and γ will become a symmetric element of $\Pi^0_2 M$, of signature $(0,+,\ldots,+)$, and satisfying $\tilde{\gamma}(\tilde{u},\ ) = 0$, that is $\tilde{\gamma}(\tilde{u},\ )_{\bar{x}} = \tilde{\gamma}_{\bar{x}}(\bar{x},\ ) = 0$ (see Part I, Sec. 4).

3. SIMULTANEITY FROM SPACELIKE METRIC

Our problem is now the discovery of a simultaneity linked to the spacelike metric γ. The process is performed in two steps: imposing both the condition of stationary spacelike volume and that of infinitely slow clock transport synchronization. In terms of Part I, we look for time forms privileged with respect to γ; let us discuss what kind of privilege it is.

At this point it is interesting to remark that the preceding description of γ is by no means restricted to a particular class; thus, since we are looking for a generalization, it would be desirable that our definition of privileged time forms could be consistently applicable to Newtonian or relativistic spacelike metrics, considered as simple and extreme examples.

In classical Newtonian spacetime or in special relativity we can verify without difficulty the following argument (we are in the special case), whose rigorous proof is the theorem in Sec. 6. Let A, B, C, D be four events determining a hyperplane. If $\bar{x}$ is a laboratory, we can measure by means of $\tilde{\gamma}_{\bar{x}}$ the volume of the tetrahedron determined by the locations of these events in the laboratory $\bar{x}$ (or $\bar{x}$ locations). Let $V(\bar{x})$ be that volume. We call $V(\bar{x})$ the spacelike volume of the (A,B,C,D)-$\bar{x}$ locations. If the field $\bar{x}\to\tilde{\gamma}_{\bar{x}}$ is smooth and A,B,C,D remain fixed, then $V:\bar{x}\to V(\bar{x})$ is a differentiable function; let dV be its differential. Suppose that dV vanishes at some laboratory $\bar{x}_0$, i.e., $(dV)_{\bar{x}_0} = 0$. This means that the spacelike volume of the (A,B,C,D)-$\bar{x}$ locations is stationary at $\bar{x}_0$. If this occurs, then in Newtonian mechanics the four events are *pairwise absolutely simultaneous*. In special relativity, we conclude that *the hyperplane A,B,C,D is spacelike and that $\bar{x}_0$ is orthogonal to that hyperplane*; *or*, equivalently, *the four events are pairwise simultaneous as viewed from the laboratory $\bar{x}_0$*. Moreover, in both spacetime theories, given the timelike vector $\bar{x}_0$, *there is a unique hyperplane whose spacelike volume is stationary at $\bar{x}_0$*, in the above sense (strictly speaking, a distribution of parallel hyperplanes). Thus we can say that such a hyperplane is privileged at $\bar{x}_0$ with respect to $\tilde{\gamma}$, since the laboratory and its corresponding stationary spacelike volume hyperplane are related by simultaneity.

The same idea serves us for defining priviledged time forms from the spacekike metric, though it should not be Newtonian or relativistic. Suppose that c is a 1-form determining a distribution of parallel hyperplanes (we keep in the special case). Choose one of them, say, H. As before, let A,B,C,D be four fixed events determining H. Let $V(\bar{x})$ be the spacelike volume of the (A,B,C,D)-$\bar{x}$ locations, as measured by means of $\tilde{\gamma}_{\bar{x}}$. Suppose $(dV)_{\bar{x}_0} = 0$; obviously this condition does not depend on the chosen four events belonging to the fixed H. Thus, we say the spacelike volume of H is stationary at $\bar{x}_0$, and that *the events belonging to H are by definition pairwise simultaneous with respect to the laboratory $\bar{x}_0$*.

Our basic requirement upon $\tilde{\gamma}$ is: Consider the subset of timelike vectors for each of them, $\bar{x}$, there is one unique (up to a multiplicative nonzero constant) 1-form $\tilde{b}_{\bar{x}}$ whose associated hyperplanes are of stationary spacelike volume at $\bar{x}$, and such that $\langle\tilde{b}_{\bar{x}},\bar{x}\rangle\neq 0$. Then, *this subset is supposed to be nonempty and open, and it constitutes our final set of timelike vectors*. Our additional demand is: there is a representant $\tilde{c}_{\bar{x}}$ of each $\{\alpha\tilde{b}_{\bar{x}}\}_{\alpha\neq 0}$ such that the field $\tilde{c}:\bar{x}\to\tilde{c}_{\bar{x}}$ is smooth and homogeneous of degree zero. This last field is called *privileged time form*, and it defines the synchronization associated to $\tilde{\gamma}$.

In the general case this question becomes rather technical; a detailed account is given in Ref. 4. A brief sketch is the following. Let $\Sigma\subset M$ be a hypersurface of M, and B a compact regular domain of Σ, contained in the domain of some chart of Σ. Let $\{\bar{f}_\alpha\}$ be the coordinate vector fields of this chart, and $\{\bar{s}^\alpha\}$ the dual base. Let $\bar{r}$ be a cross section of $\bar{\pi}$ such that $\bar{r}_m$ is not tangent to Σ for $m\in\Sigma$. Then, $\bar{\gamma} = \gamma\circ p\circ\bar{r}$ is a positive definite quadratic form when it acts upon $T^1_0\Sigma$. Thus, $\bar{\gamma}$ defines a volume form on Σ. Hence, the volume of B given by that volume form can be interpreted as the spacelike volume of B, as measured from the laboratory $\bar{r}$. It is given by $V(\bar{r}) = \int_B |\bar{\gamma}(\bar{f}_\alpha\bar{f}_\beta)|^{1/2}\bar{s}^1\wedge\cdots\wedge\bar{s}^{n-1}$, where $|\ |$ stands for determinant. If B remains fixed, this integral defines a functional on the field $\bar{r}$. Let us put $\tilde{\gamma} = \gamma\circ p$ and $\tilde{f}_\alpha = \bar{f}_\alpha\circ\tilde{\pi}$. By applying usual variational techniques, we find that $V(\bar{r})$ is stationary at $\bar{r}$ if we have $(i\tilde{v})_{\bar{r}_m}(|\tilde{\gamma}(\tilde{f}_\alpha\tilde{f}_\beta)|) = 0$ for every $m\in B$ and $\tilde{v}\in\tilde{\Pi}^1_0 M$.

This is a point-by-point condition, and it does not depend on the choice of $\{\bar{f}_\alpha\}$. In other terms, it only depends on the inclination of Σ at each point. Thus, we can build, at each $m \in M$ the set $T_m \subset \tilde{\mathcal{F}}_m$ of vectors as $\bar{x}_m$, for each of them there is unique n-1-dimensional subspace $\{\bar{f}_\alpha\}$ of M_m, not containing $\bar{x}_m$, such that $(i\tilde{v})_{\bar{x}_m}(|\tilde{\gamma}(\tilde{f}_\alpha \tilde{f}_\beta)|) = 0$ for every $\tilde{v} \in \tilde{\Pi}^1_0 M$ (note that $i\tilde{v}$ is a derivation along the fibres of $\tilde{\pi}$). Our abstract model of spacetime, the *time-elements space* (TES), consists of a manifold M, a time bundle $\pi: \mathcal{F}M \to M$, a spacelike metric field γ such that $p(T_m) = \pi^{-1}(m)$ for every $m \in M$, and an element $c \in \Pi^0_1 M$ such that c_{r_m} determines the unique subspace of M_m of stationary spacelike volume at $r_m \in \pi^{-1}(m)$. We call c a privileged time form, and c_{r_m} the simultaneity associated to r_m.

4. CLOCK TRANSPORT SYNCHRONIZATION

If c is a privileged time form, then qc also is a privileged time form, whenever $q \in \Pi^0_0 M$ is everywhere nonvanishing. Thus, the choice of q defines the time function $\langle e_a(qc), \tilde{u} \rangle$, that is the time length scale at each laboratory. Now, can this function q be arbitrarily picked without contradiction?

Let us return to the special case. If $\bar{x}$ is a laboratory, we will call the $\bar{x}$ clock an apparatus, at rest in $\bar{x}$, which computes time intervals among events of its history by means of $\tilde{c}_{\bar{x}}$. Equivalently, $\langle \tilde{c}_{\bar{x}}, \bar{x} \rangle$ is the time interval measured by the $\bar{x}$ clock between two events of its history, detached each other by the vector $\bar{x}$.

Consider two laboratories, $\bar{x}$ and $\bar{x}'$. Suppose the $\bar{x}'$ clock lying at the spacelike origin of $\bar{x}'$, passes, at some event, next to the $\bar{x}$-clock of the origin of $\bar{x}$. At that event, both clocks are set to zero. Suppose that all $\bar{x}$ clocks are synchronized among them by the condition of stationary spacelike volume, that is, through $\tilde{c}_{\bar{x}}$. Now, does *that* $\bar{x}'$ clock point to the same hour as the $\bar{x}$ clocks it is passing by? There are few chances for getting this agreement by a suitable choice of q. With a Newtonian spacelike metric, the agreement is possible; thus, *absolute universal time is, from our viewpoint, a consequence of Newtonian spacelike metric*! In special relativity, the answer is no.

However, in relativity an intermediate thing can be achieved, the agreement when the clock transport is "infinitely slow," the limit case when relative speed approaches zero (I believe this is Eddington's idea). So, could we require this weaker agreement with all generality? The answer is affirmative. In fact, suppose that the $\bar{x}'$ clock starts from the event A. Both this clock and the $\bar{x}$ clock at A, point to zero at A. After a while, the $\bar{x}'$ clock reaches another $\bar{x}$ clock at the event B. Let $\bar{z}$ be the vector joining A with B. Then $\bar{z}$ can be decomposed as $\bar{z} = \bar{y} + \alpha\bar{x}$ where $\langle \tilde{c}_{\bar{x}}, \bar{y} \rangle = 0$ and $\alpha \in \mathbb{R}$. If all $\bar{x}$ clocks are synchronized, at B the $\bar{x}$ clock points to $\langle \tilde{c}_{\bar{x}}, \alpha\bar{x} \rangle$, and the $\bar{x}'$ clock, to $\langle \tilde{c}_{\bar{y}+\alpha\bar{x}}, \bar{y} + \alpha\bar{x} \rangle$. Then, we demand that $\lim_{\alpha \to \infty}(\langle \tilde{c}_{\bar{y}+\alpha\bar{x}}, \bar{y} + \alpha\bar{x} \rangle - \langle \tilde{c}_{\bar{x}}, \alpha\bar{x} \rangle) = 0$. Now, $\tilde{c}$ is h (0); thus that expression becomes $\lim_{\alpha \to \infty} \langle \tilde{c}_{\bar{x}+\beta\bar{y}} - \tilde{c}_{\bar{x}}, \bar{y} + \alpha\bar{x} \rangle = \lim_{\beta \to 0} \langle (\tilde{c}_{\bar{x}+\beta\bar{y}} - \tilde{c}_{\bar{x}})/\beta, \bar{x} \rangle$, where we have put $\beta = 1/\alpha$. Then, since $\lim_{\beta \to 0}(\tilde{c}_{\bar{x}+\beta\bar{y}} - \tilde{c}_{\bar{x}})/\beta = \langle \tilde{c}_{\bar{x}}, \bar{x} \rangle^{-1}(D_y c)_{p\bar{x}}$, we conclude that infinitely slow clock transport agrees with stationary volume synchronization iff $\langle D_y c, k \rangle = 0$. Now, it is not difficult to see that this condition is equivalent to that of $i\bar{y}(\langle \tilde{c}, \tilde{u} \rangle) = \langle \tilde{c}, \bar{y} \rangle$.

In the general case, infinitely slow clock transport is nonsense. The best we could do is the following (this new process is equivalent to infinitely slow clock transport in the special case). Let $\sigma: [a,b] \to M$ be a world line, i.e., $\dot{\sigma}(t)$ belongs to $\tilde{\mathcal{F}}M$. Let $\bar{r}$ be a cross section of $\tilde{\pi}$, and c a privileged time form. Then, the time elapsed from $\sigma(a)$ to $\sigma(b)$, as measured by the synchronized laboratory $\bar{r}$ is $\int_a^b \langle \tilde{c}_{\bar{r}\circ\sigma}, \dot{\sigma} \rangle dt = \tau(\bar{r})$. If σ remains fixed, this integral is a functional on $\bar{r}$. Infinitely slow clock transport here means that $\bar{r}$ approaches $\dot{\sigma}$ on σ; the forementioned agreement translates into the condition that τ would be stationary *when* $\bar{r} \circ \sigma = \dot{\sigma}$. By requiring this for every world lines we easily find that $\langle D_v c, k \rangle = 0$ is the necessary and sufficient condition. If it is fulfilled, we can say that the "proper time" is an extremum for every world lines (in comparison with the time lapse measures performed from other laboratories); or, in the special case, that c gives the same synchronization as infinitely slow clock transport. In both cases, we say that c is a *fundamental time form*.

5. THE GAUGE

As for existence of fundamental time forms, see Sec. 6.

Suppose that c is fundamental. Then, if $\bar{a} \in V^0_0 M$ is everywhere nonzero and we put $a = \bar{a} \circ \pi$, then we have $\langle D_v(ac), k \rangle = 0$ because $D(\bar{a} \circ \pi) = 0$. Therefore, ac also is fundamental. All these fundamental time forms will be regarded as equaly valid for describing geometric features of spacetime. Then, the gauge for deciding if a geometric object is physically consistent must be its invariance under the transformation $c \to ac$, where $a = \bar{a} \circ \pi$ is everywhere nonvanishing.

Since $\bar{a}$ does not depend on directions, that transformation simply means certain change of time unities on each event. But $\bar{a}$ could depend on events of spacetime. Therefore, *we cannot get an absolute comparison among time scales at different events of spacetime; however, at the same event, time scales for clocks with different speed can be absolutely compared with respect to each other.* Now assume that two *real clocks* depart from an event A and travel along different paths, so that they meet at B. Someone might ask if the relative tick rythm of both clocks in B is different from that on A. Whereas this question makes sense in itself, it is not relevant here, because my time is a spacewise time, and I do not know whether the time of the *real* clocks agree with it. As it has been suggested to me, perhaps this means that this theory embodies in some nonquantic manner the following quantic assertion: The uncertainty principle prevents one from knowing both the metric of a spacelike slice and its respective extrinsic curvature.

6. THE TIME ELEMENTS SPACE

Now, suppose the γ defines a TES and c is a fundamental time form. Then, $g = \gamma - c \otimes c$ defines a symmetric element of $\Pi^0_2 M$ of signature $(-, +, \ldots, +)$. After rather long computations[4] we can characterize TES's through the following:

Theorem: Consider a given time manifold $\mathcal{F}M$, and let

$g \in \Pi^0_2 M$ be symmetric, of signature $(-,+,\ldots,+)$, and such that $\tilde{g}(\tilde{u},\tilde{u}) < 0$. Define $k = e_b(\tilde{u}/[-\tilde{g}(\tilde{u},\tilde{u})]^{1/2})$, $c = -g(k,)$, $\gamma = g + c \otimes c$. Thus, if $\langle D_v c, k \rangle = 0$ and $D_v|g(f_i f_j)| = 0$ for every $v \in \Pi^1_0 M$ and $f_i = \bar{f}_i \circ \pi$ with $\bar{f}_i \in V^1_0 M$ $(i:0,1,\ldots,n-1)$, then γ defines a TES on $\mathcal{F}M$, and c is a fundamental time form. Conversely, if γ defines a TES on $\mathcal{F}M$ and c is a fundamental time form, the field $g = \gamma - c \otimes c$ satisfies the above requirements.

The theorem enables us to build fundamental time forms. If c is a privileged time form, it is enough to multiply c by a function $q \in \Pi^0_0 M$ in order to make $|g|$ constant along each fibre. Also, it facilitates the construction of TES models. Besides the relativistic one, which trivially satisfies the theorem, the generalized Newtonian spacetime (see Part I, Sec. 4) defines another TES.

We also can alloy relativistic and Newtonian theories in the following way. Suppose M admits a Lorentz metric $\bar{g} \in V^0_2 M$, whence also a timelike one-dimensional distribution. We put $\bar{b} = \bar{g}(\bar{x},)/(-\bar{g}(\bar{x},\bar{x}))^{1/2}$ and $\bar{q} = \bar{g} + \bar{b} \otimes \bar{b}$, where $\bar{x}$ lies in that distribution. Let $\bar{N}, \bar{R} \in V^0_0 M$ be scalar fields such that $\bar{N} + \bar{R} = 1$ (the alloy ratios). By means of $\bar{g}$ we build the relativistic spacelike metric γ_R, and by means of $\bar{q}$ and $\bar{b}$ the Newtonian one, γ_N (see Part I, Sec. 4). Define the time bundle by $p(\bar{y}) \in \bar{\mathcal{F}}_m$ if $\bar{y} \in M_m$, $\langle \bar{b}_m, \bar{y} \rangle \neq 0$, $\bar{g}_m(\bar{y},\bar{y})(\bar{N}_m \bar{g}_m(\bar{y},\bar{y}) - \bar{R}_m \langle \bar{b}_m, \bar{y} \rangle^2) > 0$. Then $\gamma = N\gamma_N + R\gamma_R$, where $N = \bar{N} \circ \pi$ and $R = \bar{R} \circ \pi$, defines the mixed Newtonian relativistic TES on $\mathcal{F}M$. The proof of this assertion is rather long, and for the sake of brevity I prefer to not write it down. This TES has some bizarre properties: For example, its time bundle admits speeds greater than light.

The preceding theorem excludes from our scheme the old theories with an interval given by $ds = (-g_{ij}dx^i dx^j)^{1/2} + (e/m)A_i dx^i$, and a metric field defined through the Cartan technique. In fact, that metric field would have its determinant constant along each fibre iff $A_i = 0$.

7. GAUGE INVARIANCE AND CONNECTIONS

We are interested on connections that should be compatible with the geometric structure given by γ and fundamental time forms. The gauge invariant properties of this structure are: spacelike metric, simultaneity that is the condition $\langle c,v \rangle = 0$, and fibre constancy of $|\gamma - c \otimes c|$.

Suppose that j is the vertical homomorphism associated to c and that ${}^2\nabla$ is a vertical connection such that ${}^2\nabla\gamma = 0$, ${}^2\nabla c = 0$, ${}^2T = 0$, ${}^2\nabla_v b = jv(b)$ if $b \in \Pi^0_0 M$. If $c \to c' = ac$ is a gauge transformation, we have ${}^2\nabla c' = {}^2\nabla ac = a\,{}^2\nabla c = 0$ because $a = \bar{a} \circ \pi$. Thus, our definition of vertical metric connections goes without changes. That is, to each fundamental time form c we attach a vertical metric connection ${}^2\nabla$ on ΠM, the one satisfying ${}^2\nabla c = 0$, ${}^2\nabla\gamma = 0$, ${}^2T = 0$, ${}^2\nabla_v b = jv(b)$. This connection is unique and defined by the formula which appears in Part I, Sec. 8; but it is not gauge invariant because the connection attached to c' is ${}^2\nabla' = a\,{}^2\nabla$. This is not a bad feature, as we shall see in the following section.

The definition of horizontal metric connections requires more care. We will say that ${}^1\nabla$ is a horizontal metric connection if it is torsionless and:

(a) ${}^1\nabla\gamma = 0$, that is, ${}^1\nabla$ preserves spacelike length;

(b) $\langle {}^1\nabla_v c, w \rangle = 0$ if $\langle c,w \rangle = 0$; hence, ${}^1\nabla$ preserves simultaneity;

(c) $D_v(g^{-1} \lrcorner {}^1\nabla_z g) = 0$ whenever $v \in \Pi^1_0 M$ and $z = \bar{z} \circ \pi$ ($\lrcorner$ stands for double contraction). Equivalently, ${}^1\nabla$ preserves the fibre constancy of $|g|$.

Condition (a) tells that ${}^1\nabla_v k$ must be a multiple of k because $\gamma(k,) = 0$ and ${}^1\nabla\gamma = 0$. Thus, there must be some $\phi \in \Pi^0_1 M$ such that ${}^1\nabla_v k = -\langle \phi, v \rangle k$. Condition (b) implies that ${}^1\nabla_v c$ must be a multiple of c; but $\langle c,k \rangle = 1$. Therefore, ${}^1\nabla_v c = \langle \phi, v \rangle c$. Then ${}^1\nabla_z g = {}^1\nabla_z(\gamma - c \otimes c)$ $= -2\langle \phi, z \rangle c \otimes c$. Since $g^{-1}(c,)' = -k$, then $g^{-1} \lrcorner {}^1\nabla_z g = 2\langle \phi, z \rangle$ and $D_z(g^{-1} \lrcorner {}^1\nabla_z g) = 2\langle D_v \phi, z \rangle = 0$, because $z = \bar{z} \circ \pi$. Therefore, ϕ must be associated to some element $\bar{\phi} \in V^0_1 M$, that is, $\phi = \bar{\phi} \circ \pi$.

In Appendix A we prove that such a connection exists and is unique.

Now, let $c \to c' = ac$ be a gauge transformation. Then, if ${}^1\nabla_v c = \langle \phi, v \rangle c$, we have ${}^1\nabla_v c' = \langle (\bar{\phi} + d \ln \bar{a}) \circ \pi, v \rangle c'$ because ${}^1\nabla$ is horizontal. Thus, if we require that ${}^1\nabla$ be gauge invariant, then ϕ must change into $\phi + (d \ln \bar{a}) \circ \pi$ under a gauge transformation. Therefore, *if $\bar{\phi}$ is associated to c in such a manner that $\bar{\phi} + d \ln \bar{a}$ is associated to $(\bar{a} \circ \pi)c$, then there is a unique gauge invariant horizontal metric connection.*

We will think of ϕ, thogther with γ, as the fundamental data of spacetime geometry. We tentatively call ϕ the electromagnetic potential, though its true meaning must be disclosed only after disclosing field equations. Its operational definition is the following. Let $m \in M$ be fixed. Take $r_m \in \mathcal{F}_m$, and extend r_m to a cross section r of π in a neighborhood U of m in such a manner that r be experimentally stationary at m; in other words, *we suppose there is an operational definition for the relative rest of close particles with respect to a given one.* Now, restrict γ and ac to r, that is, take the values of these fields at the laboratory r for having the ordinary fields $\gamma \circ r$, $(ac) \circ r$; build the Lorentz metric $\bar{g} = (\gamma - a^2 c \otimes c) \circ r$, where $a = a \circ \pi$ is to be determined. Compute the Riemann standard connection of $\bar{g}$. Check if the normalized laboratory field $k \circ r$ is stationary at m, i.e., if the covariant derivative of $k \circ r$ is zero at m. If this is not so, pick $\bar{a}$ in a suitable manner in order to have an affirmative answer. Then $\bar{\phi}_m = -(d \ln \bar{a})_m$ is the value at m of the electromagnetic potential associated to c. Therefore, the electromagnetic potential associated to ac is zero. In other words, *the value at m of the electromagnetic potential associated with c is minus the differential at m of the deviation of c from the time form, which correctly gives the observable stationary (at m) character of a laboratory with normalized speed.*

If $\bar{\phi}$ is the differential of a function, it can be globally removed by a suitable election of $\bar{a}$. That is, in such a case we would have an absolute comparison among time unities at different places of spacetime. If $\bar{\phi}$ is not so, that comparison does not globally exist; we only can compare clocks at the limit when they join together. The proof of this interpretation requires additional techniques; it can be found in Ref. 4.

8. THE PHYSICAL METRIC CONNECTION AND THE CONNECTION ON $V\mathcal{F}M$

As in Part I, we now lift the pair ${}^1\nabla, {}^2\nabla$. So, we obtain a gauge invariant physical metric connection $\Delta_{\mathbf{w}}h = {}^1\nabla_{B_1\mathbf{w}}h + {}^2\nabla_{B_2\mathbf{w}}h$. Since ${}^1\nabla$ is gauge invariant and $B_1\mathbf{w} = \pi_* \circ \mathbf{w}$, then the first term is gauge invariant. As for the last term, we have seen that in a gauge transformation, the relation ${}^2\nabla' = a\,{}^2\nabla$ holds. But j changes into $j' = aj$. Hence $B_2|V^1_0\mathcal{F}M$ changes into B_2/a. Therefore, the last term and, as a consequence, Δ are gauge invariant.

Now, we can lift Δ as in Part I, Sec. 9, getting the linear connection $\mathbf{D}$ on $V\mathcal{F}M$, which is defined by $\mathbf{D}_{\mathbf{w}}\mathbf{v} = A_1\Delta_{\mathbf{w}}B_1\mathbf{v} + A_2\Delta_{\mathbf{w}}B_2\mathbf{v}$. Note that $\mathbf{D}$ is not gauge invariant. In fact we have

$$\mathbf{D}'_{\mathbf{w}}\mathbf{v} = A_1\Delta_{\mathbf{w}}B_1\mathbf{v} + A'_2\Delta_{\mathbf{w}}B'_2\mathbf{v} = \mathbf{D}_{\mathbf{w}}\mathbf{v} - \mathbf{w}(\ln a)\mathbf{V}\mathbf{v}.$$

Nevertheless, *the curvature field of* $\mathbf{D}$, *and therefore its Ricci field, are gauge invariant*, as it is easily proved.

9. CONCLUSION

The departure point of this paper is the Pythagorean spacelike metric, a principle which permeates every significant theory, experiment, and technology. The electromagnetic potential appears later, in the study of connections. I believe this point is very coherent in a tentative unified theory. In others, the electromagnetic field appears in the construction of the static geometric description—the metric—under the form of light signals; but it also appears, as a geometric object, in the dynamic description—connections or field equations. Thus, field equations must imply that the electromagnetic field propagates along null directions; otherwise, the theory would be meaningless. In our theory, this objection does not go.

We have reached a number of geometric objects enabling one to study the time manifold geometry. The main remaining problem is to suggest and justify field equations. In my opinion, it is a very difficult one:

(a) because of the horrific computations, even in simple models that perhaps could serve as a guideline for generalization;

(b) because the energy–momentum field depends on the geometry; thus, it must be reinterpreted under our basic assumptions;

(c) our manifold is now $\mathcal{F}M$; hence, usual patterns of field equation techniques cannot directly be translated here.

A naive field equation would be $\delta\int_\Omega |\mathbf{K}|^{1/2}d\tau = 0$, where Ω is a domain of $\mathcal{F}M$, $d\tau$ is the coordinate standard volume form on $\mathcal{F}M$, and $|\mathbf{K}|$ is the Ricci field determinant of $\mathbf{D}$. I have computed this integral for the relativistic model and my results are:

(a) if $\phi = 0$, then $|\mathbf{K}| = 0$, and this field equation is meaningless;

(b) if $\phi = 0$ but the Lorentz metric is constant (special relativity), then $|\mathbf{K}|$ also vanishes;

(c) I have studied a static spherical model (one-charged body problem), and there is no solution for the field equation. I believe this result is general for a relativistic model, but I do not have a proof.

I find three answers to these troubles. First, we must add to the integrand a term (or factor), depending on directions, standing for a mass-energy density. Second, the dependence of fields on directions is an essential property of spacetime, whence it precludes the assumption of a Lorentz metric; or, equivalently, that relativity is not compatible with local mass or charge anisotropy. Third, that the field equation is not appropriate.

I feel this last is the correct answer. So, it seems that this way will be around for a while.

APPENDIX A: EXISTENCE AND UNIQUENESS OF HORIZONTAL METRIC CONNECTIONS

We look for horizontal metric connections, in the sense that ∇ is horizontal, $\nabla c = \phi \otimes c$, $\nabla k = -\phi \otimes k$, $\nabla\gamma = 0$, $T = 0$, where $\phi = \bar{\phi}\circ\pi$. In Part I, ϕ is supposed to be zero.

It seems to me that this problem must be treated through local expressions, at least in a first attempt. But this way gives raise to another difficulty: The charts of the manifold $\mathcal{F}M$ are awkward to handle. So, we shall develop a technique enabling us to translate the problem to ordinary Finsler fields and Laugwitz connection (cf. Ref. 2), whose local expressions are simpler. Analogous techniques can be applied in other computations, for example the curvature or Ricci fields of $\mathbf{D}$.

If ∇ is a solution, and $\tilde{v}$, $\tilde{w}$ are $h(0)$, we can put $\widetilde{\nabla}_{\tilde{w}}\tilde{v} = e_a\nabla_{e_a\tilde{w}}e_b\tilde{v}$ for defining a Laugwitz connection; we also need to know the action of $\widetilde{\nabla}$ upon fields of $\widetilde{\Pi}^0_0M$, that is, the associated homomorphism $\widetilde{A}:\widetilde{\Pi}^1_0M \to V^1_0M$. If $\tilde{w}$ is $h(0)$ and $\tilde{\mathbf{s}}\in V^0_1\mathcal{F}M$, we define $\widetilde{A}\tilde{w}$ through $\langle\tilde{\mathbf{s}},\widetilde{A}\tilde{w}\rangle = \langle\tilde{\mathbf{s}} - \langle\tilde{\mathbf{s}},i\tilde{u}\rangle d\ln\langle\tilde{c},\tilde{u}\rangle, p_*^{-1}Ae_b\tilde{w}\rangle$, where A is the homomorphism associated with ∇ and $p_*^{-1}Ae_b\tilde{w}$ is any element of $V^1_0\widetilde{\mathcal{F}}M$ such that $p_*\circ(p_*^{-1}Ae_b\tilde{w}) = (Ae_b\tilde{w})\circ p$. This definition is consistent because $\tilde{\mathbf{s}} - \langle\tilde{\mathbf{s}},i\tilde{u}\rangle d\ln\langle\tilde{c},\tilde{u}\rangle, i\tilde{u}\rangle = 0$ and $i\tilde{u}$ spans $\ker p_*$.

Proposition: With the above notation, $\widetilde{\nabla}$ is a torsionless horizontal Langwitz connection such that $\nabla\tilde{\gamma} = 0$, $\widetilde{\nabla}\tilde{u} = -\tilde{\phi}\otimes\tilde{u}$, $\widetilde{\nabla}\tilde{c} = \tilde{\phi}\otimes\tilde{c}$, where we have put $\tilde{\gamma} = e_a\gamma$, $\tilde{c} = e_a c$, $\tilde{\phi} = e_a\phi$.

Proof: that $\widetilde{\nabla}$ is a Laugwitz connection is a trivial matter. It is horizontal because if $\tilde{v}$ is $h(0)$, then $(\widetilde{A}\tilde{v})_{\tilde{x}}(\bar{a}\circ\tilde{\pi}) = \langle d:(\bar{a}\circ\tilde{\pi}), p_*^{-1}Ae_b\tilde{v}\rangle_{\tilde{x}} = \langle p^* d(\bar{a}\circ\pi), p_*^{-1}Ae_b\tilde{v}\rangle_{\tilde{x}} = \langle d(\bar{a}\circ\pi), Ae_b\tilde{v}\rangle_{p\tilde{x}} = (e_b\tilde{v})_{p\tilde{x}}(\bar{a}) = \tilde{v}_{\tilde{x}}(\bar{a})$. Now, the torsion of $\widetilde{\nabla}$ is given by $\widetilde{T}(\tilde{v},\tilde{w}) = \widetilde{\nabla}_{\tilde{v}}\tilde{w} - \widetilde{\nabla}_{\tilde{w}}\tilde{v} - \tilde{\pi}_*\circ[\widetilde{A}\tilde{v},\widetilde{A}\tilde{w}]$. Then, if $\tilde{v}\tilde{w}$ are $h(0)$, we have $\widetilde{T}(\tilde{v},\tilde{w}) = \pi_*\circ[Ae_b\tilde{v},Ae_b\tilde{w}]\circ p - \tilde{\pi}_*\circ\{\widetilde{A}\tilde{v},\widetilde{A}\tilde{w}\}$, because $T = 0$. Taking account of the definition of $\widetilde{A}$, it is not difficult to prove that $p_*\circ\widetilde{A}\tilde{w} = (Ae_b\tilde{w})\circ p$. Thus $p_*\circ[\widetilde{A}\tilde{v},\widetilde{A}\tilde{w}] = [Ae_b\tilde{v},Ae_b\tilde{w}]\circ p$. Hence $\widetilde{T} = 0$ because $\tilde{\pi} = \pi\circ p$. Now, $\nabla_{\tilde{w}}\tilde{u} = \nabla_{\tilde{w}}\langle\tilde{c},\tilde{u}\rangle e_a k = \widetilde{A}\tilde{w}(\langle\tilde{c},\tilde{u}\rangle)e_a k + \langle\tilde{c},\tilde{u}\rangle e_a\nabla_{e_a\tilde{w}}k = -\langle\phi,\tilde{w}\rangle\langle\tilde{c},\tilde{u}\rangle e_a k = -\langle\tilde{\phi},\tilde{w}\rangle\tilde{u}$. The proof of $\widetilde{\nabla}\tilde{\gamma} = 0$ and $\widetilde{\nabla}\tilde{c} = \tilde{\phi}\otimes\tilde{c}$ is trivial.

Proposition: Let $\widetilde{\nabla}$ be a torsionless horizontal Laugwitz connection such that $\widetilde{\nabla}\tilde{c} = \tilde{\phi}\otimes\tilde{c}$, $\widetilde{\nabla}\tilde{u} = -\tilde{\phi}\otimes\tilde{u}$, $\widetilde{\nabla}\tilde{\gamma} = 0$. Then $\Delta_{\tilde{v}}\tilde{w} = e_b\widetilde{\nabla}_{e_a\tilde{v}}e_a\tilde{w}$ defines a horizontal metric connection on IIM.

Proof: The formula defining ∇ makes sense if $\widetilde{\nabla}_{\tilde{v}}\tilde{w}$ is $h(0)$ whenever $\tilde{v},\tilde{w}$ are, and this is supposed to be true in our case, as we can verify by means of the formula (1) below. The proof that $\nabla k = -\phi\otimes k$, etc., is straightforward. Now, $\widetilde{\nabla}_{\tilde{v}}\tilde{w} = e_a e_b\widetilde{\nabla}_{e_a\tilde{v}}e_a e_b\tilde{w} = e_a\nabla_{e_a\tilde{v}}e_b\tilde{w}$; therefore, ∇ and $\widetilde{\nabla}$ do induce each other in the sense of our previous proposition if A and $\widetilde{A}$ are related as before. We have $\widetilde{\nabla}\langle\tilde{c},\tilde{u}\rangle = 0$, hence $\langle\tilde{\mathbf{s}},\widetilde{A}\tilde{v}\rangle = \langle\tilde{\mathbf{s}} - \langle\tilde{\mathbf{s}},i\tilde{u}\rangle d\ln\langle\tilde{c},\tilde{u}\rangle,\widetilde{A}\tilde{v}\rangle$. But $\langle\tilde{\mathbf{s}} - \tilde{\mathbf{s}},i\tilde{u}\rangle d\ln\langle\tilde{c},\tilde{u}\rangle$ belongs to the $V^0_0\widetilde{\mathcal{F}}M$-module spanned by $p^*(V^0_1\mathcal{F}M)$; thus, if $a\in II^0_0M$, then $\langle d(a\circ p),\widetilde{A}\tilde{v}\rangle = \langle p^*da,\widetilde{A}\tilde{v}\rangle = \langle(da)\circ p, p_*\circ\widetilde{A}\tilde{v}\rangle$; but if $\tilde{v}$ is $h(0)$, then $\langle d(a\circ p),\widetilde{A}\tilde{v}\rangle = \widetilde{\nabla}_{\tilde{v}}(a\circ p) = e_a\nabla_{e_a\tilde{v}}a = \langle(da)\circ p,(Ae_b\tilde{v})\circ p\rangle$, whence $p_*\circ\widetilde{A}\tilde{v} = (A\,e_b\tilde{v})\circ p$. Then, as before, we can prove the torsion of ∇ is zero.

Thus, our problem of existence and uniqueness can be equivalently stated on $\widetilde{\mathit{IIM}}$. Let $\{\bar{x}^i\}$ be a coordinate system on $U\subset M$. We take for $\widetilde{\mathcal{F}}U$ the coordinate functions $\{q^i,p^j\}$ defined by $q^i(\bar{x}_m) = \bar{x}^i(m), p^i(\bar{x}_m) = \langle(d\bar{x}^i)_m,\bar{x}_m\rangle$. Since $\widetilde{A}$ is horizontal, we can write $\widetilde{A}\tilde{e}_i = \partial/\partial q^i + A^j_i(\partial/\partial p^j)$, where $\tilde{e}_i = (\partial/\partial\bar{x}^i)\circ\tilde{\pi}$. We put $\widetilde{\nabla}_{\tilde{e}_i}\tilde{e}_k = \Gamma^j_{ik}\tilde{e}_j$. Then $\widetilde{\nabla}_{\tilde{e}_i}\tilde{u} = \widetilde{\nabla}_{\tilde{e}_i}(u^k\tilde{e}_k) = \Gamma^j_{ik}u^k\tilde{e}_j + \widetilde{A}\tilde{e}_i(u^j)\tilde{e}_j = -\phi_i u^j\tilde{e}_j$. Therefore, $A^j_i = -\Gamma^j_{ik}u^k - \phi_i u^j$. If we put $\tilde{g} = \tilde{\gamma} - \tilde{c}\otimes\tilde{c} = g_{ij}(d\bar{x}^i\circ\tilde{\pi})\otimes(d\bar{x}^j\circ\tilde{\pi})$, then the matrix (g_{ij}) is everywhere regular. Thus, after some standard computation, we find there is a solution on $\widetilde{\mathcal{F}}U$ if the following linear system has it:

$$2g_{ir}\Gamma^r_{jk} + \frac{\partial g_{ik}}{\partial p^r}\Gamma^r_{jm}u^m + \frac{\partial g_{ij}}{\partial p^r}\Gamma^r_{km}u^m - \frac{\partial g_{jk}}{\partial p^r}\Gamma^r_{im}u^m = \frac{\partial g_{ij}}{\partial q^k} + \frac{\partial g_{ik}}{\partial q^j} - \frac{\partial g_{jk}}{\partial q^i} - 2\phi_i c_j c_k + 2\phi_j c_i c_k + 2\phi_k c_i c_j, \tag{1}$$

where $\tilde{c} = c_i(d\bar{x}^i\circ\tilde{\pi})$. If we contract (1) with u^k, it becomes

$$2g_{ir}B^r_j + c_{ir}B^r_j + \frac{\partial g_{ij}}{\partial p^r}B^r_k u^k - c_{jr}B^r_i = M_{jik}u^k, \tag{2}$$

where we have put $B^r_j = \Gamma^r_{jk}u^k$, $c_{ir} = (\partial g_{ik}/\partial p^r)u^k$, and M_{jik} is the right-hand side of (1). Since (g_{ij}) is regular, there is a solution for (1)if it occurs for (2). We can verify without difficulty that $\gamma_{ir} + \eta_{ir} = -c_{ir}$. Hence $c_{ir} = c_{ri}$ and $c_{ir}u^r = 0$. Suppose we write (2) taking at $\bar{x}_m\in\widetilde{\mathcal{F}}U$ the values of the different quantities. Then, since $\tilde{\gamma}$ has signature $(0,+,\ldots,+)$, we can choose the coordinates in such a maner that, at $\bar{x}_m$, we would have $\gamma_{\alpha\beta} = \delta_{\alpha\beta}, c_{\alpha\beta} = s_\alpha\delta_{\alpha\beta}, \gamma_{k0} = c_{k0} = 0$(Greek indexes from 1 to $n-1$). Now (2) has a unique solution at $\bar{x}_m$ if $2 + s_\alpha + s_\beta\neq 0$ for every $\alpha,\beta\in\{1,\ldots,n-1\}$. Thus taking into account that $\gamma_{ir} + \eta_{ir} = -c_{ir}$, we have after some obvious steps:

Theorem: Let $r\in\mathcal{F}M$, and suppose that $(0,1,\ldots,1)$ and $(0,\eta_1,\ldots,\eta_{n-1})$ are the diagonal elements of γ_r and η_r when they are simultaneously diagonalized. Then, if $\eta_\alpha + \eta_\beta\neq 0$ for every $\alpha,\beta\in\{1,\ldots,n-1\}$, $r\in\mathcal{F}M$, there is one unique horizontal metric connection on IIM.

Corollary: If η is supposed to have signature $(0,-,\ldots,-)$ everywhere, there is one unique horizontal metric connection on IIM.

APPENDIX B: *Existence and uniqueness of vertical metric connections*

Our conditions are: $\nabla_v a = jv(a)$, $\nabla c = 0$, $\nabla\gamma = 0$, $T = 0$.

We put $\nabla_w v - D_w v = G(w,v)$; since $\nabla_w a = D_w a = jw(a)$, G is a bilinear operator and it defines an element of II^1_2M. If $\nabla c = 0$, then $0 = \langle\nabla_w c,v\rangle$

$$= \nabla_w\langle c,v\rangle - \langle c,\nabla_w v\rangle$$

$$= D_w\langle c,v\rangle - \langle c,D_w v\rangle - \langle c,G(w,v)\rangle = \eta(w,v) - G(w,v).$$

Thus, $\langle c,G(w,v)\rangle = \eta(w,v)$ for every $v,w\in II^1_0M$.

Now, if $\langle c,v\rangle = \langle c,w\rangle = 0$, then $B_2[A_2w,A_2v] = D_w v - D_v w$, as it is easily proved. Then, we have in general that $B_2[A_2w,A_2v] = D_w v - D_v w - \langle c,v\rangle w + \langle c,w\rangle v - \langle c,D_w v\rangle k + \langle c,D_v w\rangle k$.

If $T(w,v) = B_2A_2(\nabla_w v - \nabla_v w - B_2[A_2w,A_2v]) = 0$, then $B_2A_2(G(w,v) - G(v,w) + \langle c,v\rangle w - \langle c,w\rangle v + \langle c,D_w v\rangle k - \langle c,D_v w\rangle k)$

$$= B_2A_2(G(w,v) - G(v,w) + \langle c,v\rangle w - \langle c,w\rangle v).$$ But if $\langle c,G(w,v)\rangle = \eta(w,v)$, then

$\langle c,G(w,v) - G(v,w) + \langle c,v\rangle w - \langle c,w\rangle v\rangle = 0$. Thus, our second condition upon G is

$$G(w,v) - G(v,w) = \langle c,w\rangle v - \langle c,v\rangle w.$$

If $\nabla\gamma = 0$, by a standard computation we have:

$$2\gamma(G(v,w),z) = (D_v\gamma)(w,z) + (D_w\gamma)(z,v) - (D_z\gamma)(v,w) + 2\gamma(v,w)\langle c,z\rangle - 2\gamma(v,z)\langle c,w\rangle.$$

But

$$2g(G(v,w),z) = 2\gamma(G(v,w),z) - 2\langle c,G(v,w)\rangle\langle c,z\rangle = 2\gamma(G(v,w),z) - 2\eta(v,w)\langle c,z\rangle.$$

Hence

$$2g(G(v,w),z) = (D_v\gamma)(w,z) + (D_w\gamma)(z,v) - (D_z\gamma)(v,w) + 2(\gamma(v,w) - \eta(v,w))\langle c,z\rangle - 2\gamma(v,z)\langle c,w\rangle.$$

This formula tells us that if the vertical metric connection exists, it is unique. It can be equivalently written

$$2g(G(v,w),z) = (D_v g)(w,z) + (D_w g)(z,v) - (D_z g)(v,w) + 2g(v,w)\langle c,z\rangle - 2g(v,z)\langle c,w\rangle.$$

Hence

$$\nabla_v w = D_v w + \tfrac{1}{2} g^{-1}((D_v g)(w, \ ) + (D_u g)(v, \ ) - (Dg)(v,w), \ ) - g(v,w)k - \langle c,w\rangle v.$$

This formula gives the vertical metric connection, as it is easily proved.

[1]J.I. Horváth, Suppl. Nuovo Cimento **9**, 444–96 (1958), see the Appendix.
[2]A. Montesinos, "On Finsler connections," to be published in Rev. Mat. Hisp. Amer.
[3]R. Grassini, Boll. U.M.I. **11**, 507–17 (1975).
[4]A. Montesinos, "Geometria del espacio-tiempo a partir de la métrica espacial," thesis, Universidad Complutense, Madrid, 1976.

Koornwinder's polynomials and representations of the conformal group

Klaus Ringhofer

Fachbereich 5 der Universität Osnabrück, Federal Republic of Germany, Germany
(Received 7 July 1978)

The tensor product of two massive, spinless, positive-energy ray representations of the conformal group of spacetime $SO_0(4,2)/\mathbb{Z}_2$ is reduced in a momentum basis. The basis vectors for the irreducible subspaces (the "Clebsch–Gordan coefficients") are found to be intimately connected with Koornwinder's polynomials in two dimensions.

1. SUMMARY AND INTRODUCTION

Within the last two years $SO_0(4,2)/\mathbb{Z}_2$, the conformal group of space–time, has found new interest by a paper of Lüscher and Mack[1] who proposed a globally conformal invariant quantum field theory. Furthermore, all irreducible unitary positive-energy ray representations of the conformal group have been constructed (Mack[2]) as induced representations on (the compactified) Minkowski space. They are either massless with a definite helicity (that is, they become an irreducible mass zero representation with positive energy and a definite helicity when restricted to the Poincaré subgroup) or they are massive and contain spin multiplets with spin values $l = l_0, l_0 + 1, \cdots, L$ (that is, they contain all the irreducible representations with positive mass, positive energy, and spin value either l_0 or $l_0 + 1$ or $\cdots, L$ when restricted to the Poincaré subgroup, each of them once). Besides, a dimension parameter d appears which is determined by the helicity for massless representations, but which is less restricted for massive representations.

There has been widespread opinion that only massless representations are of importance for physics. However, in Ref. 3 an argument originally due to Castell[4] has been given which shows that in the framework of partial-wave expansions massless representations occur only in exceptional cases. As a rule, massive representations enter the partial-wave expansions, and therefore a further study of massive representations is worthwhile from a physical point of view. On the other hand, a study of massive representations is also interesting from a mathematical point of view, since it will be shown that they are intimately connected with a class of orthogonal polynomials in two dimensions which have only recently been investigated by Koornwinder,[5] and since very often the study of the connections between a polynomial system and a group leads to a deeper understanding of the polynomial system.

In this paper, a momentum basis is used to obtain massive representations with spin values $0,1,\cdots,L$ by reduction of the tensor of two massive representations with spin zero. The reduction is done by decomposing the tensor product of the corresponding Lie algebra representations, using Casimir operator techniques. However, each of the resulting Lie algebra representations obtained by the decomposition is identified as a Lie algebra representation of one of the group representations obtained in Ref. 2.

In the process of reduction it turns out that the basis vectors for the description of the spin states within an irreducible representation will not simply be given by the spherical harmonics Y_{lm}, but to allow for the possibility of spin multiplets the basis vectors will have the form $P^{(l)}\mathcal{Y}_{lm'}$, where the $P^{(l)} = P^{(l)}(x,y)$ are Koornwinder's polynomials of two variables and where the $\mathcal{Y}_{lm'}$ are harmonic polynomials (actually, the $P^{(l)}(x,y)$ will carry some more indices to distinguish different irreducible representations). As a consequence, differential operators $E^{\pm}(l)$ ("ladder operators") will appear which map spin states l to spin states $l \pm 1$. They are known from a paper by Sprinkhuizen–Kuyper.[6]

The basis functions $P^{(l)}\mathcal{Y}_{lm'}$ are of considerable physical interest by themselves, since they play the role of Clebsch–Gordan coefficients which are of importance already for a purely kinematical interpretation of the conformal group in physics as proposed by Castell.[7] For instance, a very satisfying explanation for the t/s scaling behavior for elastic proton–proton scattering has been obtained by this method.[8]

2. COMMUTATION RELATIONS FOR THE GENERATORS AND CASIMIR OPERATORS

Notation: Greek indices $\rho,\sigma,\cdots,\omega$ run from 0 to 3, and the metric $(-1,1,1,1)$ is chosen for Minkowski space. Greek indices $\alpha,\beta,\cdots,\pi$ take the values 0,1,2,3,5,6, and the metric $(-1,1,1,1,1,-1)$ is chosen for the six-dimensional space with SO(4,2) as group of symmetry. The sum convention is used.

A basis for the Lie algebra of the conformal group of Minkowski space is given by the four-momentum P^μ (generators of translations), the angular momentum tensor $M_{\mu\nu}$ (generators of rotations and pure Lorentz transformation), D (generator of dilations), and K_μ (generators of special conformal transformations). This is the basis of physical interest. However, for the purpose of calculations, it is much more convenient to use the generators

$$M_{\mu\nu}, \quad M_{\mu 5} = -\tfrac{1}{2}(P_\mu + K_\mu), \qquad (2.1)$$

$$M_{\mu 6} = \tfrac{1}{2}(P_\mu - K_\mu), \quad M_{56} = -D.$$

As is well known, the $M_{\alpha\beta}$ fulfill the commutation relations of SO(4,2),

 0022-2488/79/050966-06$01.00

$$[M_{\alpha\beta},M_{\gamma\delta}] = i\{g_{\alpha\gamma}M_{\beta\delta} + g_{\beta\delta}M_{\alpha\gamma} - g_{\alpha\delta}M_{\beta\gamma} - g_{\beta\gamma}M_{\alpha\delta}\}, \tag{2.2}$$

and the Casimir operators of SO(4,2) are

$$C_{\mathrm{II}} = \tfrac{1}{2}M_{\alpha\beta}M^{\alpha\beta},$$
$$C_{\mathrm{III}} = \tfrac{1}{48}\epsilon_{\alpha\beta\gamma\delta\kappa\lambda}M^{\alpha\beta}M^{\gamma\delta}M^{\kappa\lambda}, \tag{2.3}$$
$$C_{\mathrm{IV}} = \tfrac{1}{2}C_{\alpha\beta}C^{\alpha\beta}, \quad C_{\alpha\beta} = \tfrac{1}{8}\epsilon_{\alpha\beta\gamma\delta\kappa\lambda}M^{\gamma\delta}M^{\kappa\lambda}$$

(the ϵ tensor is defined by $\epsilon_{012356} = 1$).

3. MASSIVE SPINLESS REPRESENTATIONS

The generators of a massive representation $D^{(d,0,0)}$ of the conformal group with dimension parameter d and spins $l_0 = L = 0$ have been obtained in a momentum basis by Castell,[9] who reduced the quasiregular representation of $SO_0(4,2)$ on a hyperboloid in six dimensions, by Castell and the present author,[10] who reduced the tensor product of zero-mass representations, and by Mack,[2] who constructed induced representations. They are (square brackets mean antisymmetrization)

$$P^{\mu} = p^{\mu},$$
$$M_{\mu\nu} = -ip_{[\mu}\partial_{\nu]} = -i(p_{\mu}\partial_{\nu} - p_{\nu}\partial_{\mu}), \quad \partial_{\mu} = \frac{\partial}{\partial p^{\mu}}, \tag{3.1}$$
$$D = i(p^{\rho}\partial_{\rho} + d), \quad K_{\mu} = p_{\mu}\partial^{\rho}\partial_{\rho} - 2(p^{\rho}\partial_{\rho} + d)\partial_{\mu},$$

and from this the values of the Casimir operators are, obtained as

$$c_{\mathrm{II}} = (d-2)^2 - 4, \quad c_{\mathrm{II}} = c_{\mathrm{IV}} = 0. \tag{3.2}$$

For representations of the conformal group, d takes the values 2, 4, 6,··· (see Ref. 9) and for ray representations $d > 1$ (see Ref. 2).

The generators act on the space $\mathscr{S}(M)$ of Schwartz test functions on Minkowski space with scalar product

$$(\varphi,\psi) = \int \varphi^{*}(p)\psi(p)m^{2(d-2)}\theta(m^2)\theta(p^0)\,d^4p, \tag{3.3}$$
$$m = (-p^{\rho}p_{\rho})^{1/2}.$$

$\mathscr{S}(M)$ can be completed to a Hilbert space after dividing out norm zero functions, that is, the functions vanishing on the forward light cone (latin indices run from 1 to 3),

$$L^{+} = \{p|(p^0)^2 - p^ip_i > 0, p^0 > 0\}. \tag{3.4}$$

4. THE TENSOR PRODUCT $D^{(d_1,0,0)} \otimes D^{(d_2,0,0)}$

In this paragraph, the generators and Casimir operators of the product reprsentation $D^{(d_1,0,0)} \otimes D^{(d_2,0,0)}$ are expressed by momentum variables and by "spin variables." Instead of d_1 and d_2, $\alpha = d_1 - 2$ and $\beta = d_2 - 2$ will be used.

Let $p^{(1)}$ and $p^{(2)}$ be the four-momenta appearing in representations $D^{(\alpha+2,0,0)}$ and $D^{(\beta+2,0,0)}$. Then, if $G^{(\alpha+2,0,0)}(p^{(1)})$ and $G^{(\beta+2,0,0)}(p^{(2)})$ are corresponding generators in $D^{(\alpha+2,0,0)}$ and $D^{(\beta+2,0,0)}$, the corresponding generator G of the product representation is $G = G^{(\alpha+2,0,0)}(p^{(1)}) + G^{(\beta+2,0,0)}(p^{(2)})$. Especially, $P = p^{(1)} + p^{(2)}$.

To work again in a momentum representation, the total four-momentum

$$p^{\mu} = p^{(1)\mu} + p^{(2)\mu} \tag{4.1}$$

has to be introduced as well as four more variables which will be referred to as "spin variables." For the moment they are chosen to be

$$q^{\mu} = p^{(2)\mu} \tag{4.2}$$

and the following form of the generators is obtained:

$$P^{\mu} = p^{\mu},$$
$$M_{\mu\nu} = -ip_{[\mu}\partial_{\nu]} + \Sigma_{\mu\nu},$$
$$D = i(p^{\rho}\partial_{\rho} + 2) + \Delta, \tag{4.3}$$
$$K_{\mu} = p_{\mu}\partial^{\rho}\partial_{\rho} - 2(p^{\rho}\partial_{\rho} - i\Delta + 2)\partial_{\mu} + 2i\Sigma_{\mu\rho}\partial^{\rho} + k_{\mu}.$$

The operators $\Sigma_{\mu\nu}, \Delta$, and k_{μ} are defined as

$$\Sigma_{\mu\nu} = -iq_{[\mu}\nabla_{\nu]}, \quad \nabla_{\nu} = \frac{\partial}{\partial q^{\nu}},$$
$$\Delta = i(q^{\rho}\nabla_{\rho} + \alpha + \beta + 2), \tag{4.4}$$
$$k_{\mu} = q_{\mu}\nabla^{\rho}\nabla_{\rho} - 2(q^{\rho}\nabla_{\rho} + \beta + 2)\nabla_{\mu}.$$

They depend only upon the spin variables and therefore they will be referred to as "operators in spin space." From the generators (4.3) the Casimir operators are obtained as

$$C_{\mathrm{II}} = p^{\mu}k_{\mu} - \Delta^2 + \tfrac{1}{2}\Sigma^{\rho\sigma}\Sigma_{\rho\sigma} - 4,$$
$$C_{\mathrm{III}} = 0, \tag{4.5}$$
$$C_{\mathrm{IV}} = \tfrac{1}{4}e^{\mu}e_{\mu} + \tfrac{1}{2}s^{\rho\sigma}s_{\rho\sigma}(C_{\mathrm{II}} - \tfrac{1}{2}s^{\rho\sigma}s_{\rho\sigma} + 2).$$

To explain the "covariant spin tensor" $s_{\mu\nu}$ and the 4-vector e_{μ}, it is convenient to first introduce the total mass

$$M = (-p^{\mu}p_{\mu})^{1/2}, \tag{4.6}$$

the 4-velocity

$$u^{\mu} = \frac{p^{\mu}}{M}, \tag{4.7}$$

and the projection operator

$$P^{\mu}_{\nu} = g^{\mu}_{\nu} + u^{\mu}u_{\nu}. \tag{4.8}$$

upon the momenta perpendicular to p^{μ}. Also the abbreviation

$$A_{\mu\nu} = \{\Sigma_{\mu\rho},\Sigma^{\rho}_{\nu}\} \tag{4.9}$$

(the curled brackets indicate an anticommutator) will be used. Then

$$s_{\mu\nu} = P_{\mu\rho}P_{\nu\sigma}\Sigma^{\rho\sigma}, \tag{4.10}$$
$$e_{\mu} = P_{\mu\rho}(Mk^{\rho} + 2\Delta\,\Sigma^{\rho\sigma}u_{\sigma} - A^{\rho\sigma}u_{\sigma}). \tag{4.11}$$

The Casimir operator C_{II} has already been obtained in formula (36) in Ref. 10. Slightly different variables have been used in this paper.

The scalar product is obtained as

$$(\varphi,\psi) = \int_{q,p-qL^{+}} \varphi^{*}(p,q)\psi(p,q)[-(p-q)^2]^{\alpha}[-q^2]^{\beta}$$

$$\times\theta\,[-(p-q)^2]\,\theta(-q^2)\,\theta(p^0-q^0)\,\theta(q^0)\,d^4p\,d^3q. \tag{4.12}$$

Having obtained the Casimir operators from the generators, the question arises if the operators $s_{\mu\nu}$ and e_μ, which are necessary to construct C_{IV}, have anything to do with the generators. They have! Introduce the displacement operator

$$d_\mu = \partial_\mu + \frac{i}{M}(\Delta u^\mu + \Sigma_{\mu\rho}u^\rho). \tag{4.13}$$

Then

$$\begin{aligned}
&P^\mu = p^\mu,\\
&M_{\mu\nu} = ip_{[\mu}d_{\nu]} + s_{\mu\nu},\\
&D = i(p^\rho d_\rho + 2),\\
&K_\mu = p_\mu\left[d^\rho d_\rho - \frac{1}{M^2}(C_{\mathrm{II}} + 4 - \tfrac{1}{2}s^{\rho\sigma}s_{\rho\sigma})\right]\\
&\qquad - 2(p^\rho d_\rho + 2)d_\mu + i\{s_{\mu\rho},d^\rho\} + \frac{1}{M}e_\mu.
\end{aligned} \tag{4.14}$$

The importance of d_μ, $s_{\mu\nu}$, and e_μ is further underlined by the fact that each of them commutes with the Casimir operators.

5. INTRODUCTION OF WIGNER FUNCTIONS

In this section the variables in spin space will be made dimensionless by dividing them by M and they will be subjected to the pure Lorentz transformation which maps p to $(M,0,0,0)$. The second step will be called "introduction of Wigner functions" (compare a corresponding step in Ref. 2). let latin indices run from 1 to 3 and let δ^{ij} be Kronecker's delta. Then the transformations described above are

$$\begin{aligned}
&Q^0 = \frac{1}{M}(u^0q^0 - u^iq_i),\\
&Q^i = \frac{1}{M}\left[-u^iq^0 + \left(\delta^{ij} + \frac{u^iu^j}{u^0+1}\right)q_j\right].
\end{aligned} \tag{5.1}$$

An extra factor $(M^{\alpha+\beta+2})^2$ appears now in the scalar product which can be multiplied into the Wigner functions. All these transformations are easy to do since the operators d_μ, $s_{\mu\nu}$ and e_μ transform simply. The result is

$$\begin{aligned}
&P^\mu = p^\mu,\\
&M_{ij} = -ip_{[i}\partial_{j]} + \sigma_{ij},\\
&M_{i0} = -i(p_i\partial_0 + p^0\partial_i) + \sigma_{i0},\\
&D = i(p^\rho\partial_\rho + 2),\\
&K_0 = p_0[\partial^\rho\partial_\rho - (1/M^2)(C_{\mathrm{II}} + 4 - \tfrac{1}{2}\Sigma^{ij}\Sigma_{ij})]\\
&\qquad - 2(p^\rho\partial_\rho + 2)\partial_0 + \{\sigma^j_0,i\partial_j + (1/M)\sigma_{j0}\} - 1/m\,u^jE_j,\\
&K_i = p_i[\partial^\rho\partial_\rho - (1/M^2)(C_{\mathrm{II}} + 4 - \tfrac{1}{2}\Sigma^{ij}\Sigma_{ij})]\\
&\qquad - 2(p^\rho\partial_\rho + 2)\partial_i + \{\sigma^j_i,i\partial_j + (1/M)\sigma_{j0}\} + \{\sigma^0_i,i\partial_0\}\\
&\qquad + \frac{1}{M}\left(\delta_{ij} + \frac{u_iu_j}{u^0+1}E^j\right),
\end{aligned} \tag{5.2}$$

for the generators with

$$\sigma_{i0} = -\frac{1}{u^0+1}\Sigma_{ij}u^j, \qquad \sigma_{ij} = \Sigma_{ij}, \tag{5.3}$$

$$E_i = k_i + 2\Delta\Sigma_{i0} - A_{i0}. \tag{5.4}$$

Here $\Sigma_{\mu\nu}$, Δ, k_μ and A_{i0} are again given by formula (4.4) and (4.9) with q replaced by Q. For the Casimir operators the result is

$$\begin{aligned}
&C_{\mathrm{II}} = -k^0 - \Delta^2 + \tfrac{1}{2}\Sigma_{\rho\sigma}\Sigma^{\rho\sigma} - 4,\\
&C_{\mathrm{III}} = 0,\\
&C_{\mathrm{IV}} = \tfrac{1}{4}E^iE_i + \tfrac{1}{2}\Sigma^{ij}\Sigma_{ij}(C_{\mathrm{II}} - \tfrac{1}{2}\Sigma^{ij}\Sigma_{ij} + 2).
\end{aligned} \tag{5.5}$$

Note that p has vanished from the Casimir operators. The scalar product becomes

$$\begin{aligned}
(\varphi,\psi) = \int &\varphi^*(p,Q)\psi(p,Q)\,[(Q^0-1)^2 - Q^iQ_i]^\alpha\\
&\times[(Q^0)^2 - Q^iQ_i]^\beta\,\theta(M^2)\,\theta(p^0)\,\theta(Q^0)\,\theta(1-Q^0)\\
&\times\theta\,[(Q^0-1)^2 - Q^iQ_i]\,\theta\,[(Q^0)^2 - Q^iQ_i]\,d^4p\,d^4Q.
\end{aligned} \tag{5.6}$$

6. ANGULAR MOMENTUM BASIS IN SPIN SPACE

The Casimir operators (5.5) are invariant under rotations in spin space, and therefore the introduction of polar coordinates Q, θ, φ for Q^1, Q^2, Q^3 will be useful. This is not quite trivial, since the operators E^i jump between neighboring spin states, as can be seen from the explicit expression

$$\begin{aligned}
E^i = Q^i\Big\{&(1-2Q^0)\Big(-\frac{\partial^2}{\partial(Q^0)^2} + \frac{\partial^2}{\partial Q^2} + \frac{2}{Q}\frac{\partial}{\partial Q}\\
&- \frac{1(1+1)}{Q^2}\Big) + 2\kappa\frac{\partial}{\partial Q^0}\Big\} + 2\{((Q^0)^2 + Q^2 - Q^0)\\
&\times\frac{\partial}{\partial Q^0} + Q(2Q^0-1)\frac{\partial}{\partial Q} + (\kappa+2)Q^0 - (\beta+2)\Big\}\\
&\times\frac{\partial}{\partial Q^i}, \qquad \kappa = \alpha+\beta+2,
\end{aligned} \tag{6.1}$$

for E^i. As is well known (see for instance Ref. 11, pp. 98 and 99, both Q^i and the gradient vector $\partial/\partial Q^i$ change the value l of the spin by ± 1. With respect to basis vectors

$$B^{(l)}_{m'} = P^{(l)}(Q^0,Q)\mathcal{Y}_{lm'}(Q^1,Q^2,Q^3), \tag{6.2}$$

$$P^{(l)}(Q^0 - Q) = P^{(l)}(Q^0,Q)$$

(where the $\mathcal{Y}_{lm'} = Q^l Y_{lm'}$ are the "harmonic polynomials"), the reduced matrix elements of E^i are

$$(l+1\|E\|l) = \sqrt{l+1}\,Q^{-1}E(1), \qquad \text{for } l\to l+1, \tag{6.3}$$

$$(l-1\|E\|l) = -\sqrt{1}\,Q^{-2l}E(l)Q^{2l+1}, \quad \text{for } l\to l-1,$$

with

$$E(l) = Q(2Q^0-1)\frac{\partial}{\partial(Q^0)^2} + 2((Q^0)^2 + Q^2 - Q^0) \times\frac{\partial^2}{\partial Q^0\partial Q} + Q(2Q^0-1)\frac{\partial}{\partial Q^2} + 2\kappa Q\frac{\partial}{\partial Q^0} + 2(\kappa Q^0 - \beta - 1)\frac{\partial}{\partial Q}. \tag{6.4}$$

Formula (6.3) shows that

$$E^+(l) = \frac{1}{8Q}E(l) \tag{6.5}$$

and

$$E^-(l) = 2Q^{-2l}E(l)Q^{2l+1} \tag{6.6}$$

are ladder operators for the functions $p^{(l)}(Q^0,Q)$, raising and lowering the spin value l by 1.

There are at least three reasons that make the ladder operators fundamental for the analysis of this paper:

(a) They are directly connected with the generators. To make this obvious, apply the pure Lorentz transformation L^μ_ν used in formula (5.1) [which transforms p to $(M,0,0,0)$] to the 4-vector K_μ:

$$\hat{K}_\mu = L^\nu_\mu K_\nu = \begin{cases} 0, & \text{for } \mu = 0, \\ \frac{1}{M}E_i, & \text{for } \mu = i+\cdots. \end{cases}$$

(Only the terms responsible for transitions $l\to l\pm 1$ are given, that is, the terms containing the E_i's.) Therefore,

$$(l\pm 1\|\hat{K}_i\|l) = \frac{1}{M}(l\pm 1\|E_i\|l) = \begin{cases} \dfrac{8\sqrt{l+1}}{M}, & E^+(l), \\ -\dfrac{\sqrt{l}}{2M}, & E^-(l), \end{cases}$$

which means essentially that the ladder operators are reduced matrix elements of the generators of the special conformal transformations.

(b) They are intimately connected with the Casimir operators, namely

$$c_{II} = \frac{1}{2l+1}E^+(l-1)E^-(l) - \frac{1}{2l+1}E^-(l+1)E^+(l) + 2l(l+1) - 4,$$

$$C_{III} = 0, \tag{6.7}$$

$$C_{IV} = \frac{1}{2l+1}E^+(l-1)E^-(l) + \frac{l+1}{2l+1}E^-(l+1)E^+(l) + l(l+1)[(C_{II} - l)(l+1) + 2].$$

(c) They are ladder operators for Koornwinder's polynomials. Once this fact is established it is quite obvious that Koornwinder's polynomials come into play. Introducing variables x,y by

$$Q^0 = \tfrac{1}{4}(x+y) + \tfrac{1}{2}, \quad Q = \tfrac{1}{4}(x-y), \tag{6.8}$$

one obtains

$$E^+(l) = \frac{1}{x-y}[(x^2-1)\partial^2 x + (\kappa x + \lambda)\partial x - (y^2-1)\partial^2 y - (\kappa y + \lambda)\partial y], \tag{6.9}$$

$$E^{-1}(l) = \frac{1}{(x-y)^{2l}}[(x^2-1)\partial^2 x + (\kappa x + \lambda)\partial x - (y^2-1) \times\partial^2 y - (\kappa y + \lambda)\partial_y](x-y)^{2l+1},$$

with

$$\kappa = \alpha + \beta + 2, \quad \lambda = \alpha - \beta. \tag{6.10}$$

$E^\pm(l)$ of formula (6.9) are identical with the ladder operators for Koornwinder's polynomials as given by Sprinkhuizen–Kuyper.[6] More accurately, one should write $E^{\pm(\alpha,\beta)}(l)$. The following expression is obtained for the scalar product,

$$(\varphi,\psi) = \frac{1}{2^{\kappa+3}}\int \varphi^*(p,x,y,\theta,\varphi)\psi(p,x,y,\theta,\varphi) \times[(1-x)(1-y)]^\alpha[(1+x)(1+y)]^\beta(x-y)^2 \times\theta(M^2)\theta(P^0)\theta(1-x)^2\theta(1-y)^2\theta(x-y) \times d^4p\,dx\,dy\sin\theta\,d\theta\,d\varphi \tag{6.11}$$

[the factor $\theta(x-y)$ enters to express that $Q\geqslant 0$].

Calculating the scalar product of two basis vectors $B^{(l)}_{m_1}$ and $B^{(l')}_{m_2}$ and, for the moment, performing only the integration over θ and φ, one is led to a factor $[(x-y)/4)^{2l}\delta_{ll'}\delta_{m_1m_2}$, so that the measure contains a factor

$$dm = [(1-x)(1-y)]^\alpha[(1+x)(1+y)]^\beta \times(x-y)^{2l}\theta(1-x^2)\theta(1-y^2)\theta(x-y). \tag{6.12}$$

This exactly the measure for Koornwinder's orthogonal set

of polynomials. It is positive within the triangle $1-x^2>0$, $1-y^2>0$, and $x-y>0$ but zero outside.

7. THE IRREDUCIBLE REPRESENTATIONS CONTAINED IN THE TENSOR PRODUCT

In Secs. 5 and 6 transformations of the spin variables have been performed with the aim to render the Casimir operators as simple as possible. This aim has been achieved by now. The Casimir operators (6.7) together with the ladder operators (6.10) depend only on the two variables x and y and cannot be separated any further. It is known from Koornwinder's work[5] that the Casimir operators possess the joint polynomial eigenfunctions

$$P_{m,n}^{(\alpha,\beta,l)}(x,y)=(xy)^m[x^n+c_1x^{n-1}y+\cdots+y^n]. \quad (7.1)$$

Here m and n are arbitrary nonnegative integers and the polynomial inside the square brackets is homogeneous of degree n in x and y, and also symmetric in x and y. The standardization here is so chosen that the coefficient of $x^{m+n}y^m$ is 1. Explicit expressions for Koornwinder's polynomials may be found in Ref. 12. The norm of $P_{m,n}^{(\alpha,\beta,l)}(x,y)$ with respect to the measure (6.12) has been calculated in Ref. 6. Attention should be paid to the notation; Koornwinder writes n, k, γ instead of $m+n, n, l+\frac{1}{2}$.

The standardization chosen in formula (7.1) is all the information about Koornwinder's polynomials that is necessary for the purpose of this paper. Applying the ladder operators one obtains

$$E^+(l)P_{m,n}^{(\alpha,\beta,l)}=n(\nu-l-1)P_{m,n-1}^{(\alpha,\beta,l+1)},$$
$$E^-(l)P_{m,n}^{(\alpha,\beta,l)}=(2l+n+1)(\nu+l)P_{m,n+1}^{(\alpha,\beta,l-1)},$$

with

$$\nu=\kappa+2m+l+n.$$

Note that the values of ν and of

$$L=l+n \quad (7.2)$$

are not changed by the application of $E^{\pm}(l)$, and therefore ν and L and possibly other indices characterize an irreducible representation. But formula (7.5) below shows that only ν and L enter the eigenvalues of the Casimir operators. ν has to do with the dimension of the representation, as will be seen later. The meaning of L can be explained immediately: Since n is not smaller than 0, L is the maximum spin value contained in the representation. Applying E^- repeatedly to $P_{m,0}^{(\alpha,\beta,L)}(x,y)$, one finally arrives at $l=0$. Therefore, any of the irreducible representations with maximum spin value L found in this paper contains also the spin values $0,1,2,\ldots,L-1$.

Introducing L in the formula for the ladder operators, one finds

$$E^+(l)P_{m,n}^{(\alpha,\beta,l)}=(L-1)(\nu-l-1)P_{m,n-1}^{(\alpha,\beta,l+1)},$$
$$E^+(l)P_{m,n}^{(\alpha,\beta,l)}=(L+l+1)(\nu+l)P_{m,n-1}^{(\alpha,\beta,l-1)}, \quad (7.3)$$

with

$$\nu=\kappa+2m+L. \quad (7.4)$$

The corresponding eigenvalues of the Casimir operators are obtained from formula (6.7):

$$c_{II}=\nu^2+L(L+2)-4,$$
$$c_{III}=0, \quad (7.5)$$
$$c_{IV}=(\nu^2-1)L(L+2).$$

What remains to be done is to identify the irreducible representation characterized by ν and L with the irreducible representation $D^{[\nu+2,(L/2),(L/2)]}$ found by Mack.[2] To this end, replace the Casimir operator C_{II} in the generators (5.2) by its eigenvalue $\nu^2+L(L+2)-4$, and split off a factor M^{ν} from the Wigner function so that ∂_μ is replaced by $M^{-\nu}\partial_\mu M^{\nu}$. The result will only be given for $p=(M,0,0,0)$. The generators different from zero are in this case

$$P^0=M,$$
$$M^{0i}=-M\partial^i,$$
$$D=i(M\partial_0+\nu+2),$$
$$K^0=M\partial^\rho\partial_\rho+(1/M)[l(l+1)-L(L+2)] \quad (7.6)$$
$$-2(M\partial_0+\nu+2)\partial^0,$$
$$K^i=(1/M)[l(l+1)-L(L+2)]$$
$$-2(M\partial_0+\nu+2)\partial^i+2i\sigma^{ij}\partial_j+(1/M)E^i.$$

Let I^i and N^i be the angular momentum and boost operators of an irreducible, finite dimensional representation $D^{(L/2,L/2)}$ of the Lorentz group. Then N^iN_i is $l(l+1)-L(L+2)$ times the unit matrix and formula (7.6) agrees with formula (6.39) of Ref. 2, describing $D^{(d,L/2,L/2)}$, if Mack's dimension parameter d is put equal to

$$d=\nu+2, \quad (7.7)$$

and if E^i is expressed as

$$E^i=2[i(d-1)N^i-\epsilon^{ikl}I_kN_l]. \quad (7.8)$$

Using the canonical basis of Naimark[13] it is not very hard to check that E^i and $2[i(d-1)N^i-\epsilon^{ikl}I_kN_l]$ have indeed the same matrix elements. This result is most astonishing: Koornwinder's polynomials have shown up in the study of the conformal group, but their ladder operators can be constructed already from the generators of the Lorentz group!

It has been found that the tensor product $D^{(d_1,0,0)}\otimes D^{(d_2,0,0)}$ contains the representation $D^{(d,L/2,L/2)}$ with L a nonnegative integer describing the spin content of the representation and with dimension parameter

$$d=d_1+d_2+2m+L, \quad (7.9)$$

where again m is a nonnegative integer. Each L and each m occur exactly once.

ACKNOWLEDGMENT

It is a pleasure to thank L. Castell and T. Koornwinder for numerous discussions and for their interest in this work.

[1]M. Lüscher and G. Mack, Commun. Math. Phys. **41**, 203 (1975).
[2]G. Mack, Commun. Math. Phys. **55**, 1 (1977).
[3]G. Mack, Commun. Math. Phys. **53**, 155 (1977).
[4]L. Castell, Talks presented at the Symposium on De Sitter and conformal groups. University of Colorado (Summer, 1970).
[5]T.H. Koornwinder, Indag. Math. **36**, 48 (1974).
[6]I.G. Sprinkhuizen-Kuyper, SIAM J. Math. Anal. 7, 501 (1965).
[7]L. Castell, Nuovo Cimento A **46**, 1 (1966); Phys. Rev. D **2**, 1161 (1970).
[8]L. Castell, Lett. Nuovo Cimento **6**, 115 (1973).
[9]L. Castell, Nucl. Phys. B **4**, 343 (1967).
[10]L. Castell and K. Ringhofer, in *Lectures in Theoretical Physics* (Colorado Associated U.P., Boulder, Colorado, 1971), Vol. XIII, p. 269.
[11]A.R. Edmonds, *Drehimpulse in der Quantenmechanik* (BI, Mannheim, 1964).
[12]T.H. Koornwinder and I.G. Sprinkhuizen-Kuyper, in SIAM J. Math. Anal. **9**, 457 (1978).
[13]M.A. Naimark, *Lineare Darstellungen der Lorentzgruppe, Hochschulbücher für Mathematik* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1963).

Properties of the Benjamin–Ono equation

K. M. Case

The Rockefeller University, New York, New York 10021
(Received 12 June 1978)

The question is raised as to whether the Benjamin–Ono equation, a nonlinear partial differential *integral* equation, is a completely integrable Hamiltonian system. The answer is almost certainly "yes." Particular solutions suggest the form that general polynomial constants must have. The structure of these and an algorithm to compute them is given. Explicit formulas are given for the first six.

I. INTRODUCTION

The Benjamin–Ono equation[1] can be written in the form

$$u_t = -uu_x - H[u_{xx}], \quad -\infty < x < \infty, \tag{1}$$

where H denotes the Hilbert transform, i.e.,

$$H[\Psi] = \frac{P}{\Pi}\int_{-\infty}^{\infty} \frac{\Psi(x')dx'}{x'-x}$$

and P signifies the principal value.

Physically, the equation describes approximately solutions of problems which have

(i) A quadratic nonlinearity, and

(ii) When the dispersion relation of the linearized problem has a long wavelength limit of the form,

$$\omega/k = 1 - \alpha|k|,$$

for some constant α.

A typical example is the evolution of long wave length internal waves in a stratified fluid.

Two opposite mathematical conjectures as to the nature of solutions of Eq. (1) can be made:

(1) Since Eq. (1) is in some formal sense intermediate between the linearizable Burger's equation

$$u_t = -uu_x - u_{xx} \tag{2}$$

and the much discussed Korteweg–de Vries equation

$$u_t = -uu_x - u_{xxx} \tag{2}$$

it should share the integrability of these.

(2) Formally, Eq. (1) is much different than Eq. (2) and (3). It is a nonlinear partial differential *integral* equation. It is nonlocal as compared to Eqs. (2) and (3) and so the solutions may be much different.

We will find that the first conjecture is more nearly correct.

The main question we wish to address here is as to whether the Benjamin–Ono equation describes a completely integrable Hamiltonian system. Our answer will be: Almost certainly yes.

The approach will be to look at solutions of Eq. (1) of a particular form. The results obtained have three uses.

(a) The explicit solutions of Eq. (1) ootained are themselves of interest.

(b) Finding these solutions is equivalent to solving a (finite dimensional) completely integrable Hamiltonian system. This suggests that the same is true for Eq. (1).

(c) Inserting the solutions into their constants of motion strongly suggests the form that general constants of motion of Eq. (1) must have.

Following this suggestion, we construct an algorithm for obtaining polynomial constants for Eq. (1) of arbitrary (fixed) order. Since the process is rather tedious, we limit our explicit results to the first six constants.

II. HAMILTONIAN FORM

Following Lax,[2] it is trivial to show that Eq. (1) is of Hamiltonian form. Thus, if we define functional derivatives $\delta F/\delta u(x)$ by

$$\frac{d}{d\epsilon}F[u+\epsilon v]\Big|_{\epsilon=0} = \int_{-\infty}^{\infty} \frac{\delta F(x)}{\delta u(x)} v(x)dx,$$

and Poisson brackets between functionals F_1, F_2 of u by

$$[F_1,F_2] = \int_{-\infty}^{\infty} \frac{\delta F_1}{\delta u(x')}\frac{\partial}{\partial x'}\frac{\delta F_2}{\delta u(x')}dx',$$

then Eq. (1) is

$$\frac{\partial u}{\partial t} = [u,\mathscr{H}],$$

where

$$\mathscr{H} = -\int\left\{\frac{u^3}{6} + \frac{u}{2}H[u_x]\right\}dx. \tag{4}$$

More generally we have seen[3] that the equation

$$\frac{\partial u}{\partial t} = u^P\frac{\partial u}{\partial x} + \frac{\partial}{\partial x}\int_{-\infty}^{\infty} G(x'-x)\frac{\partial u}{\partial x'}dx' \tag{5}$$

with $G(x) = -G(-x)$ is of Hamiltonian form with

$$\mathscr{H} = \int_{-\infty}^{\infty} \frac{u^{P+2}dx}{(P+1)(P+2)} + \frac{1}{2}\int\int \frac{\partial u}{\partial x}\hat{G}(x'-x)\frac{\partial u}{\partial x'}dx' \tag{6}$$

and

$$\frac{\partial \hat{G}}{\partial x} = G(x).$$

 0022-2488/79/050972-06$01.00

III. EXPANSIONS IN MEROMORPHIC FUNCTIONS

The success in finding solutions of the Burgers' equation[4] and the KdV equation[4,5] by means of pole type expansions suggest we look for solutions of Eq. (1) in the form

$$u = \sum_{l=1}^{N} \phi\,[x - x_l(t)] + \text{complex conjugate}, \tag{7}$$

where $\phi\,[x - x_l(t)]$ is a meromorphic function of x with no poles in the lower complex x plane. This has the effect that H acting on ϕ is very simple. Indeed $H\phi\,[x - x_l(t)] = -i\phi \times [x - x_l(t)]$. Now insert Eq. (7) in Eq. (1). Since the poles of highest order on the right do not appear on the left, they must cancel. This yields

$$\phi\phi' - i\phi'' = 0. \tag{8}$$

From this we conclude

$$\phi(x) = -ik\cot(kx/2), \tag{9}$$

where k is some complex constant. The analyticity assumption then requires us to take k real and

$$\mathrm{Im}x_l(t) > 0. \tag{10}$$

Note some simple properties of ϕ.

(1) *Antisymmetry*, $\phi(x) = -\phi(-x)$; (11)

(2) "*Reality*", $\phi\,[x - x_l(t)]^* = -\phi\,[x - x_l^*(t)]$; (12)

(3) *Addition theorem*,[4]

$$\phi(x)\phi'(y) + \phi'(x)\phi(y) = \phi(x-y)[\phi'(y) - \phi'(x)]; \tag{13}$$

(4) *Substitution theorem*,[4]

$$\phi'(x)[\phi(z) + \phi(y)] = \phi'(z)[\phi(x) + \phi(y)]. \tag{14}$$

Using Eq. (12), the ansatz of Eq. (7) becomes

$$u = \sum_l \phi\,[x - x_l(t)] - \sum_l \phi\,[x - x_l^*(t)]. \tag{15}$$

where $x + y + z = 0$.

IV. POLE EQUATIONS

If Eq. (15) is inserted in Eq. (1), we obtain on using the antisymmetry property and the addition theorem

$$-\sum_l \phi'(x - x_l)\dot{x}_l + \text{c.c.} = \sum_{\substack{ml \\ m\neq l}} \phi(x_m - x_l)\phi'(x - x_l)$$

$$+ \sum_{lm} \phi(x_l - x_m^*)\phi'(x - x_l) + \text{c.c.} \tag{16}$$

The coefficients of $\phi'(x - x_l)$ and $\phi'(x - x_l^*)$ must each be zero giving the equations

$$\dot{x}_l = -\sum_{m\neq l} \phi(x_m - x_l) + \sum_m \phi(x_l - x_m^*) \tag{17}$$

and the complex conjugate. While this reduction to a system of ordinary differential equations is a significant simplification the resulting Eqs. (17) are still rather complicated. Further, it is a little peculiar that the Hamiltonian character of our original Eq. (1) does not seem mirrored here. However, as pointed out elsewhere[6] this is readily remedied.

Differentiate Eq. (17) with respect to t and use Eq. (17) and its complex conjugate to eliminate the first derivatives. Repeated use of the antisymmetry property and the substitution theorem yields the remarkably simple result

$$\dot{x}_l = -\frac{k^2}{2}\frac{\partial}{\partial x_l}\sum_{m\neq n}\frac{1}{\sin^2 k/2(x_m - x_n)}. \tag{18}$$

Thus, we have a (complex) many body Hamiltonian system with potential $V \sim 1/\sin^2 x$.

We remark that since Eq. (18) is the time derivative of Eq. (17) that if we take a solution of Eq. (18) which satisfied Eq. (17) at some instant in time it will satisfy it at all times. Thus, we can take Eq. (18) as our fundamental dynamical equations and Eq. (17) as initial conditions. More precisely our Cauchy problem is the following: Given N different complex numbers $x_l(t_0)$ with $\mathrm{Im}\,x_l(t_0) > 0$. Then we compute $\dot{x}_l(t_0)$ from Eqs. (17). Then the $x_l(t)$ are to be found by solving Eq. (18) given $x_l(t_0)$, $\dot{x}_l(t_0)$.

Of basic importance is that Moser[7] has shown that the Eqs. (18) are completely integrable. We adapt his results into our notation. He showed that if we define

$$Z_{lm} = \frac{k}{2}\cot\frac{k}{2}(x_l - x_m) \tag{19}$$

and form the matrix L with elements

$$L_{lm} = \delta_{lm}\dot{x}_l + (1 - \delta_{lm})(-2i)Z_{lm} \tag{20}$$

and B with elements

$$B_{lm} = \delta_{lm}2i\sum_n [Z_{lm}^2 + (k^2/4)]$$

$$+ (1 - \delta_{lm})(-2i)[Z_{lm}^2 + (k^2/4)], \tag{21}$$

then Eqs. (18) are

$$\frac{\partial L}{\partial t} = [B, L]. \tag{22}$$

Thus, they are of the Lax[8] form.

Accordingly, we have very many constants of motion. Indeed $\mathrm{Tr}F[L]$ for any F is such. N linearly independent constants are

$$\mathrm{Tr}L^n, \quad n = 1,2,\ldots,N. \tag{23}$$

Before examining the possible implications of these for general constants of motion for Eq. (1), let us look in more detail at the special solutions we have arrived at.

V. THE PERIODIC CASE

If $k \neq 0$, we have solutions periodic in x with period $\lambda = 2\Pi/k$. For illustration consider first the trivial case, $N = 1$. Then

$$u = -ik\cot\frac{k}{2}(x - x_1(t)) + ik\cot\frac{k}{2}(x - x_1^*(t)),$$

where from Eq. (18) we see that $\ddot{x}_1(t) = 0$, i.e.,

$$x_1(t) = x_1(t_0) + \dot{x}_1(t_0)(t - t_0). \tag{24}$$

Here $\dot{x}_1(t_0)$ is to be found from Eq. (17). Thus,

$$\dot{x}_1(t_0) = -\phi\,[x_1(t_0) - x_1^*(t_0)] = ik\cot\tfrac{1}{2}k\,[x_1(t_0) - x_1^*(t_0)]$$

or

$$\dot{x}_1(t_0) = k\coth[k\,\mathrm{Im}x_1(t_0)]. \tag{25}$$

Note: (1) This "single soliton" solution is determined by two real numbers. The real part of $x_1(t_0)$ shich gives the initial center and the imaginary part which gives a velocity.

(2) If $\mathrm{Im}x_1(t_0) > 0$, then from Eq. (25) we see that $\dot{x}_1(t_0)$ is real. Equation (24) then shows that in accordance with our assumption we always have $\mathrm{Im}x_1(t) > 0$.

General N

Olshanetzky and Perelomov[9] have indicated a method which reduces the initial value problem for Eq. (18) to finding the eigenvalues of an $N \times N$ matrix. Unfortunately, the procedure is sufficiently complicated that it is difficult to get a clear picture of the behavior of the solution. Hence, we restrict ourselves to the first nontrivial case.

N = 2

Here we have a simple two-body problem with translationally invariant potential. With $x_1(t_0)$, $x_2(t_0)$ given we calculated $\dot{x}_1(t_0)$ and $\dot{x}_2(t_0)$ from Eqs. (17). Integrating Eqs. (18), we then obtain

$$x_{1,2}(t) = X(t) \pm x(t),$$

where

$$X(t) = \tfrac{1}{2}\{x_1(t_0) + x_2(t_0) + (t - t_0)[\dot{x}_1(t_0) + \dot{x}_2(t_0)]\}$$

and

$$x(t) = \tfrac{1}{2}[x_1(t) - x_2(t)]$$

$$= -k^{-1}\cos^{-1}[(1 - k^2/\epsilon^2)^{1/2}\cos k(\epsilon t + \gamma)].$$

Here

$$\epsilon^2 = [\dot{x}(t_0)]^2 + k^2/\sin^2 kx(t_0)$$

and

$$\gamma = -\epsilon t_0 - \frac{1}{ik}\cosh^{-1}\left[\frac{\cos kx(t_0)}{(1 - k^2/\epsilon^2)^{1/2}}\right].$$

VI. THE NONPERIODIC CASE

Particularly, simple (and suggestive) results are obtained when the period becomes infinite—$k \to 0$. The ansatz of Eq. (17) becomes

$$u = \sum_l \frac{-2i}{x - x_l(t)} + \text{c.c.} \tag{26}$$

The initial constraint on the velocities becomes

$$\frac{dx_l}{dt} = \sum_{m \neq l} \frac{2i}{x_m - x_l} + \sum_m \frac{-2i}{x_m^* - x_l}, \tag{27}$$

while the equation of motion of the poles becomes

$$\frac{d^2x_l}{dt^2} = -8\sum_{m \neq l} \frac{1}{(x_m - x_l)^3}. \tag{28}$$

First let us look at the trivial case.

N = 1

From Eq. (28)

$$\frac{d^2x_1}{dt^2} = 0,$$

and then

$$x_1(t) = x_1(t_0) + (t - t_0)\dot{x}_1(t_0). \tag{29}$$

The initial condition Eq. (27) then gives

$$\dot{x}_1(t_0) = [\mathrm{Im}x_1(t_0)]^{-1}.$$

Thus, again the "single soliton" solution is characterized by two real numbers. A velocity equal to $1/\mathrm{Im}x_1(t_0)$ and an initial position given by $\mathrm{Re}x_1(t_0)$.

General N

Previously,[10] we have shown that the solution for this case is elementary. The essential result is the following.

Theorem: Let $x_l(t_0)$, $l = 1,2,\ldots,N$ be N complex numbers which are all different and such that $\mathrm{Im}x_l(t_0) > 0$. Form the $N \times N$ matrix M with elements

$$M_{lm} = \delta_{lm}x_l(t_0) + (t - t_0)L_{lm}(t_0),$$

where

$$L_{lm}(t_0) = \delta_{lm}\left(\sum_{n \neq l} \frac{2i}{x_n(t_0) - x_l(t_0)} + \sum_n \frac{-2i}{x_n^*(t_0) - x_l(t_0)}\right)$$

$$+ \frac{(1 - \delta_{lm})(-2i)}{x_l(t_0) - x_m(t_0)}.$$

Then

$$u = -2i\frac{\partial}{\partial x}\ln\{\mathrm{Det}[M - xI]\} + \text{complex conjugate}$$

is a solution of Eq. (1).

As an example, we can write down the solution for the first nontrivial case.

N = 2

For our purposes it is convenient to write these solutions not in terms of four real numbers Re and Im parts of $x_{1,2}(t_0)$ but rather in terms of four real numbers $v_1, v_2, x_1^{(0)}, x_2^{(0)}$ whose significance will become clear. The solution is

$$u = \frac{-2i}{x - x_1(t)} - \frac{-2i}{x - x_2(t)} + \text{c.c.}, \tag{30}$$

where

$$x_{1,2}(t) = \frac{(v_1 + v_2)}{2}t + \frac{i}{2}\left(\frac{1}{v_1} + \frac{1}{v_2}\right) + \frac{1}{2}$$

$$(x_1^{(0)} + x_2^{(0)}) \pm x(t),$$

where

$$x(t) = \frac{1}{\epsilon}[1 + \epsilon^2(\gamma + \epsilon t)^2]^{1/2} \tag{31}$$

and

$$\epsilon = \frac{v_1 - v_2}{2}, \quad \mathrm{Im}\gamma = \frac{1}{2}\left(\frac{1}{v_1} - \frac{1}{v_2}\right), \quad \mathrm{Re}\gamma = \frac{x_1^{(0)} - x_2^{(0)}}{2}.$$

Without loss of generality we assume $v_1 > v_2$. Then we choose the square root so that as $t \to +\infty$

$$x_1(t) \to v_1 t + (i/v_1) + x_1^{(0)} \tag{32}$$

and

$$x_2(t) \to v_2 t + (i/v_2) + x_2^{(0)}.$$

Thus, we just obtain as solution the sum of two noninteracting single solitons. To see what happens as $t \to -\infty$, we need a little care. Continuity arguments yield different results depending on the ratio v_2/v_1. If $(3 + 2\sqrt{2})v_2 > v_1 > v_2$, then as $t \to -\infty$

$$\begin{aligned} x_1(t) &\to v_2 t + i/v_2 + x_2^{(0)} \\ x_2(t) &\to v_1 t + i/v_1 + x_1^{(0)}. \end{aligned} \tag{33}$$

Thus, the poles interchange their parameters. On the other hand, if $v_1 > (3 + 2\sqrt{2})v_2$, then as $t \to -\infty$ we obtain

$$\begin{aligned} x_1(t) &\to v_1 t + i/v_1 + x_1^{(0)} \\ x_2(t) &\to v_2 t + i/v_2 + x_2^{(0)}. \end{aligned}$$

Here the poles maintain their parameters. However, in either case the solution Eq. (30) fails to show any change in form as we go from $-\infty$ to $+\infty$. There is no asymptotic change in our solitons after they pass through each other.

The transition point $v_1 = (3 + 2\sqrt{2})v_2$ is rather interesting. With this and only this relation of parameters do we find what Calogero[11] has termed collapse. At some time the solution develops a singularity and we cannot integrate the equation further.

VII. CONSTANTS OF MOTION

We have seen that for the nonperiodic N soliton case the solution as $t \to \infty$ tends to N separated single solitons. Then the matrix L of Eq. (20) becomes simply

$$L_{lm} = \delta_{lm} v_l, \tag{34}$$

where v_l are the velocities of the single solitons. Accordingly, the N constants of motion are

$$\mathrm{Tr}L^n = \sum_{l=1}^{N} (v_l)^n. \tag{35}$$

Remark: The stability our solitons is well exhibited here. Thus, as $t \to -\infty$ we expect again to find N separated solitons with velocities v_l', $l = 1,2,...,N$. But, then

$$\mathrm{Tr}L^n = \sum_{l=1}^{N} (v_l')^n. \tag{36}$$

Since the traces are constants and this holds for all n, it is clear that the set $(v_1', v_2', ..., v_N')$ is merely some permutation of the set $(v_1, v_2, ..., v_N)$. Actually, we can say a little more. If we sum Eq. (28) over l we obtain,

$$\sum_l \frac{d^2 x_l}{dt^2} = 0,$$

$$\therefore \sum_l x_l = vt + \beta. \tag{37}$$

However, if evaluate this $t \to +\infty (-\infty)$ when we have solitons with parameters $[(v_1, x_1^{(0)}); (v_2, x_2^{(0)}); ...; (v_N, x_N^{(0)})]$ or $[(v_1', x_1^{(0)\prime}); ...; (v_N', x_N^{(0)\prime})]$, we see that

$$\sum_l x_l^{(0)} = \sum_l x_l^{(0)\prime}. \tag{38}$$

(Actually, from our $N = 2$ case we know that we only permute the v's and the $x^{(0)}$'s together.)

The existence of the many constants $\mathrm{Tr}L^n$ for all *pure soliton* solutions suggest that there may be similar constants for general solutions of the Benjamin–Ono equation. Let us now try to find them.

First note that there are three "classical" constants for quite general equations of the form like the Benjamin–Ono equation. For example, for all Eqs. (5) we have the three constants

$$I_1 = \int u\,dx, \quad I_2 = \int \frac{u^2}{2}\,dx, \quad I_3 = -2\mathscr{H}. \tag{39}$$

(These are essentially conservation of mass, momentum, and energy.) However, let us evaluate these constants for the Benjamin–Ono equation when u is a pure N soliton solution. To evaluate the constants, we wait till they are far apart and then find

$$I_1 = 4\Pi N = 4\Pi \sum_{l=1}^{N} (v_l)^0 = 4\Pi\,\mathrm{Tr}L^0 \tag{40}$$

$$I_2 = 4\Pi \sum_{l=1}^{N} v_l = 4\Pi\,\mathrm{Tr}L \tag{41}$$

$$I_3 = 4\Pi \sum_{l=1}^{N} (v_l)^2 = 4\Pi\,\mathrm{Tr}L^2. \tag{42}$$

These constants are clearly the simplest ones of the set we have seen are associated with N soliton solutions. Are there *general* constants corresponding to $\mathrm{Tr}L^n$ for $n > 2$?

Remarks: (i) By analogy with the K de V equation we might suspect that they are higher order polynomials.

(ii) Let us introduce the concept of "weight" in the following manner. We assign weight one to u and weight minus one to x. A consequence is that the operator H has weight zero.

(iii) We notice that for the Benjamin–One equation the integrands of $I_{1,2,3}$ are polynomials of order 1,2,3 and are homogeneous of weight 1,2,3 respectively.

(iv) Let $P_n(u)$ be any polynomial of order n and homogeneous of weight n. Then if in $I_n = \int P_n(u)dx$ we insert an N soliton solution and evaluate this at large times, we will obtain

$$I_n = C_n \sum_{l=1}^{N} (v_l)^{n-1}, \tag{43}$$

where the C_n are pure numbers.

From these remarks it becomes clear that the general structure of our constants must be of the form

$$I_n = \sum_{i=0}^{n-2} J_n^{(i)}, \quad n > 1, \tag{44}$$

where the integrand of $J_n^{(i)}$ is of order $(n-i)$ in u and is of weight n. Therefore, $J_n^{(i)}$ has i derivatives with respect to x.

We can normalize our constants so that

$$J_n^{(0)} = \int \frac{u^n dx}{n}. \tag{45}$$

Recursion relations connecting the $J_n^{(i)}$ are readily found. Indeed we note that if $\partial I_n/\partial t = 0$, the terms of the same order in u must be individually zero. Thus, let

$$\frac{\partial u}{\partial T} = -uu_x, \quad \frac{\partial u}{\partial t'} = -H[u_{xx}].$$

Then we must have

$$\frac{\partial J_n^{(i+1)}}{\partial T} + \frac{\partial J_n^{(i)}}{\partial t'} = 0, \quad i = 0,1,2,\ldots,n-2. \tag{46}$$

Note:

(i) $\dfrac{\partial J_n^{(0)}}{\partial T} = 0.$

(ii) $J_n^{(i)}$ must have an odd or an even number of operators H depending on whether i is odd or even. Further $J_n^{(i)}$ has *at most* i operators H. However, it can have fewer since $H^2 \equiv -1$.

(iii) The requirements almost uniquely determine $J_n^{(n-2)}$.

Example: Suppose n is even. Then $n-2$ is even. There must be 2 u's, no operators H, and $(n-2)$ derivatives with respect to x,

$$\therefore J_n^{(n-2)} = (\text{constant}) \int \left[\frac{\partial^{n/2-1} u}{\partial x^{n/2-1}}\right]^2 dx.$$

[Actually with $J_n^{(0)}$ normalized as in Eq. (45) we have found the "constant" is 2^{n-3}.]

(iv) We see that

$$\frac{\partial}{\partial t'} J_n^{(n-2)} = 0. \tag{47}$$

Now we can try to solve our recursion relation Eq. (46) starting at the top, $i = 0$, or at the bottom, $i = n-2$. (In practice it is convenient to work both ways and to meet in the middle.)

We give an example of how the procedure can be carried out. Let $n = 2m$. Then $J_n^{(1)}$ must have $(n-1)$ u's, one differentiation, and one H. The only possibilities are:

$$G_j = \int u^{2m-2-j} H[u^j u_x], \quad j = 0,1,\ldots,m-2.$$

Now if

$$F_j = \int u^{2m-2-j} u_x H[u^j u_x],$$

we find that

$$\frac{\partial J_n^{(0)}}{\partial t'} = (2m-1)F_0$$

and

$$\frac{\partial G_j}{\partial T} = -(2m-2-j)F_j + (2m-2-j)F_{j+1}.$$

Accordingly, Eq. (46) for $i = 0$ is satisfied if

$$J_n^{(1)} = \sum_{j=0}^{m-2} a_j G_j,$$

where

$$a_j = (2m-1)/(2m-2-j).$$

Clearly, the process gets complicated as n gets large. However, we have already given enough information to calculate $I_{1,2,3,4}$. We have carried the calculation through to obtain some of the lower I_n. The first three "nonclassical" ones are:

$$I_4 = \int \left\{(u^4/4) + \tfrac{3}{2}u^2 H(u_x) + 2(u_x)^2\right\} dx$$

$$I_5 = \int \left\{(u^5/5) + \left[\tfrac{4}{3}u^3 H(u_x) + u^2 H(uu_x)\right] + [2u(H(u_x))^2 + 6u(u_x)^2] - 4u_{xx}H(u_x)\right\} dx$$

$$I_6 = \int \left\{(u^6/6) + \left[\tfrac{5}{4}u^4 H(u_x) + \tfrac{5}{3}u^3 H(uu_x)\right] + \tfrac{5}{2}[5u^2(u_x)^2 + u^2(H(u_x))^2 + 2uH(u_x)H(uu_x)] \sum -10[(u_x)^2 H(u_x) + 2uu_{xx}H(u_x)] + 8(u_{xx})^2\right\} dx.$$

What are the constants I_n when evaluated for pure soliton solutions? We have indicated in Eq. (43) that they are indeed proportional to $\mathrm{Tr}L^{n-1}$. From Eqs. (40) – (42) we see

$$c_1 = c_2 = c_3 = 4\Pi.$$

An obvious conjecture is $c_n = 4\Pi$ all n. We have checked this for $n = 4$. Accordingly, we feel very confident in the following conjecture. If the constant I_n is normalized as in Eq. (45), then for pure soliton solutions

$$I_n = 4\Pi \sum_{l=1}^{N} (v_l)^{n-1}. \tag{48}$$

VIII. CONCLUSION

It seems rather certain that the Benjamin–Ono equation describes a completely integrable Hamiltonian system.

This raises a number of interesting questions. Is there a Lax[8] pair of operators B, L associated with this equation so that it can be written in the form $\partial L/\partial t = [B,L]$? If the answer is positive, we should be able to use the Inverse Scattering Transform method to discuss the general initial value problem for the Benjamin–Ono equation. Further there

should then be an analog of Lenards[12] algorithm which would enable us to calculate the I_n much more efficiently.

If the answer is negative, i.e., there is no Lax pair, this would be extraordinarily interesting. It would be the first such example known.

Are the constants I_n in involution? The ones we have given are.

APPENDIX: EXPLICIT SOLUTION OF THE CALOGERO MODEL

In Sec. VI we have given the explicit N soliton solution for the nonperiodic problem. The proof of this uses a theorem due to Olshanetsky and Perelomov[11] and Calogero.[13] The published proofs are rather complicated. Here we give an elementary one.

Theorem: Let $x_j(t)$ be the solution of Eqs. (28) subject to $x_j(t_0)$, $\dot{x}_j(t_0)$ being given. If L is the appropriate Lax matrix and $K[x(t)]$ is the matrix with elements $K_{ij} = \delta_{ij} x_j(t)$, then

$$K[x(t)] = U\{K[x(t_0)] + (t - t_0)L[x(t_0),\dot{x}(t_0)]\}U^{-1}. \tag{A1}$$

Proof: We have indicated that Eqs. (28) can be written in the form

$$\frac{\partial L}{\partial t} = [B,L]. \tag{A2}$$

Define $U(t)$ by $\partial U/\partial t = BU$ subject to $U(t_0) = 1$. Consider

$$J(t) = U^{-1}K[x(t)]U. \tag{A3}$$

On differentiating we find

$$\frac{\partial J}{\partial t} = U^{-1}\{K[\dot{x}(t)] + [K,B]\}U.$$

However, in Ref. 5 it is shown that

$$[K,B] = L - K[\dot{x}(t)],$$

$$\therefore \frac{\partial J}{\partial t} = U^{-1}LU. \tag{A4}$$

Differentiating this once more yields,

$$\frac{\partial^2 J}{\partial t^2} = U^{-1}\left\{\frac{\partial L}{\partial t} - [B,L]\right\}U = 0, \tag{A5}$$

in view of Eq. (A2),

$$\therefore J = C_1 + C_2(t - t_0), \tag{A6}$$

where C_1 and C_2 are constant matrices. Putting $t = t_0$ in Eq. (A3), shows that

$$C_1 = K[x(t_0)] \tag{A7}$$

while putting $t = t_0$ in Eq. (A4) yields

$$C_2 = L[x(t_0),\dot{x}(t_0)]. \tag{A8}$$

Putting then C_1 and C_2 into Eq. (A6) and multiplying on the left by U, the right by U^{-1} yields Eq. (A1).

[1] T.B. Benjamin, J. Fluid Mech. **29**, 559 (1967); **25**, 241 (1966); R.F. Davis and A. Acrivos, J. Fluid Mech. **29**, 593 (1967); H. Ono, J. Phys. Soc. Japan **39**, 1082 (1975).
[2] P.D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975).
[3] K.M. Case, Phys. Rev. Lett. **40**, 351 (1978).
[4] D.V. Chudnovsky and G.V. Chudnovsky, Nuovo Cimentos B **40**, 339 (1977).
[5] H. Airault, H.P. McKean, and J. Moser, Commun. Pure Appl. Math. **30**, 95 (1977).
[6] K.M. Case, Proc. Natl. Acad. Sci. **75**, 3562 (1978).
[7] J. Moser, Adv. Math. **16**, 197 (1975).
[8] P.D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).
[9] M.A. Olshanetsky and A. M. Perelomov, Lett. Nuovo Cimento **17**, 97 (1976).
[10] K.M. Case, Physica (to be published).
[11] In the proof of this theorem explicit use is made of a result of M.A. Olshanetsky and A.M. Perelomov, Lett. Nuovo Cimento **16**, 333 (1976). A considerably simpler proof of this theorem is given in the Appendix.
[12] See Sec. 6, Ref. 2.
[13] F. Calogero, Lett. Nuovo Cimento **16**, 35 (1976).

ERRATA

Erratum: A particle model based on stringlike solitons [J. Math. Phys. 19, 1304 (1978)]

U. Enz

Philips Research Laboratories, Eindhoven, The Netherlands
(Received 19 September 1978)

In this paper, Eq. (16) is in error and should be replaced by

$$\alpha = e^2/\hbar c = a^2 R_0^{-1} l I_1^{-1}. \qquad (16)$$

Equation (16) follows from (14) and (15) with $e^2 = a^2 E = A l^2 a^2$. The numbers a, R_0/l, and I_1 follow from the basic action principle and depend on the specific form of $w = w(D)$ in Eq. (6). They have not yet been calculated.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*

 0022-2488/79/050978-01$00.00